Why Money Trickles Up

Geoff Willis

gwillis@econodynamics.org

The right of Geoffrey Michael Willis to be identified as the author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

0.0 Abstract

This paper combines ideas from classical economics and modern finance with Lotka-Volterra models, and also the general Lotka-Volterra models of Levy & Solomon to provide straightforward explanations of a number of economic phenomena.

Using a simple and realistic economic formulation, the distributions of both wealth and income are fully explained. Both the power tail and the log-normal like body are fully captured. It is of note that the full distribution, including the power law tail, is created via the use of absolutely identical agents.

It is further demonstrated that a simple scheme of compulsory saving could eliminate poverty at little cost to the taxpayer. Such a scheme is discussed in detail and shown to be practical.

Using similar simple techniques, a second model of corporate earnings is constructed that produces a power law distribution of company size by capitalisation.

A third model is produced to model the prices of commodities such as copper. Including a delay to capital installation; normal for capital intensive industries, produces the typical cycle of short-term spikes and collapses seen in commodity prices.

The fourth model combines ideas from the first three models to produce a simple Lotka-Volterra macroeconomic model. This basic model generates endogenous boom and bust business cycles of the sort described by Minsky and Austrian economists.

From this model an exact formula for the Bowley ratio; the ratio of returns to labour to total returns, is derived. This formula is also derived trivially algebraically.

This derivation is extended to a model including debt, and it suggests that excessive debt can be economically dangerous and also directly increases income inequality.

Other models are proposed with financial and non-financial sectors and also two economies trading with each other. There is a brief discussion of the role of the state and monetary systems in such economies.

The second part of the paper discusses the various background theoretical ideas on which the models are built.

This includes a discussion of the mathematics of chaotic systems, statistical mechanical systems, and systems in a dynamic equilibrium of maximum entropy production.

There is discussion of the concept of intrinsic value, and why it holds despite the apparent substantial changes of prices in real life economies. In particular there are discussions of the roles of liquidity and parallels in the fields of market-microstructure and post-Keynesian pricing theory.

0.2 Introduction

"The produce of the earth -- all that is derived from its surface by the united application of labour, machinery, and capital, is divided among three classes of the community; namely, the proprietor of the land, the owner of the stock or capital necessary for its cultivation, and the labourers by whose industry it is cultivated.........To determine the laws which regulate this distribution, is the principal problem in Political Economy…"

On The Principles of Political Economy and Taxation - David Ricardo [Ricardo 1817]

"We began with an assertion that economic inequality is a persistent and pressing problem; this assertion may be regarded by many people as tendentious. Differences in economic status – it might be argued – are a fact of life; they are no more a 'problem' than are biological differences amongst people, or within and amongst other species for that matter. Furthermore, some economists and social philosophers see economic inequality, along with unfettered competition, as essential parts of a mechanism that provides the best prospects for continuous economic progress and the eventual elimination of poverty throughout the world. These arguments will not do. There are several reasons why they will not do.........However there is a more basic but powerful reason for rejecting the argument that dismisses economic inequality as part of the natural order of things. This has to do with the scale and structure of inequality……"

Economic Inequality and Income Distribution – DG Champernowne [Champernowne & Cowell 1998]

"Few if any economists seem to have realized the possibilities that such invariants hold for the future of our science. In particular, nobody seems to have realized that the hunt for, and the interpretation of, invariants of this type might lay the foundations for an entirely novel type of theory."

Schumpeter (1949, p. 155), discussing the Pareto law – via [Gabaix 2009]

This paper introduces some mathematical and simulation models and supports these models with various theoretical ideas from economics, mathematics, physics and ecology.

The models use basic economic variables to give straightforward explanations of the distributions of wealth, income and company sizes in human societies.

The models also explain the source of macroeconomic business cycles, including bubble and crash behaviour.

The models give simple formulae for wealth distributions, and also for the Bowley ratio; the ratio of returns to labour and capital.

Usefully, the models also provide simple effective methods for eliminating poverty without using tax and welfare.

The theoretical ideas provide a framework for extending this modelling approach systematically across economics.

The models were produced firstly by taking basic ideas from classical economics and basic finance. These ideas where then combined with the mathematics of chaotic systems and dynamic statistical mechanics, in a process that I think can be well summed up as 'econodynamics' as it parallels the approaches of thermodynamics, and ultimately demonstrates that economics is in fact a subset of thermodynamics.

This makes the process sound planned. It wasn't. It was a process of common sense and good luck combined with a lot of background reading.

It was suggested to me in 2006 That the generalised Lotka-Volterra (GLV) distribution might provide a good fit for income data. The suggestion proved to be prescient. The fit to real data proved to be better than that for other previously proposed distributions.

At this point, in 2006, I used my limited knowledge of economics to propose two alternative models that might fit the simplest economically appropriate terms into two different generating equations that produce the (GLV). I passed these ideas forward to a number of physicists. The history of this is expanded in section 12.

After that, nothing very much happened for three years. This was for three main reasons. Firstly, I didn't understand the detailed mathematics, or indeed have a strong feel for the generalised Lotka-Volterra model. Secondly, my computer programming, and modelling skills are woeful. Thirdly, the academics that I wrote to had no interest in my ideas.

In 2009/2010 I was able to make progress on the first two items above, and in early 2010 I was able, with assistance from George Vogiatzis and Maria Chli, to produce a GLV distribution of wealth from a simulation programme with just nine lines of code, that included only a population of identical individuals, and just the variables of individual wealth (or capital), a single uniform profit rate and a single uniform (but stochastic) consumption (or saving) rate. This simple model reproduced a complex reality with a parsimony found rarely even in pure physics.

After a brief pause, the rest of the modelling, research and writing of this paper was carried out between the beginning of May 2010 and the end of March 2011. This was done in something of a rush, without financial support or academic assistance; and I would therefore ask forbearance for the rough and ready nature of the paper.

From the first wealth-based model, and with greater knowledge of finance and economics; models for income, companies, commodities and finally macroeconomics dropped out naturally and straightforwardly. The models are certainly in need of more rigorous calibration, but they appear to work well.

The wealth and income models appear to be powerful, both in their simplicity and universality, and also in their ability to advise future action for reducing poverty.

The macroeconomic models are interesting, as even in these initial simple models, they give outcomes that accord closely with the qualitative descriptions of business and credit cycles in the work of Minsky and the Austrian school of economics. These descriptions describe well the actual behaviour of economies in bubbles and crashes from the Roman land speculation of 33AD through tulipomania and the South Sea bubble up to the recent credit crunch.

Part A of this paper goes through these various models in detail, discussing also the background and consequences of the models.

The agents in the initial models were identical, and painfully simple in their behaviour. They worked for money, saved some of their money, spent some of their money, and received interest on the money accumulated in their bank accounts.

Because of this the agents had no utility or behavioural functions of the sort commonly used in agent-based economic modelling. As such the models had no initial underlying references to neoclassical economics, or for that matter behavioural economics. There simply was no need for neoclassicism or behaviouralism.

As the modelling progressed, somewhat to my surprise, and, in fact to my embarrassment, it became clear that the models were modelling the economics of the classical economists; the economics of Smith, Ricardo, Marx, von Neumann (unmodified) and Sraffa.

With hindsight this turned out to be a consequence of the second of the two original models I had proposed in 2006. In this model wealth is implicitly conserved in exchange, but created in production and destroyed in consumption. Ultimately total wealth is conserved in the long term. This model denies the premises of neoclassicism, and adopts an updated form of classical economics.

Despite the rejection of neoclassicism, the models work.

Classical economics works.

Where the classical economists were undoubtedly wrong was in their belief in the labour theory of value. They were however absolutely correct in the belief that value was intrinsic, and embodied in the goods bought, sold and stored as units of wealth. Once intrinsic wealth, and so the conservation of wealth is recast and accepted, building economic models becomes surprisingly easy.

The re-acceptance of intrinsic wealth; and so the abandonment of neoclassicism, is clearly controversial. Given the wild gyrations of the prices of shares, commodities, house prices, art works and other economic goods, it may also seem very silly. Because of this a significant section of part B of this paper discusses these issues in detail, and the economic and finance background in general.

The other main aim of part B of this paper is to introduce the ideas of chaotic systems, statistical mechanics and entropy to those that are unfamiliar with them.

Partly because of these theoretical discussions this paper is somewhat longer than I initially expected. This is mainly because I have aimed the paper at a much larger audience than is normal for an academic paper. In my experience there are many people with a basic mathematical background, both inside and outside academia, who are interested in economics. This includes engineers, biologists and chemists as well as physicists and mathematicians. I have therefore written the paper at a level that should be relatively easy to follow for those with first year undergraduate mathematics (or the equivalent of a UK A-level in maths).

Although the numbers are much smaller, I believe there is also a significant minority of economists, especially younger economists, who are acutely aware that the theory and mathematical tools of economics are simply not adequate for modelling real world economies. This paper is also aimed at these economists.

I would not be particularly surprised if every single model in this paper has to be reworked to make them describe real world economies. It may even be the case that many of the models have to be superseded. This would be annoying but not tragic, but is beside the point.

The main point of this paper is the power of the mathematical tools. The two main tools used in this paper are chaotic differential equation systems and statistical mechanics. In both cases these tools are used in systems that are away from what are normally considered equilibrium positions.

It is these tools that allow the production of simple effective economic models, and it is these tools that economists need in order to make progress.

Comparative statics may be intellectually satisfying and neat to draw on a blackboard, but it doesn't work in dynamic multi-body systems.

For a dynamic system you need dynamic differential equation models. For systems with large numbers of interacting bodies you need statistical mechanics and entropy.

Although a minority of economists have toyed with chaos theory, and many economists claim to use 'dynamic' models, I have only encountered one economist; Steve Keen, who truly 'gets' dynamic modelling in the way that most physicists, engineers and mathematical modellers use dynamic modelling.

Indeed the macroeconomic model in this paper shares many ideas with, and certainly the approaches of, Steve Keen who has used dynamical mathematical models to follow the ideas of Goodwin, Minsky and others; and who has used the Lotka-Volterra dynamics in particular. Although Keen's models are certainly heterodox he is almost unique in being an economic theoretician who predicted the credit crunch accurately and in depth. While other economists predicted the credit crunch, almost all the others who did so did this from an analysis of repeating patterns of economic history. That is, they could spot a bubble when they saw one. Steve Keen is unusual in being a theoretical economist who is able to model bubbles with a degree of precision.

The use of statistical mechanics in economics is even more frustrating. Merton, Black and Scholes cherry-picked the diffusion equation from thermodynamics while completely ignoring its statistical mechanical roots and derivation. They then sledge-hammered it into working in a neoclassical framework. Tragically, a couple of generations of physicists working in finance have not only accepted this, but they have built more and more baroque models on these flimsy foundations. The trouble with Black-Scholes is that it works very well, except when it doesn't. This basic flaw has been pointed out from Mandlebrot onwards, to date with no notice taken. This is most frustrating. If physicists were doing their jobs properly, finance would be one of the simplest most boring parts of economics.

The only economist I have encountered who truly 'gets' statistical mechanics is Duncan Foley. He is uniquely an economist who has fully realised not only the faults with the mathematics used by most economists, but also dedicated considerable effort to applying the correct mathematics, statistical mechanics, to economics. Although primarily modelled in a static environment, Foley's work is profoundly insightful, and demonstrates very clearly how statistical mechanical approaches are more powerful than utility based approaches, and how statistical mechanics approaches naturally lead to the market failures seen in real economies. Despite this visionary insight he has ploughed a somewhat lonely furrow, with the relevant work largely ignored by economists, and more embarrassingly also by physicists.

Because chaos and statistical mechanics are unfamiliar in economics, I have spent some effort in both the modelling sections and the theory sections in explaining how the models work in detail, how these concepts work in general, and why these mathematical approaches are not just relevant but essential for building mathematical models in economics.

This extra explanation for less mathematical scientists and economists may mean that the paper is over-explained and repetitive for many physicists and mathematicians. For this I can only offer my apologies.

However, even for physicists some of the background material in the discussions on entropy contains novel and powerful ideas regarding non-equilibrium thermodynamic systems. This is taken from recent work in the physics of planetary ecology and appears not to have percolated into the general physics community despite appearing to have general applicability. The ideas of Paltridge, Lorenz, Dewar and others, along with the mathematical techniques of Levy & Solomon, may not be familiar to many physicists, and I believe may be very powerful in the analysis of complex 'out of equilibrium' systems in general.

In fact, although I was trained as a physicist, I am not much of a mathematician, and by emotional inclination I am more of an engineer. My skills lie mostly in seeing connections between different existing ideas and being able to bolt them together in effective and sometimes simpler ways. Part of the reason for the length of this paper is that I have taken a lot of ideas from a lot of different fields, mainly from classical economics, finance, physics, mathematics and ecology, and fitted them together in new ways. I wish to explain this bolting together in detail, partly because very few people will be familiar with all the bits I have cherry-picked, but also I suspect that my initial bolting together may be less than ideal, and may need reworking and improving.

I feel I should also apologise in advance for a certain amount of impatience displayed in my writing towards traditional economics. From an economics point of view the paper gets more controversial as it goes along. It also gets increasingly less polite with regard to the theories of neoclassical economics.

In the last two years I have read a lot of economics and finance, a significant proportion of which was not profoundly insightful. Unfortunately, reading standard economics books to find out how real economies work is a little like reading astrology books to find out how planetary systems work. Generally I have found the most useful economic ideas in finance or heterodox economics, areas which are not usually well known to physicists, or indeed many economists. These ideas include recent research in market microstructure, liquidity, post-Keynesian pricing theory as well as the work of Foley, Keen, Smithers, Shiller, Cooper, Pettis, Pepper & Oliver, Mehrling, Lyons and others.

Neoclassical economics, while forming an intellectually beautiful framework, has proved of limited use to me as a source of knowledge. Partly this is because the mathematics used, comparative statics, is simply inappropriate. Partly it is because some of the core suppositions used to build the framework; such as diminishing returns and the importance of investment and saving, are trivially refutable.

The only defence I can make for my impoliteness is a very poor one; that I am considerably more polite than others. If any of my comments regarding neoclassical economics cause offence, I advise you to read the work of Steve Keen and Phillip Mirowski with some caution. Both are trained economists who have the mathematical and historical skills to realise the inappropriateness of neoclassicism. Their writing has the polemical edge of a once devout Christian who has recently discovered that the parish priest has been in an intimate liaison with his wife for the last fifteen years.

Finally I would like to comment on the work of Ian Wright, Makoto Nirei & Wataru Souma and others.

Throughout this paper comparisons are made to the work of Ian Wright who describes simulated economic models in two notable papers [Wright 2005, 2009]. Wright's models are significantly different to my own, most notably in not involving a financial sector. Also, unlike the present paper, Wright takes a 'black box' and 'zero intelligence' approach to modelling which eschews formal fitting of the models to mathematical equations. Despite these profound differences, at a deeper level Wright's models share fundamental similarities with my own, sharing the basic conservation of value of the classical economists, as well as using a dynamic, stochastic, statistical mechanical approach. More significantly, the models are striking in the similarities of their outputs to my own work. Also it is important to note that Wright's models have a richness in some areas, such as unemployment which are missing from my own models.

In relevant sections I discuss detailed differences and similarities between the models of Wright and myself.

In two papers Souma & Nirei [Souma & Nirei 2005, 2007] build a highly mathematical model that produces a power tail and an exponential distribution for income. Their approach also builds ultimately on the work of Solomon & Levy. However their approach is substantially more complex than my own. Their models do however share a number of similarities to my own models. Firstly, the models of Souma & Nirei use consumption as the negative balancing term in their model in a manner almost identical to the role of consumption in my own model. Secondly, their models ascribe a strong positive economic role to capital as a source of wealth, however this is ascribed to the process of capital growth, not the dividends, interest, rent, etc that is used in my own models.

Both Wright's work and that of Souma & Nirei predate this paper. Their work also predates my original models produced in 2006. Given the process by which I came to produce the models below, I believe I did so independently of Wright, Souma & Nirei. However, I would be very foolish to discount that possibility that I was subconsciously influenced by these authors, and so I do not discount this. It is certainly clear to me that Wright, Souma & Nirei have made very substantial inroads in the same directions as my own research, and that if I had not had lucky breaks in advancing my own research, then one or other of them would have produced the models below within the near future.

Given that the work of Wright, Souma & Nirei predates my own, and so gives rise to questions of originality, I have included a brief history of the gestation of the present paper in section 12, History and Acknowledgements.

With regard to precedence, I would like to note that the general approach for the macroeconomic models in section 4 were partly inspired by the work of Steve Keen, though the models themselves grew straight out of my company and commodity models; and ultimately out of my income models.

More importantly, not a word of this paper would have been written without the work of Levy & Solomon and their GLV models. Manipulation of the GLV is beyond my mathematical ability. Although Levy & Solomon's economic explanations are naïve, their gut feeling of the applicability of the GLV to economics in particular, and complex systems in general, was correct. I believe their work is of profound general importance.

In later sections of this paper I quote extensively from the work of Ian Wright, Duncan Foley and Steve Keen, as their explanations of the importance of statistical mechanics and chaos in economics are difficult to improve on.

0.3 Structure of the Paper

Part A of this paper discusses a number of economic models in detail, Part A.I discusses a number of straightforward models giving results that easily accord with the real world and also with the models of Ian Wright. Part A.II discusses models that are more speculative.

Part B discusses the background mathematics, physics and economics underlying the models in Part A. The mathematics and physics is discussed in Part B.I, the economics in part B.II, the conclusions are in part B.III. Finally, Part C gives appendices.

Within Part A; section 1 discusses income and wealth distributions; section 1.1 gives a brief review of empirical information known about wealth and income distributions while section 1.2 gives background information on the Lotka-Volterra and General Lotka-Volterra models. Sections 1.3 to 1.5 gives details of the models, their outputs and a discussion of these outputs.

Section 1.6 discusses the effects that changing the ratio of waged income to earnings from capital has on wealth and income distributions.

Sections 1.7 and 1.8 discuss effective, low-cost options for modifying wealth and income distributions and so eliminating poverty.

Finally, section 1.9 looks at some unexplained but potentially important issues within wealth and income distribution.

Sections 2.1 to 2.4 go through the background, creation and discussion of a model that creates power law distributions in company sizes.

Sections 3.1 to 3.4 use ideas from section 2, and also the consequences of the delays inherent in installing physical capital, to generate the cyclical spiking behaviour typical of commodity prices.

Sections 4.1 to 4.4 combine the ideas from sections 1, 2 and 3 to provide a basic macroeconomic model of a full, isolated economy. It is demonstrated that even a very basic model can endogenously generate cyclical boom and bust business cycles of the sort described by Minsky and Austrian economists.

In section 4.5 it is demonstrated that an exact formulation for the Bowley ratio; the ratio of returns to labour to total returns, can easily be derived from the basic macroeconomic model above, or indeed from first principles in a few lines of basic algebra.

In section 4.6 and 4.7 the above modelling is extended into an economy with debt. From this a more complex, though still simple, formulation for the Bowley ratio is derived. This formulation suggests that excessive debt can be economically dangerous and also directly increases income inequality. The more general consequences of the Bowley ratio for society are discussed in more depth in section 4.8.

In section 4.9 two macroeconomic models are arranged in tandem to discuss an isolated economy with a financial sector in addition to an ordinary non-financial sector. In section 4.10 two macroeconomic models are discussed in parallel as a model of two national economies trading with each other.

To conclude Part A, section 4.11 introduces the role of the state and monetary economics, while section 4.12 briefly reviews the salient outcomes of the modelling for social equity.

In Part B, section 6.1 discusses the differences between static and dynamic systems, while section 6.2 looks at the chaotic mathematics of differential equation systems. Examples of how this knowledge could be applied to housing markets is discussed in section 6.3, while applications to share markets are discussed in section 6.4. A general overview of the control of chaotic systems is given in section 6.5.

Section 7.1 discusses the theory; 'statistical mechanics', which is necessary for applying to situations with many independent bodies; while section 7.2 discusses how this leads to the concept of entropy.

Section 7.3 discusses how systems normally considered to be out of equilibrium can in fact be considered to be in a dynamic equilibrium that is characterised as being in a state of maximum entropy production. Section 7.4 discusses possible ways that the statistical mechanics of maximum entropy production systems might be tackled.

Moving back to economics; in section 8.1 it is discussed how an intrinsic measure of value can be related to the entropy discussed in section 7 via the concept of 'humanly useful negentropy'.

Section 8.2 discusses the many serious criticisms of a concept of intrinsic value in general, with a discussion of the role of liquidity in particular.

Section 9.1 looks at theories of supply and pricing, the non-existence of diminishing returns in production, and the similarities between the market-microstructure analysis and post-Keynesian pricing theory. Section 9.3 looks for, and fails to find, sources of scarcity, while section 9.4 discusses the characteristics of demand.

In section 10 both the theory and modelling is reviewed and arranged together as a coherent whole, this is followed by brief conclusions in section 11.

Sections 12 to 16 are appendices in Part C.

Section 12 gives a history of the gestation of this paper and an opportunity to thank those that have assisted in its formation.

Section 13 gives a reading list for those interested in learning more about the background maths and economics in the paper.

Section 14 gives details of the Matlab and Excel programmes used to generate the models in Part A of the paper.

Sections 15 and 16 give the references and figures respectively.

Part A – Some Models

Section A.I – Heavy Duty Models

1. Wealth & Income Models

1.1 Wealth & Income Data – Empirical Information

"Endogeneity of distribution

Neoclassical economics approaches the problem of distribution by positing a given and exogenous distribution of ownership of resources. The competitive market equilibrium then determines the relative value of each agent's endowment (essentially as rents). I think there are problems looming up with this aspect of theory as well. One reason to doubt the durability of the assumption of an exogenous distribution of ownership of resources is that income and wealth distributions exhibit empirical regularities that are as stable as any other economic relationships. I think there is an important scientific payoff in models that explain the size distributions of wealth and income as endogenous outcomes of market interactions." Duncan K. Foley [Foley 1990]

Within theoretical economics, the study of income and wealth distributions is something of a backwater. As stated by Foley above, neo-classical economics starts from given exogenous distributions of wealth and then looks at the ensuing exchange processes. Utility theory assumes that entrepreneurs and labourers are fairly rewarded for their efforts and risk appetite. The search for deeper endogenous explanations within mainstream economics has been minimal.

This is puzzling, because, as Foley states, it has been clear for a century that income distributions show very fixed uniformities.

Vilfredo Pareto first showed in 1896 that income distributions followed the power law distribution that now bears his name [Pareto 1896].

Pareto studied income in Britain, Prussia, Saxony, Ireland, Italy and Peru. At the time of his study Britain and Prussia were strongly industrialised countries, while Ireland, Italy and Peru were still agricultural producers. Despite the differences between these economies, Pareto discovered that the income of wealthy individuals varied as a power law in all cases.

Extensive research since has shown that this relationship is universal across all countries, and that not only is a power law present for high income individuals, but the gradient of the power law is similar in all the different countries.

Typical graphs of income distribution are shown below. This is data for 2002 from the UK, and is an unusually good data set [ONS 2003].

Figure 1.1.1 here

Figure 1.1.1 above shows a probability density function. A probability distribution function (pdf) is basically a glorified histogram or bar chart. Along the x-axis are bands of wage. The y-axis shows the number of people in each wage band.

As can be seen this shape has a large bulge towards the left-hand side, with a peak at about £300 per week. To the right hand side there is a long tail showing smaller and smaller numbers of people with higher and higher earnings.

Also included in this chart is a log-normal distribution fitted to the curve. The log-normal distribution is the curve that economists normally fit to income distributions (or pretty much anything else that catches their attention). On these scales the log-normal appears to give a very good fit to the data. However there are problems with this.

Figure 1.1.2 here

Figure 1.1.2 above shows the same data, but this time with the y-axis transformed into a log scale. Although the log-normal gives a very good fit for the first two thirds of the graph, somewhere around a weekly wage level of £900 the data points move off higher than the log-normal fit. The log-normal fit cannot describe the income of high-earners well.

Figure 1.1.3 here

Figure 1.1.3 above shows the same data but organised in a different manner. This is a 'cumulative density function' or cdf. In this graph the wealth is still plotted along the x-axis, but this time the x-axis is also a log scale. This time the y-axis shows the proportion of people who earn more than the wage on the x-axis.

In figure 1.1.3 about 10% of people, a proportion of 0.1, earn more than £755 per week.

It can be seen that the curve has a curved section on the left-hand side, and a straight line section on the right-hand side.

This straight section is the 'power-tail' of the distribution. This section of the data obeys a 'power-law' as described by Pareto 100 years ago.

The work of Pareto gives a remarkable result. An industrial manufacturing society and an agrarian society have very different economic systems and societal structures. Intuitively it seems reasonable to assume that income would be distributed differently in such different societies.

What the data is saying is that none of the following have an effect on the shape of income distribution in a country:

- Whether wealth is owned as industrial capital or agricultural land
- Whether wealth is owned directly or via a stock market
- What sort of education system a country has

- What sort of justice system a country has
- Natural endowments of agricultural land or mineral wealth
- And so on with many other social and economic factors

Intuitively it seems reasonable that any or all of the above would affect income distribution, in practice none of them do. Income distributions are controlled by much deeper and basic processes in economics.

The big unexpected conclusion from the data of Pareto and others is the existence of the power tail itself. Traditional economics holds that individuals are fairly rewarded for their abilities, a power tail distribution does not fit these assumptions.

Human abilities are usually distributed normally, or sometimes log-normally. The earning ability of an individual human being is made up of the combination of many different personal skills.

Logically, following the central limit theorem, it would be reasonable to expect that the distribution of income would be a normal or log-normal distribution. A power law distribution however is very much more skewed than even a log-normal distribution, so it is not obvious why individual skills should be overcompensated with a power law distribution.

While Pareto noted the existence of a power tail in the distribution, it should be noted that more recently various authors have suggested that there may be two or even three power tail regions, with a separation between the 'rich' and 'super-rich', see for example [Borges 2002, Clementi & Gallegati 2005b, Souma, Nirei & Souma 2007].

While the income earned by the people in the power tail of income distribution may account for approximately 50% of total earnings, the Pareto distribution actually only applies to the top 10%-20% of earners. The other 80%-90% of middle class and poorer people are accounted for by a different 'body' of the distribution.

Going back to the linear-linear graph in figure 1.1.1 it can be seen that, between incomes of £100 and £900 per week, there is a characteristic bulge or hump of individuals, with a skew in the hump towards the right hand side.

In the days since Pareto the distribution of income for the main 80%-90% of individuals in this bulge has also been investigated in detail.

The distribution of income for this main group of individuals shows the characteristic skewed humped shape similar to that of the log-normal distribution, though many other distributions have been proposed.

These include the gamma, Weibull, beta, Singh-Maddala, and Dagum. The last two both being members of the Dagum family of distributions. Bandourian, McDonald & Turley [Bandourain et al 2002] give an extensive overview of all the above distributions, as well as other variations of the general beta class of distributions. They carry out a review of which of these distributions give best fits to the extensive data in the Luxembourg Income Study. In all they analyse the fit of eleven probability distributions to twenty-three different countries. They conclude that the Weibull, Dagum and general-beta2 distributions are the best fits to the data depending on the number of parameters used.

For more information, readers are referred to 'Statistical Size Distributions in Economics and Actuarial Sciences' [Kleiber & Kotz 2003] for a more general overview of probability distributions in economics, and also to Atkinson and Bourguignon [Atkinson & Bourguignon 2000] for a very detailed discussion of income data and theory in general.

The author has analysed a particularly good set of income data from the UK tax system, one example is shown in figures 1.1.1-3 above. This data suggests that a Maxwell-Boltzmann distribution also provides a very good fit to the main body of the income data that is equal to that of the log-normal distribution [Willis & Mimkes 2005].

The reasons for the split between the income earned by the top 10% and the main body 90% has been studied in more detail by Clementi and Gallegati [Clementi & Gallegati 2005a] using data from the US, UK, Germany and Italy. This shows strong economic regularities in the data. In general it appears that the income gained by individuals in the power tail comes primarily from income gained from capital such as interest payments, dividends, rent or ownership of small businesses. Meanwhile the income for the 90% of people in the main body of the distribution is primarily derived from wages. These conclusions are important, and will be returned to in the models below.

This view is supported, though only by suggestion, by one intriguing high quality income data set. This data set comes from the United States and is from a 1992 survey giving proportions of workers earning particular wages in manufacturing and service industries.

The ultimate source of the data is the US Department of Labor; Bureau of Statistics, and so the provenance is believed to be of the good quality. Unfortunately, enquiries by the author has failed to reveal the details of the data, such as sample size and collection methodology.

The data was collected to give a comparison of the relative quality of employment in the manufacturing and service sectors. Although the sample size for the data is not known, the smoothness of the curves produced suggest that the samples were large, and that the data is of good statistical quality. The data for services is shown in figures 1.1.4 & 1.1.5 below, the data for manufacturing is near identical.

Figure 1.1.4 here

Figure 1.1.5 here

Like the UK data, there appears to be a clear linear section in the central portion of the data on a log-linear scale in figure 1.1.5, indicating an exponential section in the raw data. Again this data can be fitted equally well with a log-normal or a Maxwell-Boltzmann distribution.

What is much more interesting is that, beyond this section, the data heads rapidly lower on the logarithmic scale. This means it is heading rapidly to zero on the raw data graph. With these two distributions there is no sign whatsoever of the 'power tail' that is normally found in income distributions.

It is the belief of the author that the methodology for this US survey restricted the data to 'earned' or 'waged' income, as the interest in the project was in looking at pay in services versus manufacturing industry. It is believed income from assets and investments was not included as this would have been irrelevant to the investigation.

This US data set has been included for a further reason, a reason that is subtle; but in the belief of the author, important.

Looking back at figure 1.1.1 for the UK income data, there is a very clear offset from zero along the income axis. That is the curve does not start to rise from the income axis until a value of roughly £100 weekly wage.

The US data shows an exactly similar offset, with income not rising until a weekly wage of $100.

This is important, as the various curves discussed above (log-normal, gamma, Weibull, beta, Singh-Maddala, Dagum, Maxwell-Boltzmann, etc) all normally start at the origin of the axis, point (0,0) with the curve rising immediately from this point.

While it is straightforward enough to put an offset in, this is not normally necessary when looking at natural phenomena.

In the 1930s Gibrat, an engineer, pioneered work in economics that studied work on proportional growth processes that could produce log-normal or power law distributions depending on the parameters. His work primarily looked at companies, and was the first attempt to apply stochastic processes to produce power law distributions.

Following the work of Pareto, the details of income and wealth distributions have rarely been studied in mainstream theoretical economics, a notable and important exception being Champernowne. Champernowne was a highly gifted mathematician who was diverted into economics, he was the first person to bring a statistical mechanical approach to income distribution, and also noted the importance of capital as a major creator of inequality, though his approach concentrated on generational transfers of wealth [Champernowne & Cowell 1998].

Despite the lack of interest within economics, this area has had a profound attraction to those outside the economics profession for many years, a review of this history is provided by Gabaix [Gabaix 2009].

In recent years, the study of income distributions has gone through a small renaissance with new interest in the field shown by physicists with an interest in economics, and has become a significant element of the body of research known as 'econophysics'.

Notable papers have been written in this field by Bouchaud & Mezard, Nirei & Souma, Dragulescu & Yakovenko, Chatterjee & Chakrabarti, Slanina, Sinha and many, many, others [Bouchaud & Mezard 2000, Dragulescu & Yakovenko 2001, Nirei & Souma 2007, Souma 2001, Slanina 2004, Sinha 2005].

The majority of these papers follow similar approaches; inherited either from the work of Gibrat, or from gas models in physics. Almost all the above models deal with basic exchange processes, with some sort of asymmetry introduced to produce a power tail. Chatterjee et al 2007, Chatterjee & Chakrabarti 2007 and Sinha 2005 give good reviews of this modelling approach.

The approaches above have been the subject of some criticism, even by economists who are otherwise sympathetic to a stochastic approach to economics, but who are concerned that a pure exchange process is not appropriate for modelling modern economies [Gallegati et al 2006].

An alternative approach to stochastic modelling has been taken by Moshe Levy, Sorin Solomon, and others [Levy & Solomon 1996].

They have produced work based on the 'General Lotka-Volterra' model. Unsurprisingly, this is a generalised framework of the 'predator-prey' models independently developed for the analysis of population dynamics in biology by two mathematicians/physicists Alfred Lotka and Vito Volterra.

A full discussion of the origin and mathematics of GLV distributions is given below in section 1.2.

These distributions are interesting for a number of reasons; these include the following:

- the fundamental shape of the GLV curve
- the quality of the fit to actual data
- the appropriateness of the GLV distribution as an economic model

Figure 1.1.6 here

Figure 1.1.7 here

With regard to the fundamental shape of the GLV curve, figures 1.1.6 and 1.1.7 above show plots of the UK income data against the GLV on a linear-linear and log-log plot.

The formula for this distribution is given by:

$$P(w) = K(e^{-(\alpha-1)/(w/L)})/((w/L)^{(1+\alpha)}) \qquad (1.1a)$$

and it has three parameters; K is a general scaling parameter, L is a normalising constant for w, and α relates to the slope of the power tail of the distribution.

It should firstly be noted that the GLV has both a power tail and a 'log-normal'-like main body. That is to say it can model both the main population and the high-end earners at the same time. This is a very significant advantage over other proposed distributions.

The second and more subtle point to note is that the GLV has a 'natural' offset from zero. It is in the nature of the GLV that the rise from zero probability on the y-axis starts at a non-zero value on the x-axis, this is discussed further in section 1.2 Below.

Finally the detailed fit of the GLV appears to be equivalent or better than the log-normal distribution.

Figure 1.1.8	Reduced Chi Squared	
	Full Data Set	Reduced Data Set
Boltzmann Fit	3.27	1.94
Log Normal Fit	2.12	3.02
GLV Fit	1.21	1.83

Figure 1.1.8 above gives results from a basic statistical analysis using the GLV, log-normal and Maxwell-Boltzmann distributions. (The values in the table are the reduced chi-squared values, using an assumed standard measurement error of 100. The actual measurement error is not known, so the values above are not absolute, however, changing the measurement value will change the values in the table by equal proportions, so the relative sizes of the values in the table will stay the same.)

It can be seen from the figures in the first column that the GLV, with the lowest value of chi-squared, gives the best fit. In itself this is not altogether surprising, as it is known that the log-normal and the Maxwell-Boltzmann have exponential tails, and so are not able to fit power tails.

More remarkably, the figures in the second column show the same analysis carried out using a truncated data set with an upper limit of £800 per week. This limit was taken to deliberately exclude the data from the power tail. Again it can be seen that the GLV still just gives the best fit to the data. This in itself suggests that the GLV should be preferred to the log-normal or the Maxwell-Boltzmann distributions.

It is also of note that in parallel to the work of Solomon et al, Slanina has also proposed an exchange model that produces the same output distribution as the GLV [Slanina 2004].

Unfortunately the modelling approaches of Solomon et al, and Slanina use economic models that are not wholly convincing, and as such have significant conceptual shortcomings.

It is the belief of the author that an alternative economic analysis, using more appropriate analogies allows a much more effective use of GLV distributions in an intuitive and simple economic formulation. This is the third main reason for preferring the GLV distribution, and forms the key content of the initial sections of this paper. As previously noted Souma & Nirei have also pursued research in this direction.

Before discussing the GLV distribution in detail I would firstly like to review some background on power law distributions.

Power laws are deeply beloved of theoretical physicists, and there are many different ways to produce power laws. Most theoretical physicists tend to have a particular affection for their pet process and it's particular mathematical derivation, and then proceed to fit their pet equations to any model that happens to have a power tail with gay abandon. Also, as is usually necessary, this requires the sledgehammer of many pages of complex mathematical derivation, in an attempt to fit a square peg into a round hole. An unfortunate consequence of this is that most of the very extensive literature on power laws is confusing, apparently conflicting, and to a great extent simply incoherent.

This is a shame, as most power laws distributions are actually produced very simply, in a restricted number of ways. For those who want more background on the formation of power laws, log-normal laws and related processes, there are three very good background papers by Newman [Newman 2005], Mitzenmacher [Mitzenmacher 2004] and Simkin & Roychowdhury [Simkin & Roychowdhury 2006].

The papers by Newman and Mitzenmacher give very good overviews of what make power law and log-normal normal distributions without being mathematically complex.

One basic point from the papers is that there are many different ways of producing power law distributions, but the majority fall into three main classes.

The first class gives a power law distribution as a function of two exponential distributions; of two growth processes.

The second class gives power law distributions as an outcome of multiplicative models. This is the route that Levy and Solomon have followed in their work, and forms the basis for the GLV distribution discussed in detail in the next section.

The third class for producing power laws uses concepts of 'self-organised criticality' or 'SOC'.

A second basic point, discussed in Mitzenmacher, is that the difference between a log-normal distribution and a power law distribution is primarily dependent on the lower barrier of the distribution, if the lower barrier is at zero, then you get a log-normal distribution, if the barrier is above zero, then the distribution gives a power tail. A non-zero barrier, provided by wage income, is an essential part of the GLV model discussed in section 1.2 below.

The paper of Simkin and Roychowdhury is illuminating and entertaining. It shows that the same basic mechanisms for producing power laws, and branching processes in general, have been rediscovered dozens of times, and that most power law / branching processes are in fact analogous. As an example, the models of Levy & Solomon follow processes previously described by Champernowne in economics, and ultimately by Yule and Simon almost a century ago. This is not to devalue the work of Solomon and Levy; their approach allows for dynamic equilibrium formation, this includes an element missing from most branching models that in my opinion makes the Solomon and Levy model much more powerful as a general model. This is returned to in section 1.2 below. It is however my belief that reading Simkin and Roychowdhury by all those involved in modelling power laws would make their lives a lot easier.

Finally it is important to note the difference between income and wealth.

Income data is relatively easy to collect from income tax returns. Pareto's original work and almost all subsequent analysis of data is based on that from income data.

Wealth data of any quality is very difficult to find. Where this data has been collected it almost exclusively pertains to the richest portion of society, and suggests that wealth is also distributed as a power law for these people.

I am not aware of any data of sufficient quality to give any conclusions about the distribution of wealth amongst the bottom 90% of individuals.

This has led to some very unfortunate consequences within the econophysics community.

Without exception all the exchange models by all the various authors above, including those of Solomon and Slanina, are wealth exchange models. I have not yet seen a model where income (trivially the time derivative of wealth) is measured.

Despite this, the output distributions from these wealth models are often judged to be successful when they map well onto data derived from income studies.

Wealth and income (and sometimes money) are used interchangeably in econophysics papers. This is most unfortunate. A paper on physics; written by an economist, that used energy and power interchangeably would be greeted with considerable scorn by physicists.

An explanation for why wealth models can give outputs that can then define income data successfully is given in section 1.4.4 below.

Before moving on to the modelling of income and wealth distributions, I would first like to discuss the derivation and mechanics of the Lotka-Volterra distribution and the GLV distribution in more detail.

1.2 Lotka-Volterra and General Lotka-Volterra Systems

1.2.1 Lotka-Volterra systems

Lotka-Volterra systems were independently discovered by Alfred Lotka [Lotka 1925] and Vito Volterra [Volterra 1926] and are used to describe the dynamics of populations in ecological systems. Ultimately this dynamic approach goes back directly to the economic growth equations of Malthus and Sismondi.

A basic Lotka-Volterra system consists of a population of prey (say rabbits) whose size is given by x, and a population of predators (say foxes) given by y.

Not explicitly given in this simple case, it is further assumed that there is a steady supply of food (eg. grass) for the prey.

When no predators are present this means that the population of the rabbits is governed by:

$$\frac{dx}{dt} = ax \qquad (1.2.1\,a)$$

where a is the population growth rate.

Left to their own business, this would give exponential, Malthusian growth in the population of the rabbits.

In the absence of any rabbits to eat, it is assumed that there is a natural death rate of the foxes:

$$\frac{dy}{dt} = -cx \qquad (1.2.1\,b)$$

where c is the population die-off rate, and the negative sign indicates a decline in the population. This would give an exponential fall in the fox population.

When the foxes encounter the rabbits, two further effects are introduced, firstly the rate at which rabbits are killed is proportional to the number of rabbits and the number of foxes (ie the chance of foxes encountering rabbits), so:

$$\frac{dx}{dt} = -\alpha x y \qquad (1.2.1\,c)$$

where α is a constant, and the –ve sign indicates that such encounters are not good for the rabbits. However these interactions are good for the foxes, giving:

$$\frac{dy}{dt} = \gamma x y \qquad (1.2.1d)$$

Where γ is again a fixed constant.

Taken together, the results above give a pair of differential equations:

$$\begin{aligned}\frac{dx}{dt} &= ax - \alpha x y \\ &= x(a - \alpha y) \qquad (1.2.1e)\end{aligned}$$

for the rabbits, and:

$$\begin{aligned}\frac{dy}{dt} &= \gamma x y - cy \\ &= y(\gamma x - c) \\ &= y(-c + \gamma x) \qquad (1.2.1f)\end{aligned}$$

for the foxes.

The most important point about this pair of equations is that x depends on y, while at the same time, y depends on x. The dependency goes in both directions, this make things fun.

While it is possible for these equations to have a single stable solution, this is often not the case. Commonly the populations of both rabbits and foxes fluctuates wildly. An example is given in figure 1.2.1.1 below for lynx preying on arctic hares [BBC]:

Figure 1.2.1.1 here

The data for the graph above comes from long-term records of pelts collected by the Hudson Bay Company. The graph shows very closely the recurrent booms and busts in population of the two types of animals. In the short term the population and total biomass of both lynx and hares can increase or decrease substantially. The population of lynx can be large or small in proportion to that of the hares. The populations of both are highly unstable.

A subtlety to note is that the population of the lynx follows, 'lags', the population of the hares. It is also worth considering, even at this early stage, the behaviour, or indeed the 'behaviouralism' of the lynx in particular.

Following a previous collapse, the population of hares can expand rapidly as there are very few lynx to hunt them.

As the population of hares increases rapidly, the lynx behave 'rationally' (at least given the absence of long-term, liquidly tradable, hare futures) in both eating lots of hares, and also giving birth to lots of new lynx to feed on the excess of hares.

Eventually, of course there are too many lynx for the population of hares, and ultimately there are too many lynx and hares for the underlying amount of grass available.

At the peaks of hare and lynx populations there is simply too much biomass wandering around for the land to support.

Despite the substantial fluctuations seen in figure 1.2.1.1 above, the populations of both lynx and hares show stable fluctuations around long term averages; roughly 40,000 or so for the hares and 20,000 or so for the lynx, though note that the populations pass through these average values very quickly.

In fact the values of the two populations are confined to a band of possible values. The population can move round in a limited set of possible options, this is shown for example in the two figures from simulations below.

Figure 1.2.1.2 here

Note also the figure 1.2.1.2 shows the same leads and lags in predator and prey populations as the real data. The populations of wolves and rabbits can be displayed on one graph, this then produces the phase diagram in figure 1.2.1.3 below showing how the population of wolves and rabbits vary with each other, and how they are constrained to a particular set of paths.

Figure 1.2.1.3 here

These diagrams are taken from the website of Kumar, [Kumar 2006], which gives a very good brief introduction to the maths and modelling of Lotka-Volterra systems.

It can be seen that the simulated population of wolves and rabbits wanders continuously around average values of approximately seventeen rabbits and six wolves.

In contrast, figures 1.2.1.4 & 5 below show the same system with minor changes to the rates of growth. In this model the oscillations slowly die down to stable long-term values. Another alternative is that the oscillations can grow in size unstably and explode to infinity.

Figure 1.2.1.4 here

Figure 1.2.1.5 here

One of the important things to note about non-linear dynamic systems such as these is that relatively minor changes in parameters can result in dramatic differences in system behaviour.

All the talk of predators and prey can give rise to emotive, and wholly inappropriate, language and modelling. It is an easy, but foolish, course to represent one group of actors (financiers say) as predators, and others (workers) as prey. This is flawed for two reasons. Sometimes the mathematics works the other way, so for example, the Marxian inspired models of Goodwin actually model workers as predators. More importantly, the maths and models are impersonal; they are totally unconnected to the motives of the actors.

In fact you don't need both predators and prey, a solitary animal population that grows too quickly can also suffer from population booms and crashes. An example is that of Soay sheep on the island of Soay (in this case the grass can be considered to be the prey, though a better solution would be to use the logistic equation or a similar carrying capacity based approach).

1.2.2 General Lotka-Volterra (GLV) systems

As the name implies, the General Lotka-Volterra system (GLV) is a generalisation of the Lotka-Volterra model to a system with multiple predators and prey. This can be represented as:

$$\frac{dx_i}{dt} = x_i r_i + \sum_{j=1}^{N} a_{i,j} x_i x_j \qquad (1.2.2a)$$

$$= x_i(r_i + \sum_{j=1}^{N} a_{i,j} x_j) \qquad (1.2.2b)$$

here, dx_i/dt is the overall rate of change for the i-th particular species, out of a total of N species. This is made up of two terms.

The first term is the natural growth (or death) rate, r_i, for the species, where x_i is the population of species i. This rate r_i is equivalent to the growth rate 'a' in equation (1.2.1e) or the death rate '-c' in equation (1.2.1f).

The second term gives the sum of all the interactions with the j number of other species. Here $a_{i,j}$ is the interaction rate defining the relationship between species i and j.

$a_{i,j}$ is negative if species j is a predator feeding on species i, positive if species i is a predator feeding on species j, or can be of either sign for a heterotroph. $a_{i,j}$ is equivalent to the α of equation (1.2.1e) or the γ of equation (1.2.1f).

Hopefully it is clear that equations (1.2.2a) and (1.2.2b) are generalisations of equations (1.2.1e) and (1.2.1f) for many interacting species.

For each species in the system, potentially N-1 interaction rates $a_{i,j}$ are needed, while N! separate differential equations are needed to describe the whole system. This makes direct solution of the equations for the system somewhat problematic.

Fortunately in many systems it is possible to make simplifying assumptions. As an example Solomon [Solomon 2000] proposes the following difference equation as a possible explanation for the power law distribution of city population sizes. This equation describes changes in the distribution in terms of discrete time-steps from time t to time t+1:

$$w_{i,t+1} = \lambda_t w_{i,t} + a_t \bar{w}_t - c_t \bar{w}_t w_{i,t} \qquad (1.2.2c)$$

The terms on the right hand side, in say the year 2003, the year t, add up to give the population w of city i in the year 2004 on the left hand side, which is at time t+1.

Such equations are typically used in simulations, one after the other, to give a model of how populations change. Sometimes, though often not, clever mathematicians can derive output population distributions from the underlying difference equations.

In equation (1.2.2c), λ is the natural growth rate of the population w of city i, but is assumed that λ is the same for each city.

a_t is the arrival rate of population from other cities, which is multiplied by the average population $\bar{w}$ of all the cities.

The final term gives the rate of population leaving each city, which is due to the probability c_t of an individual meeting a partner from another city. This is given by multiplying the average population $\bar{w}$ with the population of city i.

Leaving aside the detail of the model, important generalisations have been made to produce a more tractable model.

In this case λ, a and c are universal rates, applicable to all members of the system.

λ and a both give 'positive autocatalytic' (positive feedback) terms which increase the population w of each city. While the negative value of c ensures that the population of each city has an element of decrease.

In the absence of the negative feedback term, the populations of the cities can increase indefinitely to infinity without reaching a stable solution.

In the absence of the positive autocatalytic growth of the λ in the first term on right hand side, the second and third terms will cause all of the population to end up in a single city.

Normally one or more variables are assumed to be stochastic; that is they can vary randomly. In Solomon's example above, all three of λ, a and c are assumed to be stochastic. This stochasticity need not be large; it can be small fluctuations around a long-term mean, but it ensures that a locally stable solution is not reached, and that the system evolves into a single long term equilibrium solution.

While the above may seem complex, it will be argued later in section 7.3 that this model can be seen as a very general model across many different real world complex systems.

It is possible to show (though not by me) that the above system can give a stable resultant probability distribution function of populations over the various cities of the form:

$$P(w) = (e^{-(\alpha-1)/w})/(w^{(1+\alpha)}) \qquad (1.2.2d)$$

Which is the general form of the GLV distribution. Or more specifically:

$$P(w) = K(e^{-(\alpha-1)/(w/L)})/((w/L)^{(1+\alpha)}) \qquad (1.2.2e)$$

As has been shown above in section 1.1 this formula gives a very good fit to income data.

As well as the quality of fit there are three other reasons that suggest that the GLV may be appropriate for wealth and income distributions.

The first two reasons are technical and are discussed below, the third is more subjective and forms the core of this paper.

A first reason for preferring the GLV is that this distribution is notable in that the distribution has a main body that is similar to a Maxwell-Boltzmann distribution or log-normal Maddala etc distribution, while the tail follows a power law distribution.

While other theories, from both economics and physics, are able to explain one part of the distribution well, it is generally necessary to invoke complex assumptions to explain the remaining part of the distribution, if such an explanation is even attempted. The GLV kills both the birds of income distribution with a single theoretical stone.

The second reason for preferring the GLV is that the autocatalytic terms in the GLV give the GLV an automatic offset from zero.

As noted above in section 1.1 both the UK and US income data show this offset.

While it is perfectly straightforward to put an offset into a log-normal or Maxwell-Boltzmann and other distributions, systems commonly found in nature modelled by the above distributions typically have their origin at zero.

The third reason is that the GLV naturally describes complex dynamic flow systems that have reached a maximum entropy production equilibrium. Economics is such a complex dynamic flow system, and it will be seen that the straightforward models described below model real economic outcomes surprisingly well.

Solomon further proposes a similar model as an explanation for income distribution:

$$w_{i,t+1} = \lambda_t w_{i,t} + a_t \bar{w}_t - c_t \bar{w}_t w_{i,t} \qquad (1.2.2f)$$

In this case λ is proposed to be positive gains by individuals with origins on the stock market, 'a' is assumed to represent wealth received in the form of *'subsidies, services and social benefits'*, while 'c' is assumed to represent competition for scarce resources, or *'external limiting factors (finite amount of resources and money in the economy, technological inventions, wars, disasters, etc.) as well as internal market effects (competition between investors, adverse* *influence of bids on prices such as when large investors sell assets to realize their value and prices fall as a result'.*

While it is the author's belief that a form of the GLV is appropriate for modelling wealth and income distributions, it is believed that the above economic mechanisms are not realistic.

At heart the models of Levy & Solomon remain pair exchange models, with random movements of wealth between individuals. As a realistic description of an economic system this falls short of reasonable requirements.

As noted previously, Souma & Nirei [Souma & Nirei 2005, Nirei & Souma 2007] have uniquely moved forward from Levy & Solomon's work in a way that gets closer to meaningful economic fundamentals, however their models include a high degree of complexity.

It is also noteworthy that Slanina has produced a pair exchange model that generates an identical output distribution to the GLV output, again it is contended that simple pair exchange is not appropriate as an economic model [Slanina 2004].

In the next section an economic model is proposed that I believe much more closely represents real life economic mechanisms.

1.3 Wealth & Income Models - Modelling

Figure 1.3.1 here

Figure 1.3.1 above shows a simple macroeconomic model of an economy. This model is taken from figure 1 of chapter 2 of 'Principles of Economics', by Gregory Mankiw [Mankiw 2004].

Figure 1.3.2 below shows a modified version of the diagram. The two 'markets' between the firms and households have been removed, investment and saving streams have been added, as well as the standard economics symbols for the various flows.

Figure 1.3.2 here

All standard economics textbooks use similar diagrams to figures 1.3.1 and 1.32 for macroeconomic flows; I have chosen that of Mankiw as his is one of the most widely used.

Flows of goods and services are shown in the black lines. The lighter broken lines show the flows of money. (As a simple-minded engineer I prefer diagrams that include flows of goods as well as cash, as I find them easier to follow.)

Note that Mankiw shows households owning 'factors of production' such as land and capital, which the households are then shown as selling to firms. This is indicated as a flow of land and capital (along with labour) from households to firms.

I personally have never actually sold any machine tools to a manufacturing company, and I have never met any householder who has done so. We will return to this particular 'flow' later.

Note also that the total system shows a contained circularity of flow, with balances between supply and demand of goods and services.

In this circular flow model economic textbooks assume some basic equalities:

$$C = G \qquad (1.3a)$$

$$C = Y \qquad (1.3b)$$

Equation (1.3b) state that the total income gained from firms adding value is equal to the total consumption of goods and services.

[Nb. In writing this paper I have attempted to use standard notation from economics wherever possible. This occasionally results in confusion. It should be noted that the capital letter Y is used as standard in (macro) economics for income, while small y is used as standard in (micro) economics for outputs from companies.

This is not normally a problem, as the two are rarely discussed at the same time in standard economic models.

In the discussions of income that follows y is not actually necessary for the analysis, and Y invariably refers to income in the equations of the mathematical model and is normally subscripted.]

Figure 1.3.3 here

In figure 1.3.3 above I have modified this standard model to reflect what I believe is something closer to reality.

Firstly in this model households have been changed to individuals, this is simply to bring the model more in line with the standard analysis of statistical physics and agent based, modelling techniques. This amounts to little more than pickiness. This distinction can be made irrelevant by simply assuming that all households consist of a single individual.

Much more importantly, the flow pattern has been changed and the circularity has been disturbed.

In the real world most goods and services are consumed in a relatively short period of time. To show this, Consumption C, has been changed to represent the actual consumption of goods. This is a real flow of goods, and represents a destruction of value. Note that this is a change from the standard use of C in economics textbooks.

That which was previously shown as consumption is now shown as 'y' the material output of goods and services, which are provided to consumers from the firms operating in the economy.

The money paid for these goods and services is shown by My.

As can be seen in figure 1.3.3 above, the income stream Y has been split into two components, one, e is the earnings; the income earned from employment as wages and salaries, in return for the labour supplied.

π is the 'profit' and represents the payments made by firms to the owners of capital, this can be in the form of dividends on shares, coupons on bonds, interest on loans, rent on land or other property, etc.

The flow of capital has been shown as a dotted line. This is because, as pointed out previously, capital doesn't flow. Householders do not hold stocks of blast furnaces in their backyards in the hope of selling them to firms in exchange for profit or interest on their investments.

Capital, such as machine tools and blast furnaces, is normally bought in by firms from other firms, sometimes using money provided by households, but mostly by retained earnings.

In fact in all the various models that follow in this paper we are going to ignore both investment I, and saving S. In the income models it is always assumed that the overall economy is in a steady state and so, firstly, that all funds required for wear & repair are taken from internal flows. More importantly, in later models; both for companies and macroeconomic modelling, it is also assumed that all new capital is produced from retained earnings within companies.

For many economists, somewhat oddly, this will be seen as a serious flaw. Since at least the time of Keynes, investment and saving have been at the heart of macroeconomic modelling, and this is true of neo-classical and other heterodox modelling, not just that in the Keynesian tradition. The reasons for this are not understood by the author; given that:

"Most corporations, in fact, do not finance their investment expenditure by borrowing from banks." [Miles & Scott 2002, 14.2]

As examples, Miles & Scott give the following table for proportions of investment financing for four countries averaged over the years 1970-1994.

Figure 1.3.4 here

[Miles & Scott 2002 / Corbett & Jenkinson 1997]

As can be seen the maximum possible proportion of external financing (the IS so beloved of economists) is 36.8% for Japan. For the UK it doesn't even reach 20%. This financing is small to negligible in importance. Most financing is taken from cash flow. Companies that have spare cash buy new toys to play with. Companies that don't, don't. In the whole of this paper the economic models follow reality rather than hypothesis. They are built by modelling capital created and destroyed through imbalances in cash flow.

External investment is ignored as the sideshow that it is. Why the whole of macroeconomics should build their models directly contrary to observed data evidence remains a profound mystery.

Going back to capital; real capital, in the form of land, machine tools, computers, buildings, etc will be represented in the diagram as fixed stocks of real capital K, held by the companies.

All of this real capital is assumed to be owned by households, in the form of paper assets, W, representing claims on the real assets in the form of stocks or shares. In the following discussions bonds and other more complex assets will be ignored, and it will be assumed that all the wealth of K is owned in the form of shares (stocks) in the various firms.

This paper wealth will be represented as W in total, or w_i for each of i individuals.

For the income models in the first part of this paper it will further be assumed that the paper wealth of the households accurately represents the actual physical capital owned by the companies, so:

$$\text{total W} = \text{total K} \qquad (1.3c) \qquad \text{or:}$$

$$\sum w_i = W = K \qquad (1.3d)$$

the total real capital invested in the firms is equal to the total value of financial assets held by individuals.

The dotted line in the figure 1.3.3 indicates the assumed one to one link between the financial assets W and the real assets K. It is dotted to show that it is not a flow, it simply indicates ownership.

This mapping of real and financial assets assumes that the financial assets are 'fairly' priced, and can be easily bought and sold in highly liquid markets.

In the models below it is assumed that there is a steady state, so the totals of W and K are both constant. This means that the model has no growth, and simply continues at a steady equilibrium of production and consumption. There is no change in population, no change in technology, no change in the efficiencies of the firms. The example of Japan over the last two decades has shown that economies can continue to function in a normal manner with extended periods of negligible growth. For a modern economy the difference between the creation and the destruction is economic growth of the GDP, and at 2%-4% or so per annum is pretty close to being stable.

This assumption of equality between W and K will be relaxed in later models, with interesting results; but for the moment we will assume the market operates efficiently with regard to asset pricing.

It is important to note that the capital discussed here is only the capital vested in productive companies. Other personal capital is excluded, the most important of these is housing. I have ignored the role of housing in these early models, though clearly this is a major simplification. This is discussed further in section 1.9.1 below. For the moment all wealth held is assumed to be financial assets. All other personal assets such as housing, cars, jewellery, etc are ignored.

There are some other important base assumptions of the model. These are discussed briefly below:

The economy is isolated; there are no imports or exports.

There is no government sector, so no taxation or welfare payments, government spending, etc.

There is no unemployment; all individuals are employed, with a given wage, either from a uniform distribution or a normal distribution depending on the model.

Labour and capital are assumed to be complementary inputs and are not interchangeable at least in the short term. It turns out, much later, that this assumption is not only true, but of profound importance, this is discussed at some length later in this paper.

There is no investment and saving, the economy is stationary, and depreciation is made good from earned profits.

The role of money is ignored in these models, for the sake of argument, it can be assumed that payments are made in the form of units δW of the paper assets held by the individuals, say in units of DJI or FTSE all share trackers.

Finally there is no debt included in the income models.

Figure 1.3.5 below shows some of the assumptions above, it also adds in some more flows to help bring the model closer to the real world.

Figure 1.3.5 here

There are two main reasons for changing the diagram in this manner. One reason is to bring the diagram into line with the ideas of the classical economists such as Smith, Ricardo, Marx and Sraffa. The second is to help the model comply with some of the more basic laws of physics.

Starting with the classical economics. It has previously been defined that consumption by the individuals means the destruction of value in the form of using up resources. This consumption could be food eaten in a few days, clothes which wear out in a few months or cars and furniture that take years to wear out, but which ultimately need to be replaced periodically. The consumption can also be services such as meals in restaurants, going to see films, receiving haircuts, going on holiday, etc. All value destruction is assumed to take place within households as consumption.

In physics terms, this destructive process is characterised as a local increase in entropy.

To balance this destruction, it is assumed that all value is created in the processes of production, and that all this value is created within firms.

I am going to follow in Schrödinger's footsteps and describe this increase in value as the creation of something called 'negentropy'. For physicists a better term might be 'humanly useful free energy'. For non-physicists, it is asked that detailed understanding of the meanings of 'entropy', 'negentropy' or 'humanly useful free energy' are postponed to part B, where it is discussed at length. For the moment the important thing to grasp is that negentropy is equivalent to economic value, the more negentropy something has, the more you are willing to pay for it.

Although the discussions in these models use production of manufactured goods as an easily understandable example; it should be noted that 'production' is any process that adds value, and produces higher value inputs than the outputs. So agriculture, mining, power generation, as well as distribution, retail, personal and financial services are all forms of production.

Indeed, almost any process that is done within a company is production. That is why companies exist, so that the value added is kept securely within the company.

In general, exchange processes don't create value, they are simply a means for swapping goods from different points along the supply chain leading up to the final point of consumption. Exchanges are simply a result of the division of labour between different companies or individuals who have particular sets of skills and abilities.

Whether it is the sale of 'lemon' used cars, or the manipulative momentum trading of high-frequency traders, if value is created for one party during an exchange process then this is usually a consequence of an inadequately regulated market that lacks proper informational transparency.

The model in figure 1.3.5 above essentially goes back to the ideas of the classical economists; of Smith, Ricardo, Marx, Sraffa and others. It assumes that goods and services have meaningful, long term, intrinsic values, and that long-term prices reflect these values. Short-term prices may move away from these values, primarily to allow generation of new capital.

In the models in this paper it is always assumed that value is created in production and that normally exchanges are 'fair' and so there is not net gain of value for either party in an exchange process, again this discussed at more length later in the paper.

This paper explicitly rejects the marginalist view that value is exogenously set by the requirements and beliefs of individuals, and that exchange between such individuals creates value.

Figure 1.3.6 here

Figure 1.3.6 above figure demonstrates these assumptions for a more complex model of linear flows of value added.

In figure 1.3.6, all the horizontal flows (flows through the side walls) are direct exchanges of actual goods for monetary tokens. Assuming a free market with fair pricing, and that the currency is a meaningful store of value, then all the horizontal exchange flows have zero net value.

$$x1 + Mx1 = 0 \qquad \text{or:}$$

$$x1 = -Mx1, \qquad x2 = -Mx2, \qquad \ldots \qquad xk = -Mxk, \qquad \text{etc}$$

Vertical flows, through the top and bottom of the boxes, involve changes; increases or decrease in negentropy.

In economic terms this is stated as value being added or wealth being created. In figure 1.3.6 above the values of the final output y and the series of inputs x are related by:

$$y > x3 > x2 > x1 \quad \text{and clearly}$$

$$My > Mx3 > Mx2 > Mx1$$

The differences between these values represents the wealth created by the employees and capital of the firm acting on the inputs to create the outputs. The employees are rewarded for this wealth creation via their wage earnings, while the owners of the capital are rewarded with returns on their capital.

Figure 1.3.7 here

Figure 1.3.7 above gives another layout that shows that the whole system doesn't have to be linear, but that the same assumptions regarding adding value still hold.

Finally to satisfy the physicists reading; waste streams are included so that the 2nd law is not violated. The total entropy created by the waste streams from the firms, principally low grade heat, is greater than the negentropy created in the products of the firms.

Essentially figures 1.3.5 to 1.3.7 bring together the economic and physical diagrams discussed in Ayres & Nair [Ayres & Nair 1984]; so that the circulation of wealth and money complies with the laws of physics as well as the laws of finance. The discussions of Ayres & Nair clearly have strong antecedents in the theories of Georgescu-Roegen [Georgescu-Roegen 1971].

Figure 1.3.5 here

So, going back to figure 1.3.5, we are now at a point where we can move into the detail of the mathematical model.

Firstly we will assume that x = Mx and that both are irrelevant to the rest of the debate.

We will also assume that L = e, ie that labour is fairly rewarded for the value of its input. In later sections this is discussed in more depth, but becoming bogged down in a tedious Marxist debate at this stage of the modelling would be particularly unhelpful.

Next we will assume y = My, ie that 'fair' prices are being paid for the goods sold to the consumers. We will eventually relax this assumption in later models.

In this model it will further be assumed that:

$$\text{total } C = \text{total } y = \text{total } My$$

at steady state equilibrium.

It will be seen later that this is actually a natural outcome of the model, and doesn't need to be forced. Note that although the totals of C and y are the same, they may not be the same for individuals. Some individuals may consume less than they earn, or vice versa.

In these earlier models, we are not interested in the detail of the firms so we are going to ignore the difference between the capital K and it's financial equivalent W.

We will assume that total K = total W, and so assume that companies are fairly and accurately priced in the financial markets. These assumptions will be relaxed later, again with interesting consequences.

The paper wealth W will be split between N individuals, so from individual i = 1 to individual i = N.

Going back to figure 1.3.5 and equation 1.3d above; although the total capital and wealth is fixed, individual wealth is allowed to vary, so:

$$\sum w_{i,t} = \sum w_{i,t+1} = W = K = \text{constant} \qquad (1.3e)$$

Where w_i is the wealth of individual i.

This is economics at a statistical level; a level below microeconomics, nanoeconomics perhaps.

Looking at a single individual in the box on the right of figure 1.3.5, in one time unit, from t to t+1, the change in wealth is given by the following equation:

$$w_{i,t+1} = w_{i,t} + y_{i,t} - My_{i,t} + e_{i,t} + \pi_{i,t} - C_{i,t} - \text{labour}_{i,t} - \text{capital}_{i,t} \qquad (1.3f)$$

This equation states that the wealth for a single individual at time t+1, on the left hand side, is equal to the wealth at time t, plus the contributions of the seven arrows going into or out of the box on the right hand side of figure 1.3.5.

However equation (1.3f) is not meaningful as it is trying to add apples and oranges. The items y, C, labour and capital are real things, while w, My, e and π are all financial quantities. Adding the non-financial things is not appropriate, however all the financial flows must ultimately add up.

So looking then at the financial flows, we have the following equation:

$$w_{i,t+1} = w_{i,t} - My_{i,t} + e_{i,t} + \pi_{i,t} \qquad (1.3g)$$

This now counts things that are the same (remember that the currency used for our cash flows were units of δW).

As stated above, although the totals of My = y = C some individuals can consume less than y, and so accumulate more wealth W, others can consume more than y and so reduce their total W.

To make this process clearer, I am going to use – $C_{i,t}$ in place of – $My_{i,t}$ in equation (1.3g).

In this case $C_{i,t}$ is now a monetary unit, and effectively reverts to standard economics usage. To keep the units correct, it is assumed that in practice heavy consumers exchange part of their wealth W with some heavy savers, in return for some of the savers real goods y. This may seem a little confusing but is hoped this will become clearer as the model is more fully explained.

Substituting and rearranging, this then leaves us with the following equation:

$$w_{i,t+1} = w_{i,t} + e_{i,t} + \pi_{i,t} - C_{i,t} \qquad (1.3h)$$

This then is the difference equation for a single agent in this model.

In a single iteration, the paper wealth w of an individual i increases by the wages earned e plus the profits received π. The individual's wealth also reduces by the amount spent on consumption C.

A moment's reflection suggests that this is trivially obvious.

We now need to investigate the mechanics of this in more detail.

Looking at the second, third and fourth terms on the right hand side of (1.3h) in order, we start with earned income; e.

In the first model, Model 1; it is assumed that all agents are identical, and unchanging in their abilities in time, so:

$$e_{i,t} = e = \text{constant}; \qquad (1.3i) \quad \text{for all i agents.}$$

The assumption above effectively assumes that the economy as a whole is in dynamic equilibrium (the difference between static and dynamic equilibria is discussed at length in section 6 below), there is no technological advancement, no education of employees, etc. It assumes that all individuals have exactly the same level of skills and are capable of producing the exact same level of useful output as one another; and that this is unchanging through time.

We move next to π, the income from returns. We assume that the economy consists of various companies all with identical risk ratings, all giving a uniform constant return; r on the investments owned, as paper assets, by the various individuals. Here r represents profits, dividends, rents, interest payments, etc to prevent confusion with other variables, r will normally be referred to as the profit rate.

This gives:

$$\pi_{i,t} = w_{i,t} r \qquad (1.3j) \quad \text{for each of the i agents.}$$

Given r as constant, then:

$$\sum \pi_i = r \sum w_i \qquad (1.3k) \quad \text{so:}$$

$$r = \frac{\sum \pi_i}{\sum w_i} \qquad \text{and}$$

$$r = \frac{\sum \pi_i / N}{\sum w_i / N} \qquad \text{giving:}$$

$$r = \frac{\bar{\pi}}{\bar{w}} \qquad (1.3l)$$

where $\bar{\pi}$ and $\bar{w}$ are the average values of π and W respectively.
Note that r, $\bar{w}$ and $\bar{\pi}$ are all fixed constants as a consequence of the definitions.

So for an individual:

$$\pi_{i,t} = w_{i,t} \frac{\bar{\pi}}{\bar{w}} \qquad (1.3m)$$

For the final term consumption; C is assumed to be a simple linear function of wealth. As wealth increases, consumption increases proportionally according to a fixed rate Ω (a suggested proof that this might be reasonable a assumption is given in Burgstaller [Burgstaller 1994], the constancy of Ω is discussed in depth in section 4.5).

So:

$$C_{i,t} = w_{i,t}\Omega \qquad (1.3n)$$

This final assumption gives the conceptual reason for using C rather than My for this final term.

Clearly a linear consumption function is not realistic, and a concave consumption function would reasonably be expected, with the rate of consumption declining as wealth increased. For most of the modelling, this simple consumption function is sufficient to demonstrate the required results, this is examined further in section 1.9.1 below.

In model 1A, Ω is made to be stochastic, with a base value of 30% multiplied by a sample from a normal distribution which has a variance of 30% of this base value.

By stochastic it is meant that the value can vary randomly up and down about a central average.

Consumption is chosen as the stochastic element, as being realistic in a real economy. While earnings are usually maximised and fixed as salaries, choosing to save or spend is voluntary. It should be noted that all agents remain fully identical. While the proportion consumed by each agent changes in the model in each iteration, on average each agent spends exactly 30% of its wealth. This is critically important, and I will not tire of repeating it, in model 1A all the agents are identical and have the same long-term average saving propensity, as well as earning ability.

Taken together and substituting into (1.3h) this gives the difference equation for each agent as follows:

$$w_{i,t+1} = w_{i,t} + e + w_{i,t}\frac{\overline{\pi}}{\overline{w}} - w_{i,t}\Omega \qquad \text{or simply:}$$

$$w_{i,t+1} = w_{i,t} + e + w_{i,t}r - w_{i,t}\Omega \qquad (1.3o)$$

Equation (1.3o) is the base equation for all the income models.

Although this is a little different to the standard GLV equations quoted in section 1.2 above, it shares the same basic functions.

Firstly it is worth noting how simple this equation is. Here w is the only variable. e, r and Ω are all constants of one form or another, depending on the modelling used. Note that equation (1.3o) is for a single individual in the model.

In future models e, r and Ω may be different constants for different individuals. However, in this first model, e and r are constant, and the same for all individuals.

Ω is slightly different. It is the same for all individuals, and is constant over the long term, but varies slightly bigger and smaller over the short term due to stochastic variation.

The second term on the RHS, the earned income e, provides a constant input that prevents individual values of wealth collapsing to zero. Note that this is additive, where in the models of Levy & Solomon in section 1.2 above this term was multiplicative.

The third term on the RHS is a multiplicative term and gives a positive feedback loop. The fourth term is also multiplicative and gives negative feedback.

In all the income models studied, the total income Y per time unit was fixed, and unless otherwise specified, the earned income was fixed equal to the returns income. So:

$$Y = \sum e_i + \sum \pi_i = \text{constant}, \quad \text{always} \qquad (1.3p) \quad \text{and}$$

$$\sum e_i = \sum \pi_i = \frac{Y}{2} \qquad \text{usually} \qquad (1.3q)$$

So unless otherwise specified, the total returns to labour are equal to the total returns to capital. This last relationship; that total payments in salaries and total profits are similar in size is not outlandish. Depending on the level of development of an economy, the share of labour earnings out of total income can vary typically between 0.75 and 0.5.

Although the value appears to vary cyclically, in developed economies the value tends to be very stable in the region of 0.65 to 075. This was first noted by a statistician, Arthur Bowley a century ago, and is known as Bowley's Law, and represents as close to a constant as has ever been found in economics, figure 1.3.8 below gives an example for the USA. In developing economies, with pools of reserve subsistence labour, values can vary more substantially. Young gives a good discussion of the national income shares in the US, noting that the overall share is constant even though sector shares show long-term changes [Young 2010]. Gollin gives a very thorough survey of income shares in more than forty countries [Gollin 2002].

Figure 1.3.8 here
[St Louis Fed 2004]

We will come back to Bowley's Law in some depth in sections 1.6 and 4.5-4.8 as it turns out that Bowley's law is of some importance. Because of this importance, it is useful to define some ratios. We already have:

$$\text{Profit rate} \qquad r = \frac{\sum \pi}{\sum w} \qquad (1.3r)$$

Where profit can refer to any income from paper assets such as dividends, rent, coupons on bonds, interest, etc.

To this we will add:

$$\text{Income rate} \quad \Gamma = \frac{\sum Y}{\sum w} \qquad (1.3s)$$

which is the total earnings over the total capital. Here total earnings is all the income from wages and all the income from financial assets added together.

To these we add the following:

$$\text{Bowley ratio} \quad \beta = \frac{\sum e}{\sum Y} \qquad (1.3t)$$

$$\text{Profit ratio} \quad \rho = \frac{\sum \pi}{\sum Y} \qquad (1.3u)$$

These two define the wages and profit respectively as proportions of the total income. Following from the above, the following are trivial:

$$\beta + \rho = 1 \qquad (1.3v)$$

$$\text{Profit ratio} \quad \rho = \frac{r}{\Gamma} \qquad (1.3w)$$

Finally, in most of the following models, unless otherwise stated $\beta = \rho = 0.5$

Going back to equation (1.3o), at equilibrium, total income is equal to total consumption, so:

$$\sum w_{i,t+1} = \sum w_{i,t} \qquad \text{so:}$$

$$\sum Y_{i,t+1} = \Omega \sum w_{i,t}$$

where $\sum Y_i$ is the total income from earnings and profit.

$$\bar{w} = \frac{\bar{Y}}{\Omega} \qquad (1.3x)$$

so the average wealth is defined by the average total income and the consumption rate.

There is an important subtlety in the discussion immediately above. In the original textbook economic model the total income and consumption are made equal by definition. In the models in this paper, income is fixed, but consumption varies with wealth. The negative feedback of the final consumption term ensures that total wealth varies automatically to a point where consumption adjusts so that it becomes equal to the income.

This automatically brings the model into equilibrium. If income is greater than consumption, then wealth, and so consumption, will increase until C=Y.

If income is less than consumption, the consumption will decrease wealth, and so consumption, until again, C=Y.

1.4 Wealth & Income Modelling - Results

1.4.1 Model 1A Identical Waged Income, Stochastic on Consumption

In the first model, Model 1A, the model starts with each agent having an identical wealth.

The distribution of earning power, that is the wages received e, is completely uniform. Each agent is identical and earns exactly 100 units of wealth per iteration.

The split between earnings to labour and earnings to capital are fifty-fifty, ie half to each.

The consumption of each agent is also identical, at an average of 30% of wealth. So 70% of wealth is conserved by the agent on average through the running of the model.

However the consumption of the agents is stochastic, selected from a normal range so that almost all the agents have a consumption rate between zero and 60% on each iteration.

So although the consumption of each agent is identical on average, consumption varies randomly from iteration to iteration. So an agent can consume a large amount on one iteration, followed by a small amount of consumption on the next iteration.

It is restated, in the very strongest terms, that all these agents are identical and indistinguishable.

The models were run for 10,000 iterations, the final results were checked against the half-way results, and this confirmed that the model quickly settled down to a stable distribution.

The results in figure 1.4.1.1 show the probability density function, showing the number of agents that ended up in each wealth band. This is a linear-linear plot. Also shown is the fit for the GLV function.

Figure 1.4.1.1 here

It can be seen that the data has the characteristic shape of real world wealth and income distributions, with a large body at low wealth levels, and a long declining tail of people with high levels of wealth.

As expected, the GLV distribution gives a very good fit to the modelling data.

Figure 1.4.1.2 shows the cumulative distribution for wealth for each of the agents in the model on a log-log plot. The x-axis gives the amount of wealth held by the agent, the y-axis gives the rank of the agents with number 1 being the richest and number 10,000 Being the poorest.

So the poorest agent is at the top left of the graph, while the richest is at the bottom right.

Figure 1.4.1.3 shows the top end of the cumulative distribution. It can be seen from figure 1.4.1.3 that there is a very substantial straight-line section to the graph for wealth levels above 1000 units. It can also be seen that this section gives a very good fit to a power law, approximately 15% of the total population follow the power law.

Figures 1.4.1.2 here

Figures 1.4.1.3 here

The earnings distribution for this model is uniform, so the Gini coefficient for the earnings is strictly zero.

The Gini coefficient for wealth however is 0.11. In this wealth distribution, the wealth of the top 10% is 1.9 times the wealth of the bottom 10%. The wealthiest individual has slightly more than four times the wealth of the poorest individual.

So the workings of a basic capitalist system have created an unequal wealth distribution out of an absolutely equal society.

This model, gives probably the most important result in this paper.

A group of absolutely identical agents, acting in absolutely identical manners, when operating under the standard capitalist system, of interest paid on wealth owned, end up owning dramatically different amounts of wealth.

The amount of wealth owned is a simple result of statistical mechanics; this is the power of entropy. The fundamental driver forming this distribution of wealth is not related to ability or utility in any way whatsoever.

In the first model, the random nature of changes in consumption / saving ensure that agents are very mobile within the distribution; individual agents can go from rags to riches to rags very quickly.

As a consequence, income changes are very rapid as they depend on the amount of wealth owned. So individual incomes are not stable. For this reason the distribution for income is not shown for model 1A.

1.4.2 Model 1B Distribution on Waged Income, Identical Consumption, Non-stochastic

In model 1B, the characteristics of the agents are changed slightly.

Firstly, the agents are assumed to have different skills and abilities, and so different levels of waged income (it is also assumed the are being fairly rewarded for their work).

It is still assumed that all agents has an average earning power of 100, and the total split of earnings to capital is still 50%-50%.

However, prior to starting the model, each agent is allotted an earnings ability according to a normal distribution so earning ability varies between extremes of about 25 units and 175 units.

The worker retains exactly the same working ability throughout the model.

Meanwhile the saving propensity in this model is simplified. Throughout the running of the model, each agent consumes exactly 20% of its wealth. There is no longer a stochastic element for the saving, and all agents are identical when it comes to their saving propensity.

It should be noted that, although there is a random distribution of earning abilities prior to running the model, because this distribution is fixed and constant throughout the simulation, the model itself is entirely deterministic. This is not a stochastic model.

It turns out this model is in fact very dull. With equal savings rates the output distributions for wealth and income are exactly identical in shape to the input earnings distribution. All three distributions have exactly the same Gini coefficient.

1.4.3 Model 1C Identical Waged Income, Distribution on Consumption, Non-stochastic

In model 1C, the characteristics of the agents are reversed to those in model 1B.

As with model 1A, the agents are assumed to have absolutely identical skills and abilities, and so identical levels of waged income.

It is again assumed that each agent has an earning power of exactly 100, and the total split of earnings to capital is still 50%-50%.

However, prior to starting the model, each agent is allotted a consumption propensity according to a normal distribution so average consumption rates are 20%, but vary between extreme values of 12% and 28%, while 95% of values fall between 16% and 24%. This is a much narrower range of consumption rates than model 1A with rates only varying plus or minus 20% from the normal rate for the vast majority of people. The big difference to model 1A is that each worker retains exactly the same saving propensity throughout the model, from beginning to end.

Again it should be noted that, although there is a random distribution of saving propensity prior to running the model, because this distribution is fixed and constant throughout the simulation, the model itself is entirely deterministic. This is not a stochastic model.

Figures 1.4.3.1 here

Figures 1.4.3.2 here

Figure 1.4.3.1 and 1.4.3.2 show the distributions of the wealth data. Figure 1.4.3.1 is the probability density function in linear-linear space while figure 1.4.3.2 is the cumulative density function in log-log space.

Again it can be seen that the GLV distribution fits the whole distribution, and that the tail of the distribution gives a straight line, a power law.

The fit to the GLV distribution is now less good, especially when compared with figure 1.4.1.1 for model 1A. This is because model 1C is not a 'true' GLV distribution. In the original GLV model described in sections 1.2 and 1.3, and modelled in model 1A, the consumption function was stochastic, and balanced out to a long-term average value. All the agents were truly identical. In model 1C the distribution of consumption is fixed at the outset and held through the model, the agents are no longer identical. As a result the underlying consumption distribution can influence the shape of the output GLV distribution. This is explored in more detail in section 1.4.4 and 1.9.1.

In this model, because the consumption ratios are fixed and constant throughout, the hierarchy of wealth is strictly defined. The model comes to an equilibrium very quickly, and after that wealth, and so income, remain fixed for the remainder of the duration of the modelling run.

This allows a meaningful sample of income to be taken from the last part of the modelling run.

Figures 1.4.3.3 and 1.4.3.4 below show the pdf and cdf for the income earned by the agents in model 1C.

Figures 1.4.3.3 here

Figures 1.4.3.4 here

Figure 1.4.3.4 shows a very clear power law distribution for high earning agents. However figure 1.4.3.3 shows that a fit of the GLV distribution to this model distribution for income is very poor. This income distribution does not match the real life income distributions seen in section 1.1 above. There is a very good reason for this. This is most easily explained by going on to model 1D.

Not withstanding this, it is worth looking at some of the outputs of the model, compared to the inputs. The inputs are exactly equal earning ability; so a Gini index of zero, and a consumption propensity that varied between 0.16 and 0.24 for 95% of the population – hardly a big spread.

The outputs are a Gini index of 0.06 for income and 0.12 for wealth. The top 10% of the population have double the wealth of the bottom 10%, and the richest individual has more than six times the wealth of the poorest individual.

As with model 1A, near equality of inputs results in gross wealth differences on outputs.

1.4.4 Model 1D Distribution on Consumption and Waged Income, Non-stochastic

In model 1D the distribution of wages is a normal distribution as in model 1B, however the distribution is narrower than that for model 1B. The average wage is 100 and the extremes are 62 and 137. 95% of wages are between 80 and 120. The Gini coefficient for earnings is 0.056 and the earnings of the top 10% is 1.43 times the earnings of the bottom 10%.

The distribution of consumption is exactly as model 1C.

Importantly the distributions of wages and consumption propensity are independent of each other. Some agents are high earners and big savers, some are high earners and big spenders, similarly, low earners can be savers or spenders.

As in models 1B & 1C, the earning and consumption abilities are fixed at the beginning of the model run and stay the same throughout. Again the model is deterministic, not stochastic.

Figures 1.4.4.1 here

Figures 1.4.4.2 here

Figures 1.4.4.1 and 1.4.4.2 show the distributions of the wealth data. Figure 1.4.4.1 is the probability density function in linear-linear space while figure 1.4.4.2 is the cumulative density function in log-log space.

Again it can be seen that the GLV distribution fits the whole distribution, and that the tail of the distribution gives a power law section. Again, as with model 1C, there are small variations from the GLV due to the influence of the input distributions.

In this model the hierarchy of wealth is strictly defined. The model comes to an equilibrium very quickly, and after that wealth, and so income, remain fixed for the remainder of the duration of the modelling run.

Figure 1.4.4.3 and 1.4.4.4 below show the pdf and cdf for the income earned by the agents in model 1D.

Figures 1.4.4.3 here

Figures 1.4.4.4 here

It can be that the GLV distribution gives a good fit to the curve, much better than that for model 1C. On the face of it the curve for income distribution appears to be a GLV and the power law tail is also evident. (In fact it is possible that two power tail sections are present, this will be returned to in section 1.9.1 below.)

However these assumptions are not quite correct.

The power law tail is a direct consequence of the income earned from capital. For the individuals who are in the power tail the amount of income earned from capital is much higher than that earned from their own labour, and the capital income dominates the earned income. So the power tail for income is directly proportional to the power tail for capital.

In the main body, things are slightly different. This is not in fact a GLV distribution. The income distribution is actually a superposition of two underlying distributions.

The first element of the income distribution is the investment income. This is proportional to the wealth owned. The wealth owned is a GLV distribution; as found above, so the distribution of investment income is also a GLV distribution.

The second element of income distribution is just the original distribution of earned income. This input was defined in the building of the model as a normal distribution. By definition the graph is a sum of the two components of Y that is e for wage earnings, and π for payments from investments. The full distribution of income is the sum of these two components.

This then explains why the income graph in model 1C fitted reality so badly. In model 1C the underlying earnings distribution was a flat, uniform distribution. This is highly unrealistic, so reality shows a different distribution.

In fact there are reasons to believe that the underlying distribution is a 'pseudo-Maxwell-Boltzmann' or 'additive GLV' distribution, which would show a longer, exponential, fall. This is discussed in section 1.9.2 below.

Finally this model represents a more realistic view of the real world, with variations in both earning ability and consumption propensity. It is again worth looking at the outcomes for different individuals. Earnings ability varies by only plus or minus 20% for 95% of individuals in this model. Similarly consumption propensity only varies by plus or minus 20% for 95% of people.

Despite this the top ten percent of individuals earn more than twice as much as the poorest 10% and the most wealthy individual has 11 times the wealth of the poorest. The outputs give a Gini index of 0.082 for income and 0.131 for wealth.

1.5 Wealth & Income Modelling - Discussion

To start a discussion of the results above, it is worth firstly looking back at figure 1.4.4.2 above. There is a changeover between two groups in this distribution. The bottom 9000 individuals, from 1000 to 10,000 (the top quarter of the graph) are included in the main, curved, body of the distribution. The top 1000 individuals are included in the straight-line power tail. In this, very simple model, class segregation emerges endogenously.

The distribution has a 'middle class' which includes middle income and poor people; 90% of the population. This group of individuals are largely dependent on earnings for their income. Above this there is an 'upper class' who gain the majority of their income from their ownership of financial assets.

As discussed in 1.4.1 above, the rewards for this group are disproportionate to their earnings abilities, this is most obvious in model 1A where earnings abilities are identical.

In economic terms this is a very straightforward 'wealth condensation model'. The reason for this wealth condensation is due to the unique properties of capital. In the absence of slavery, labour is owned by the labourer. Even with substantial differences in skill levels, assuming approximately fair market rewards for labour, there is a limit to how much any single person can earn. In practice only a very limited number of people with special sporting, musical, acting or other artistic talents can directly earn wages many times the average wage, and in fact, such people can be seen as 'owning' monopolistic personal capital in their unique skills.

Capital however is different.

Crucially, capital can be owned in unlimited amounts.

And with capital, the more that is owned, the more that is earned. The more that is earned, then the more that can be owned. So allowing more earning, and then more ownership.

Indeed, in the absence of the labour term providing new wealth each cycle, the ownership of all capital would inevitably go to just one individual.

(Trivially, this is demonstrated in the game of Monopoly, where there is negligible consumption and insufficient provision of new income (via passing Go, etc) to prevent one agent accumulating all the capital.)

In the various income models above, the new wealth input at the bottom (due solely to earnings not capital) prevents the condensation of all wealth to one individual, and results in a spread of wealth from top to bottom. But this still results in a distribution with a large bias giving most of the wealth to a minority of individuals.

Going back to the Lotka-Volterra and GLV models discussed in section 1.2, it is better to abandon the predator-prey model of foxes killing rabbits, and instead think in terms of a 'grazing' model where the 'predators' are sheep and the 'prey' is grass. In this model the prey is not killed outright, but is grazed on, with a small proportion of its biomass being removed.

The wealth condensation process can then be thought of in terms of a complex multi-tier grazing model, a little analogous to the tithing model in medieval Europe.

In a simple tithing system, the peasants don't own the land, but are tied to the land-owners. They are allowed to work the land and keep a proportion of the crops grown. However they are obliged to pay a portion of the tithes to the lord of the manor, and also some to the church. The tithes form the rent payable for being allowed to use the land. The lord of the manor may be obliged to pay taxes to the local noble. The noble will be obliged to pay taxes to the king. As national institutions the church and king can gain substantial wealth, even with a relatively low tax, as they can tax a lot more people.

In a modern capitalist system things are similar but the payments are now disintermediated. People supply their labour to employers, and receive payments in wages as compensation. Payments to capital are returned in the form of interest on the owners of the capital. The more capital you have, the more return you get. The more capital you have, the bigger grazer you are in a near infinite hierarchy of grazers. The higher up you get the grazers get bigger but fewer.

So, to take an example, Rupert Murdoch is a fairly high level grazer as he owns many national newspapers and television stations, so many people make use of his business, and reward him with a small percentage of profit.

At the time of writing, Bill Gates is the apex grazer, because even Rupert Murdoch's companies use lots of computers with Windows software.

The more capital you have got, the more grazing you get to do.

That capital causes wealth to condense at high levels in this way is in fact a simple statement of the obvious. To the man on the street it is clear that the more money you have, the easier it is to make more, and the question of whether money that is gained by investment is 'earned' or justified remains open to debate.

The fact that paying interest unfairly benefits the rich has of course been noted by Proudhon, Marx, Gesell and other economists and philosophers. For the same reasons usury was also condemned by the writers of Exodus, Leviticus and Deuteronomy. Other critics of usury include Allah, Thomas Aquinas, and all the popes from Alexander III (1159 to 1181) to Pope Leo XII (1823 to 1829); not to mention writers in Hinduism and Buddhism.

In these circumstances, the failure of mainstream economists to notice this basic problem with capitalism is puzzling.

As an aside, this may explain the common emergence of civilisation in river valleys that run through deserts; such as Mesopotamia and Egypt. What these areas have in common is good fertile land, but land that is limited in supply.

If there is a bad year, a farmer with excess food, due say to different balance of crops, could offer assistance to another farmer with no food, in return for a portion of land. After a while, some farmers will end up with excess land, others with insufficient land. Those with insufficient land will be obliged to labour for those with excess. This then starts off the multiplicative process of accumulation that ends up with Pharohs who own very large amounts of land, and can afford to luxuriate in the arts. For evidence of the existence of power laws in ancient Egypt see [Abul-Magd 2002].

This would not have worked in for example the Rhine or Danube valleys, because while both these rivers have fertile land, there is also plenty of surrounding, rain-fed land, which is also available. A person who became landless would simply move up the side of the valley and create some new personal capital by changing forests into fields with an axe.

The actual details of how the wealth is shared out is a consequence of entropy.

An understanding of entropy provides standard methodologies of counting possible states that a multi-body system can occupy. In the case of the GLV, this appears to be a consequence of 'path entropy' the number of different routes through a system that can be taken.

One of the profound things about entropy, and one of the reasons why it can be so useful, is that the statistical power of entropy can make microscopic interactions irrelevant. So important macroscopic properties of multi-body systems can be calculated without a knowledge of detailed microscopic interactions.

It is not proposed to discuss this in detail here; the second part of this paper discusses the concept and consequences of entropy in much more detail.

The essential point that needs to be understood at this point is that the GLV distribution is the only possible output distribution in this model because of simple statistical mechanical counting. No other output distribution is possible given the restraints on the system.

The invisible hand in this system is the hand of entropy.

As has been repeatedly noted, a GLV, complete with power tail, and gross inequality, can be produced from model 1A which uses absolutely identical agents.

In this regard, it is worth noting; and this is extremely important, some of the many things which are not needed to produce a wealth distribution model that closely models real life.

It is clear that to produce such a model, you don't need any of the following:

- Different initial endowments
- Different saving/consumption rates
- Savings rates that change with wealth
- Different earning potentials
- Economic growth
- Expectations (rational or otherwise)
- Behaviouralism
- Marginality
- Utility functions
- Production functions

In this equilibrium, utility theory is utterly irrelevant. In fact there is no need for utility in any form whatsoever; and, sadly, in an act of gross poetic injustice; you don't need Pareto efficiency to produce a Pareto distribution.

The GLV distribution is a direct consequence of the power of entropy combined with the simple concept of a rate of return on capital. It is a full equilibrium solution, a dynamic equilibrium, but an equilibrium nonetheless.

In economic systems utility is not maximised. In fact it appears that there is an alternate maximisation process controlling economics, the maximisation of entropy production, and that this is of profound importance, this is discussed in 7.3 below.

The non-maximisation of utility of course has important consequences; the distributions of wealth and income dictated by the GLV are neither efficient or rational, never mind fair.

In real life human beings are not rewarded proportionally for their abilities or efforts.

I would like to end this discussion by noting the similarities and differences between my own models and those of Ian Wright.

Superficially Wright's models are very different to my own. Wright does not include a financial sector, or interest rate payments. So clearly Wright's models can not follow my own mathematical definitions. (Wright's approach does not discuss mathematical modelling formally in general.)

In Wright's models, the workforce is split into owner manager 'capitalists' who each own an individual company, and 'workers' who are employed by the capitalists. Importantly, Wright allows movement between the capitalist and worker class, through new company formation and dissolution.

In practice this results in the same fundamentals as my own models. The capitalists pay the workers for their labour, which is identical to my own models. The capitalists are then rewarded with income according to the size of their own company. So although wealth is not disintermediated, stochastic effects allow wealth to concentrate in the hands of individual capitalists to form a power law identical to my own models. As a result the distributions of wealth and income are similar in Wright's models to my own.

While I believe that my own models are more realistic in using the disintermediation of interest/dividend payments. Wright's models are 'purer' and demonstrate the fundamental power of statistical mechanics. Wright demonstrates that you don't even need a financial sector to produce the same income distributions that are seen in the real world.

1.6 Enter Sir Bowley - Labour and Capital

All the income models above were carried out using a 50%/50% split in the earnings accrued from capital and labour. So in all the previous models the profit ratio ρ and the Bowley ratio β are both equal to 0.5. In this section the effects of changing these ratios is investigated.

It was noted in model 1B that the input wage distribution, of itself, has no effect on the output distribution. That is to say; the input wage distribution is copied through to the output distribution. It is the consumption/savings ratios that generate the power tails and make things interesting. To keep things clearer, model 1C was therefore chosen, as this has a uniform wage distribution. This is less realistic, but makes analysis of what is happening in the model easier.

Reruns of the simulations were carried out for model 1C with varying proportions of returns to capital and labour. The profit ratio ρ; the ratio of returns to capital over total returns, was varied from 0 to 1, ie from all returns to labour to all returns to capital.

From the resulting distributions it was possible to calculate the Gini coefficients and the ratio of wealth/income between the top 10% and the bottom 10%.

The poverty ratio, the proportion of people below half the average wealth/income is also shown.

The data for this model is included in figure 1.6.1. The variation of Gini coefficients and poverty ratios with profit ratio are shown in figure 1.6.2. Figure 1.6.3 shows how the ratio of the top 10% to the bottom 10% changes with profit ratio.

The results are dramatic.

Figure 1.6.1											
Profit Ratio	**0.00**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**1.00**
Bowley Ratio	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
Gini coefficient wealth	0.06	0.06	0.07	0.08	0.10	0.12	0.15	0.37	0.63	0.84	1.00
Gini coefficient total income	0.00	0.01	0.01	0.02	0.04	0.06	0.09	0.26	0.50	0.75	1.00
decile ratio wealth	1.43	1.49	1.57	1.68	1.84	2.09	2.58	7.81	22.68	67.31	Inf
decile ratio income	1.00	1.04	1.10	1.17	1.28	1.45	1.78	4.60	12.46	36.04	Inf
poverty ratio wealth	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.76	0.99	1.00
poverty ratio income	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	0.99	1.00

Figure 1.6.2 here

Figure 1.6.3 here

The model used is model 1C In which the earnings potential is a uniform distribution and so is equivalent for all individuals, that is all the agents have equal skills. However in model 1C savings rates are different for different agents. Clearly when all earnings are returned as wages ρ = 0, β = 1, and the Gini index is zero. In contrast, when all earnings are returned as capital, one individual, the one with the highest saving propensity, becomes the owner of all the wealth, and the Gini index goes to 1.

(From a profit ratio of 0.65 upwards, the Gini coefficient for wealth appears to vary linearly with the profit ratio, though the mathematics of this were not investigated.)

Figures 1.6.4 and 1.6.5 show the variation of the power exponent (which describes the power tail of the distribution) with the profit ratio.

Figure 1.6.4							
Bowley Ratio	1.00	0.90	0.80	0.70	0.60	0.50	0.40
Profit Ratio	**0.00**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**
Power Tail Slope Wealth	na	-17.42	-14.81	-12.20	-9.59	-6.97	-4.23

Figure 1.6.5 here

For very low and very high values of the profit ratio the power tail is not well defined, but for a range of values in the middle the results are mathematically very interesting.

For model 1C The relationship between alpha and the profit ratio ρ is strikingly linear. If the plot is limited to the thirteen data points between 0.05 and 0.65 the R^2 value is 0.9979. If the plot is further restricted to the eleven points between 0.1 and 0.6 the R^2 value rises to 0.9999.

It appears that in this case there is a direct mathematical relationship between the Bowley Ratio and the α that defines the power tail in the GLV equation.

This relationship was investigated further by rerunning the model and varying the various parameters in the model systematically. The value of α was calculated in the model using the top 400 data points and the formula:

$$\alpha = 1 + n\left[\sum \ln(x_i/x_{min}) \right]^{-1} \qquad (1.6a)$$

where n is 400, and the sum is from 1 to n.

The parameters available to change are as follows. Firstly the ratio of total income to total capital; that is the total income to both labour and capital (wages plus dividends) as a proportion of total capital, this was defined as the income rate, Γ, in equation (1.3s).

Secondly relative returns to labour and capital; that is either the profit ratio ρ or the Bowley ratio β. Either can be used as they sum to unity.

Thirdly the average value of the consumption rate Ω, and fourthly the variance of this consumption rate.

The first interesting thing to come out of this analysis was that the income rate, the ratio of total returns to total capital Γ had no effect on α whatsoever. Indeed the author reran the models a number of times believing an error had been made in the coding – eventually the presence of very small differences at multiple decimal places demonstrated that the models were in fact working correctly.

The second attribute to drop out of the model was that seen in figure 1.6.5 above; for fixed values of the other parameters there was a substantial central section of the profit ratio ρ for which (absolute) α declined linearly with increasing ρ.

Like the total returns, varying the absolute value of the consumption rate Ω had no effect whatsoever on the value of α.

Although the absolute value of Ω had no effect on α, changing the variance of Ω had a significant effect. In this model Ω is distributed normally, and v is used to denote the matlab variance (σ^2) parameter compared to the total value of Ω.

In this model the value of α appears to vary as a power law of v. It should be noted that the value of v could only be increased from 0 to roughly 0.25. Around this value of 0.25 the outliers in the distribution of Ω become similar to the average size of Ω. This creates negative values of Ω for some individuals which results in no consumption, and so hyper-saving for these individuals. This is both unrealistic and results in an unstable model. (a better model would treat this as a new boundary condition.)

A first attempt at fitting of the data gave very good fits across the range of ρ and v using the following equation for (absolute) α:

$$\alpha = \frac{1.5}{v^{1.30}} - \frac{1.9\rho}{v^{1.07}} \qquad (1.6b)$$

The presence of power laws for v under both terms, with similar powers, was too tempting. So a second fit was attempted using a common denominator. This gave the equation below which gave a fit to the data almost as good as equation (1.6b):

$$\alpha = \frac{(1.37 - 1.44\rho)}{v^{1.15}} \qquad (1.6c)$$

now the two constants had moved suspiciously close together, so a further fit was carried out using a common constant, again this gave a data fit almost as good as (1.6b) and (1.6c):

$$\alpha = \frac{1.36(1 - \rho)}{v^{1.15}} \qquad (1.6d)$$

Of course (1.6d) can more simply be written as:

$$\alpha = \frac{1.36\beta}{v^{1.15}} \qquad (1.6e)$$

Where β is of course the Bowley ratio.

Equations (1.6d) and (1.6e) are deceptively simple and appealing, and their meaning is discussed below in more detail.

Before this is done, it is worth stressing some caveats.

Firstly the two equations (1.6d) and (1.6e) have been extracted empirically from a model. They have not been derived mathematically. Neither have they been extracted from real data. Although it is the belief of the author that the equations are important and are sound reflections of economic reality, this remains solely a belief until either the equations are derived exactly or supporting evidence is found from actual economic data; or, ideally, both.

Secondly the nature of the two variables β and v are different. The Bowley ratio is well known in economics and is an easily observed variable in national accounts. In contrast v is the variance in an assumed underlying distribution of consumption saving propensity. In real economics the shape of such a distribution is highly controversial and is certainly not settled.

Thirdly, the two equations are limited by the parameters included in a highly simplified model. In real economies it is likely that other parameters will also effect α.

Finally, the two equations are for wealth, and do not fit the income data. A similar investigation was carried out to look at the variation of the α for the income distribution power tails. The results were much more complex, and beyond this authors mathematical abilities to reduce to a single equation. As with the wealth distributions, neither the total returns or the average value of the consumption ratio Ω had any effect on the value of α for income.

For any fixed value of v, the absolute value of α declined with increasing ρ, however the decline appeared to be exponential rather than linear. Similarly for any fixed value of ρ the value of α appeared to decline exponentially with v. Attempts to combine these facts together necessitated introductions of increasing numbers of terms and proved fruitless. Hopefully somebody with greater mathematical skills than myself should be able to illuminate this.

Despite this failure to extract a meaningful formulation, it is clear that increasing the value of the profit ratio ρ, or reducing the Bowley ratio β has a direct causal relationship on α resulting in reducing the absolute value of α for income, just as it does for the α for wealth.

This is of the utmost importance for the welfare of human beings in the real world.

It is of course trivially obvious that decreasing the Bowley ratio and increasing the profit ratio is bad for wealth and income distribution. If more income is moved to the small numbers of capital holders, at the expense of the much larger number of wage earners, then income distribution as a whole is going to get worse.

But equation (1.6d) shows that it is in fact much worse than that.

The α of the GLV defines the log law of differences in wealth for people in the power tail. As the absolute value of α decreases, inequality increases. Because α is the 'slope' of an inverse law curve (rather than say the slope of a straight line), small changes in α produce very large changes in distribution of wealth. Also by moving wealth around in the main body of the GLV, the α has a profound effect on the wealth and income of all people, not just the rich. The clear link between the Bowley ratio and the α's of the wealth and income distributions means that the changing value of the Bowley ratio has profound effects on the Gini index, relative poverty levels etc. Increasing returns to capital, at the expense of labour produces substantial feedback loops that increases poverty dramatically.

All of this of course begs the question of what exactly controls the values of the profit ratio ρ, the Bowley ratio β and the shape of the consumption rate distribution, so giving v. I intend to return to the source of the Bowley ratio in detail in sections 4.5 to 4.8 below with what appears to be a straightforward derivation.

My answer to the source of v is more tentative and more subjective, this will be introduced briefly below, but will be returned to in more depth in section 7.3 under the theoretical part below.

Before discussing the source of the consumption rate distribution, I would first like to return to equations (1.6d) and (1.6e):

$$\alpha = \frac{1.36(1 - \rho)}{v^{1.15}} \qquad (1.6d)$$

$$\alpha = \frac{1.36\beta}{v^{1.15}} \qquad (1.6e)$$

Although equation (1.6e) is simpler, equation (1.6d) is the key equation here. Indeed the more diligent readers; those who boned up on their power law background material, may have noted the strong resemblance of equation (1.6d) with the exponent produced from equation (45) in Newman [Newman 2005], which gives a general formula for α as:

$$\alpha = 1 - a/b \qquad (1.6f)$$

Where a and b are two different exponential growth rate constants.

This is of course exactly what we have in equation (1.6d) where ρ is the ratio of two different growth constants, r and Γ.

Going all the way back to equations (1.3h, 1.3p, 1.3v, 1.3s and 1.3w) ρ is the ratio of the different components of Y, which are e and π.

The total income produced by capital, the amount of value created in each cycle, is given by the sum of wages and profits:

$$\text{Total Income} \qquad \sum Y = \sum e + \sum \pi \qquad (1.3p)$$

$$\text{Income rate} \qquad \Gamma = \frac{\sum Y}{\sum w} \qquad (1.3s)$$

The direct returns to capital; that is the returns to the owners of the capital, is given by the profit rate:

$$\text{Profit rate} \qquad r = \frac{\sum \pi}{\sum w} \qquad (1.3r)$$

but ρ is defined by:

$$\text{Profit ratio} \qquad \rho = \frac{\text{direct returns to capital}}{\text{total income from capital}}$$

$$\text{Profit rate} \qquad r = \frac{\sum \pi / \sum w}{\sum Y / \sum w} \qquad \text{so:}$$

$$\text{Profit ratio} \qquad \rho = \frac{r}{\Gamma} \qquad (1.3w)$$

The value of ρ is simply the growth rate that capitalists get on capital, divided by the growth rate that everybody (capitalists and workers) gets on capital.

It is the combination of these two growth rates that creates and defines the power law tail of the wealth and income distributions. This is the first, and simplest class of ways to generate power laws discussed in Newman [Newman 2005].

And a curious thing has happened here.

There are many different ways to produce power laws, but most of them fall into three more fundamental classes; double exponential growth, branching/multiplicative models, and self-organised criticality.

The models in this paper were firmly built on the second group. The GLV of Levy and Solomon is a multiplicative model built along the tradition of random branching models that go back to Champernowne in economics and ultimately to Yule and Simon [Simkin & Roychowdhury 2006].

Despite these origins we have ended up with a model that is firmly in the first class of power law production, the double exponential model.

It is the belief of the author that this is because the first two classes are inherently analogous, and are simply different ways of looking at similar systems.

Much more tentatively, it is also the belief of the author that both the first two classes are incomplete descriptions of equilibrium states, and further input is need for most real systems to bring them to the states described by the third class; that of self organised criticality (SOC).

Going back to the wealth and income distributions, equation (1.6d) can define many different possible outcomes for α. Even with a fixed Bowley ratio of say 0.7, it is possible to have many different values for α depending, in this case, on the value of v.

It is worth noticing that there is a mismatch between the values for α given by the models and economic reality. The models give values of α of 4 and upwards for both wealth and income. In real economies the value of alpha can vary in extreme cases can between 1 and 8, but is typically close to a value of 2 see for example Ferrero [Ferrero 2010]. While the model clearly needs work to be calibrated against real data, it is the belief of the author that the relationship between α and ρ or β is valid and important.

It is the belief of the author that in a dynamic equilibrium, the value of α naturally tends to move to a minimum absolute value, in this case by maximising v to the point where the model reaches the edge of instability. At this point, with the minimum possible value of α (for any given value of ρ or β) there is the most extreme possible wealth/income distribution, which, it is the belief of the author is a maximum entropy, or more exactly a maximum entropy production, equilibrium. This belief; that self-organised criticality is an equilibrium produced by maximum entropy production, is discussed in more detail in section 7.3 below.

It is the suspicion of the author that the unrealistic distribution for Ω used in the modelling approach above results in a point of SOC, that is artificially higher than that in real economies. Indeed, it is a suspicion that movement towards SOC may of itself help to define underlying distributions of earnings and consumption. This is returned to in section 7.4.

1.7 Modifying Wealth and Income Distributions

The modelling above shows that grossly unequal distributions of wealth and income are produced as a natural output of statistical mechanics and entropy in a free market society.

In particular, the ownership of capital and the function of saving are key to the formation of inequality in wealth and income distributions.

In communist states strict, and active, microeconomic control was the normal way of attempting to prevent large discrepancies in wealth. In democratic countries this has generally been avoided, partly because of the stunting effects on economic growth, but also because of the restrictions on liberty. Instead these countries have instituted substantial systems of taxation and welfare in an attempt to transfer income from the rich to the poor. Meanwhile trade unions and professional societies also attempt to modify wealth distributions for their own members.

From an econodynamics point of view the above methods of attempting to influence income distribution are deeply flawed. In a system of a large number of freely interacting agents the GLV distribution is inevitable and methods of exchange, even ones such as tax and welfare, are largely irrelevant.

One approach that does make some sense is that of the trade unionists and professional societies. By tying together the interests of thousands, or even millions, of individuals their members are no longer "freely interacting" and are able to release themselves from the power of entropy to a limited extent. (Monopolistic companies attempt to subvert entropy by similar means).

Traditional methods of taxation and welfare have much less justification. This solution attacks the income flows directly, and does not address the issues of capital. Also by attempting to directly micromanage the income distribution, taxation and welfare attempts to impose a non-equilibrium statistical outcome at a microscopic level. This approach is doomed to failure.

It is common experience that such transfers give little long-term benefit to the poor. Transfers need to be massive and continuous to be effective, and there is a wealth of data to suggest that many welfare programmes result in the giving of benefit to those of medium and high incomes, rather than to the poor, see section 1.8 below for a discussion of this. This is of course exactly what an econodynamic analysis would predict.

Given the power of entropy to force the overall distribution regardless of different sorts of microeconomic interactions, it would initially seem that attempts to modify income distribution will be futile. This is not necessarily the case.

As discussed above trying to fight entropy head on is a pointless task.

However in the following two sections alternative approaches look at how wealth and income distributions might be modified, given the knowledge that these distributions are formed in a statistical mechanical manner. The first approach looks at imposing boundary conditions on a model of society, the second looks at modifying the saving mechanism feedback loop.

1.7.1 Maximum Wealth

The author has previously proposed that the imposition of a maximum wealth level should, by symmetry, produce a symmetrical distribution of wealth and income [Willis 2005].

This proposed solution was based on the (mistaken) assumption that wealth and income distributions were formed in a static exchange equilibrium.

Model 1D was rerun to test this theory.

Two different versions were rerun, a lazy version and a greedy version. Both versions included an additional rule that came into play when any agent reached a wealth level of more than 50% greater than the average wealth level.

In the first rerun; the lazy version, any agent that reached the maximum wealth level duly had their incentives reduced, and reduced their work rate by 5% (5% of its current value). If the agent repeatedly hit the maximum wealth limit, then they repeatedly had their work rate reduced.

In the second rerun; the greedy version, any agent that reached the maximum wealth increased their consumption by 5% of current.

Figure 1.7.1.1 shows the cdf outcome for the increasing consumption model, the graph for the decreasing work model is almost identical.

Figure 1.7.1.1 here

Contrary to the expectations of the author, the maximum wealth model fails dismally in achieving it's hoped for aims. The resulting distribution merely flattens off the unconstrained distribution.

This has the effect of bunching a large minority of agents at near equal wealth levels close to the maximum permitted wealth. It is worth noting that, in the real world, this particular group of agents would include most of the ambitious, clever, innovative, entrepreneurial, well educated and politically well connected.

This model also has the notable non-effect of not assisting the impoverished at the bottom of the distribution in any noticeable way. This model makes the rich poorer, but doesn't make the poor richer.

Taken together, this social model would seem to present a highly effective way of precipitating a coup-d'état.

While the author remains romantically attached to the concept of maximum wealth limits, and believes that they may form the basis for interesting future research, this approach is not currently proposed as a basis for tackling inequality in a real economy.

1.7.2 Compulsory Saving

The second approach for changing income distributions focuses on the crucial role of saving in the GLV equation. From models 1B and 1C it appears that rates of consumption and saving are critical to the formation of the power tail and so large wealth inequalities. If saving is the problem, it seems sensible to use saving as the solution.

Again model 1D was used as the base model.

In this model a simple rule was introduced. If any agent's current wealth was less than 90% of the average wealth, that agent was obliged to decrease their consumption rate by 20 percent. This could be thought of an extra tax on these individuals, which is automatically paid into their own personal savings plan. It should be noted that this increase, though significant, is not enormous, and is comparable say to the rate of VAT/income tax in many European countries.

Figure 1.7.2.1 here

Figure 1.7.2.2 here

Figures 1.7.2.1 and 1.7.2.2 show the log-log and log-linear cumulative distributions for the model, with and without the compulsory saving rule.

It can be clearly seen in figures 1.7.2.1 and 1.7.2.2 that the number of poor people is much smaller with compulsory saving. For the bottom half of the agents (the top half of figure 1.7.2.2), the distribution is very equal, though it retains a continual small gradient of wealth difference.

The top half of society retains a very pronounced power-law distribution, with approximately the same slope. Each individual in the top half is less wealth by an amount that varies from roughly 5% for those in the middle to roughly 10% for those at the top. Despite this they remain far richer than the average. This drop in wealth seems a very slight price to pay for the elimination of poverty, and the likely associated dramatic reduction in crime and other social problems. The power tail structure would leave in place the opportunity for the gifted and entrepreneurial to significantly better themselves. Retaining the group of high earners in the power tail would also have the useful secondary effect of providing an appropriate source of celebrity gossip and target for quiet derision for the remaining, now comfortable bottom half.

Figure 1.7.2.3 shows various measures of equality with and without the saving rule.

Figure 1.7.2.3	No Compulsory Saving	Compulsory Saving
Gini Earnings	0.056	0.056
Gini Wealth	0.131	0.077
Gini Income	0.082	0.058
Earnings Deciles Ratio	1.429	1.429
Wealth deciles ratio	2.268	1.617
Income deciles ratio	1.686	1.451

The results are dramatic and also very positive.

Without compulsory saving the input earnings distribution was magnified through saving in the GLV into a more unequal distribution for wealth and income. This can be seen in both the Gini indices and also the ratio of the wealth or income of the top 10% to the bottom 10%.

With compulsory saving the output distribution for income has almost the same inequality values as the original earnings distribution for both the Gini index and deciles ratio. Wealth is more unequal, but much less so than in the model without compulsory saving.

In fact the shapes of this output income distribution (in figures 1.7.2.1 & 2 above) is significantly different in shape to the input earnings distribution, which in this case is a normal distribution. But by smoothing out the rough edges of the GLV, compulsory saving provides an output that is similar in fairness to the skill levels of the inputs. This is probably a distribution that society could live with.

In practice poverty has been eliminated for all except those that combine a very poor earnings ability with a very poor savings rate – individuals who in real life would be necessarily be candidates for intervention by the social services.

Rather than being purely equitable distributions, the output distributions could be better described as pre-Magrathean: *"Many men of course became extremely rich, but this was perfectly natural and nothing to be ashamed of because no one was really poor……"*

It is also worth noting the form in which this transfer of wealth takes place.

In this model the rich are not taxed.

In this model the poor are compelled to save.

The rich would only notice this form of financial redistribution in the form of increased competition for the purchase of financial assets.

In practice a compulsory saving scheme would be highly effective once the new, more equal, distribution was in place. However expecting people who are currently very poor to save their way out of poverty is not reasonably realistic.

Section 1.8 below discusses extensions of these ideas in more detail.

1.8 A Virtual 40 Acres

In this section more detailed proposals are made for modifying wealth and income distributions; based on the outcomes of the models above. It is hoped that these proposals will provide solutions that are more practical, effective and far less costly than current mechanisms such as welfare and subsidised housing.

Before continuing with these discussions, I believe it is worth stating some of my own personal political beliefs. This paper uses theoretical ideas from Marx, though the classical economics is equally attributable to Adam Smith. In addition the discussion below is substantially about the reallocation of capital. However I emphasise that I disagree in the very strongest terms with Marx's proposed methods for redistributing capital. I strongly believe that the creation of wealth by market capitalism, within a democratic state, must remain at the core of any effective economic system.

I believe that redistribution of capital can be achieved in an effective manner within a democratic, capitalist state, in ways that are much cheaper and more effective than methods currently used in democracies. My aim is not to take from the rich and give to the poor. My aim is to achieve a property owning democracy, where all members own sufficient property to guarantee a basic standard of living (and where the word property does not refer just to housing).

In sentiment, though not in policy particulars, I am much closer emotionally to the followers of binary economics and their Capitalist Manifesto, than I am to the ideas in the Communist Manifesto.

In the previous section, I proposed that redistribution is carried out by forcing the poor to save, rather than taxing the rich. It is hoped that this makes clear that, while I am very sympathetic to some Marxian insights into economic theory, I am wholly opposed to traditional Marxist proposals to deal with inequality.

In many ways I believe the ideas represented in this section are improvements on ideas first proposed by Milton Friedman. Although staunchly right wing, unlike most laissez-faire free-market economists, Milton Friedman recognised that capitalist economies did not ensure a distribution of income that allowed all citizens to meet their basic needs. In his book 'Capitalism and Freedom' [Friedman 1962], he proposed the introduction of a 'negative income tax', a policy that now exists in the form of 'earned income tax credit' in the USA, and which has been copied successfully in other countries. As a form of income redistribution, Friedman's ideas suffer from needing continuous flows. I believe my own proposals achieve the same aims of those of Milton Friedman, at much less cost.

I would ask that readers consider these proposals to be more neo–Friedmanite than neo-Marxian.

If, however, my ideas are incorrect, then I would rather live with freedom and inequality than equality and injustice. Civil rights are more important than economic rights.

To briefly review the conclusions on income models discussed above in sections 1.3 to 1.5, it is possible to conclude the following:

Income and wealth distributions are defined by entropy.

Income and wealth distributions are not defined by utility, marginality, ability in general or entrepreneurial ability in particular.

Income and wealth is gained in a reinforcing circular flow, the more money you have, the more money you will receive.

Income and wealth distributions are strongly skewed, giving disproportionate wealth to a small minority.

Income and wealth distributions are strongly biased in favour of those who inherit wealth.

Despite the above conclusions there is still a question, that needs to be answered, as to why it is felt necessary to change income distributions at all. Some of the arguments are discussed briefly below.

The first thing to note is that recognising that wealth and income distributions are caused by entropy, rather than say utility or ability, changes the whole nature of the political debate on redistribution.

At present, it is normally assumed within economics that income and wealth distributions are 'natural' and caused by maximisation of utility and/or rewards for entrepreneurial or other ability. It is further assumed that moving away from this 'natural' equilibrium will have bad effects; interfering with the market, reducing overall utility, removing incentives for wealth creation, etc.

Under these assumptions, economists and many politicians take the view that any case for changing existing income distributions must be very strong, and movement from the 'natural' position must be justified.

Once it is realised that income and wealth distributions are caused by entropy, then things become very different. The entropy equilibrium position may be 'natural' in the scientific sense, but it does not maximise utility. It specifically punishes hardworking people, the majority of individuals, who are effectively debarred from the ownership of capital. This is despite the fact that the labour of these people form the main supply of new wealth that allows capital formation.

In this sense the current system of ownership of capital works as a private taxation system acting on the majority of individuals, transferring the majority of the wealth to a small minority of individuals. This 'taxation' is far more iniquitous then any standard taxation system used in a normal democracy.

Under these circumstances, failing to modify income distributions becomes a highly political decision. It becomes a decision to support and entrench a system that takes from the poor and middle classes to reward the rich.

If this is what the public in a democracy choose to do, then that is fine; but the political debate needs to be made absolutely clear.

Two recent papers suggest that the understanding of the deep seated nature of this injustice is very deep. In their paper, Griffiths and Tenenbaum [Griffiths & Tenenbaum 2006] demonstrate that ordinary people, lacking in a mathematical education, are capable of accurately judging whether data fit different mathematical distributions such as the normal or power law. Given that most skills are based on a normal or log-normal distribution, and that wealth is distributed as a power law, this would suggest that people intuitively, and reasonably, realise that distributions of wealth are unfair. In another paper Norton and Ariely [Norton & Ariely 2010], show that Americans, even rich Americans, believe that the United States would ideally have a distribution of wealth more like that of Sweden.

Given the political nature of the decision discussed above, the first obvious reason for modifying income distributions is simply common decency.
Or, alternatively, basic obedience to spiritual teachings. All major religions recognised the inequities of usury; the bible clearly prohibits usury in Deuteronomy 23:20.

For many, particularly the wealthy and those that remain wedded to neo-classical ideas, an appeal to common decency or divine guidance may not be sufficient. So it is worth considering two other, more selfish, reasons for modifying income distributions.

The first issue to consider is that strongly skewed income distributions negatively affect rich people as well as poor people, though clearly they affect poor people more than the rich.

There are two main reasons that the rich are disadvantaged by skewed wealth distributions, the obvious one is crime, the other, less obviously, is in overall health levels.

I will review these very briefly below, for more information; the arguments are discussed at length, in great detail, with much supporting evidence, in the book 'The Spirit Level' by Richard Wilkinson and Kate Pickett [Wilkinson & Pickett 2009].

The issues of crime are easily understood. More unequal societies have much more crime, and higher general levels of aggression and violence. In unequal societies rich people have material wealth, but may have their quality of life significantly reduced through fear of crime. This includes the fear of being attacked in the street or having their homes broken into, and may result in not being able to move about freely or being obliged to live in isolated, highly secure accommodation.

The data on health is much more counter-intuitive. It is of course obviously plausible that average life expectancy and health outcomes correlate very closely with fairer wealth distributions, and the statistical data supports this.

Critically, and quite surprisingly, these statistical benefits are not just due to outcomes in the poorer parts of the populations. Rich people live longer and are healthier in countries like Sweden or Japan that have more equal wealth distributions. In fact often poor people in more equal countries have better health outcomes than rich people in countries with unequal wealth distributions, see for example figures 1.8.1, 1.8.2 and 1.8.3 below from Wilkinson & Pickett. The reasons are not fully clear, but appear to be due to increased levels of stress throughout the whole of society.

Figure 1.8.1 here

Figure 1.8.2 here

Figure 1.8.3 here

The second 'selfish' reason for using statistical theory for changing income distribution is that in practice all democracies attempt to carry out income redistribution. Such efforts, by fighting entropy head on, are normally expensive and of limited effectiveness. Ultimately such efforts must be paid for out of taxation, whether they are effective or not.

In Europe of course, the welfare state and high taxation are used in attempts to redistribute income. The workings are obvious, as is the expense. Such systems are generally looked down on by individuals from 'free-market' countries such as Hong Kong, Singapore and the US.

In fact, even in the most avowedly free-market democracies, leaving things completely to the market has never been acceptable. All democracies put in some sort of support for the poor. Hong Kong famously has very poor benefits for unemployment, but few people realise that about half of the population of Hong Kong live in subsidised public housing. Those that are purchasing property are allowed to offset up to 100% of home loan interest payments against tax up to a maximum of $100,000 per year. The proportion of the population living in subsidised housing in Singapore is even higher than that in Hong Kong [Telegraph 2010b]

The US of course publicly repudiates the horrors of providing public housing. Instead for many years they have given covert subsidies to housing of the poor and middle classes indirectly. Americans, though presumably not particularly poor ones, can receive mortgage tax relief on up to $1,000,000 worth of debt on their homes. Also, very large housing subsidies have been provided through the underwriting policies of the GSE's primarily Freddie Mac and Fannie May. The effects of these gross distortions to the market have been disastrous, not just to the US but to the whole world, as the credit-crunch was triggered by the sub-prime mess created by these back door subsidies. Remarkably, the US appears not to have understood the lessons of this recent disaster. I don't know of any country in 'socialist' Europe that uses government backed mortgage insurance, but in the US the future of the GSE's is still under discussion.

The big problem with all the current forms of welfare, whether overt or covert, is that they don't work. The welfare systems currently used by states around the world fall into one of two classes, either they provide income, in the form of benefits, or they provide subsidies to housing.

What poor people actually need is capital. If they had capital, they would have their own income, and if they had sufficient income they would be able to provide their own housing.

Simply providing income directly doesn't work. This is because the income will be spent immediately and so the income stream needs to be continuous, and even then will not lift people out of poverty.

It is also iniquitous. As the British MP Frank Field has pointed out; effectively, in the UK, welfare claimants are stuck in a poverty trap because the income streams they receive mean they 'own' the equivalent of very substantial capital which amount to 'lifetime pensions'.

Subsidising housing is better, but is not ideal. Housing is not real capital (see discussions below in section 6.3) and does not give good long-term gains, and again providing housing at less than its cost means that subsidies are continuous. Housing provided by the state also badly affects freedom of choice, allows social stratification and creates ghettos for the poor with associated problems of crime and restricted economic opportunities.

The aim of the proposals in this section is to make the process of aiding the poor much easier, by understanding and so using the statistical mechanics of the economic system. The main aim is to transfer capital to poorer people and ensure that they retain that capital. This would make transfers one-offs rather than continuous. In the longer term this in itself would reduce taxes significantly. If secondary effects include less crime and better health, then total tax takes should reduce even further.

From the analysis and modelling in sections 1.4 to to 1.6 above it is clear that there is a fundamental near-fixed nature to the ratio of returns to labour and capital (this is discussed in much greater depth in section 4.5 below). This fixed ratio of returns to labour and capital then gives fixed parameters for the GLV distribution, which in turn gives a fixed proportion of people in poverty, as discussed in sections 1.5 and 1.6.

The fixed nature of the ratio of labour/capital returns, and the fixed shape of the GLV distribution necessarily mean that the only way that the elimination of poverty can be achieved is by moving capital into the hands of poorer people.

Without changes of ownership of capital, poverty will remain fixed. Other methods of attempting to alleviate poverty will necessarily fail. If these methods involve taxation, then they will fail expensively.

As discussed above, I believe the key to eliminating poverty is increasing the amount of capital owned by poorer individuals.

One solution to this problem would be to encourage employee ownership much more strongly. For example it would be possible to increase the use of employee share ownership plans (Esops) by giving greater tax advantages to them.

A better alternative is to encourage full-scale ownership of companies. In the UK employee-owned organisations currently include companies such as John Lewis, a major retailer and Arup and Mott-Macdonald, both of which are major engineering consultancies. Such companies have been very successful in the service sector where capital costs are relatively low and quality of service is key to success. In these companies, profits are normally distributed to employees as bonuses, which are typically paid out in proportion to annual salaries. In 2010 John Lewis staff received bonuses equal to 15% of basic salary, in 2009 they received 13%, in 2008, pre-recession, it was 20%. Although this still results in an unequal distribution of capital, it is a much more equal distribution than that found through the normal pattern of distribution via shares owned by private individuals, which of course is a GLV distribution.

Stronger encouragement of employee owned organisations, by the use of tax advantages might in itself be very successful in producing a more equal distribution of wealth.

In practice though, it is difficult to see how such organisations could easily raise the capital needed for extractive industries, heavy manufacturing industry, or for that matter companies involved in scientific research or large-scale finance. (Clearly, if such companies use external debt financing for capital investment this just recreates the problem of paying out profits to external capital owners, so recreating the GLV).

There can also be very severe problems when people's personal capital is tied up in their employer. In the case of bankruptcy, individuals lose twice over, losing their investments as well as their jobs.

Additionally, employee owned organisations do not solve the problem of balancing saving of individuals over the lifecycle. If all companies were employee owned, middle-aged people would not have suitable places to invest their savings for their pensions. (And Robert Maxwell showed that investing your pension in your employer is a profoundly unwise thing to do.)

Realistically, for much of the economy there will need to remain a separation of ownership of capital from employment.

In practice, I believe the target must be to create a 'virtual 40 acres' of capital for all members of society.

The phrase '40 acres and a mule' is 150 years old. In 1865 at the end of the American Civil War, it was the policy of the Northern army to provide freed slaves with 40 acres of fertile land and an ex-army mule to provide a draft animal. At the time it was recognised that this combination was enough to provide a family with a self-sufficient homestead. In practice the policy was not carried out except in parts, and was mostly rescinded even then.

As shown in the model in 1.7.2 above, one way of ensuring that people have extra capital is simply to introduce compulsory saving. The main reason for using compulsory saving in this model is simply because it is very easy to model mathematically.

In real life such a model would have a big problem starting up. Once it was up and running, and income was already well distributed, then it would be easy to enforce compulsory saving. However trying to enforce compulsory saving, which will feel like an extra tax, on people who are currently poor would be very difficult. It would also have the perverse short-term effect of making people significantly poorer in terms of day-to-day income.

A more realistic model for starting the system up would be to introduce assisted saving, where governments allowed tax rebates and/or paid subsidies to people who were saving money.

To make such a scheme work effectively, the easy bit is giving assistance to poorer people. The difficult bit is ensuring that the money is not spent as income; to ensure that it is in fact saved.

Fortunately there are well-established precedents for schemes of this type, most notably pension systems. In most democracies, people who save for pensions are given tax relief and even assistance with their savings. As a quid pro quo for this assistance, governments lay down strict rules as to when and how the money can be withdrawn in old age.

From country to country many other forms of government assistance are given, such as tax relief on mortgage payments, tax-free savings accounts or tax-free share ownership (ISA's in the UK), and even assisted savings such as Government Gateway in the UK.

Unfortunately such schemes tend to have grown up historically on an ad hoc basis, without any theoretical underpinnings. As such the results have been, at best, haphazard.

Taking the UK as an example, a review of who benefits from such schemes, is quite enlightening.

Firstly in the UK, individuals are allowed to invest in tax free savings accounts or 'ISA's'. Any individual is allowed to pay in £5,100 per year if the investment is in cash, or £10,200 per year if the investment is in shares. Money can be left in as long as is wanted. If money is removed, it can't be put back in; the ISA allowance is lost. Any dividends or capital growth achieved are completely tax-free. It is rumoured that some successful stock-pickers have managed to accumulate millions of pounds in their ISA's, and are allowed to receive income from these investments tax-free. It is not clear exactly how this contributes to social equity and cohesion. Clearly the ISA system is much more advantageous to the rich who can both save regularly, and are less likely to need to raid their ISA's in the short term. Also tax-free savings are of no benefit to people who are so poor that they pay little taxes.

Policy on pensions provision in the UK is even more interesting, though profoundly confusing. (UK pension and tax policy is very complex, if I have made errors in the brief summary below, I would welcome correction.)

Individuals in the UK can pay income into a personal pension fund free of tax. If you are a basic rate taxpayer (a poor person), the maximum you can save is 20%. If you are a higher rate taxpayer (a rich person), then the amount of tax relief you can earn can increase up to a maximum of 40% total.

Contributions to your private pension scheme are capped each year to your maximum income. So if you are a poor person, you are only allowed to put a small amount in, and receive a small amount of tax relief. If you are a rich person, you are allowed to put a lot in to your pension, and earn a lot of tax relief. This is an important restriction, as it prevents people with variable income from paying money saved from a good year in during a bad earnings year.

Sensibly, there is a maximum limit to how much you can save in your pension tax-free each year. The current maximum is £50,000 per year. (This was recently reduced from £255,000 - I am not making this up.) So the maximum subsidy, per rich person, is nearly £20k per year.

The average salary in the UK is approximately £25,000 per year.

In addition to the above, there is also a 'lifetime allowance' on the total notional size of the pension fund, and pension receipts from the part of the fund above this allowance are subject to income tax. The lifetime allowance is currently £1.5 million. Even on an interest only basis, assuming no draw down of the fund, at 3% real interest rates this would allow a tax-free pension of roughly double the average UK salary.

The 'aim' of all these subsidies to the rich is to avoid people being dependent on state pensions in their old age. The current maximum UK basic state pension is £97.65 a week, so if a person retired at an age of 65 and lived for thirty years, the cost to the state would be roughly £150k. Even including for housing benefit in rented accommodation the cost would be less than £300k. It is not clear to me that the 'aim' of saving money for the state is being successfully achieved.

All the above system was put into place and managed under the Labour government of 1997 to 2010, notionally a social democratic, if not socialist party.

Perhaps due to a concern with the above largesse lavished on the rich, the same government also introduced an assisted saving scheme called the Savings Gateway.

To qualify for the Savings Gateway you must earn less than £16,040 per year, and must also be claiming some sort of benefit.

The maximum payment into the scheme is £25 per month. For every £1 that a participant saves, the government will add a further 50p. So the maximum subsidy, per poor person, per year is £150. Whether the Savings Gateway proves to be successful in helping to reduce poverty remains to be seen. I, for one, am not holding my breath.

This disparity in assistance for the rich and the poor is not restricted to the UK, this from the Economist in 2005:

Politicians' main method for boosting thrift is a swathe of tax-advantaged retirement accounts. This year these accounts will cost some $150 billion in foregone tax revenue. Most of this subsidy goes to richer Americans, who have higher marginal tax rates and who are more likely to save anyway. Only one saving incentive—the Saver's Credit—is targeted at poorer Americans. It is worth only about $1 billion in forgone tax revenue and is due to expire in 2006. And even that offers no incentive to the 50m households who pay no income tax. [Economist 2005].

The report 'Upside Down' gives a detailed analysis of how the majority of assistance given to working families in the USA ends up in the hands of the rich [Woo et al 2010].

While the efficacy of the many different policies used above can rightly be questioned, the important point is that the financial tools and institutions needed for creating private capital for all members of society are already available.

Interestingly perhaps the best example of such a system is one initiated by a group of radically right-wing free market economists.

The Chilean pension system that the 'Chicago boys' created for dictator Augusto Pinochet in Chile works in exactly this manner.

In Chile, all salaried workers are forced to pay 10% of their salary into one of a number of strongly regulated pension funds. The pension funds in turn invest in private companies through the stock market, bond purchases, etc. The pension funds are strictly regulated, and individuals are allowed to switch easily between different suppliers.

The major difference between the Chilean pension scheme, and my proposed 'virtual 40 acres' (henceforth 'v40') is that part or all of the interest from the capital, and some of the capital, would be made available during the normal working life of an adult.

A rough outline of the 'v40' is as follows.

The v40 would consist of a pot of money, held with an officially sanctioned investment fund exactly like those that operate in Chile. The funds would have controls on appropriate investments and proportions of investment in different assets, as is normal with regulated pension funds.

At any one time there would be a maximum amount that could be held in the v40, for the present discussions the maximum amount will be assumed to be £50k. This is approximately twice the average annual wage, and as an investment sum it is not particularly large. There is an important reason for this small proposed size, this is discussed later.

All people who are in paid employment would be obliged to pay into their v40 at a minimum rate of say 10% of salary. This would apply to all people who had not got a full pot of £50k invested in their v40.

Note that people who had the full £50k invested would not be obliged to pay into their v40 pot; in fact people with a full v40 pot would be specifically prohibited from paying further into their v40.

To make this compulsory saving more palatable, all payments into the v40 would be before tax and any other payments such as social security. Similarly all interest payments, and eventually capital repayments out of the v40 would also be free of income and capital gains or any other taxes, provided they had been invested for a minimum period of say five years.

There would be no limit to the amount paid into the pot each year, up to the total limit of £50k, and all payments up to this amount would be tax-free. (In the UK for example, all current ISA holdings, up to £50k, could be transferred over into the v40 tax-free. ISA's would then be discontinued as a tax-free vehicle.)

For poorer people, two further regimes are proposed. Here poorer can mean one of two things. Firstly it can mean people who have low levels of savings in their v40, and so low income from the v40. Secondly it can mean people who have poor employment income, either through low skills level or because of intermittent employment. In practice either or both of these definitions may apply to the 'poor' and 'very poor' discussed below.

For 'poor' people further assistance can be given by allowing payments to the v40 account to be counted as an alternative to taxation. So if a poor person is paying 10% of their salary into a v40, then they would have their 'normal' taxation reduced by the same amount of money.

For 'very poor' people the government would follow the ideas of the 'Savings Gateway' and other similar schemes, and pay matching amounts to give assisted saving, so helping the very poor move into the category of simply poor.

With regard to withdrawals, a portion of interest payments could be withdrawn immediately, but on a sliding scale with strict rules. So the percentage of interest earned that could be withdrawn each year would vary as the percentage of the total v40 allowance held.

To take some examples. Assume that the real interest is at 3% per annum (halfway between long-term US and UK rates, see section 4.5 below). Assume also that the v40 limit is £50k.

If somebody had a full pot of £50k invested in their v40, then they would earn £1,500 interest per year, and would be allowed to take the full amount out each year as tax-free income. In fact they would be obliged to remove this interest, and any capital accumulation above the £50k, from the account.

If somebody had saved half of their v40 allowance or £25k, then they would earn £750 in a year, and would be allowed to remove half of this interest, or £375. The remaining £375 would be automatically reinvested as capital in the v40. Clearly there would be no compulsion to remove any of the interest.

If somebody had only £10k in their v40, or 20% of the allowance, then they would earn £300 interest. They would only be allowed to remove 20% of this interest, or £60, with the remaining £240 of interest being reinvested as capital.

Finally, to further discourage early removal of interest, a minimum five-year period should be included with punishment of taxation if the interest is removed within five years of it being earned, for 'normal' investors. Or, in the case of 'poor' investors, a reward if the accrued interest is held in the account for a minimum of five years, similar to the 'Savings Gateway' scheme. Note that this punitive taxation would not apply to those who have reached the maximum of the v40 pot.

While the above may seem somewhat complex, the aim of all the detail is the same. All the incentives, for rich or poor, are to encourage people to save as much money in their v40 as they can, as quickly as possible.

It is hoped that in this manner the v40 will be seen as a sensible way to build up capital by all members of society, even the poorest.

While the v40 is being built up, a portion of the accrued interest will be available for removal, as emergency funding, in the case of a financial crisis. But the incentives should encourage such use only in genuine emergency.

Once the v40 allowance has been fully reached, then the fund becomes a useful additional income support. At this point, removal of interest and capital gains would become compulsory, and would need to be spent as consumption or moved into private investments that do not attract tax exemption.

With regard to removal of capital, it is suggested that rules along the lines of the following are used.

Firstly no capital can be withdrawn until a minimum age of say forty years. After that age, capital can be withdrawn according to a set rate depending on the notional length of time that the v40 account will be held.

A notional date for the end of the account is assumed, this effectively being a notional date of decease of the account holder. This could be say the age of 80 years old, or ten years older than the current age, whichever is the larger.

The amount of capital that could then be withdrawn would be the reciprocal of the number of years between the current age and the notional end date.

So if the owner of the v40 was forty, and the notional end date was 80, the difference would be 40 years, and the holder would be allowed to remove 1/40th of the value of the v40's capital, in addition to the allowed interest.

At sixty years old the holder would be allowed to remove 1/20th of the value of the v40. From age seventy onwards the holder would be allowed to remove 1/10th of the value of the v40. This would be the maximum amount of capital that could be removed from the account at any time. Removal of capital would not be compulsory.

Following the decease of the v40 account holder, all the value of capital would be inheritable. This would be fully tax free, including free of death duties, providing that the v40 money was passed to other individuals, with sufficient spare allowance, for transfer into their own v40's. If the capital was brought out of the protection of the v40 system, it would be taxed, and subject to death duties, as normal capital.

Finally there is one subtlety that needs to be controlled if the v40 scheme is to be effective. It is not sufficient simply to prevent people running down the capital in the scheme and using it as income. It is also essential that people be fully prevented from using the capital in the v40 as collateral against which they can borrow money. This would destroy the v40 scheme by allowing savings to be converted into income.

The best way to do this is to allow relatively lax personal bankruptcy laws and to specifically exempt money invested in a v40 from being included in bankruptcy cases. That is, even a person who has been made bankrupt is allowed to keep the full value of their v40 intact. If this is put into place, then it will not be possible to secure loans made to an individual against their v40, as such loans will be extinguished in the bankruptcy. In such circumstances individuals should not be able to get loans against v40's. Protection in this manner will also have the advantage of encouraging use of the v40 as a savings vehicle.

The net result of this is to have something that works in very similar manner to a pension scheme, but also has characteristics similar to that of an employment insurance scheme. It is aimed to meet basic and/or emergency needs throughout a working life.

As such it can be seen as a 'personalised' welfare scheme, and at least in part, can form an effective 'personalisation' of welfare. By handing the main responsibility for management of this 'welfare' to individuals it should be much more effective than state run welfare schemes that lose the link between contributions and benefits.

Despite this 'personalisation' it has to be stated in the strongest possible terms that the iron law of the GLV means that some form of government action will always be necessary if such a personalised form of welfare is to succeed. As an absolute minimum, a government would need to strictly enforce compulsory saving to ensure that such schemes operate. It seems more realistic that general tax advantages, assistance for the poorest and a backstop of enforcement will be the most effective policy mix to ensure the v40 operates effectively.

To give an example of how this could work, I would like to take Norway as an example, though as will be seen later this is not quite a reasonable choice.

Norway is of course very rich. Not only does it have a very well run Scandinavian social and political system, it has also enjoyed four decades of oil production.

Despite this, Norwegians still have problems of relative poverty, where depending on definitions, between 4% and 10% of the population have less than 60% of median earnings [EWCO 2010]. Given the very high costs of living in Norway this relative poverty can be debilitating. Poverty in Norway was seen as a priority for the incoming government in 2005.

As a result of careful saving, by successive governments, Norway now has a sovereign wealth fund of more than three trillion Norwegian crowns, equivalent to about 500 billion US dollars. The population of Norway is 4.7 million, which must mean there are roughly 3.5 million adults. Using these figures the sovereign wealth fund is worth about $130,000 per person.

So trivially, the Norwegian government could simply create 3.5 million v40 accounts tomorrow and give each Norwegian adult $140,000 worth of assets to hold in the account.

This isn't actually very sensible, as many Norwegians are already quite wealthy and don't need to be given all that money.

Let's assume that say 20% of Norwegians are quite rich and have many assets to hand which they will be happy to transfer into a tax free v40 account given the opportunity. Let's assume 20% of Norwegians are comparatively poor and need to be given their full v40 allowance by the state.

Finally we will assume that the remaining 60% of Norwegians are middle income and that they will only need an incentive to transfer their savings and/or income to their v40's. Suppose this is a tax-free incentive of equivalent to 30% of the v40 investment.

This means the Norwegian government can make its sovereign wealth fund go much farther, actually more than two and a half times farther. So now the v40 allowance can be set at about $375,000 per head.

If we again assume that long-term real interest rates are 3%, then this gives each and every Norwegian adult an independent income of $11,000 per year.

Just for comparison, a quick look on the internet suggests that rents in Oslo for a 3-bed apartment are currently about $1000 per month, so such an income would pay most housing costs. But then if you lived in a beautiful country like Norway, and you had an independent income, why on earth would you live in Oslo. From my own limited knowledge of Scandinavian culture, a surprising proportion of Scandinavians have second homes hidden away as rural retreats.

With private income like this, if Norwegians moved to the countryside; apart from childcare, hospital care and care for the elderly; the whole of Norway could pretty much retire, and live, a little frugally, on their investment income.

There is, of course, no reason to stop at this point. The Norwegian government could still oblige all Norwegians to continue investing a portion of their earnings in their v40's. By enforcing some short term frugality, and maybe even working a couple of days a week, Norwegians could be forced to further increase the value of their v40's, making the whole country richer and richer.

Although this should work for Norway, there is a significant problem with expanding such schemes on a global basis.

Going back to the UK example given above, I set the v40 allowance at £50k per year. Using long-term UK interest rates, this gives an investment income of £1500 per year. A typical rent in the midlands of the UK would be in the region of £500 per month for a two bed flat. Even with two adults, £3000 a year would only cover half a year's rent, never mind other living costs.

While this money would be very helpful, it would fall far short of being truly a 'virtual forty acres'. Even sharing housing costs, and living very frugally, it is not possible to survive in the UK on £1500 per year. In fact £30 a week would hardly cover food and utility costs even if you owned your own home.

I chose the value of £50k for an important reason. The stock market capitalisation of the top forty UK companies is in the region of £1000 billion, if we assume the total capitalisation is double this, a brave assumption, then the total wealth available for investment in the UK is £2000 billion. The population of the UK is 61 million, or say roughly 50 million adults. So the available capital on the UK stock market for investment in v40's is about £40k per head. This assumes no other investment use for this capital, such as, for example pensions.

Alternatively in 2009 UK gdp per head was roughly $35,000 per head [Economist 2010c]. Assuming that total non-residential capital per head is roughly 2.5 times gdp per head [Miles & Scott 2002, 5.1 or 14.1], this gives $88,000 capital per person, or roughly £57k per person.

Another calculation; the Halifax Building Society [BBC 2010a] estimates that total UK personal wealth amounts to £6.3 trillion or £237,000 per household, however more than a third of this is in the form of housing. A large part of the rest will be in pension funds.

If one third is in housing, that leaves £158k per household. Assuming 2 adults per household this gives £80k per adult, which gives ball-park agreement with the figures for stock market capitalisation above.

This leaves us with a basic problem. If UK capital is used for UK savings, there simply isn't enough wealth per person, even if it is shared out absolutely equally, to give a modest investment income for every person. And of course a major part of the current capitalisation is already tied up in pension funds and is committed to future retirement needs.

This actually is obvious if you go back to Bowley's rule as discussed in sections 1.3 and 1.6 above. Historically, in capitalist societies, total returns on capital are roughly equal to half of the total returns to labour. So even if capital was shared absolutely equally to all individuals, it would only be equivalent to half their wages. With present levels of capital it would not be enough money to live on.

Norway's sovereign wealth fund represents a special case. Most of the investments in Norway's sovereign wealth fund are invested in companies outside Norway. So most of the investment income accruing to Norway comes from other countries. Interestingly this means that egalitarian, liberal Norway, with it's generous high per capita spending on foreign aid, is probably the world's most effective, and most discrete, neo-colonialist nation.

This general problem of insufficiency of capital will be returned to in depth in section 4.8 below.

1.9 Wealth & Income Distributions - Loose Ends

Before leaving discussions of income modelling I would like to briefly discuss two areas of income distribution that I have not been able to model successfully, but which I think are of importance.

1.9.1 Double Power Laws

Back in section 1.1 above, it was noted that some researchers have noted that there appears to be a split in the power tail of income distribution into two or even three separate sections. This appears to give a split between the 'rich' and the 'super-rich'. Some models have been proposed for this, of varying plausibility.

It is possible that this arises simply from the basic models above.

Figure 1.9.1.1 here

For example, figure 1.9.11 above for model 1E is simply a rerun of model 1D but with larger spreads on the normal distributions for consumption. Figure 1.9.1.1 is a log-log graph, with a long power tail that shows two or possibly three different straight line zones. It is likely that a more realistic log-normal distribution would exaggerate this effect.

Another possible source of different power laws is the consumption function. All the models in this paper have used a savings/consumption function that is strictly proportional to wealth. This has the value of simplicity, but may not be realistic.

Common sense suggests that the more wealth people have the smaller the proportion of their wealth they consume and the greater the proportion they will save. Note that rich people are assumed to spend more as they get richer, just that the extra spend is not as big as the extra wealth.

It should be noted however that this assumption is controversial, though recent research findings tend to support this assumption [Dynan et al 2004].

The idea that consumption functions are concave in this manner seems so obvious that it has in fact been proposed as a source of wealth condensation effects. Clearly this paper has demonstrated that this mechanism is not necessary.

During modelling for this paper, an attempt was made to run income models that included concave consumption functions.

The results suggested that concave consumption functions did indeed produce a two-section power law. However the results were highly unstable; small change in parameters could result either in a return to a single power law, or collapse of the distribution to a single wealthy individual.

The results were not sufficiently strong to justify presentation here, but they do suggest that this is a possibly useful area for future research, given access to better data to calibrate the models with.

Finally, while discussing the role of consumption and savings functions, it is worth noting that there is little role for being judgemental with regard to savings.

It is very easy to suggest that it is the fault of poor people for being poor if they do not save for the future. But as has been seen in previous income models the rewards for saving are disproportionate.

While it the form of savings functions are still up for debate, it is clearly easier to save a portion of your income if your income is higher.

Indeed, in the exact opposite of the '40 acres' model, in normal life people face a 'compulsory spending' world. People are obliged to spend a minimum amount of money on food, clothing, housing, heating costs, transport, etc. This compulsory spending will have exactly the reverse effect of the compulsory saving of section 1.7.2 above; it will make inequality worse. Rich people have more discretionary spending, which makes saving easier. On top of this, as Champernowne pointed out, the role of inherited wealth gives an enormous advantage to the better off.

1.9.2 Waged Income

The second loose end is potentially much more interesting, and relates to the payment of income in the form of wages and salaries.

In all the models in this paper, wage distributions have assumed to be either uniform or normal distributions.

The uniform distributions are clearly very unrealistic. They were used primarily for simplicity, and also to demonstrate very clearly that gross inequalities of wealth could be produced with absolutely identical individuals.

The normal distribution was used in the more realistic models primarily to avoid controversy, and to provide a useful comfort blanket to any economists still reading the paper. In fact a log-normal would probably have been a more realistic choice, as per figures 1.1.1, 1.1.2, 1.1.4 & 1.1.5. The author has looked at a comparison of the log-normal and the Maxwell-Boltzmann distribution for describing income distributions applied to high quality data sets from the UK and US [Willis & Mimkes 2005]. From this I am firmly of the belief that waged income is distributed as a Maxwell-Boltzmann, or rather a Maxwell-Boltzmann like distribution.

The main reason for this is that the Maxwell-Boltzmann distribution is inherently a two-parameter distribution, unlike the log-normal which is a three parameter distribution. So the Maxwell-Boltzmann is inherently simpler than the log-normal. Another way of thinking about this is that the log-normal can take many different shapes, the Maxwell-Boltzmann only has one. It is an extraordinary coincidence that two completely separate sets of data from the US and UK can be fitted by the only log-normal, out of all possible log-normals, that can fit a Maxwell-Boltzmann distribution exactly.

There is however one small fly in the ointment for these Maxwell-Boltzmann distributions (and also for the equivalent log-normal distributions). The Maxwell-Boltzmann distributions in income distribution show a significant offset from zero, something that is not normally seen in physics applications. Or indeed in physics theory; which in these models usually uses pure exchange processes subject to conservation principles (much more on this below in section 7.3).

With their offsets and their exponential mid-sections, these 'Maxwell-Boltzmann' distributions in fact look very like GLV distributions, but of course without the power tails.

It is my belief that these distributions are in fact the product of a dynamic equilibrium process that produces an 'additive GLV' distribution, in contrast to the normal 'multiplicative GLV' distributions, that have been seen throughout this paper.

A possible explanation for this is discussed in section 7.4 below, though this is highly speculative. Although speculative, I believe that this might be an important line of research. It also raises some important philosophical questions on the nature of inequality.

If the distribution of income is a log-normal, then it could reasonably be suggested that the distribution arises from the inherent skills possessed by the individuals, which following the central limit theorem, could reasonably be distributed as a log-normal. This would make the distribution of wages exogenous to the models, as in fact they have been modelled in this paper.

I personally am not convinced that the log-normal found in income distributions is exogenous. My personal experience of human skills is that the majority of human beings fall into a narrow band of skills and abilities; more like a normal than a log-normal, with a very large offset from zero. Fig 1.9.2.1 below shows my assumption of how skills might reasonably be distributed. Figure 1.9.2.2 gives the example of height.

Figure 1.9.2.1 here

Figure 1.9.2.2 here

[Newman 2005]

Intuitively, intelligence and other employment skills seem likely to be distributed in a similar manner.

If the distribution of income is in fact a Maxwell-Boltzmann-like 'additive GLV', this would put a very different light on things. Such a GLV would be an outcome of a dynamic equilibrium process and would be created endogenously within the economic model.

The consequences of income distribution being an endogenously created GLV are simple. It means that poor people are being underpaid for the labour, and better off people are being overpaid. It means that capitalism doesn't reward people fairly, even at the level of waged income.

Clearly before such a bold statement can be made, it would be appropriate to produce a meaningful model for producing an 'additive GLV'.

Notwithstanding these loose ends, we have effectively dealt with the problems of poverty. Time now to investigate some other problems in economics.

2. Companies Models

Going back to figure 1.3.5, having looked in detail at the wealth and income distributions, we will now move our interest from the wealth owning individuals on the right hand side of the figure 1.3.5 over to companies, the source of wealth, on the left hand side.

2.1 Companies Models - Background

The theory of the firm has long been recognized as a weak point of neoclassical theory. The paradigmatic case for neoclassical theory is the competitive industry, in which a large number (how large is open to considerable discussion) of similar firms coexist. Neoclassical theory roots its explanations in properties of resources, technology and preferences that are independent of the organization of economic activity itself (that is, are exogenous from the point of view of economic theory). What technology could give rise to the coexistence of many similar firms in an industry with free entry? If there are diminishing returns to scale, the industry should be atomized by the entry of ever-smaller rivals. If there are constant returns to scale, the theory cannot explain the actual size distribution of firms except as an historical or institutional datum. If there are increasing returns to scale the theory predicts the emergence of a few large firms, not the competitive market originally posited. [Foley 1990]

As discussed previously, it is the belief of the author that firms exist to protect their value-increasing property, their sources of negentropy.

Firms buy goods that have well defined prices such as raw materials, components, electricity and labour.

They then use these inputs to go through a series of intermediate goods stages with, at best, indeterminate prices, at worst, very low prices. As an obvious example think of a car body shell, which has its engine and transmission installed, but hasn't yet had its electrics, glassware, finishes etc, installed. To the manufacturer it probably has more than two-thirds of its true value installed, in terms of components and labour supplied. However if it were sold on the open market it would have very low value, even to another car manufacturer, as the cost to completion for another company, or an individual, would be very high.

To complete the process of production successfully, a company has to finish the goods to a well-defined point, where they can be easily priced in the market and sold to consumers or to other companies as intermediate goods.

The company, with its plant, trained workforce, patents, designs and trademarks, exists to protect this wealth creation process.

In neo-classical economic theory, as discussed above by Foley, the sizes of the companies should either be very small if entry to markets is easy, or very big and monopolistic, depending on the returns to scale.

In fact it is well documented that company sizes, whether measured by number of employees or capitalisation follow well defined power law distributions. For background see Gabaix [Gabaix 2009] or [Gaffeo et al 2003].

These power law distributions are of course similar to the power law distributions of wealth for property owning individuals that we have seen in the discussions of wealth and income above.

The model for companies in this paper builds on the income models introduced in section 1.3 above. The modelling looks at company sizes in terms of total capitalisation K of the companies. To extend these models, three basic assumptions are made.

Firstly, in a break with the previous models, it is no longer defined that the valuation of the paper assets W matches the real capital of the company K. That is to say the short-term stock-market price W is allowed to vary significantly from the 'fundamental' value of a company's real capital K.

As well as introducing this degree of freedom, three further important assumptions are introduced.

Firstly, it is assumed that shareholders are myopic, and judge expected company results simplistically on previous dividend returns.

Secondly, it is assumed that managers of companies act to preserve the stability of dividend payouts,

Thirdly, and more importantly it is assumed that managers act to preserve the capital of their companies.

Justifications for these assumptions are given below.

Until a few years ago, despite the wealth accumulated by Warren Buffet and other acolytes of the Benjamin Graham school of investing, the concept of companies having fundamental value was highly controversial. In recent years, these views have become more acceptable for discussion, firstly following the dramatic changes in value during the dotcom and housing booms of the last decade, and secondly because of the detailed research of Shiller, Smithers and others that both disprove a purely stochastic basis for stock market movements and also give substantial evidence for long term reversion to mean for stock market prices when measured by Tobin's q or by CAPE; the 'Cyclically Adjusted Price to Earnings ratio'. This is discussed at length in Smithers [Smithers 2009] for example, and is looked at in more detail in section 8.2.1 on liquidity, below. Following the credit crunch and the dramatic changes in prices associated with liquidity problems, ideas of fundamental values have become more acceptable.

Following the recent work of Smithers, Shiller and others, and also the beliefs of the classical economists, this section takes as it's starting point the viewpoint that economic companies do have 'fundamental' values, and that these are frequently at odds with their stock market valuations.

With regard to myopic behaviour the book 'How Markets Fail' by John Cassidy [Cassidy 2009], gives an extensive discussion of data that gives evidence for short term pricing behaviour. This is discussed in depth in chapter 14.

It appears that this naïve behaviour is not restricted to naïve investors. Recent work by Baquero and Verbeek for example [Baquero & Verbeek 2009] suggests that pension funds, private banks and wealth individuals all commonly invest based on short term returns.

In their paper 'The Cross-Section of Expected Stock Returns' [Fama & French 1992] Fama and French, originators of the efficient market hypothesis, carried out econometric analysis that confirmed that four empirical factors appear to be involved in the pricing of stocks. The first of these is the risk associated with stocks, in line with the original capital asset pricing model (CAPM). The second is the size of the company. The third is the book to market value of the company. The fourth factor identified to fully explain stock market valuation is the presence of short-term momentum in pricing based on recent returns of the stock.

The work of Korajczyk and Sadka [Korajczyk & Sadka 2005] also suggests that momentum is important in company valuations and arises from liquidity considerations.

Recent academic work suggests that both size and book to market effects can be explained by changes in liquidity. This is potentially a very important topic, and is discussed at some length in section 8.2.1 below. For the companies model, liquidity, and so company size and book to market values are assumed to be irrelevant. It is assumed that liquidity is constant throughout the modelling process.

As modelled by the CAPM, risk is peculiar to individual companies. In this model it is assumed that risk is identical, and in fact zero, for all companies in the model.

Given the above assumptions of zero risk and high liquidity; following Fama & French, this leaves short term returns as the only factor that investors use to value companies.

So, using basic finance theory, then the present value of a company is given simply by:

$$\text{Present Value} = \frac{\text{Dividend}_1}{\text{r}}$$

Where r is the relevant market interest/profit rate; Dividend_1 is the latest dividend payment, and capital growth is ignored. See for example [Brealey et al 2008, chapter 5].

This is the naïve neo-classical approach to valuing capital for aggregation; simply divide by the profit rate. We will simply take this naïve approach as it stands and follow the consequences through the model.

With regard to management behaviour, research from Brav, Graham, Harvey and Michaely [Brav et al 2005] support the contention that maintenance of a constant dividend stream is an important priority for managers of corporations.

Finally, with regard to the retention of capital within companies, the history of the defence company General Dynamics, gives a very interesting case study. General Dynamics (GD) are interesting in that GD formed a casebook example of how companies are supposed to behave, according to finance textbooks, by working solely to enhance the value of shareholder's stock.

In the real world, GD are notable in their exceptionalism, in that their deliberate downsizing to enhance profitability was not only unique in the defence industry, but pretty much unique in corporate history.

In contrast to GD, other defence contractors in the 1990s followed deliberate policies of acquisition or diversification in order to maintain their size. This despite the obvious collapse of the defence market following the end of the Cold War.

The following are quotations from 'Incentives, downsizing and value creation at General Dynamics' by Dial and Murphy:

In the post-Cold War era of 1991, defense contractor General Dynamics Corporation (GD) faced declining demand in an industry saddled with current and projected excess capacity. While other contractors made defense-related acquisitions or diversified into non defense areas, GD adopted an objective of creating shareholder value through downsizing, restructuring, and partial liquidation. Facilitating GD's new strategy were a new management team and compensation plans that closely tied executive pay to shareholder wealth creation, including a Gain/Sharing Plan that paid large cash rewards for increases in the stock price. As GD's executives reaped rewards amid announcements of layoffs and divestitures, the plans became highly controversial, fueling a nationwide attack on executive compensation by politicians, journalists, and shareholder activists. Nonetheless, GD managers credit the incentive plans with helping to attract and retain key managers and for motivating the difficult strategic decisions that were made and implemented: GD realized a dividend-reinvested three year return of 553% from 1991 to 1993—generating $4.5 billion in shareholder wealth from a January 1991 market value of just over $1 billion.1 In the process, GD returned more than $3 billion to shareholders and debtholders through debt retirement, stock repurchases, and special distributions. [Dial & Murphy 1994]

In contrast to the explicit strategy of creating shareholder value initiated by General Dynamics, this was the behaviour followed by their competitors:

Table 7 summarizes the strategies selected by GD and eight other defense contractors from 1990 through 1993, based on an analysis of quantitative financial data as well as our qualitative interpretation of annual reports, press releases, and news articles. The table includes the nine largest domestic defense contractors (ranked by cumulative 1989-1992 defense contracts). Exceptions are General Electric and Boeing, excluded because their defense operations account for less than 10% of total firm revenues. Some of the strategic options adopted by these firms include: Acquisitions to achieve critical mass; diversification into non defense areas, or converting defense operations to commercial products and services; globalization, i.e., finding international markets for defense operations; downsizing and consolidation; and exit.

Diversification and commercialization. *A 1992 survey of 148 defense companies sponsored by a defense/aerospace consulting firm found that more than half of the respondents report past attempts to "commercialize" (i.e., applying defense technologies to commercial products) and more than three-quarters predict future commercialization. Martin Marietta CEO Norman Augustine, however, cautioned his industry counterparts about wandering too far from their areas of expertise:*

> "Our industry's record at defense conversion is unblemished by success. Why is it rocket scientists can't sell toothpaste? Because we don't know the market, or how to research, or how to market the product. Other than that, we're in good shape."

…Globalization. *A number of firms are retaining a defense focus, attempting to bolster sales through globalization, selling U.S. built weapons abroad. This strategy is unlikely to yield* *dramatic growth, since the demand for weapons is declining world-wide and many foreign countries have their own national producers who are also faced with excess capacity*

Downsizing, consolidation and exit. *Table 7 shows that while most contractors adopted a combination of strategies, all adopted some form of downsizing or consolidation to reduce excess capacity. However, while a few contractors (including GM Hughes, Grumman, and McDonnell Douglas) have divested unprofitable non core businesses where they had little chance of building strategically competitive positions, only General Electric (not included in table 7) followed GD in exiting key segments of the defense industry. Interestingly, it was General Electric (where Anders held his first general management position) that pioneered the "#1 or #2" criterion as a strategic assessment for the composition of its portfolio of business units…*

…Goyal, Lehn, and Racic (1993) also analyze investment policies in the defense industry. They report evidence that defense contractors began transferring resources from the industry as early as 1989-1990 through increased leverage, dividends, and share repurchases. Our complementary evidence suggests that although other contractors also espoused and eventually adopted consolidation and downsizing, GD's response in moving resources out of the industry was quicker and more dramatic. To draw an analogy: While other defense contractors engaged in a high-stakes game of musical chairs—hoping to be seated when the music stopped—GD pursued a strategy of offering its chair to the highest bidder. [Dial & Murphy 1994]

Despite the obvious and dramatic decline of the defence industry following the end of the Cold War, and even despite the example of General Dynamics, the managers and directors of some of the largest and most important companies in the world's largest economy followed a clear pattern of attempting to maintain the size of their companies, without regard to the value of their shareholders investments.

It is the belief of the author that this pattern is widespread throughout the management of limited companies, and so this will be used as a base assumption of the companies model that follows.

2.2 Companies Models - Modelling

Figure 2.2.1 here

Figure 2.2.1 above is a slightly modified version of figure 1.3.5.

A few changes have been made, though the overall process is the same. We are now looking at the financial assets from a company point of view, and we are not interested in the individuals. So we now have a total of N companies, which we count from j=1 to j=N.

The big difference with previous models is that we removed the assumption that K = W or that $k_j = w_j$.

So here we differentiate between the fundamental value of the real capital k_j formed of the firms buildings, plant, patents, etc and the market valuation of the company w_j. w_j represents the sum of the stock market value of paper share certificates held by the owners of the company.

(Note here that w_j is the total wealth represented by all the shares in company j held by various different individuals – w_j is not the same as w_i.)

At the beginning of each simulation we start with $\Sigma k_j = K$ for all the companies, and also $\Sigma w_j = K$ initially.

That is, to start with, all the companies are the same size, and all are valued fairly by the stock market, with the fundamental value of each company equal to its market capitalisation.

It is assumed that each of the j companies has a standard rate of growth r_j. The average $\bar{r}_j$ will be 0.1, that is each company produces value roughly equal to 10% of its capital each year. So each of the companies is identically efficient in the use of their capital.

However, to introduce a stochastic element, we will allow a normal distribution in the values of r_j with a variance which is 20% of the value of r. So r varies typically between 6% and 14%.

Effectively this assumes that although companies return the same on capital over the long term, they may have short-term good and bad years which allow returns to fluctuate slightly around the long term average.

It is assumed that the market is not well informed about the fundamental value of individual companies. Following the research of Fama & French and others, it is assumed that investors simply use the average market rate of returns (0.1 or 10%) as their guide for valuing companies.

So the new market capitalisation w_j for each iteration of the model will simply be the last actual real returns $\pi_{j,t}$ divided by the long-term rate of returns.

$$\text{so:} \qquad w_{j,t+1} = \frac{\pi_{j,t}}{r}$$

Then the expected returns for the next year will be the market capitalisation Wj multiplied by the average market rate of return.

$$\text{so:} \qquad \hat{\pi}_{j,t+1} = w_{j,t} r$$

Which is an unnecessarily complicated way of saying that next year's expected returns will be the same as the previous years actual returns.

As in the previous models, we will assume that labour is fairly rewarded for the amount of added value that it is supplies.

So L = e exactly, and both L and e can be ignored in the mathematical model.

The loop of the simulation was carried out as follows:

The amount of production is calculated by multiplying the capital of each company by the relevant production rate, so:

$$\text{production} = k_{j,t} r_{j,t}$$

After a round of production all of the companies will receive cash from purchasers of its manufactured goods. This cash value will represent the value added in the production process.

Each of the companies will have a value of expected returns ($\hat{\pi}_{j,t}$) based on its current market capitalisation.

In the simulations carried out actual payouts of profit π were varied by using different payout ratios.

If the value added; the production, is greater than the expected returns then the managers might pay out 90% of the earnings, retaining 10% of the extra value, so allowing a buffer to be built up against future problems, also to allow expansion of the company, empire building, etc. This extra value is added to the total capital.

If the managers only pay out 90% of the earnings, this is defined from now on as an 'payout ratio' of 90%. The model allows different payout ratios on the upside and downside. So managers may have an upside payout ratio of 90% and a downside payout ratio of 80%. This would mean that the management would pay out 90% of the earnings if earnings were greater than market expectations, but would only pay out 80% of earnings if earnings were less than market expectations.

For example in model 2B both the upside and downside payout ratios were 90%.

These actual payouts then give the market its new information for resetting the market value w_j of the various companies.

The capital k_j of each company is then recalculated as follows:

$$k_{j,t+1} = k_{j,t} + \text{production} - \text{actual\_returns}$$

Finally at the end of each round the values of the company capitalisations have to be normalised. The reasons for this are as follows.

This model assumes a stationary economy with a fixed total amount of capital K.

This capital can be bought and sold between different companies, as they are required to give earnings in requirements of market expectations.

All of the companies will receive cash from purchasers of its manufactured goods. This cash value will represent the value added in the production purpose.

Some companies will receive more cash than they are expected to payout, some will receive less.

It is assumed that the cash rich companies will purchase real capital off the cash poor companies, so allowing the cash rich to expand, and the cash poor to pay their earnings.

At each round of the modelling process, the sum of the capital is renormalised to the original K. This is because asymmetric retention of funds allows excess growth or decline for the whole economy.

Ideally a more realistic model would automatically adjust these processes. However, this is problematic, there are deeper, and interesting, instabilities at work, these are the subject of models in section 4 below.

2.3 Companies Models - Results

2.3.1 Model 2A Fully Stochastic on Production, No Capital Hoarding

Model 2A is the simplest model, so simple that it inevitably fails.

Firstly the model is completely stochastic. Each company produces output worth exactly 10% of its capital on a long-term average. However the value of 10% varies up and down stochastically according to a normal distribution.

In model 2A the payout ratio is deliberately set at 1. This means that the managers of the companies payout the full amount expected by the market. They do this no matter how well, or how badly the companies perform.

Figure 2.3.1.1 shows the full log-log distribution of all the (non-negative) companies. Figure 2.3.1.2 shows the power tail with the trend line fit for the power tail.

Figure 2.3.1.1 here

Figure 2.3.1.2 here

Companies that lose money, due to poor production, still pay out to market expectations, so they slowly drain their capital and lose it to other companies that have above average production. Because of this the model is not stable, and the distribution changes as the model progresses.

Despite this, it is noticeable that the model quickly generates a stable power tail with an exponent close to –1; close to the value seen in real life. The power tail remains stable from 10k to 50k iterations. Above 50k iterations the number of companies being eliminated (going negative) becomes very large and the transfer of capital to the larger companies starts to change the exponent of the power tail.

The important thing to note here is that a very simple model, using the standard valuation system of capitalism, quickly generates a power tail of companies of vastly different sizes. In the

50k, run power tail companies vary in their capital between 80k units and 80,000k units. But all the companies are absolutely identical in their earning ability, effectively the companies have identical managements making identical products with identical inputs. The differentiation in size has only occurred through the stochastic forces of chance.

2.3.2 Model 2B Fully Stochastic on Production, Capital Hoarding

Model 2B is identical to model 2A in that the companies are identical in average earnings, but these earnings vary stochastically from model to model.

Model 2B is different in that the payout ratios were changed in an attempt to create a stable model. Unfortunately this proved difficult. The only values that prevented 'washout' of smaller companies were payout ratios of 0.9 on both the upside and downside. Initial investigations suggest that this is related to the production rate of 0.1.

The results are shown in figures 2.3.2.1 and 2.3.2.2.

Figure 2.3.2.1 here

Figure 2.3.2.2 here

Unfortunately this model is a bit too stable. Although it shows a very clear power law, still with identical companies, the exponent of the power law is very different to that seen in the real world.

It appears that the retention is too great and is forcing a high minimum value for companies, so preventing the formation of the power tails with slopes seen in model 1A.

2.3.3 Model 2C Deterministic on Production, Capital Hoarding

In model 2C the production rates of the companies was set prior to running the model, and were again drawn from a normal distribution. So in this model some companies produced more than 10% all the way through the model, some produced less than 10% all the way through the model.

Note that model 2C is not stochastic, it is deterministic.

In this model some companies are more efficient than others with their use of capital.

Again the payout ratios were adjusted to prevent elimination of companies from the bottom of the distribution. It was found that any downside payout ratio of less than 0.5 or so prevented this washout. Figures 2.3.3.1 and 2.3.3.2 below are for a downside payout ratio of 0.5 and an upside payout ratio of 0.9.

Figure 2.3.3.1 here

Figure 2.3.3.2 here

Intriguingly the power law exponent of –0.68 is close to the value of –1 seen in real life. However the fit is poor, and it turns out that the value of the exponent is highly sensitive to the value of the upside payout ratio and can change to high tens or low decimals for small changes in the upside payout ratio. Initial modelling suggests that the value of 0.9 is closely related to the production ratio of 0.1. As the production ratio is changed, an upside payout ratio of one minus the production ratio gives a power tail close to one.

Again, the important thing to note is that relatively small changes in relative efficiency of the companies produces a power tail with very large, multiple factors of ten, differences in size for the companies.

2.4 Companies Models - Discussion

As can be seen from the results, using a very simple combination of classical economics and dynamic statistical mechanics allows the building of simple models that give power law distributions for company sizes similar to those found in real life economies.

As with the income models it noticeable that there are many things that are not needed to produce such a model, these include:

- Economic growth
- Population changes
- Technology changes
- Different initial endowments (of capital)
- Shocks (exogenous or endogenous)
- Marginality
- Utility functions
- Production functions

The issue of marginality, utility, production functions will be returned to in a moment, before that I would like to discuss the roles of shocks, expectations and behaviouralism.

It is notable that the models do not include for exogenous shocks, which are often found in explanations of company size.

Models 2A and 2B are stochastic, and do therefore model minor endogenous shocks to productivity. These could be issues such as a variation in breakdown rates of machinery, management efficiency, etc from period to period. What is notable about models 2A and 2B is that the average productivity of all companies over the long term is identical; and yet a power law still results.

Model 2C is effectively deterministic. The initial productive efficiencies of the companies are determined prior to the simulation. The simulation then rapidly reaches an equilibrium with a power law distribution. There are no shocks in model 2C; external or internal.

Expectations and behaviouralism do enter into the model in two different ways, firstly with regard to the pricing of stocks, and secondly with regard to the retention of capital within companies.

In both cases these are very obvious forms of behaviour and are supported by economic research.

With regard to returns, the assumption is simply to take the pricing of financial assets as strictly based on their recent returns. This is in fact the "traditional" naïve neo-classical form of pricing capital and is supported by the research of Fama & French and other work discussed in section 2.1 above. This assumption that prices of assets are defined by simplistic projections of present earnings is also at the heart of Minsky's theories.

The assumptions on capital retention are more subjective than the assumptions on returns, and more arbitrary in the specific amounts of returns chosen, and is the weakest part of my company modelling. This is discussed in more detail below, when comparing with the work of Ian Wright.

However the work of Dial & Murphy regarding General Dynamics and other companies make the assumptions very plausible.

What is important to note is that the above assumptions on expectations are the only assumptions needed. No detailed assumptions about the understanding of the economy, interest rates, growth, technology, etc are needed.

The only 'behaviourism' that we need to assume is that, firstly investors are deeply short sighted, and secondly that managers don't like sacking themselves.

It is clear from the models that neither utility nor marginality are relevant.

Much more importantly, the output distribution for the models is demonstrably not 'efficient' in the normal neo-classical usage.

To take models 2A and 2B as examples, capital is rapidly shifted between companies according to short-term results, and companies with equal long-term efficiencies end up being sized very differently. In a neo-classical version of model 2A of 2B, either one company would dominate, or all companies would be equally sized.

Model 2C is far more realistic, and much more interesting. It also shows how profoundly free markets fail to allocate capital effectively.

Model 2C has a range of production efficiencies. Some companies make better use of their capital than others.

In a neo-classical outcome (or indeed in the classical models of Smith, Ricardo, etc) the outcome of such a model should be crystal clear. The most efficient company should continually be rewarded with more capital until it ends up being a monopolist, owning all the capital in the economy. Despite the best efforts of managers to cling on to their capital, investors should continually remove their capital from all the less efficient companies until these companies have no capital left and go out of business.

This is not what happens.

In model 2C, and as Graham, Buffet and others have discovered, also in real life, poorly performing capital is simply written down.

Companies are allowed to retain some of their real, book value, capital K. But part of their financial wealth is written off. Once an under-performing company's financial wealth W is small enough to make the (poor) returns from the actual K equal to the normal market rate, then the company is allowed to continue under-performing, and under-utilising its capital, indefinitely.

So it is noticeable that moderately bad companies are only downgraded, they are not driven out of business as economic theory suggests they should be.

This represents an enormous misallocation of real capital.

In model 2C the top company has a capitalisation/capital ratio of 1.37, the bottom company has a capitalisation/capital ratio of 0.62. The bottom company is half as efficient as the top company, but once it has been written down, it is allowed to limp on inefficiently.

That this happens in real life is supported by the effective long term investing models of Benjamin Graham, Warren Buffet and others. The accumulated wealth of Warren Buffet has always been one of the most pertinent criticisms of the efficient market hypothesis.

In an economy such as model 2C above, the Graham/Buffet approach is straightforward.

Finding companies with under-valued physical assets is straightforward; you simply look at the book value of assets compared to the stock price.

Generally it is poorly performing human capital that has driven companies into under-performance. The quality of human capital is something that can change very quickly. As General Dynamics showed, a change of CEO can be sufficient.

The Graham/Buffet approach uses various measures to identify increases in the efficiency of human capital. These include qualities such as paying down debt and good recent dividend history.

By this process, investors such as Graham and Buffet can identify companies that are undervalued, with under-performing capital, and that are also likely to move quickly to over-valuation.

In practice this failure of capitalism may not be as bad as painted above.

Firstly it is likely that other processes will ensure that capital gets redeployed more quickly. Despite the best efforts of capital retaining managers, many companies do go bankrupt; many more get merged or taken over. Newer, more efficient companies also enter the market and take market share from existing non-performing companies.

It may also be the case that the power law distribution is, accidentally, highly effective in preventing monopoly or oligopoly in the market place.

Indeed, looking at deviations from power law distributions, in industry sectors as well as whole economies, may well be a very useful way of identifying monopolistic behaviour. If a company is bigger than its place on a power law suggests, then it is probably behaving in a monopolistic or oligopolistic manner and should either be split up or subject to a super tax of some sort.

It is the belief of the author that this modelling approach is generally applicable. Although the model focuses specifically on dividends, a simplistic Modigliani & Miller assumption of the irrelevance of forms of payout would allow that the model would work when capital growth was substituted for, or used in addition to, dividend payments.

Even in the non-listed sector the same basic arguments hold. If a small business goes to a local bank for a loan, the bank may look at the size of the business assets as collateral for the loan, but the calculations of loan size will be based on estimates of the future revenue streams of the business, based on recent historic revenue streams.

The general applicability of this type of model can be seen by looking at the shortcomings of my own model, and also by comparing the model with those of Wright.

The workings of the model above are straightforward, and similar to the other GLV models. The companies have a positive feedback loop which means that the more companies earn, the more capital they get.

There is also a negative feedback loop, so the bigger companies get the more income they have to pay to investors.

If these were the only two rules, then the most efficient company would grow explosively into a monopoly. A true power law distribution can not go down to zero, so to be stable, a power law always needs some other distribution to 'support' it. That is why power law distributions are normally 'tails' to other distributions.

As Levy & Solomon make clear, there needs to be a 'reflective barrier' above zero.

The assumption of retention of capital assures a continuous, if minimal income to all companies, however small. This prevents collapse of the distribution to a single point, and allows the generation of the power tail distribution.

This is the weakest part of the model above, with factors 'selected' (fixed, if you prefer) to ensure the distribution does not collapse.

While these assumptions are somewhat contrived, the work of Wright shows that different, but similar assumptions are just as effective.

In the modelling of companies the models of Ian Wright are significantly different to, and significantly better than, my own, but detailed analysis shows strong similarities.

Wright does not model a financial sector, and the mathematical modelling above is not therefore relevant.

In Wright's models, each company is owned by a single 'capitalist', and there is no distinction between the capital of the company and the wealth of the owner. Wright models the expenditure of the capitalist and the income of the company as both being stochastic, and crucially, independent of each other. So the capitalist spends at a set, but stochastic, rate, which depends only on the wealth of the capitalist. So the capitalist is spending his 'expectation' of the future wealth of his company, which is implicitly assumed to be the same as the present wealth of his company (which is identical to his personal wealth).

Meanwhile the income of the capitalist's company is set stochastically in the market, and may not match the expenditure of the company. Any mismatch then results in an expansion or contraction of the wealth of company/capitalist. This consequently results in a power law of company sizes that is analogous to my own model.

It should be noted that in at least two ways Wright's models of companies are superior to my own. Firstly, Wright models employment directly which my own models ignore, substituting capitalisation. Secondly, Wright allows for the extinguishing of companies as they become too small to trade, and the creation of new start-up companies as individuals become sufficiently wealthy to employ other individuals.

This avoids the somewhat artificial 'capital hoarding' approach that is used in my own model, which maintains all companies as operational entities, however severe their losses.

In real life clearly both mechanisms operate, with bankruptcy and new company formation happening alongside poorly performing companies that limp on for years without giving good returns on their capital. A third mechanism of corporate takeover, divestment and splitting of companies also takes place. Detailed research would be needed to determine the relative importance of the different mechanisms. Personally I believe that Wright has identified the most important factor in new company formation and extinction.

The main point is that, as long as you have a means of supporting the base of the distribution, the basic pricing mechanisms of capitalism produce a power law tail as seen in reality.

The differences between the models of Wright, and my own, underline a much more important point. If you use the basic ideas of the classical economists, combined with statistical mechanics, it is in fact very easy to get the same power law distributions that are seen in real life. If you use neoclassical theory, efficient markets, and static equilibria, it is pretty much impossible to give convincing reasons for power law distributions. Neither Wright's or my own models may be fully correct, but they are both clearly closer to the truth than anything produced by neoclassical theory.

Another area that needs further investigation is the exponent of the power tail. Data from real economies suggest that this has a value close to 1 in all cases whether measured by employees, capitalisation or other variables. This suggests that a deeper underlying equilibrium is being formed, with a 'self-organising criticality' (SOC) as previously suggested for income distribution.

My first model produces this exponent well, but is not stable over the long term. My stable models can reproduce this value, but only by 'fixing' the parameters of the model, a solution that is neither universal nor acceptable. Wright's model does produce this exponent, and without any apparent 'tuning'. As such Wright's model appears to be superior to my own, but as a non-mathematised model, it is not fully clear why his model does this. This is a suitable area for further investigation.

3. Commodity models

The following is a brief model, mainly to introduce some concepts and demonstrate the importance of a dynamic modelling approach to markets.

This paper has taken a classical economics approach that assumes that all goods and services have a meaningful intrinsic value that ultimately relates through to basic concepts of entropy in physics and biology.

It is immediately obvious that the prices of some goods; land, housing, gold, artworks, cabbage-patch dolls, etc, show wild fluctuations in price that appear to contradict the assumptions of fundamental value in classical economics.

To investigate this further a simple dynamic model of a commodity market is constructed, largely following the lines of the previous company model.

The intention is to model the behaviour of a commodity such as copper, platinum or coffee. For such commodities prices can fluctuate wildly, and this is often blamed on external factors such as demand, weather, war etc.

In the model below it is demonstrated that the main source of price fluctuations are endogenous and relate to the provision of capital by financial markets.

3.1 Commodity Models - Background

The model aims to model the behaviour of mining or agricultural commodities such as copper, aluminium, nickel, platinum, coffee, tea, cocoa, sugar, etc.

Such commodities have wildly fluctuating prices, normally characterised by long periods of low prices punctuated by severe spikes. The figure 3.1.1 below for copper shows a typical example.

Figure 3.1.1 here

This pattern is also seen in other commodities such as oil or natural gas, land, housing, etc.

While it is believed that similar forces operate in the markets for oil and houses, these commodities are sufficiently important that they can in turn have large impacts on the economy as a whole.

For simplicity the model below chooses to model something like copper or sugar that can have large price spikes without having a significant effect on the economy as a whole. This allows important simplifying assumptions to be made in the model.

Although at first glance copper, aluminium, nickel, platinum, coffee, cocoa and sugar would seem to have little in common; in fact they share three important factors.

Firstly, in a stable economy demand for these things is quite stable and relatively insensitive to price.

Cables are made from copper, and if you build a house you need cables and you pay the price necessary. Similarly, most planes are made from aluminium. Even in poor countries people tend to drink a certain number of cups of tea or coffee each day, with their usual number of spoons of sugar. The total costs are small compared to other outgoings such as food or rent, and the pleasure obtained, so people tend not to cut back even if prices increase significantly.

The second factor these commodities have in common are non-substitutability. Copper is both an excellent conductor and corrosion free, and is also relatively cheap compared to other metals with these properties. It is slowly being displaced by plastics for plumbing and aluminium for electrical use, but the substitution process is very slow. While Boeing are beginning to build airliners out of composites, the process has not been easy and demand for aluminium seems likely to remain high for decades. While some people swap between tea and coffee, most have a favourite brew, and there is no other easy substitute for hot caffeinated drinks. I don’t know of anything that can effectively substitute for chocolate.

The third factor is that all the above commodities take a long time to increase their output by installing new capital. Mines are large, complicated, and often isolated. To bring a new mine into production can easily take three to five years, even expanding an existing mine can take two to three years. Unlike say wheat or rice; coffee, tea and cocoa grow on trees or bushes, and there is a limit to how much you can rush nature.

For commodities such as these, price signals take a long time to result in increased output.

It is this delay that changes the problem from one of comparative statics to one of dynamics, so it a dynamic model that is needed.

3.2 Commodity Models - Modelling

This model follows on from the companies model above, and in one way is much simpler. So simple that the model was moved to a spreadsheet. For anybody who is interested this can be copied and installed into Excel from appendix 14.8.

Although the same basic model is used as that in the companies model above, in this case one section of the economy is modelled as a single unit, so there is only a single set of equations running in the model.

For the sake of the argument, assume the commodity is copper.

In this model, along the lines of classical economics, the production cost of copper is fixed and related directly to its inputs, a mix of energy, machines and various types of labour. We have assumed that the price of copper, even if it varies dramatically, has very little effect on the economy as a whole.

This means that the prices of the inputs of energy, machines, labour and any other inputs vary negligibly with the price of copper.

So the cost of producing copper is a simple linear function of the amount of copper produced.

As with the companies and incomes models, the total amount produced is a fixed ratio of the capital installed.

Taken together this means that the marginal price of extra copper is zero. This model ignores marginality, because its importance is marginal, to the point of irrelevance.

The price of copper is a different matter. It is assumed that total demand for copper is almost constant with a 'normal' amount required in the market place. In this model 100 units of copper.

When this amount, or more, is available, copper companies charge the costs of production. Also they lower their output by closing down excess capacity. This gives a base price, a classical economics price, for copper of 1.0 in this model.

If production drops below that required, then price increases very rapidly and demand is choked off very slowly, the demand is highly inelastic. Figure 3.2.1 below shows the price volume curve used in the model.

Figure 3.2.1 here

This is of course a completely unrealistic, hypothetical demand curve of the type beloved by economists.

In a comparative statics analysis an economist would then draw one or more hypothetical supply curves across the same graph and predict a static equilibrium based on marginal outputs of the different mines.

This is not a meaningful approach. The effects of delays in installing capital, and/or the retention of wealth by companies mean that a static equilibrium is not possible.

In this model, just as in the companies model, the standard market interest rate defines the expected returns, based on the previous market capitalisation w.

Again, as in the previous model, payouts are predicated on the expected returns using payout ratios, with companies hoarding capital or returning it to shareholders as appropriate.

When supply is low, and prices jump up, the mining companies find themselves with much higher receipts than costs. In these circumstances the excess cash is used to provide more capital.

As discussed above, this capital is added to the productive capital, but only after a lag of a number of iterations. This lag can be adjusted in the model from zero to ten cycles.

Once the new capital has been added after the lag in time, then production can be increased. Eventually this allows supply to meet demand and prices can drop again.

3.3 Commodity Models - Results

The results are fairly straightforward.

Figure 3.3.1 below gives the output for Model 3A, this shows the prices for copper with no lag on capital installation and payout factors of one; ie no capital hoarding.

Figure 3.3.1 here

Even with this very simple model the system is unstable and produces wide cyclical variations in prices (this was something of a surprise, I had thought the model might be stable with instant installation of capital and no capital hoarding). The real price of copper, based on inputs, should be 1 unit; note that the system is only at its true input price for short periods of time. Left to itself the market charges an average price slightly over 50% of the input cost price. The extra 50% being caused by the cyclical over-production and destruction of capital, and consequent rent taking.

Figure 3.3.2 below for model 3B shows a capital lag of two periods, but still with a payout ratio of one.

Figure 3.3.2 here

This shows a pattern closer to reality; long periods at 'classical' prices are interrupted with intermittent spikes. Even in this simple model it is notable that the spikes have a variable pattern showing chaotic (not stochastic) behaviour. With this capital lag, the average price is raised to 1.7 times input cost, as the cycles of capital creation and destruction become more aggressive, and rent taking becomes larger.

Finally figure 3.3.3 shows model 3C with zero capital lag, but with up and downside payout ratios of 0.9.

Figure 3.3.3 here

This figure demonstrates that capital hoarding alone can produce complex cyclical chaotic behaviour. As with figure 3.3.1, cycling only results in 50% price gouging.

3.4 Commodity Models - Discussion

I intend to keep the discussion of the commodity model quite brief. The main issues raised are dealt with in more depth elsewhere. Some of the main points of note are as follows.

Very simple dynamic economic models can result in complex chaotic behaviour. Behaviour that mimics real life surprisingly well.

The behaviour is chaotic, not stochastic.

The random changes are generated endogenously. There is no stochastic generator in this model. This distinction is very important, and is discussed at length in section 5 below.

This is a Lotka-Volterra model, not a General Lotka-Volterra model. This model is very similar to the lynx and hares model first discussed back in section 1.2, in fact it is closer to the Soay sheep and grass model. The build up of excess capital in the mining companies is analogous to the build up of excess sheep biomass on the island of Soay. The build up of capital is too much for the economy to support, as the build up of sheep is too much for the island to support. While the GLV models were stable, like many Lotka-Volterra models, the build up of capital in the commodity sector is inherently unstable. The problems are deep in the maths of the system. Blaming investors or speculators for misjudging their investments is as sensible as blaming the sheep for procreating.

Diminishing returns and marginality are conspicuous by their absence.

Diminishing returns are not needed for the model to work. Neither is marginality, and any costs associated with marginality are of an order smaller than those associated with dynamic effects.

Using comparative statics to analyse a dynamic process is simply not appropriate. It is the wrong tool for the job. Using comparative statics to analyse dynamic problems is about as sensible as trying to do long division with roman numerals.

Using classical economics within a dynamic framework works. It produces output prices that can be at substantial variance with input prices, and can vary substantially with time.

It should also be noted that the model does not average to the correct input prices even over the long term. The correct input prices are instead associated with the bottoms of the cycles, and are only touched for short periods of time.

Due to problems associated with the way assets are priced, the time taken to install capital, and (financial) capital hoarding by companies, the market is profoundly inefficient. Average prices are substantially higher than they would be if they had the opportunity to settle to long-term static equilibrium prices.

The form of this over-pricing is interesting. Above I referred to it as associated with capital appreciation and destruction, but the process is more subtle than this.

In a boom period, customers are substantially overcharged compared to the input costs. Extra capital is created, but the nominal capitalisation increases much faster than the real value of the capital installed. In short the companies become grossly overvalued. As a consequence they pay excessive dividends. In a boom most of the over-pricing passes straight through to shareholders as excess profits.

In the following crash, the company is still expected to match dividends at the market rate. It does so by drawing down capital to pay dividends.

Over the cycle as a whole customers are forced to overpay, with the payments transferred direct to excess profits.

Allowing dynamic cycling of economic variables in this way allows large-scale rent-taking by the owners of resources.

For most markets these effects are not so important, with the very notable exception of oil, commodities are not a critical price input to the world economy. The price of manufactures and services are much less prone to bubble behaviour, partly due to the speed with which ordinary factories and offices can be built, and also to the fungibility of most non-commodity goods.

The problems with oil have been largely mitigated in Europe with very high taxation of petroleum products. This makes the variable element much smaller, and also encourages the reduction of oil energy intensity in the economy.

There are two other commodities for which these effects are of great importance. The first is housing, which seems particularly prone to destructive bubbles, this is returned to later in section 6.3.

The other commodity is much more interesting, and is unique and of great importance to the analysis of the economy as whole.

This commodity is labour.

4. Minsky goes Austrian à la Goodwin – Macroeconomic Models

4.1 Macroeconomic Models - Background

So far in this paper three basic models have been developed using the tools of classical economics and the mathematics of the Lotka-Volterra and General Lotka-Volterra models (GLV's). The first set of models looked at the consumption side of the economy and the resulting distribution of income, the second series of models looked at the production side, and the resultant distribution of company sizes. The third, looking at commodities, introduced a very simple supply and demand based model.

Although the GLV has not previously been used significantly in economics, some non-linear modelling work has been carried out at a macroeconomic level by Kalecki, Kaldor, Desai and others. Most notably Goodwin used the Lotka-Volterra predator-prey system to model a qualitative cycle described by Marx (though true-blooded Marxists will be disappointed to learn that in these models the workers are modelled as predators; the capitalists are the prey). Keen has extended the Goodwin model to model a Minskian business cycle [Keen 1995].

Despite (or possibly because of) these heterodox Marxian origins there is significant evidence to suggest that these cycles exist in real economies. Barbosa-Filho & Taylor [Barbosa-Filho Taylor 2006] have carried out a detailed study of business cycles in the US. Harvie [Harvie 2000] has carried out a similar study for ten OECD countries. In both cases the evidence is qualitatively strongly suggestive of cyclical changes in labour share of return and employment that match the patterns predicted by Goodwin. In both case though there are significant difficulties in fitting the data quantitatively.

In addition to the work above there have also been substantial qualitative studies of business cycles in other schools of non-orthodox economics.

In the Austrian school, it has long been proposed that the build up of excess capital has been a fundamental cause of business cycles, with the blame for this generally put on government mishandling of credit availability.

In parallel with this Minsky, coming primarily from the post-Keynesian school, but also following the work of Fisher, has also studied the build up of economic cycles, though with the blame being primarily placed with speculation and the unsustainable endogenous creation of debt.

The Austrian and Minskian models share significant common features, the most obvious being their beliefs that booms and busts are natural features of economics. Another, unfortunately, is their shared disdain for formal mathematical modelling.

In the modelling that follow a very simple macroeconomic model is built, that combines the Lotka-Volterra approach of Goodwin with the basic ideas of the Austrian / Minskian business cycles.

The main ingredients for this model, including many simplifications, are already available in the proceeding models above.

4.2 Macroeconomic Models - Modelling

In this section a simple macroeconomic model is introduced, based on most of the same variables as the company and income models above.

The main assumptions of this model are as follows:

In line with classical economic theory, produced goods have real values, but market prices can vary from these values in short time periods due to insufficient or excess demand.

Consumption is a fixed proportion of consumers' perceived wealth, held in the form of paper assets, as in the income models above.

Companies have real capital which can produce a fixed amount of output, and needs a proportional supply of labour, as in all the models above.

The price of paper wealth assets is defined by the preceding revenue stream; as in the myopic companies model above.

The management in companies can be capital preserving, as in the companies model above.

There can be delays in installing capital as seen in the commodities model above.

The price of labour is non-linear according to supply. That is real wage rates go up when there is a shortage of labour, and go down when there is a surplus of labour. Labour is a genuinely scarce resource.

It should be noted that, unlike the Goodwin models, both population and technology are fixed.

Although this macroeconomic model will be more complex, as it has more variables, in other ways it will be simpler, as we will not look at individual consumers or companies, but look at the aggregated whole of supply and demand, in the same manner as the commodities model.

With the macro economic model there will also be a much stronger interest in the behaviour of the model as a function of time.

The big new assumption in this model is that labour costs vary with employment and unemployment.

It is assumed that labour costs vary as a concave function of employment, ie labour costs will increase as the employment ratio increases, and will increase at an increasing rate.

Figure 4.2.1 here

In this model I have used a simple square law function, shown in figure 4.2.1 above. This is not a particularly realistic function, more realistically it should be asymptotic to the vertical on the right hand side as there is a realistic maximum somewhere around 6000 hours per year. However this basic function is sufficient for the needs of the model.

It is also worth noting, this is not an inflation Phillips curve. This curve is a simple supply-price Phillips curve for labour in real terms. In this model, prices of goods and labour both go up and down, just as they did in the commodities model, but they move around stable long-term values. The analogy is with the cyclical price changes seen in a Victorian economy with a gold standard. There is no long-term monetary inflation. For a pithy study of the misinterpretation of the Phillips curve see Hussman [Hussman 2011].

Again, an element of marginality has been introduced. Over short to medium terms, the supply of labour is fixed, while demand can change. Because of this labour prices can change significantly through business cycles.

In these models, it is assumed that individuals always spend 40% of their income at all times, $\Omega = 0.4$.

It is possible that the consumption spending will exactly balance the amount of production capacity available in the companies, however this will not always be the case. It is also possible that there will be too much or too little capital available to match the consumption demand.

Looking firstly at the case of too little demand; if the 40% spending provides insufficient demand, then excess capital will be available and some of that capital will be unused. As a consequence of this there will also be a reduction in labour employed.

Also, following exactly the same logic as the companies models above, if companies create insufficient wealth to meet the payout targets set by their market capitalisation, then they will be obliged to convert some of their capital to wealth for payout.

Clearly in this model such a conversion of capital to returns is less realistic than the companies model. In the companies model capital was swapped for cash between the successful and unsuccessful companies.

In this macroeconomic model, all companies are shrinking in size at the same time. This would mean that first stocks of goods and then fixed capital would need to be converted into payouts. This would normally mean substantial losses on the value of the capital, especially the fixed capital. In this simple model, this problem is ignored, and capital is assumed to be converted into payments at par. This assumption is returned to in the discussion in section 4.4.

It is also possible that there may be insufficient capital available. In these circumstances it is assumed that consumption is still maintained at the full 40% of current wealth, even though insufficient capital available, and so insufficient goods are produced. In this case the consumption funds available for purchasing are simply divided amongst the goods that are available to be purchased, so increasing the nominal market price of the goods above their long-term natural prices. Consequently this results in short-term consumer price inflation.

It is implicitly assumed that consumers judge value by price and continue to spend a fixed proportion of their wealth, even though they actually receive less real value for that wealth.

When this happens super-profits are then earned by the corporate sector. If employment and so wage levels are low, then the income retained by the companies is converted into new capital to allow the production of more commodities. In this manner, super-profits are converted into new capital and new production until supply rises to meet the new demand, and the prices of consumer goods then drop back to their 'natural' values based on input costs. This is closely analogous to the commodities model.

It is important to note that, in the company models, the total amount of capital was fixed; however in this macroeconomic model, the amounts of capital and labour employed can vary, though labour is still needed in a fixed proportion to capital used.

In this macroeconomic model the capital and labour are still used in a fixed ratio to give a given output.

The amount of capital can vary freely, in line with the demand of goods from consumers.

The total supply of labour is fixed however, with the amount of the labour pool employed varying in fixed proportion to the amount of capital. Labour costs vary non-linearly with the amount of labour employed, which means that labour costs vary non-linearly with the amount of capital employed. So returns to labour and capital can vary.

It is still assumed that the proportion of labour required to capital is fixed over the whole period of time being modelled. This means that there is no technological progress, and also that it is not possible to substitute capital for labour.

Each iteration of the model operates as follows:

The expected returns are defined as 10% of the current market capitalisation.

The consumption, and so the payments made for consumer goods are defined as 40% of total wealth.

If these payments are less than 20% of the available capital, then the amount of goods produced is equal to the value of the consumer payments.

If the payments for consumer goods are greater than 20% Of the available capital, then the goods produced are equal to 20% of the total capital, ie, the maximum production possible is 0.2 times the capital K that is in existence.

The income accruing to labour is calculated, according to the amount of capital used, and so the proportion of labour employed, according to the square law.

The surplus revenue that the company generates is then the value of the consumer payments received, less the earnings income paid out.

The new value of the total real capital is then the old capital, plus the payments received for goods, less the labour earnings paid out, less the actual returns paid out.

Finally, the consumers receive their dividends from the companies and revalue the market capitalisation according to the actual returns paid out.

At this point, the cycle starts again.

As in the companies model, the actual returns paid to the owners (shareholders) that is the payout ratios can depend on whether the surplus revenue generated is greater than the expected returns or less than the expected returns.

For example in model 1D the actual returns paid out are always 70% of the revenue generated. However in models 4A to 4C the actual returns paid out are equal to the real returns produced.

It is noted that these payout factors are different to the ones in the companies model above, clearly these models are preliminary and in need of future calibration to real economies.

As with the commodities model, it is also possible to put a variable lag in to model the time it takes to install capital.

A further important ingredient in this model is the existence of a 'cash balance' for the householders. This is needed in their role as owners of capital and spenders of money. This cash balance can result as an imbalance of spending outgoing against income received as a consequence of these being dynamic models. If the cash balance is positive then this represents spare cash in the bank. The householders have received more in wages and dividends than they have spent in consumption.

If the cash balance is negative, then this represents a debt to the bank, due to the consumers spending more than they earn.

In the notes following, the cash balance is referred to as H to differentiate it from the capital owned which is now labelled Q. The consumers are assumed to be sensible, so they carry out their consumption based on their total wealth W which is the sum of Q and H, so:

$$C = W\Omega \qquad (4.2a) \qquad \text{or:}$$

$$C = (Q+H)\Omega \qquad (4.2b)$$

So, for example, if H is negative because the consumers have net debt, then consumption is reduced below that judged by the size of Q only.

This model was carried out in Excel, those who wish to go through the maths in detail can past the model into Excel from appendix 14.9.

4.3 Macroeconomic Models - Results

As expected this model can show different sorts of behaviour, some examples are given below:

Model 4A is the base model, with all the numbers designed to be nice and round. This model has payout ratios of 1 for both the upside and downside. It also allows capital to be added instantly, without any lags. It can be seen from figure 4.3.1 that the output is very stable, and so very dull.

Figure 4.3.1 here

Model 4B, shown in figure 4.3.2 has exactly the same parameters as model 4A, the only difference is that the initial values were different.

Figure 4.3.2 here

This shows just how stable this model is, with the model quickly settling down to equilibrium values. Though even in this stable model it is notable that model 4B needs to go through a number of fluctuations before it arrives at stability (cf figure 1.2.1.4).

But there is a more important difference to note between model 4A and 4B. The parameters of the model are exactly the same, but the equilibrium points are very different. Model 4A started with real capital of 100 units, and settled to an equilibrium at 100 units. Model 4B started with real capital of 400 units, and settled to an equilibrium at about 184 units.

As a consequence, total capital employed at equilibrium in model 4B is much higher than that in model 4A, and more importantly, total employment is higher in model 4B than model 4A. Also the ratio of returns to labour to returns to capital is significantly higher in model 4A.

This is Keynes writ large.

Unlike static equilibria, dynamic equilibria can have multiple points of stability. The point of equilibrium that is reached depends on the parameters of the model, but also on the initial conditions. Different initial conditions can give different equilibria even with the same parameters. Once it has reached its equilibrium, the model can stay at that point indefinitely. To change the equilibrium an exogenous force is needed. The model will not rebalance itself to a particular point; a point such as full employment for example. Mass unemployment can continue indefinitely without positive external action.

Model 4C is the most interesting, and most realistic, model.

In this model a time lag has been introduced between capital being purchased and being brought into use. This is identical to the way capital is installed in the commodities models in section 3. Note that the payout ratios are still at unity.

Figure 4.3.3 shows the long term behaviour of the model.

Figure 4.3.3 here

As can be seen the model shows regular cycles of capital being created and destroyed. Again it is important to note that this is a chaotic model, not a stochastic one. There is no stochasticity in this model. All fluctuations in the model are created endogenously from the Lotka-Volterra like differential equations in the model.

Figure 4.3.4 shows the detail of couple of cycles.

Figure 4.3.4 here

These are real live Minskian / Austrian business cycles. But with one big exception.

It can be seen that real capital K builds up in advance of the total wealth (in this simple model paper wealth; capitalisation is constant), this build up of capital is unsustainable, and so leads to a fall in real capital. Interestingly, although debt (negative cash wealth) is present, this is a lagging variable. In this model debt creation is fuelled by capital growth, not the other way round. The chaotic, bubbly behaviour is not caused by excess credit, it caused by the basic pricing system of capitalism.

Model 4D, shown in figure 4.3.5 below has no lag in the installation of capital. Instead this model has payout ratios of 0.7 on both the upside and the downside.

Figure 4.3.5 here

It is believed that this is a less realistic model, however it does demonstrate how highly chaotic behaviour can be generated in even a very simple model.

Finally model 4E is shown in figure 4.3.6 below. This has a just a small lag of 1 unit for the installation of capital, and payout ratios of 0.8.

Figure 4.3.6 here

Interestingly, it seems that similar results can be achieved without a lag. If both interest rates and payout factors are reduced, an explosive result is also seen.

As can be seen these minor changes in the model are sufficient to create explosive behaviour. This is a true bubble, similar to that of Japan in the 1980s, or the US in the 1920s or in the last decade. Again the cash wealth (debt) is a lagging indicator. It is possible to create explosive bubbles just from the basic pricing system of capitalism.

There is finally one important thing worth noting about the models. The value of the Bowley ratios, β, for the first four models were as follows:

Figure 4.3.7	β
Model 4A	0.75 (exactly)
Model 4B	0.92
Model 4C	0.78
Model 4D	0.85

The Bowley ratio is the ratio of returns to labour to the total returns. The values for models 4C and 4D are averages; the Bowley ratio varies wildly over the course of a cycle in these models.

The numbers above are close to the 'stylised facts' for the Bowley ratio, and are of considerable importance. This is returned to at length in section 4.5 onwards.

4.4 Macroeconomic Models - Discussion

As with the previous models, the results above show that a simple combination of classical economics and a dynamic analysis gives interesting results that mirror real economies.

The author expected that such a model would be easily capable of producing boom and bust business cycles, and this is discussed in some detail in this section.

The production of a suitable Bowley ratio was a surprise, though a pleasant and very important one. This is discussed further in sections 4.5 to 4.7.

Leaving aside the Bowley ratio, the most interesting result of this model is that the booms and busts are generated internally via an endogenous spiral of creation of wealth. In the model real capital is installed, which generates more paper wealth, which generates more consumption, so feeding into another cycle of wealth creation. The upswing is finally constrained by rising wages making the capital unproductive.

This then generates a downswing of declining wealth, consumption and wages.

This is the normal cycle of capitalism as described by Minsky and the Austrians. Booms and busts are endogenous. Free markets are not inherently stable.

Again, as with the income and company models it noticeable that there are many things that are standard elements of neo-classical or Keynesian economic theory which are simply not needed to produce this macroeconomic model, these include:

- Economic growth
- Population changes
- Technology changes
- Productivity growth
- Investment
- Saving
- Accelerators
- Multipliers
- Shocks (exogenous or endogenous)
- Stochasticity (in any form)
- Different initial endowments (of capital or wealth)
- Utility functions
- Production functions

It has been noted that marginality has worked it's way into the modelling in the form of the pricing curve for labour, this is a reasonable argument, as labour is a commodity that is truly unchangeable in it's supply. Although marginality might be a mathematically useful way to address this, the history of entropy and information suggests there may be better ways to address this. More importantly, the results of the model show that the detailed form of curve are completely irrelevant to the model. The curve simply needs to be concave, to ensure that labour costs eventually choke the growth. Within reason, any concave curve will do this. So the actual detail of the calculations of marginality are irrelevant and do not have any influence on the long-term equilibrium, the cycle frequency or the distributions of wealth and income. This is discussed further in section 4.7 below.

It is also worth considering the 'efficiency' of the economy in this model. This model again creates chaotic behaviour endogenously. There is no stochastic noise in this model. It is politely suggested by the author that a system that endogenously creates booms and busts, with short term creation of excess capital, and far worse; short term destruction of the very same capital, may not, in fact, be allocating capital in a particularly efficient manner.

Investment and saving has been deliberately ignored in this model, as it has been in all previous models. This is because, as the data given from Miles and Scott in section 1.3 show, saving and investment is a minor part of the economic cycle. The core driver of business investment is the availability of cash streams. When firms have more money coming in from revenue than they need to pay out as dividends they use it for investment. When they don't have spare money they don't invest. The mechanics of saving and investment are a side-show and diversion from the base model of macroeconomics.

Similarly the general public is assumed to simply consume a fixed proportion of their wealth. In the real world it seems much more reasonable to assume that people who gain more wealth will divert a greater portion of this to saving, particularly in an environment, as here, in which companies appear to be showing increasing profits on their capital. I believe this is a simplification rather than a flaw. The point of the model is that endogenous business cycles arise at the heart of the system of pricing financial assets. Allowing transfers of excess savings in booms to investment rather than consumption would clearly exacerbate these booms. Indeed it is possible that the effects of saving and investment multipliers might be significant, but that is not the issue, the issue is that saving and investment is a multiplier rather than the root cause of the instability.

In identical fashion to the companies models above, expectations and behaviouralism do enter into the model in two different ways, firstly with regard to the pricing of stocks, and secondly with regard to the retention of capital within companies.

Again these are obvious forms of behaviour and are supported by economic research as discussed in section 2.1 above.

It can be seen from the model results that economies can behave very differently according to relatively small changes in input parameters.

This is because a system like this can show different regions of behaviour, a general property of Lotka-Volterra and other similar non-linear differential equation models.

Depending on the settings of the variables in the model, there can be three different cases for the outputs.

Firstly, the outputs can be completely stable, quickly going to constant values, this was seen in models 4A and 4B.

Secondly, the outputs can be locally unstable with values constantly varying, but hunting round within a prescribed range of values, this is similar to the lynx and hares Lotka-Volterra model discussed back in section 1.2. This appears to be the way that most normal economies behave. This effect can be caused by the behaviour of capital, either by deliberate hoarding of capital by company managers, or by the time it takes for capital to be installed. The cyclical rise and fall of capital in business cycles is analogous to the cyclical rise and fall of biomass in a biological Lotka-Volterra system. Just as the hares and lynx respond rationally to the available grass, so business investors and speculators react rationally to the opportunities in the economy.

Finally, the outputs can be explosive, moving quickly off to $\pm$ infinity.

In models 4A and 4B these values were 'fixed' to ensure a stable model, in 4C and 4D the parameters were fixed to give a quasi-stable cyclical model, in model 4E they were changed to get explosive models. In the real world it appears that economies operate largely in zones 4C/D, with occasional excursions into zone 4E.

Model 4E suggests that if both interest rates and payout rates are too low then the company sector is too profitable and capital expands exponentially before finally wrecking the whole economy in a glut of capital, see figure 4.3.6 above.

It seems plausible to argue that this reflects what actually happened in the US during the late nineteen twenties, and Japan in the late eighties. Following each of these bubbles the respective economies failed to return to a self-regulating pattern of booms and busts, but appear to have been moved to new equilibrium with much less productive economic patterns. So the economies moved very quickly from a 4E to a poorly performing 4A/B.

It is the belief of the author that keeping interest rates and payout ratios too low allows a second common form of macroeconomic suicide. (The first form of economic suicide is introduced in section 4.6. Both forms of suicide are discussed in more detail in section 4.10.)

A very important point to emphasise in the models above is the absolute lack of stochasticity. While there is certainly a significant element of stochasticity in real markets, the macroeconomic model above contains no stochasticity. The model is not stochastic, it is merely chaotic. Chaotic models like this are common in physics, astronomy, biology, engineering, and in fact all of the sciences other than economics, where determinism has hunkered down for a very effective last stand. The failure of these models to penetrate into mainstream economics, given the obvious turbulence of stock, commodity, housing and other financial markets, is puzzling.

This endogeneity of chaos in business cycles is of profound importance. Standard economic theory, whether Keynesian lack of demand or the impacts of technology in 'Real Business Cycle theory', never mind neoclassical economics, seems incapable of believing that chaotic short term behaviour can be anything but externally driven.

Exogenous drivers are simply not needed for quasi-cyclical, or explosive chaotic behaviour; all that is needed is the use of the correct modern mathematics, where 'modern' means post 1890. This mathematics, and chaotic systems in general, is discussed in section 6 below.

As discussed above, Lotka-Volterra models have been used in Marxian analysis by Goodwin and others, though the models can be somewhat complex.

The models presented above seem more efficacious than the Goodwin type Lotka-Volterra models, as they don't need:

- population change
- growth in labour force
- technology change
- productivity growth
- inflation (long-term)
- accelerators

all of which are used as standard in the Goodwin and descendant models.

A central problem in the thinking of Goodwin and the researchers that followed Goodwin is in the idea of growth. It appears to have been assumed that to model short-term cycles of growth and decline it was necessary to include long-term economic growth rates. So these models include growth in the labour force, productivity, money supply, etc.

This is a bit like trying to model waves on the ocean's surface by including things that cause changes in sea level, such as the tidal effects of the sun and moon, evaporation, precipitation, glacier melt rates, etc.

This brings in a lot of irrelevancies into the basic model, and make it very hard to build the basic model.

Even without any of the things listed above, natural cycles can occur that build up too much capital.

That is not to argue against the secondary importance of any of the above factors, especially in long-term economic cycles.

Going back to the evidence of Harvie [Harvie 2000] and Barbosa-Filho & Taylor [Barbosa-Filho Taylor 2006], the cycles for the mainland European countries appear to be long term, on a decadal scale; which would suggest a strong role for technology change and productivity growth (though very little for population change). However the cycles for the US and UK appear to show much faster oscillations; of only two to three years. Intuitively it is difficult to see how technology change could operate significantly on such short timescales, and this is more suggestive of the operation of the normal business cycle modelled above.

Indeed the simple model proposed above may be more appropriate for modelling the regular short period cycles of booms and crashes seen in Victorian times.

The important thing to note is that the basic instability in financial markets is much deeper than that proposed by Goodwin. Goodwin style feedbacks may exaggerate this basic cycle, or add longer super-cycles, however in this regard it appears that the basic insight of Minsky and the Austrians with regard to the essential instability of capitalism was correct.

However, although I believe this basic Minskian/Austrian insight is valuable, it is also notable that to build model 4A to E, and create dramatic business cycles, you don't actually need any of the following:

- governments
- fiat money
- fractional reserve banking
- speculators
- Ponzi finance
- debt deflation

or other common elements of the Austrian school or the work of Fisher and Minsky.

Debt, in the form of a negative cash balance, certainly does appear in the cyclical and explosive models. But models 4C and 4D show that the debt follows the cyclical instability of capital rather than the other way round.

I would not wish to understate the importance of debt in exacerbating business cycles, indeed the role of debt appears to be very interesting and important, and is discussed further in 4.6 below. However debt itself is not the prime cause of the business cycles.

Again, it is not suggested that any of the factors listed above are unimportant, however it appears that all the other factors are just potential magnifiers of an underlying inherent instability

The instability is very basic, and, in the short term at least, perfectly rational. The instability arises, as Minsky noted, from the fundamental fact that paper prices of assets are based on projected future cash flows, not on costs of production. This is Minsky's crucial insight, of much greater importance than his analysis of the debt cycle.

This is the same assumption originally proposed in the companies model in section 2.2 above.

This instability naturally produces a growing cycle of apparent wealth, which is turned into excess capital as predicted by Hayek [Hayek 1931] in Austrian business cycles. But contrary to the Austrians, and in line with research data [Kydland & Prescott 1990], the liquidity or excess paper wealth is initially generated within the valuation system of capitalism, not by lax government policy.

Creation of liquidity and monetary growth are endogenous to the basic pricing mechanisms of the finance system. Endogenous creation of financial wealth then feeds back into the creation of more real capital, so creating more financial wealth.

This endogenous creation of financial wealth then gives apparently secure paper assets against which debt can be secured, and of course this debt allows yet more capital creation.

Clearly, if the underlying system is unstable, with endogenous liquidity production à la Minsky; then other factors such as excessive debt, speculation, fractional reserve banking and inappropriate central bank intervention policies will all magnify the size and damage of the underlying cycles. But it is not excessive debt, speculation, fractional reserve banking or poor central bank policy that causes the boom and bust cycles. The cycles are caused by the basic pricing system of capitalism.

Governments may of course fail to calm the markets by extracting liquidity in a timely manner, but it is scarcely the fault of governments that most investors are momentum chasers rather than fundamental analysts.

Just as central banks are expected to control changes in the money supply caused by fractional reserve banking, it seems appropriate that they also need to control liquidity growth caused by Minskian asset pricing. This is discussed in more depth in section 8.2.1 on liquidity below.

As noted previously, Minsky, although a follower of Fisher and Keynes, shared the Austrians' disdain for mathematics. It is the author's belief that bringing in a dynamic mathematical approach, on the lines of Lotka-Volterra modelling, to Minskian and Austrian ideas might not only give more weight to both these approaches, but also show them to be very comfortable bedfellows.

Essentially the company, commodity and macroeconomic models are all simple composites of ideas from Minsky and the Austrian school, though my producing them in this way happened more by accident than design. The models have Minsky's basic split between 'normal' assets such as goods and services that are priced on a mark-up basis, and financial assets which are priced on the basis of expected future cash flow. Following Minsky, and ultimately Keynes, the expectations of future flows are simplistic projections of present flows [Keen 1995].

Unlike Minsky the models use simple known behaviour of capital to explain the source of instability. In the companies model this was company managers hoarding incoming spare cash, and using it to build more capital. In the commodity model the instability was caused by the time actually taken to build and install new capital. In the macroeconomic models, either or both of these factors could cause instability. In this sense the models follow Austrian ideas. This has the advantage over the Minsky models that you don't need a complex financial system; speculators, Ponzi finance, etc, to form the instability. You can get the instability in pretty much any system where financial assets can be overvalued; this can be Industrial Victorian Britain with its savage business cycles, or even the Roman Empire (see 4.10 below).

The critical insight of Minsky, in contrast to the Austrians, and seen in these models is that liquidity and new credit are generated endogenously in even the most basic of financial systems. You don't need governments to create excess credit, though certainly they can make things worse. In fact, faced with endogenous credit creation, you do need governments to actively remove credit and liquidity when financial assets become overpriced.

In defining this macroeconomic model, a number of assumptions were made. I would like to briefly review these here:

Note that the assumption of conversion of capital to equity at par in a downturn does not undermine the arguments. The losses incurred in a fire sale of assets to meet investor demands would simply exaggerate the viciousness of the cycles downwards.

It was assumed that the ratio of capital to labour is fixed over the time of the business cycle, and that it is not possible to substitute capital for labour. There are two parts to discuss with this assumption. Firstly, in the short term, going into a boom, replacing labour with capital would simply allow further excess capital to be installed before wage inflation would kick in, so making the booms even larger. The resultant larger overhang of capital would then make the following slump more severe. So relaxing this assumption would simply make the business cycles worse. More importantly, the model shows that, in the long term, at the level of the economy as a whole, there is in fact a fixed ratio of capital to labour at any given set of market conditions. So it is not actually possible to substitute one for the other. Much more on this in sections 4.6 to 4.8 below.

Note that allowing the market interest rate to float, say by making it the moving average of real returns over the previous few periods, would also have a large magnifying effect. As more capital was employed, overall interest rates would go down, making previously unprofitable capital investment profitable. Again this would encourage further excess capital creation in the booms.

Finally I would like to return to a major assumption of the companies model in section 2.2. In this model capital was deliberately, and artificially, renormalised in each of the model iterations to keep a constant value of K. As I hope is now clear, this was a necessary fix in the company model to prevent the introduction of severe cycling in the model output.

Directly comparing my own macroeconomic models with those of Wright is not straightforward. My own models include a financial sector which is clearly more realistic, as Wright acknowledges in the Social Architecture of Capitalism [Wright 2005], where he notes that conflation of capital concentration with firm ownership may distort modelling results. So clearly Wright's models can not show cycles of debt build up and draw down.

Despite this Wright's models do show recurrent booms and recessions, with much more complex behaviour than my own. Although Wright's business cycles are debt free, he models individual companies/owners, where my own model models the business sector as a whole. As a result recessions in Wright's models are of differing length and are quasi-periodic, this is clearly superior to my own models.

Wright's models are also superior to my own in that they include for unemployment. My models just measure total over-employment and under-employment against a nominal full employment.

Despite these substantial differences both Wright's and my own models produce cyclical endogenous business cycles from simple models based on statistical mechanics and classical economics.

4.5 A Present for Philip Mirowski? – A Bowley-Polonius Macroeconomic Model

"I mean the stability of the proportion of national dividend accruing to labour, irrespective apparently of the level of output as a whole and of the phase of the trade cycle. This is one of the most surprising, yet best-established, facts in the whole range of economic statistics……… Indeed…the result remains a bit of a miracle." [Keynes 1939]

"…no hypothesis as regards the forces determining distributive shares could be intellectually satisfying unless it succeeds in accounting for the relative stability of these shares in the advanced capitalist economies over the last 100 years or so, despite the phenomenal changes in the techniques of production, in the accumulation of capital relative to labour and in real income per head." [Kaldor 1956]

"FUTURE ISSUES - Theory

1. Is there a deep explanation for the coefficient of 1/3 capital share in the aggregate capital stock? This constancy is one of the most remarkable regularities in economics. A fully satisfactory explanation should not only generate the constant capital share, but some reason why the exponent should be 1/3 (see Jones 2005 for an interesting paper that generates a Cobb-Douglas production function, but does not predict the 1/3 exponent). With such an answer, we might understand more deeply what causes technological progress and the foundations of economic growth." [Gabaix 2009]

Whenever economists hit a bad patch, it is inevitable that outsiders will begin to sneer how it is not a science and proceed to prognosticate how "real science" would make short work of the crisis. This is such a tired Western obsession that it is astounding that it has not occurred to critics that such proleptic emotions must have occurred before, and are thus themselves a part of a chronic debility in our understanding of economic history. As I have shown elsewhere in *detail, neoclassical economics was born of a crude attempt to directly imitate physics in the 1870s, and American orthodoxy was the product of further waves of physicists cascading over into economics in the Great Depression and WWII….*

…Actually, it is understood among the cognoscenti that physicists have again been tumbling head over heels into economics since the 1980s, as their own field experienced severe contraction at the cessation of the Cold War. And where did most of them end up? Why, in the banks, of course, inventing all those ultra-complex models for estimating and parceling out risk. Some troubled to attain some formal degree in economics, while others felt it superfluous to their career paths. In any event, the exodus of natural scientists into economics was one of the (minor) determinants of the crisis itself—without "rocket scientists" and "quants," it would have been a lot harder for banks and hedge funds to bamboozle all those gullible investors. So much for the bracing regimen of a background in the natural sciences.

If anything, responses to critics that tended to pontificate upon the nature of "science" were even more baffling than the original calls for deliverance through natural science in the first place. Economists were poorly placed to lecture others on the scientific method; although they trafficked in mathematical models, statistics, and even "experimentation," their practices and standards barely resembled those found in physics or biology or astronomy. Fundamental constants or structural invariants were notable by their absence. Indeed, one would be hard pressed to find an experimental refutation of any orthodox neoclassical proposition in the last four decades, so appeals to Popper were more ceremonial than substantial. Of course, sometimes the natural sciences encountered something commensurable to a crisis in their own fields of endeavor—think of dark matter and dark energy, or the quantum breakdown of causality in the 1920s—but they didn't respond by evasive manoeuvres and suppressing its consideration, as did the economists.

In retrospect, science will be seen to have been a bit of a red herring in coming to terms with the current crisis. In the heat of battle, economists purported to be defending "science," when in fact, they were only defending themselves and their minions. [Mirowski 2010]

As a physicist myself, I am somewhat embarrassed to admit that physicists as a class stand guilty as charged when accused of unnecessarily increasing the complexity and opacity of finance. This is the more embarrassing as the behaviour is so far from the norm in physics, where careful investigation and gaining of understanding is the general aim, and true kudos is gained by discovering neat and beautiful solutions to seemingly complex and insoluble problems. The entry of quants into finance seems not only to have been marked by a joy in the deliberately complex, but also a wilful desire to avoid any understanding of what is really happening in an economic or financial system. As previously noted, physicists seem very comfortable in using wealth and income interchangeably, some even conflate these two concepts with money. From my own conversations, I am led to doubt whether a majority of physicists working in finance could successfully define the difference between a real and a financial asset.

As a penitence, on behalf of a profession behaving badly; I had hoped in this section to present to Philip Mirowski the explanation of a basic 'constant' in economics. Sadly for me, the constant turns out not to be constant at all but merely a humble ratio; an indicator of an underlying equilibrium. Unfortunately it cannot be described as either 'fundamental' or 'invariant'.

On the bright side this at least allows for changing of the 'constant', and indeed it is one of the aims of later sections to change this 'constant' to the benefit of the population in general.

Even more worryingly this constant may simply be seen by many as a trivial accounting identity, a red herring at best.

I do not believe this is the case and, however humble this ratio may be, I believe it is the first 'constant' to be explained in economics, and as such is worthy of note.

The constant in question is the ratio of earnings received by labour to those received by labour and capital, the Bowley ratio β that was first introduced in section 1.3 above. Before looking at the derivation of the Bowley ratio, it is worth considering this 'constant' in more detail.

For most mature economies the constant varies between about two-thirds and three-quarters and can be very stable, as discussed in section 1.3 above. Young gives a good discussion of the national income shares in the US, while Gollin gives a very thorough survey of income shares in more than forty countries [Young 2010, Gollin 2002].

In emerging economies β can be much lower, as low as 0.5. Currently, and exceptionally, in China it may be as low as 42% [Bai et al 2006, Subramanian 2008]. Arthur Lewis [Lewis 1954] has explained this as being due to wages being artificially depressed by the reserve of subsistence workers simultaneously with the wealthy being able to save more due to low living costs caused by low wage rates.

Once economies absorb this spare rural labour, and pass their 'Lewisian turning point', then the ratio of returns to labour to total income stabilises and moves only slightly. In the UK, the first country in the world to absorb its rural labour force, the ratio has been fairly stable for a century and a half.

The thing about this stability is that the more you consider it, the more bizarre it seems.

In the last 150 years Britain has changed from a nation of factories powered by steam engines to a modern service economy. The amount of capital currently installed in the UK is many times greater than that of 150 years ago, labour intensive industry has all but disappeared. Wealth levels have changed incredibly. In the 1850s gdp in the UK was comparable to current gdp in Indonesia or the Philippines, however life expectancy in the UK in the 1850s was roughly half that of Indonesia or the Philippines today [gapminder].

It is quite extraordinary that the Bowley ratio has remained roughly constant throughout this period.

In fact it is counter-intuitive.

For somebody in Victorian Britain, as in modern day Indonesia, the majority of income would have been spent on food and basic housing, with little left over for anything else, most money is paid to other people carrying out labouring duties.

As incomes rise it would naturally be expected that more money would be spent on manufactures and property, and that more spare cash would be available for investing in capital of one form or another, so increasing the returns to capital. Also, as wages rise it would also seem sensible for capital to substitute for labour, and again for returns to capital to increase at the expense of labour. In the long-term total factor productivity should increase, reducing the returns to labour and increasing those to capital.

Indeed futurologists have been predicting for most of a century that as capital gets more efficient and productive the need for labour should slowly decline to nothing. To date these predictions have been conspicuously wrong. Working weeks have barely declined in the last forty years, huge numbers of women have entered the labour markets and people continue to complain of the problems of the work/life balance. Indeed at the time of writing this section France is currently paralysed by strikes trying to prevent an increase in retirement ages.

In the long run it seems logical that mechanisation and the increasing use of capital would result in the Bowley ratio slowly moving towards zero.

In fact if you analyse the data on a sectoral basis, this is exactly what is happening. Young [Young 2010] shows clearly that for agriculture and manufacturing, returns to labour have declined significantly while returns to capital have increased. In the US returns to labour in agriculture have dropped from nearly 0.8 of total income in 1958 to less than 0.6 by 1996. In manufacturing, the change has been from 0.75 to two-thirds.

This has happened because labour has been slowly displaced by machines in these industries. The fascinating thing is that despite the changes in the Bowley ratios for these two (very large) sectors, the national value of the Bowley ratio has stayed near constant between 0.69 and 0.66 using the same measures.

The reason for this is that the labour intensive service sector has grown dramatically in size through the same period, and this has kept the national balance of returns to labour and capital very nearly constant.

In the discussions that follow it is hoped that these puzzles will be explained.

As shown in section 4.3 above, the output from a fairly randomly chosen model 4A produced an output with a Bowley Ratio, of waged earnings to total earnings, of exactly 0.75 with zero debt (It is to be noted, that Wright found similar results with β equal to 0.6 and 0.55 in his two papers). This was the subject of further modelling.

A first problem with the models used in section 4.3 above is that they have too many degrees of freedom. Depending on the parameters and the starting values of a model run, different zones of stability can be encountered, and even if the model is restricted to options that end in stable, stationary outputs, different end points can be reached with the same parameters, but different starting positions.

A second problem is the role of the 'cash balance', H, which can either be a positive surplus or a negative debt.

In many of the models the stable output can have very large positive or negative cash balances, with an order of size of the capital wealth Q.

As is often the way with debt, an item that was used as a minor temporary convenience ends up taking on a major unlooked for negative role.

Having been introduced as a simple method of ensuring that the sums add up; the role of this cash balance is not clear, and it is not obvious that it is a meaningful item. There are problems as to exactly who or what this money is borrowed from / lent to, and also why interest is not charged on the lending or borrowing.

Firstly, to remove these problems, the models were rerun in Excel, deliberately choosing parameters that stabilised into stationary outputs.

A second condition used was that the payout ratios, both positive and negative, were set to 1.0. This makes for an immediate simplification of the model, as company payouts are just to the market expectations and make no reference to the profits produced by the companies. In this model payout ratios are not necessary, because although the total capital can increase and decrease, other mathematical limitations prevent the capital from shrinking to zero, at least in the stationary and periodic zones.

Thirdly, using 'solver', the range of stationary outputs was then restrained to the single solution that satisfied the requirement that there be no net borrowing or lending, ie the cash balance was always constrained to zero. This gives the single 'Bowley-Polonius' equilibrium point. With net borrowing and lending fixed to zero, the philosophical problem of what exactly the cash balance is becomes irrelevant.

By changing the parameters of the model systematically some very interesting results arose.

The first interesting thing was the role of the pricing of labour. As discussed in section 4.2 above, this model assumes that labour can be a scarce supply, and that the price of labour depends on the amount required.

As such the concept of marginality has introduced it's way into the modelling in the form of a pricing curve for labour, this is a reasonable argument, as labour is a commodity that is truly unchangeable in it's supply.

However, investigating the model shows that the actual form of the curve is not relevant to the model. If you change the parameters of the labour curve, then the model values change, with an offsetting increase or decrease in the cash balance. But if you reoptimise the model and force the cash balance back to zero, then the model returns to an equilibrium point with exactly the same value for the Bowley ratio. This is looked at again in section 4.7.

Within reason, the parameters of the labour supply curve are simply not relevant to the ratio of wages to profits. The curve simply needs to be concave, to ensure that labour costs eventually choke the growth of the economy with higher costs. Any reasonable concave curve will do this. So the actual detailed calculations of marginality are utterly irrelevant and do not have any influence on the long-term equilibrium.

('Within reason' means that there are some labour curves that prevent the model coming to an appropriate equilibrium; that is they don't allow an equilibrium at zero cash balance. But as long as the curve allows an equilibrium, the parameters of the curve do not effect the location of the equilibrium).

The second interesting thing is that, at the B-P equilibrium, the Bowley ratio is influenced by only two things; the consumption rate and the profit rate.

Moreover, the ratio is given by the very simple form as follows:

$$
\begin{aligned}
\beta &= \text{Bowley Ratio} \\
&= \frac{\text{waged income}}{\text{total income}} \\
&= \frac{\Omega - r}{\Omega} \\
&= 1 - (r/\Omega) \qquad (4.5a)
\end{aligned}
$$

It is straightforward to check equation (4.5a) against reality. A suitable long-term profit rate could be anywhere between long-term interest rates and long-term real stock-market returns. Long-term real interest rates are generally in the region of 2% to 5% [Homer & Sylla 1996, Measuring Worth], see also figure 4.5.1 below. Long-term stock-market returns appear to be in the region of 7% to 8% [Campbell 2003, Ward 2008] see also figure 4.5.2 below.

Consumption is typically about 60% of gdp [Miles & Scott 2002, section 2.2, fig 2.3]. While non-residential capital stock is typically 2.5 to 3 times gdp [Miles & Scott 2002, section 5.1 & 14.1]. Taken together this would give Ω, the consumption rate as a proportion of capital a range of about 0.2 to 0.25.

Substituting into equation (4.5a) this then gives a possible range of values for the Bowley ratio of between 0.60 and 0.92

Clearly this range is a little on the high side when compared with the 'stylised facts' of observed Bowley Ratios in the real world varying between the values of 0.5-0.75.

We are however in the right ballpark. (The figures also confirms the common sense notion that stock-market returns are more appropriate than interest rates for 'r'.)

As discussed above, intuitively it is not obvious why Bowley's law holds and the ratios of returns to capital are not much higher than the returns to labour. Using the basic ideas of classical economics we would expect the returns to have increased significantly as machines have got steadily more productive over the last two hundred years. Neoclassical ideas of utility and marginality have no theory to explain this.

What equation (4.5a) says clearly is that Bowley's ratio will always be less than one, and given that rates of return are generally much lower than consumption rates, the value will be closer to one than zero. This agrees in general with the stylised facts, if not in detail.

In section 4.6 below possible reasons for the mismatch between the values produced in the model and the real world models are discussed. These reasons are speculative, so before moving on to this I would first like to discuss the equation (4.5a) and its consequences in a little more detail.

Firstly it should be noted that this equation was discovered by experimenting with the parameters of the model. The results from the simulations give results that match the formula above to multiple decimal places.

With a little playing it turns out that it is in fact quite straightforward to derive formula (4.5a) from first principles.

Firstly, when the model is at equilibrium, all values of flows and stocks are constant (in this part of the modelling, only models giving stable time outputs were used, the models suggest that the periodic models move around this point on average, as would be expected in a Lotka-Volterra model).

At this equilibrium point, if the total capital Q is to be constant, then the total income must equal the total outgoings, so the algebra works as follows (note that for simplicity the summations have been dropped, all variables are assumed to be summed over the whole economy):

$$\text{Consumption} = \text{Income}$$

$$\begin{aligned} C &= Y \\ &= e + \pi \qquad (4.5b) \end{aligned}$$

Here, at the Bowley-Polonius equilibrium, H = 0 and W = Q.
Also, the consumption ratio Ω is defined by:

$$\Omega = \frac{C}{Q} \qquad (4.5c)$$

Trivially, the profit rate is defined by:

$$\frac{\pi}{Q} = r \qquad (4.5d)$$

If we multiply equation (4.5b) by equation (4.5d), then we get:

$$\frac{\pi}{Q} C = rY \qquad (4.5e)$$

Substituting from (4.5c) into the left hand side gives:

$$\pi \Omega = rY \qquad (4.5f)$$

Rearranging gives:

$$\frac{\pi}{Y} = \frac{r}{\Omega} \qquad (4.5g)$$

substituting from (1.3u) gives the profit ratio:

$$\rho = \frac{r}{\Omega} \qquad (4.5h)$$

Subtracting both sides from unity gives:

$$1 - \rho = 1 - \frac{r}{\Omega} \qquad (4.5j)$$

or, substituting from (1.3v):

$$\beta = \text{Bowley ratio}$$
$$= 1 - \frac{r}{\Omega} \qquad (4.5k)$$

The base equation here is (4.5h) which is the ratio of returns from capital, to total returns. This equation looks suspiciously like an equation of state, discussion of which will be postponed to section 4.7. Whether equations (4.5h) and (4.5k) are sufficiently 'fundamental' to satisfy Phillip Mirowski remains to be seen; I would ask judgement to be reserved until the end of section 4.7.

Multiplying consumption by interest rates isn't an 'obvious' thing to do, and clearly I discovered this derivation by reverse engineering my model output.

At this point, more observant readers may have noticed something familiar about equation (4.5k). Equation (4.5k) gives:

$$\beta = 1 - \frac{r}{\Omega} \qquad (4.5k)$$

while back in section 1.3 equations (1.3v) and (1.3w) defined the Bowley ratio as:

$$\beta = 1 - \frac{r}{\Gamma} \qquad (4.5l)$$

This is made simpler by looking at the profit ratio ρ, then (4.5h) and (1.3w) give:

$$\rho = \frac{r}{\Omega} = \frac{r}{\Gamma} \qquad (4.5m)$$

which clearly means:

$$\Omega = \Gamma \qquad (4.5n)$$

from the definitions of Ω and Γ it then follows that:

$$\frac{C}{W} = \frac{Y}{W} \qquad (4.5o)$$

Where C is the consumption and Y is the total income from wage earnings and profits/dividends, etc. From which trivially we arrive at:

$$C = Y \qquad (4.5p)$$

which we have seen a very long time ago as (1.3b).

This is of course a basic assumption of all traditional macroeconomics, and so is something of an anticlimax; like setting out across the Atlantic to find the Indies, and instead discovering Rockall.

It is however firstly worth noting that while this identity is an assumed equality in traditional economics, it is a self-balancing outcome of the GLV and L-V models used in this paper. Consumption is not defined as equal to income or vice versa, consumption of individuals rises and falls with wealth, wealth changes with income and consumption, income depends on consumption. In the models in this paper the dependencies go round in circles, hence the Lotka-Volterra outputs, the equality of total income and consumption naturally falls out at the equilibrium of the model.

This leads to a much simpler derivation of the Bowley ratio:

$$\beta = 1 - \rho \quad \text{and} \quad \rho = \frac{r}{\Gamma} \qquad \text{by definition,}$$

$$\text{so} \quad \beta = 1 - \frac{r}{\Gamma}$$

$$\text{also:} \quad \Omega = \frac{C}{W} \quad \text{and} \quad \Gamma = \frac{Y}{W} \qquad \text{by definition,}$$

$$\text{but} \quad C = Y \qquad \text{by definition,}$$

$$\text{so:} \quad \Omega = \Gamma \qquad \text{and so:}$$

$$\text{so} \quad \beta = 1 - \frac{r}{\Omega} \qquad \text{QED.}$$

Of course the definition above does not require a single line of my modelling, theorising or pontificating.

And for most economists it will appear to be a trivial and unimportant accounting identity.

But it isn't.

It is all a question of directionality. Of cause and effect.

For most people it is 'obvious' that consumption follows income, ie that people earn then spend, or that:

$$C = Y$$

Actually it is the other way round:

$$Y = C \qquad \text{or more accurately:}$$

$$\Gamma = \Omega$$

It is the consumption rate Ω that defines Γ; the ratio of total income to capital.

Trivially this is the case in my models, where r and Ω are fixed and Γ is allowed to float. But of course this is not sufficient justification.

The problem with the economic literature with regard to the Bowley ratio is that economists have first defined the profit ratio and Bowley ratio as:

$$\rho = \frac{r}{\Gamma}$$

$$\beta = 1 - \frac{r}{\Gamma}$$

They have then spent the last hundred years or so trying to explain the two ratios above by attempting to look at the microeconomic structure of industry that could affect r and Γ. This has almost entirely revolved around the analysis of 'production functions', the supposed microeconomic relations between capital and labour.

The Cobb-Douglas production function has become a particular focus of attention, as its form gives rise to constant shares of returns to labour and capital. (I am somewhat reluctant to criticise Gabaix, as he is one of the few economists who has recognised the importance of power-laws and other 'anomalous' invariants in economics. However his quote at the start of this section shows how deeply ingrained within economics this approach has become. Gabaix defines the solution to the problem of the Bowley ratio as the finding of a theory that not only produces the Cobb-Douglas production function, but also gives certain fixed exponents for the Cobb-Douglas function).

There are however very major problems with this approach.

Firstly, real analysis of companies suggests that any meaningful production function needs to be based on high fixed costs and increasing returns, and is far away from the Cobb-Douglas or other standard production functions used in neoclassical economics.

Secondly, as the data from Young [Young 2010] shows the relative shares accruing to labour and capital can change quite significantly within individual sectors such as agriculture and manufacturing. This shows that production functions are not giving the required output on a sector-by-sector basis. (Casual inspection of company accounts shows that returns to labour and capital can vary dramatically from company to company.)

The third and most important reason is the problems following the logical steps.

Firstly, traditional economics states that production functions define the relationship between r, the rate of return to capital, and Γ, the rate of total income to capital.

Secondly, traditional economics states that total income is equal to total consumption, so, logically, $\Omega = \Gamma$.

Putting these two statements together logically means that production functions, the microeconomic structure of the commercial sector, define the saving rate Ω. (This leaves aside r for the moment, we will return to r shortly.)

This is very difficult to swallow.

Squirrels save. As do beavers. And also some woodpeckers and magpies.

Laplanders build up their reindeer herds as a form of saving, as also Arab pastoralists build up their herds of camels and goats, and the Masai and BaKgalakgadi build up their cattle herds. Almost all agricultural societies store grains and other foods to tide them from one harvest to the next. And whether you live in the tropics with alternating wet and dry seasons, or a temperate climate with warm and cold seasons, saving is a biological necessity genetically selected in human beings for its beneficial outcomes.

From a behavioural point of view saving is a deeply ingrained human behaviour that borders on the compulsive. Most people put money away for a rainy day. While Bill Gates and Warren Buffet have shown extraordinary benevolence, they both continue to hoard wealth far beyond their possible needs.

Leaving biology aside, traditional economics has well-established logical theories for saving. Lifetime cycles make it logical for young, and especially middle-aged people to save to ensure support in their old age.

Whether you look at biology or economics, savings rates are largely exogenous to the economic system. They are defined by people's assessment of, and fear of, an unknown future.

Clearly my use of Ω as a consumption function is simplistic. Ω uses only total wealth as a definer of consumption. In reality consumption and saving decisions are going to depend on current income and projected earnings in a complex manner. In particular, individual consumption and spending decisions will vary significantly with age and family circumstances.

Indeed an interesting paper by Lettau and Ludvigson [Lettau & Ludvigson 2001] suggests that there is a constant rebalancing of asset wealth to ensure long-term consumption, and that this feeds back predictably into asset prices.

In reality, as people are born and die at roughly the same rates, the total pattern is relatively fixed, and over the long-term national consumption rates are relatively steady.

Clearly consumption and savings rates are affected by economic fundamentals. Savings rates go down, and consumption goes up in booms, when returns look good and fear of unemployment is low. In recessions savings rates go up, and consumption goes down, as returns go down and fear of unemployment is high. But these reasons simply reinforce the hypothesis of exogenous drivers of biology and economic lifetime planning for consumption and saving.

Despite the changes with economic cycles, over the long-term, savings rates show consistent trends linked to the relative wealth of a society, as originally described by Lewis [Lewis 1954].

The point here is that Ω can be explained by long-term societal trends such as age, sex, family size, amounts of spare labour in a society and the state of a country's social-security system. Short-term trends can be explained by return rates of investments, unemployment rates, etc. While Ω is not an absolutely fixed exogenous variable, it is a slow-changing variable that can be calculated from mostly long-term variables.

It stretches credulity to breaking point, to believe that saving and consumption behaviour is ultimately defined by the microeconomic production functions of commercial companies.

The causality works the other way, the systems of capitalism are set up in such a manner that the consumption rate Ω defines Γ, the rate of total income to capital.

When viewed in this way the data of Young makes sense [Young 2010].

In the period Young analysed, consumption rates stayed approximately constant, as did rates of return.

During the same period, both agriculture and manufacturing increased their returns to capital and reduced returns to labour.

Given fixed Ω, to keep things balanced, the economy as a whole was obliged to create new, labour-intensive, industries to ensure that returns to labour were maintained as a whole.

All those cappuccino bars and hairdressers were created by the economy; by entropy, to ensure that the Bowley ratio remained equal to $1-(r/\Omega)$.

In fact the consumption rate Ω, the Bowley ratio β, and the profit rate ρ are not very interesting pieces of economics at all. Ω is already well defined by life-time planning and/or behaviouralism. The Bowley ratio and profit ratio are trivial outcomes from Ω and r.

I find it difficult to believe that I am the first researcher to propose that the Bowley ratio should be defined by:

$$\beta = 1 - \frac{r}{\Omega} \qquad \text{rather than :}$$

$$\beta = 1 - \frac{r}{\Gamma}$$

However, I have not been able to find any other proposal of this relationship, and the recent writings of Gabaix, Young and others suggest that this is the case. If I am the first to do so I am happy to take the credit. If not I would be happy to update this manuscript appropriately.

The interesting economics is in r; the rate of returns. To date I have generally been vague about the meaning of r and have included dividends and interest payments as well as rents in r.

In fact there are three near economic constants which all show very stable long-term behaviour. In all three cases the behaviour is counter-intuitive and I believe likely to be related. The three variables are long-term real interest rates, long-term stock returns and long-term gdp growth rates.

Figure 4.5.1 below shows the long-term cumulative returns due to real interest rates for the UK and the US. For the UK this starts with a value of 1.0 in 1729, for the US the start is at a value of 1.0 in 1798. The returns are calculated by multiplying the successive value from each year by the interest rate less the inflation rate.

Data for these graphs, and also for the gdp graphs below were taken from the website 'Measuring Worth', for a very full discussion of historic interest rates see Homer and Sylla [Homer & Sylla 1996, Measuring Worth].

Figure 4.5.1

As can be seen, although there is significant variation around the trend, there is a very clear long-term trend, which is slightly over 2% for the UK and slightly over 4% for the US.

Figure 4.5.2 below shows long-term stock-market returns for the USA, from 1800 to 2008.

Figure 4.5.2 [Ward 2008]

Again, although there are significant short-term variations, the long-term trend of 7% is clear.

Finally figure 4.5.3 below shows real GDP in 2005 dollar for the United States from 1790 and 2005 pounds for the United Kingdom from 1830. The same long-term trend can be seen. This time the trend is slightly below 2% for the UK and slightly below 4% for the US. The match of long-term gdp growth trends to long term interest rates is striking.

Figure 4.5.3

In the discussions above, I have chosen r as an exogenously given constant. I have been vague about whether r should be the 2-4% of interest rates or the 7% of stock-market returns, or somewhere in between. This is, of course, because I don't know. I suspect it is somewhere between the two.

I do think the assumption of exogeneity, at least for the level of discussions in this paper, are reasonable. Like the Bowley ratio, both interest rates and stock-market returns show long-term constancy. The Bowley ratio is the dull one, as it is simply a result of the regularity of returns r and consumption propensity Ω.

(As an aside, a quick note on the changes of the Bowley ratio in recessions. It is well known that returns to labour increase in recessions, and so that the value of β increases. It is also well known that saving increases and consumption decreases in recessions. If consumption decreases, then equation (4.5k) would mean that β would decrease, which appears to be a contradiction. However in recessions both interest rates and stock-market returns also decrease, and the proportional decrease in interest-rates and stock-market returns is usually much larger than the decrease in consumption. So, overall, β does increase in recessions despite falling consumption.)

The interesting thing is where the constancy of interest rates, stock-market returns and gdp growth all come from.

Traditional economics has tended to look at technology change and microeconomic factors as the drivers, again this seems difficult to justify.

Firstly, technology tends to come in bursts; steam power, electrification, motorised transport, electronics, the internet, etc. This would suggest that both gdp growth and stock-market returns would come in bursts, and not necessarily bursts with the same rate of growth.

Secondly, the rate of change of technology, from casual observation, appears to be accelerating, with the bursts of new technology becoming more frequent and wide-ranging.

Thirdly, the growth of economies appears to be back to front. For the UK, growth started with the industrial revolution somewhere around 1800 and has continued at a regular rate of 2-2.5% for the last two centuries.

Almost all the other rich countries have followed a different path. In the first phase of the catch-up they generally had high rates of growth; typically between 5% and 10%. Until they caught-up or slightly over-took the UK. From that point on they then slowed down to a similar 2-4% rate as the UK.

For a very good visualisation of the process go to gapminder [gapminder].

This is counter-intuitive, as common sense says that as countries get wealthier they should be able to devote more and more capital to investment, and so they should be able to grow more rapidly, not less.

The constancy of the values of interest rates, returns and gdp suggest a much deeper equilibrium is present, a simple mathematical equilibrium. An equilibrium that is actually restraining growth significantly below that possible as a consequence of technology.

It is the source of these three constants, and the relations of the three to each other, that is the most pressing mystery of economics. A possible, though highly speculative, proposal for the source of this equilibrium is suggested in section 7.4.

Before moving on, I would like to discuss the parallels with Wright's models. In the Social Architecture of Capitalism Wright's model produces a value of β of 0.55, while in Implicit Microfoundations for Economics β is 0.6.

Wright's models are not formally mathematical, so it is not fully clear how these values are generated. In both these papers the expenditure is drawn randomly from a uniform distribution of an agent's wealth, which I believe makes Ω equal to 0.5 in both models. The way that excess wealth is generated in Wright's models is much more complex, and possibly recursive, and it is not clear (at least to me) how the equivalent to interest rate in these models would be calculated. If equation (4.5a) proves to be correct then Wright appears to have defined the interest rates for the two papers above at 22.5% and 20% respectively.

Finally, it should be noted that equation (1.6d) for the exponent of the wealth distribution power-law tail should now read as:

$$\alpha = \frac{1.36(1 - (r/\Omega))}{v^{1.15}} \qquad (4.5q)$$

Part A.II - Speculative Building

At this point in the discussion of the modelling, I believe it is appropriate to give a clear and unambiguous health warning.

Up to this point in the paper; although both the economics and the mathematical approaches of the modelling have been heterodox, I believe that the models built accord with basic common sense, most notably with the various variables and constants matching, at least approximately, measurable quantities in real life economics.

In the remainder of the first section of this paper, this no longer remains the case. For one reason or another the models and policy proposals in the rest of this section are speculative. The models have been included because they give results which may be interesting or plausible, and that may allow future building of alternate, more realistic, models in the future.

The conclusions produced from these models must also therefore be presumed to be highly speculative. I fully expect that some or all of the models and conclusions below will prove to be wrong. It is my hope that they will however prove to be informative for further work.

4.6 Unconstrained Bowley Macroeconomic Models

In section 4.5 above, we looked at Bowley models that deliberately constrained the net cash / debt balance to zero.

In this section these models are explored further by changing the net value of the cash balance so it is positive or negative and seeing what happens. As previously discussed, I have a profound philosophical problem with this approach. It is not clear to me who is holding this balance or debt, where it is held, etc. Because of this no interest is paid on the balance, or interest charged on the debt, for the simple reason that I do not know where in the model I should debit the interest from, or pay the interest to.

Despite this I am presenting the results because, firstly they are mathematically interesting, and secondly the outcomes are beguilingly plausible. I find this worrying, as it characterises some of the attitudes I have found most frustrating in my reading of much mainstream economics; the triumph of interesting equations and common sense over meaningful models related to underlying data.

The first model run was simply to put in typical parameters, from real economies of:

Returns rate	r	0.03
Consumption rate	Ω	0.2
Bowley ratio	β	0.7
Along with a Capital Wealth,	Q	100

And let the model reach an equilibrium, the resulting cash balance is:

Cash Wealth	H	-50

There are two things to note here. Firstly, allowing a negative cash balance; that is allowing the use of debt, allows the Bowley ratio to drop. This means that the returns to labour are reduced and the returns to capital are increased.

So, in short, allowing the use of debt allows more returns to capital.

It should be noted however that using an returns rate of 0.07, based on stock market returns, gives a positive cash balance of +17.

To investigate this further, the parameters of the cash/debt balance were changed systematically, along with changes to other variables, to investigate the results on the model.

As with the Bowley-Polonius model, the model was surprisingly easy to parameterise, and gives an equation as follows:

$$\beta = \frac{\Omega + \Omega(H/Q) - r}{\Omega + \Omega(H/Q)}$$

$$= \frac{1 + (H/Q) - (r/\Omega)}{1 + (H/Q)} \qquad (4.6a)$$

where H is the cash balance (wealth held in the form of cash or negative debt) and Q is the wealth held as capital.

Again, this equation has been derived 'experimentally' by investigating the model, but the equation fits the modelling exactly.

As in the previous section it is fairly trivial to derive equation (4.6a) from first principles.

As before, when the model is at equilibrium, all values of flows, stocks and debts are constant. At this point, if the values of capital Q and cash H are to be constant, then the total income must equal the total outgoings, so, as before:

$$C = Y$$

$$= e + \pi \qquad (4.6b)$$

However this time, in the original model, in equation (4.2b), we defined the consumption ratio Ω as:

$$\Omega = \frac{C}{(Q + H)} \qquad \text{so,}$$

$$\Omega(Q + H) = C \qquad \text{or, substituting from (4.6b):}$$

$$\Omega(Q + H) = Y \qquad (4.6d)$$

again, the profit rate is defined by:

$$\pi = rQ \qquad (4.6e)$$

If we multiply equation (4.6d) by equation (4.6e), then we get:

$$\pi\Omega(Q + H) = rQY \qquad (4.6f)$$

Rearranging gives:

$$\frac{\pi}{Y} = \frac{rQ}{\Omega(Q + H)} \qquad \text{or:}$$

$$\rho = \frac{rQ}{\Omega(Q + H)} \qquad (4.6g)$$

Subtracting both sides from unity gives:

$$1 - \rho = 1 - \frac{rQ}{\Omega(Q + H)} \qquad (4.6h) \qquad \text{or from (1.3v):}$$

$$\beta = \frac{\Omega(Q + H) - rQ}{\Omega(Q + H)} \qquad \text{or,}$$

$$\beta = \frac{\Omega Q + \Omega H - rQ}{\Omega Q + \Omega H} \qquad \text{dividing by } \Omega \text{ and Q;}$$

$$\beta = \frac{1 + (H/Q) - (r/\Omega)}{1 + (H/Q)} \qquad (4.6a)$$

Once again the base equation here is (4.6g) which is the ratio of returns from capital, to total returns. In the next section I would like to discuss the overall meaning of equation (4.6g) in more detail, but before that I would like to look at some consequences of varying the debt value H.

It can be seen from equation (4.6a) that the Bowley ratio can be manipulated by changing the value of the cash balance H.

If the cash balance is positive and increasing, Bowley's ratio just heads closer and closer to unity, good for workers, bad for capitalists.

More interestingly, if H is negative, a debt, and the size of the debt is increased, then the size of both the numerator and denominator reduce, however the value of the numerator reduces more rapidly than the size of the denominator, and the Bowley ratio slowly decreases. At least at first.

If debt is allowed to continue increasing, then a rather dull function suddenly becomes more interesting. Firstly the Bowley ratio drops rapidly to zero, and then shortly afterwards heads off to negative infinity.

In the model itself it isn't possible to reach these points; as the Bowley ratio heads to zero the model becomes unstable, and explosive – the economy blows up in an entertaining bubble of excess real capital and even more excess debt.

This may sound familiar.

This brings us to the first, more traditional, form of macroeconomic suicide; allowing too much debt in an economy. Again this is discussed in more detail later in the international model in section 4.10 below.

Unfortunately the model gives no indication of the policies to be followed post explosion, though it does suggest that sensible limits on total debt (or debt ratios) in a well run economy might be a good idea.

There is a further consequence of this model that is intriguing. In this model the role of debt gives a direct output to the Bowley ratio.

As was found in section 1.6 above, the Bowley ratio in turn gives a direct output to the parameters of the GLV income distribution.

So, if the above models hold, there is a direct link from levels of debt in the economy to the levels of inequality. Specifically, increased levels of debt lead to increased levels of inequality.

Intuitively this seems plausible. Looking back over the last century, especially at the US, the first part of the century was associated with high levels of inequality, and high levels of leverage, which ultimately resulted in the Wall Street crash and the depression. In reaction to this, from the 40's to the 70's, leverage was strictly controlled, and also income distribution was much more equitable. From the 70's to the end of the 20th century, increased financial deregulation, and increased leverage, went hand in hand with increased inequality.

Given the mathematical simplicity of equations (4.6g) and (4.6a) it should be straightforward to check these relationships both historically for individual countries as well as across different countries. It seems highly likely that the complexity of economics means that there are other factors that need to be included in equation (4.6g), for example, all the above has been carried out with payout factors fixed at one. However, with luck the errors might be systematic and relationships may appear.

As a minimum it should be noted that a more realistic version of (4.6g) would include net returns based on returns from investments, based on say 7% [Ward 2008] less returns on debt at 3%; representing long term interest rates. I would guess that this would give something like:

$$\rho = \frac{(r_k - r_f)Q}{\Omega(Q + H)} \qquad (4.6i)$$

Where r_k is the typical return on investments in companies and r_f is a long term risk free interest rate. I emphasise that equation (4.6i) is merely a supposition and has neither been derived nor modelled.

If actual economic data give support for the relationship in (4.6g) above, then this would give some support to the fact that the debt in equation (4.6g) was in fact a meaningful value.

If economic data does support equation (4.6g), or a variant of it, then this raises interesting discussions on the role of debt in a national economy. The history of the last forty years has been one in which neoclassical economists have argued forcefully for the liberalisation of financial markets under the assumption that deregulation would allow deeper and cheaper financial markets and that self-regulation would ensure a natural balancing of an equilibrium. Equation (4.6g) begs to differ.

Equation (4.6g) dictates that persuading governments to allow greater leverage merely allows benefits to the owners of capital, while simultaneously moving towards a more unstable equilibrium that coincidentally increases overall wealth inequalities.

In fact this is the second form of rent-seeking we have seen exposed. If they were true to the core values of their religion, neoclassical economists would condemn this rent-seeking for what it is, and support strict controls on leverage. In practice neoclassical economists have consistently supported the 'freeing' of credit markets in the mistaken belief that greater access to funding will reduce prices and increase overall 'welfare'. In the real world any practical cost benefits are negligible compared to the disadvantages. The disadvantages are a substantial shift of funds from the productive sector of the economy to rent-seeking financiers, and a large transfer of 'welfare' from the poor to the rich.

Equation (4.6g) suggests that control of the national level of leverage can provide three separate economic benefits. Firstly for the working of the economy there will be a optimum level of debt that allows liquidity and provides capital for genuine economically productive investment. Secondly, by preventing extreme levels of debt financial instability can be prevented. Thirdly, the level of debt may be reduced to achieve reduced levels of inequality.

If the third item above is tackled successfully then the second becomes irrelevant, so the debate regarding the appropriate level of debt becomes a trade off between the first and third items.

While the income distribution requirements suggest an elimination of debt, this is clearly not practical for a well functioning economic system. While much investment is funded directly from cashflow, if the economy is to grow successfully non-financial firms clearly need access to debt financing for major capital investments.

Similarly, while it is always fashionable to attack 'speculation' a significant proportion of speculation is clearly useful. Neither farmers nor bakers are experts at predicting weather patterns. Both use derivatives on grain production to hedge their prices. It is the entrance of speculators into the grains futures markets, speculators who are able to look at weather patterns across the different grain producing countries of the world, who keep these markets working effectively, so benefiting both farmers and bakers. The same is true of speculators in any derivative market when they are functioning correctly.

However there are clearly points where derivative markets fail to be efficient finders of future prices and start to be used by uninformed momentum chasers as apparent sources of financial growth in their own right.

Although the work of Minsky is not quantitative in nature, his characterisation of the phases of debt build up is clear and easy to relate to real economic cycles. If equation (4.6g) above is found to be applicable, it should be possible to look through past economic cycles and note where debt moved from a useful point; of providing funds for investment and price finding speculation, to turning into a self-sustaining provider of bubble finance. This would then provide central banks with a guide to controlling financial markets for the benefit of the economy as a whole.

I would now like to look at the character of equations (4.5h) and (4.6g) in more detail.

4.7 A State of Grace

It has been previously stated that equation (4.5h):

$$\rho = \frac{r}{\Omega} \qquad (4.5h)$$

for non-debt economies, and equation 4.6g:

$$\rho = \frac{rQ}{\Omega(Q + H)} \qquad (4.6g)$$

for economies with debt, look suspiciously akin to what physicists call 'equations of state'. This is a very brave statement and time will tell if this proposition is accepted. However it is clear that the equations work in ways similar to equations of state, and this is important for understanding what these equations signify, especially with regards to economic equilibrium.

Firstly I would like to give a little background of other equations of state in physics. Historically, the study of thermodynamics; things such as the expansion of gases, heat engines, heat production from chemical reactions, etc, was problematic because there were large numbers of macroscopic and microscopic variables. Changing one of the variables generally resulted in simultaneous changes in many other variables and it was very difficult to work out what was actually happening. In this regard, classical thermodynamics was similar to present day economics.

In the study of gases a series of pioneering scientists carried out various carefully controlled experiments that resulted in various relationships being established.

So Boyle's law states that, at constant temperature, the volume of a gas varied inversely with the pressure. Charles law states that, at constant pressure, volume is proportional to temperature, and so on.

Finally it was found that all the different laws could be put together to give the 'ideal gas law' in the form of an equation:

$$PV = nRT \qquad (4.7a)$$

where P is the pressure, V is the volume, T is the Temperature, n is the amount of substance in moles, and R is a fundamental constant of the sort wished for by Mirowski.

In fact the 'fundamental' nature of R is an accident of history. The concepts and measurement units of pressure, volume and temperature were generated independently with idiosyncratic units. Here R is just a method of adjusting the different measurement systems so that the units fit together.

Later microscopic theory showed that that the equation could be changed to a more fundamental form of:

$$PV = NkT \qquad (4.7b)$$

where N is the number of molecules, and k is another much more fundamental constant (Boltzmann's constant) that once again mops up all the different unit systems. If physicists were allowed to start from scratch they would change all the units so that the constants were all dimensionless '1's, which would make things easier for physicists but harder for butchers, bakers and shoppers.

The point about equation (4.7a) is that for an ideal gas (and the 'ideal' is very important) equation (4.7a) defines all possible equilibrium points for the volume of gas you are looking at. With the three variables of p, V and T there are an infinite number of points of equilibrium on a two-dimensional sheet in a three-dimensional space that can be occupied. However, any equilibrium must be on this sheet.

So if you double the pressure of the gas, you will either halve the volume or double the temperature, or simultaneously change both volume and temperature so that equation (4.7a) balances.

Other thermodynamic systems are characterised by similar equations They are interesting for a number of reasons.

Firstly, despite the complexity of the underlying system, equations of state are often surprisingly simple.

Secondly, the way the variables fit together can be non-obvious or even counterintuitive. Familiarity with equation (4.7a) means that people are used to it, but for the pioneers in the field, there was no obvious reason why these three variable should fit together in this way, and in fact it wasn't until many years later that the equation was independently explained at an atomic level by Maxwell and Boltzmann.

Thirdly, the equations do not refer to underlying microscopic mechanisms or variables. In equation (4.7a) there are no references to elasticities of collision, the masses of the gas molecules, etc, in fact the equation should be the same for any perfect gas.

Fourthly, it is common to find that many of the variables in an equation of state are intensive, that is the properties do not depend on the amount of material present.

So in equation (4.7a) pressure and temperature are both intensive parameters, you can measure pressure and temperature locally at different points throughout the system as long as it is at equilibrium. Volume on the other hand is an extensive parameter that depends on the amount of stuff present.

Finally, by reducing a complex system to a simple equation, equations of state are extraordinarily useful for defining and analysing systems.

Going back to equation (4.6g):

$$\rho = \frac{rQ}{\Omega(Q + H)} \qquad (4.6g)$$

this equation appears to fill all the above characteristics fully.

Firstly it can be noted that both ρ (returns/total-returns) and (Q/(Q+H)) can be seen as macroeconomic ratios.

Then equation (4.6g) becomes a formula incorporating just four intensive variables and could be expressed as:

$$\Omega \rho(1 + G) = r \qquad (4.7c)$$

Where ρ is the profit ratio and G is a cash-debt gearing ratio H/Q, and none of Ω, ρ , G or r depend on the size of an economy.

This meets conditions one and four.

Condition three is certainly met; there are none of the microscopic foundations beloved of economists in equation (4.6g).

Condition two would appear to be the case, given that this equation has followed Bowley's original discovery by over a century.

The fifth condition remains to be proved.

Just as an aside, an accident of history means that I am unable to present Phillip Mirowski with his fundamental constant, something similar to the R of (4.7a) or the k of (4.7b). Luckily for economists almost all variables in economics have been defined in terms of money, people or money per person. As a result the equations of state fit together automatically and the balancing constant is simply unity. Unfortunately for naming conventions, persuading people that the dimensionless number 'one' is a fundamental constant rather than a lucky accident is a little tricky.

Why equation 4.6g (or (4.7c)) is important is that it says that you can't change the Bowley ratio without changing the savings ratio, the gearing ratio or long term returns. Or vice versa for any of the savings ratio, gearing ratio or long term returns.

Which means that you can't change the Bowley ratio by changing things like the tax system, the education system, trade union bargaining rights, monopolistic behaviour, reducing friction in capital markets, affirmative action, inheritance laws, or a thousand and one other things that people believe will make incomes better for ordinary folks. None of the above will have any effect on the Bowley ratio unless they change one of the other factors in equation (4.6g).

In extremis, as the Russians discovered and the Chinese are discovering, you can't even get more money into the pockets of the workers by introducing state ownership and a workers paradise. Ultimately, if your economy becomes technologically advanced, the factories become informally 'owned' by a nomenklatura or similar business class linked to the elite, and Bowley's law and the appropriate matching unequal GLV distribution reasserts itself. Sadly for Marx, his perceptive insights prove so powerful that they work their wonders even in 'Marxist' economies.

It is for these reasons that my own proposals for solving poverty look at redistributing wealth rather than redistributing earnings.

Going back to equation (4.6g), it is worth focusing again on the underlying model in section 4.6. There are very important economic factors in the model that do not appear in equation (4.6g). This includes the amount of physical capital K, or the proportion of this capital that is used. It includes the productivity of this capital. It also includes the function of the compensation of the workers, and so in a real economy, the level of employment and unemployment.

All of these things have no relevance to the overall, macroeconomic balance of the model. All these things have secondary functions in the model.

The overall model has an infinite number of equilibrium points that balance to equation (4.6g) even when the solutions are stationary. This is the prime equilibrium that is being sustained. The equilibrium that the system automatically and inevitably returns to.

When the model moves into unstable zones, the equilibrium hunts around an equilibrium with the parameters in (4.6g) changing cyclically. There is an infinite number of points the cycles can pass through, but within a constrained zone, much like the foxes and rabbits of the original Lotka-Volterra model.

Within each of these infinite solutions the values of capital, capital productivity and waged earnings all adjust to a give a solution that satisfies equation (4.6g).

To take a trivial example, suppose that the amount of labour needed to service the real capital K is exactly halved for all values of K. This can be modelled in model 4A, or the other models, in appendix 14.9. by changing the parameter 'labour_required' from 1 to 0.5.

If you simply change the value of labour_required from 1 to 0.5 then all the various parameters in equation (4.6g) will change to new values. Most notably the value of the cash/debt balance will change. If the model is then returned to it's original overall parameters, by using solver to return the debt to its original value by adjusting K, then a new equilibrium is achieved, with a higher value of K.

A comparison is shown below, column A is the first equilibrium, column B shows the result of changing the value of labour_required, finally column C shows the result of returning the cash balance to zero.

Figure 4.7.1	**A**	**B**	**C**	
interest_rate	0.10	0.10	0.10	
production_rate	0.20	0.20	0.20	
consumption rate (Ω)	0.40	0.40	0.40	
labour_required	**1.00**	**0.50**	**0.50**	Halved A to B
goods_payments	40.00	32.39	40.00	
earnings_income	30.00	22.39	30.00	
actual_returns	10.00	10.00	10.00	
capital (K)	100.00	119.03	135.61	
capital_wealth (Q)	100.00	100.00	100.00	
cash_wealth (H)	**0.00**	**-19.03**	**0.00**	Forced to zero B to C
total_wealth (W)	100.00	80.97	100.00	
total_returns	40.00	32.39	40.00	
Bowley Ratio (β)	**0.75**	**0.69**	**0.75**	Reverts to 0.75 A to C

In this case an increase in labour productivity has been balanced by decreasing employment. A new equilibrium has been achieved, and at this point there is no need for any further adjustment in the model.

In the case of the change of labour_required from 1 to 0.5, the new equilibrium at zero cash balance is 136 units of capital. The requirements of labour per unit of capital has halved, but the amount of capital has increased by only a third. The actual labour required to be employed has reduced by nearly a third. The new equilibrium has rebalanced by sacking workers. The marginality of labour is not relevant to the model, the model simply moves to ensure that equation (4.6g) is balanced, it does this without any reference to the underlying labour supply curve. Model 4A, and all the other models, can create mass unemployment as a consequence of improved technology, and can then sustain that mass unemployment indefinitely.

Indeed one of the main conclusions of models of section 4 and equation (4.6g) is that labour and capital, because of their different forms of ownership are not substitutable at a macroeconomic level. This is discussed at length in section 4.8 below.

There are many different ways that the model can be rebalanced, and many different ways that the equilibrium can be achieved. The key for the model and equation (4.6g) is that the total earnings; wages plus dividends, must balance the total consumption, which must be Ω times the wealth. Which equilibrium point will be achieved will depend on other factors, but the model won't naturally rebalance to full employment of its own volition. To get a clearer understanding, I urge readers to load the model in excel from appendix 14.9 and experiment for themselves.

This demonstrates that Keynes' fundamental insight was correct; that such a system could be stable even though it was not at the level of full employment, and that deliberate demand management would be needed to move it back to full employment. Unfortunately, Keynes avoided detailed mathematics in his main works, also his theories have been developed almost exclusively using the concepts of saving and investment as drivers, even when, as discussed in section 1.3 above, it has become clear that the IS paradigm is a secondary part of the economic cycle.

Returning to the discussions of an equation of state it is worth noting that equation (4.6g) does not mean that other relationships can not affect the variables in equation (4.6g), just that if one factor of (4.6g) is changed, then the others must vary to compensate. Similarly it is possible that other relationships could cause one variable in 4.6g to affect another variable.

It is also worth noting that the original gas model, shown in equation (4.7a) was that for an 'ideal' gas. While some gases, such as the noble gases, are close to ideal, most gases divert from the behaviour of (4.7a) under certain circumstances, most notably as temperatures drop.

Water vapour, for example, obeys (4.7a) fairly closely at atmospheric pressure above 100C. However if water vapour is cooled to 100C at atmospheric pressure, the volume of the gas drops dramatically as the gas condenses into a liquid.

To cope with such problems, instead of using equations of state, scientists and engineers use phase diagrams that show the relations between the state variables (p, V, T, etc) as the substance under observation changes between different states. Sometimes changes in state can be large and instantaneous. For example, superheated liquid can suddenly boil off explosively, or supercooled water can freeze instantaneously. Both these changes can be precipitated by for example a minor contaminant, or small movement.

Casual observation suggests that similar phase changes may be encountered with national economies. Looking at the bubble behaviour in Japan in 1989 or the US in 1929 or 2008, in all three cases it looks like a superheated, apparently stable, system suddenly made a dramatic shift to another, very distant equilibrium point accompanied by dramatic changes in debt level, consumption level and the ratios of nominal capital (Q) to real capital (K). The example of Argentina between 2000 and 2005 suggests that income distributions can also change dramatically in the short term during major economic shocks [Ferrero 2010].

Such system changes also typically involve hysteresis so it is not possible to simply reverse conditions and return to the start point.

Such phase change behaviour can be modelled within non-linear dynamics and chaotic systems, see Strogatz for example [Strogatz 2000].

It remains the case that claims that equations (4.5h) and (4.6g) are equations of state, rather than simple accounting conventions, could merely be an act of pretension. It is of course possible that the modelling, and so the equation is simply wrong. However the models and equations remain the only ever effective attempt to model theoretically the stylised facts that

Bowley observed a century ago, and the values produced are uncannily close to the observed data. If this approach is in fact wrong it does suggest that a similar approach may be one that finally clarifies this mystery of economics.

4.8 Nirvana Postponed

In the previous section it was explained how a Bowley type model could produce an equilibrium that resulted in persistent long-term unemployment. This in itself gives severe poverty problems for the least able in society, as well as a significant tax burden for those in employment, who have to provide the welfare.

A second problem for a Bowley type model is that, with interest rates, consumption rates and debt ratio generally stable over the long term; equation (4.6g) (shown again below), gives a fixed value for the Bowley ratio, and so, as we saw in section 1.5 a fixed value for alpha in the GLV distribution.

The fixed value of alpha then gives a fixed ratio of inequality and means that a significant minority of the population receives substantially below the average income.

Taken together these two elements mean that the bottom third or so of society in a modern economy can get a very raw deal; moving between long-term unemployment and intermittent low wage employment.

There are however deeper and much more important reasons why all individuals, including the rich, suffer from poor life quality in a Bowley type economy.

Going back to equation (4.6G):

$$\rho = \frac{rQ}{\Omega(Q + H)} \qquad (4.6g)$$

Again given that the profit rate, consumption rate, and debt gearing are all fairly constant in a mature economy, then the Bowley ratio tends to be close to constant, and the stylised facts show that the returns to labour are typically two-thirds to three-quarters, while the returns to capital are one third to a quarter.

To all intents and purposes, at the level of the economy as a whole, this means that the ratio of returns to capital and labour is pretty much close to invariant. At a macro level at least, the basic neo-classical, Walrasian assumption of substitutability of labour and capital is simply wrong.

In this respect, the Austrian school is fundamentally correct, there is a ‘natural balance’ between capital and labour.

And, in the absence of severe epidemics or genocide, the quantity of labour cannot easily be changed.

While it is possible to build up capital in the short term this is not sustainable, and a boom in capital above the long-term trend is followed by a bust, with at best stagnation in capital growth. If too much capital has built up, then there is the danger of capital destruction.

Interestingly, in the models in section 4, the amount of financial capital Q can increase dramatically for small increases in actual capital K, especially when debt is allowed to increase.

In these circumstances, the Austrian remedies for bubbles seem very sensible. As well as reducing debt back to sensible levels, the nominal value of capital, Q, needs to be reduced quickly via bankruptcies, wiping out the value of share and bond holders, etc. If this is done quickly then the economy can rebalance financial flows easily so that employment can be maintained and the fullest use of the real capital can be achieved. This was the approach used successfully in the 1990's by Sweden and other Nordic countries.

In recent crises in Japan and the US, fear of hurting owners of financial assets; ultimately mostly politically important holders of pension funds, has resulted in deliberate government policies of attempting to maintain the value of financial assets in 'zombie' institutions, or to bail out asset holders altogether by nationalising debts. While this may seem sensible in the short term, the effect of delaying a return to the natural equilibrium of equation (4.6g) above may result in unexpected consequences of deflation or inflation, and the long-term destruction of real (as against financial) capital.

Clearly a much better plan is simply to prevent excess debt, and so inappropriate capital building up in the first place.

One thing that should be clear from a fixed ratio of returns to capital and labour, is that attempting to 'rebalance' the economy by cutting wages and 'pricing workers back into jobs' is a course of great foolishness, and would guarantee a spiral of reducing returns to both labour and capital, so reducing employment and utilisation of capital. This was one of Keynes's central insights.

In one sense this 1/3rd – 2/3rd split of returns to capital and labour can be seen as a good thing. It is caused by the shortage of surplus labour past a Lewisian turning point, and prevents Marx's prediction of ever increasing returns to capitalists and ever further impoverishment of workers.

However, in a deeper sense this is also a very negative thing.

As has been discussed above in section 4.7, when the productivity of machines increases, one way the system can reach equilibrium is simply by using less human input.

As capital becomes more productive, to get the same returns you just use less of it.

What equation (4.6g) means, in fact what any formulation of Bowley's law means, is that because the balance of returns to labour and capital is fixed, to get any progress, to get any growth in gdp; to get more wealth, you must get more returns to labour.

Historically this generally been achieved by increasing the output from labour.

If the returns ratio of labour to capital is fixed at 2:1, then it is the amount and efficiency of labour that has to be improved to get gdp growth.

Progress is constrained by the amount and productivity of labour, not capital. Increasing the amount and efficiency of capital is relatively easy. But doing this alone has no useful effect.

Although Western economies are now highly mechanised, the workings of the financial system dictate that two-thirds of the earnings that are produced by capitalism are paid directly to people in the form of wages. Also, as discussed in section 1 of the paper, for 80% of people, payment for labour forms almost all their income. This necessarily demands the full time presence of people at work.

We have been enslaved by the machines.

In the second half of the 20^{th} century, for most Western countries, increasing the amount of production provided by labour was very easy. It was achieved very simply by moving women out of the home and into the workforce. This one change in itself was probably the most important source of economic growth through the fifties to the seventies.

Once this step has been completed, increasing the size of human capital becomes much more problematic. So the next stage is to increase the efficiency of human capital, however this is also problematic.

Human capital is primarily restricted to the skills and abilities that human beings have, and carry around with them in their brains. There are a few obvious skills such as driving, using basic word processing software, or other basic computer skills that can be easily learnt by almost all people. But beyond that things get difficult.

Information Technology is a good example. Computers are generally owned by companies, so returns on their wealth generated are taken by the companies. As we have seen above, if this improves returns to companies, it just results in less capital being needed overall. By replacing many basic clerking and administrative duties computers have actually taken skills that used to be in the hands of human beings and moved them to the owners of capital.

Some people of course have made a great deal of money out of their personal capital in the IT revolution. Computer programmers and mathematical modellers are two examples. But to get the returns to the humans, the human capital needed is knowledge of VBA, C++, Excel, etc as well as advanced mathematics. This is human capital that is only available to a minority of people with the requisite logical and mathematical abilities.

Another way to benefit from IT is to be a good and effective manager. However most would agree that this is also a minority skill.

This may explain some of the apparent problems of the modern world.

Firstly it might account for the non-visibility of IT in productivity despite the amount spent on it.

It might also account for the imbalance in work requirements between different skill groups. Unskilled labour is now of marginal assistance to serving machines, and has been largely replaced by the machines themselves. This is as true for clerking and administrative work as it is for labour. Spreadsheets and stock control systems have replaced the clerks. Forklift trucks and containers have replaced the labourers. In contrast skilled professionals, from plumbers and technicians to programmers and managers, people who have the abilities to serve the machines, find themselves under continuous pressure to increase their working hours.

Taken all together this might account for the fairly acrid taste that is seen in political debate in most Western societies.

On one side there is large population of the unskilled who find it difficult to find and hold decent work of any sort. These people face unemployment, poor wages, no opportunities for advancement and semi-permanent dependence on welfare. They often have stretches of involuntary inactivity. Despite their subsidies and enforced leisure, for these people hard work is not rewarded and life lacks hope of betterment.

On the other side there are skilled trades people, professionals and managers, who work longer hours and pay higher taxes than their parents, primarily, as they see it, to support the idle poor.

This is not a happy recipe.

Futurologists have been predicting for decades that once basic needs have been satisfied, human beings would be able to relax into a life of leisure. To date, futurologists have been wrong.

And it is not for the want of suitable capital, the progress of automated technology continues at an extraordinary rate. In section 9.3 examples such as fruit picking machines, automated hospitals and personal rapid transport systems are discussed. All of these examples share the common features of being able to replace large amounts of unskilled labour and also being technologies that are being brought into use.

Despite this, in real life, almost the opposite is happening, working weeks have been steady, and in some cases increasing. In Europe and the US retirement ages are being revised upwards rather than downwards.

In the west we have achieved enormous personal wealth, but through an accident of mathematics, we have been required to sacrifice our time to the mechanism of wealth production.

Nirvana has been postponed.

As an amateur futurologist, it is possible to conceive of a world where the main inputs of human labour could be reduced to direct care for the young, the sick, the elderly and the provision of entertainment and spiritual needs.

Which is what, biologically, human beings are designed to do. Other animals that dance, sing and make art works; such as birds of paradise for example, are generally animals that do not face significant predation and that have more than enough resources available, and so time on their hands. In the absence of predators to compete with, or resources to fight over, they turn to competition in the arts. Almost certainly prior to the agricultural revolution, human beings fell into this class of animal.

Human beings were simply not designed to work forty hours a day five days a week. Both hunter-gatherers and most agricultural societies are characterised by underemployment. Historically this was true in the West until recently.

The second half of the twentieth century is almost unique in being one in which the well off are characterised by having full time employment. In the past the rich were notable by not working, they lived off their capital and looked down on paid work.

This labour capital split of the Bowley ratio might also explain the bizarre behaviour of growth.

As has been discussed in section 4.5 above, when they start growing, economies typically follow a path of rapid expansion to use up surplus subsistence labour. Casual observation suggests that this can be associated with growth rates of up to 10%. The 10% restraint appears to be due to the difficulties of building infrastructure fast enough. China has been following this path for the last two decades, the Asian tigers did so before this; now India appears to be following the same route.

Once the surplus labour has been used up then growth generally drops to a slow continuous growth rate of about 2-4%. The UK has been expanding like this for over 200 years, the US for over 150 years see figure 4.5.3 above, or gapminder for some very pretty graphics [gapminder].

In theory this is very odd, once economies are mature, why do they just not continue increasing the capital stock at 10% per annum to provide for all people's needs and eliminate the need for labour? This should be easy, as countries, and people, get richer, more of their basic needs should be provided for, so diverting revenue (in the most general sense – not just public taxation) for provision of capital should become easier to do.

If however, growth is restrained by the productivity of labour, then a growth rate of 2-4% seems more sensible. Once reserves of subsistence labour have been exhausted, human capital cannot quickly be increased in the same way that physical capital can be.

I suspect that this might be only part of the explanation. As discussed previously, I find the growth rate of 2-4% suspiciously regular. It also goes hand in hand with suspiciously constant real interest rates at 3% or so, and suspiciously regular stock market returns, at 7%, see figures 4.5.1 to 4.5.3 in section 4.5.

The 'stylised facts' of these three growth rates are very suggestive of a deeper underlying process equilibrium.

The presence of a fixed ratio of returns to capital and labour also gives a very big problem that there is a general shortage of 'real' assets. As we have seen in section 1.8 above, there simply aren't enough real assets available to provide even for everybody's retirement needs.

This in itself could be a source of the search in the finance industry to create new and exotic assets that appear to solve this problem. Unfortunately, Bowley's law dictates that the underlying 'real' economy is fixed, so the total real returns are fixed. Trying to create new assets out of old is no more possible than other more traditional forms of alchemy. You can't create real new revenue streams simply by repackaging assets.

Similarly, this may explain the hunger for government bonds in the financial markets, especially given their apparent safety. But ultimately, government bonds are dependent, via taxation, on revenue earned in the private sector.

The most obvious example in the shortfall of capital is the example of housing. Other public goods such as health, education and pensions have obvious market failure reasons for not being provided fully.

Housing should be simple to provide for in a wealthy society. Simply build enough of it for everyone, then all you need to do is maintain it. In practice many societies have attempted to do this, through mass council (public) housing in the UK to the recent disaster of state subsidised mortgages in the USA. The problem of course has always been that the poor have rarely been able to afford the maintenance of the housing, never mind the capital payments.

So, the key question here is whether this system can be changed so that more capital can be accumulated to carry out more work on behalf of labour.

Interestingly, history suggests that the system can be changed significantly, and especially as a result of the scarcity of labour.

The trick is not to change the efficiency of labour but to fully remove the surplus labour and turn it into an increasingly scarce resource that is over compensated for its efforts.

Back in section 1.3 I made the assumption that labour was 'fairly' paid for its inputs to the production process. I kept this assumption through all the income models, though it was then discretely abandoned in the macroeconomic modelling.

Actually, because labour is a uniquely non-adjustable factor input, it is the only truly scarce, non-substitutable resource. Also, because of Bowley's law, labour is very rarely paid it's true worth. It is usually significantly under or overpaid.

Following the theories of WA Lewis [Lewis 1954], or for that matter Marx, in a society with excess subsistence labour, capital can 'under-pay' labour employed in the commercial sector, as pay rates are held down at subsistence level by the presence of under-utilised rural labour.

This has been the normal state for most countries for most of history, and has provided the main critique of capitalism until at least the end of the Second World War.

In such an economy, with surplus labour, the economy doesn't reach a true equilibrium for the Lotka-Volterra / GLV approaches described above. The subsistence farmers are outside the equilibrium, and they also hold down the wages of those employed. In such a society the rich are overcompensated for their ownership of capital, and also have low living costs due to the low labour costs. In these societies the Bowley ratio can be as low as 0.5, this can be seen in China today, even as it approaches its Lewisian turning point.

Things are much more interesting in a 'normal' industrialised country; one that has passed it Lewisian turning point and has absorbed the majority of its cheap labour. In such an economy labour is generally over-rewarded; returns to labour are in excess of the value actually provided by labour. This was actually the case in the macroeconomic model in section 4 Where labour generally gained through the economic cycle, being 'overpaid' in exactly the same way that suppliers of commodities were overpaid in the commodity cycle in section 3. In this case the employees are successfully extracting 'rents' from the capitalists. And a good thing too.

I believe that, in the second half of the 20th century, parts of the world moved, for a period, fully into the zone described in this model.

Following the Second World War; all the communist countries, most of the de-colonised countries, and most of Latin America voluntarily withdrew from the world trade system. The communists followed their own socialist paths; almost all of the rest followed a route of import substitution behind high tariff barriers.

Following rapid post war growth, most of Western Europe and North America went through a period in the fifties and sixties with full employment and ongoing labour shortages. Meanwhile the few poor or poorer countries that remained in the world trading system; countries such as Japan, Italy, South Korea, Taiwan, Hong Kong, Singapore and Malaysia, saw breakneck growth, moving from subsistence agriculture to industrialisation in a generation.

In the West full employment artificially increased returns to labour. Through the Bowley ratio this then forced investment in capital to increase returns to capital. Over the longer term, expensive labour forced investment in labour saving production, so increasing the efficiency of capital.

This period resulted in a virtuous circle with high wages and full employment forcing rapid growth. Returns to both labour and capital kept increasing in lockstep.

It is worth remembering that labour was so scarce in this period that large-scale immigration was allowed into the UK, and guest workers were invited to Germany, to do the menial work that Britons and Germans were unwilling to do.

From the nineteen-seventies onwards many poorer countries, most notably China, re-entered the world economic system, providing alternate supplies of cheap labour, and competition for labour in industrialised countries.

The portion of the world's economy that is integrated into the trade system moved back to a pre-Lewisian state with excess subsistence labour in Asia, Africa and South America competing with Western labour.

It is the belief of the author that, at the time of writing, the richer, industrialised, countries are currently simultaneously in a complex pre- and post-Lewisian state. Pre-Lewisian for unskilled labour, and post-Lewisian for skilled labour. This is due to an accident of history caused by the third world's absence from, and then re-entry to, the global economy.

These conclusions appear to have some support from data. As well as showing smaller cycles, many of the country graphs in Harvie [Harvie 2000] show a much longer term cycle of change in the compensation to labour, starting with lows in 1956 going to high points in the 1970s, then returning to lower points by 1994 (the last points in the data sets, all of which were for industrial economies).

It will be interesting to see what happens in the near future. China appears to be passing through it's Lewisian turning point. Already China's low-cost manufacturing base is relocating to poorer countries such as Vietnam and Bangladesh. That is the manufacturing base that supplies cheap toys, shoes and clothes to richer countries. This in itself will spread wealth, and labour shortages, to these countries as they start exporting to the West.

Simultaneously China will also need to start importing cheap manufactures from poorer countries to supply its own population. Given that India is already close to peak expansion rates, primarily through providing information services to the West, the worldwide supply of surplus cheap labour could dwindle very quickly.

It is possible that we are close to seeing a repeat of the full employment boom of the 50s and 60s, but this time repeated on a worldwide scale.

Even without waiting for this process to happen naturally, it is possible that the proposed '40 acres' compulsory saving process proposed in section 1.8 above might also be able to produce the same effect artificially in single countries.

Although some people are natural workaholics, most would choose to 'downsize' and have more leisure time if they could.

But they can't.

It is common in neoclassical economics to see discussions of individuals choosing between spending and leisure. Because of the workings of the GLV, most individuals have no such choice.

To seriously consider reducing working hours; a family needs to own their own house, have a good pension plan in place, have enough money coming in to cover day-to-day expenses and be sure of access to a decent health service and a good education system for their children.

Even in the richest of Western countries few people have all, or even most, of these things. Primarily because they have insufficient capital.

If a 'forty acres' style system is used it would give more returns from capital to all members of society, it would reduce reliance on earned income.

It could slowly start a virtuous circle like that seen in the 50s and 60s.

By ensuring that all individuals move up to the point that they have sufficient wealth and income to meet their day to day needs, compulsory saving would allow people to move into voluntary saving and allow faster investment in decent housing and sufficient pensions. This would then allow a much more genuine choice between work and leisure. As individuals begin to withdraw from the labour market, this would then start a virtuous circle of rising labour costs and full employment. In the longer term this would then also encourage a drive to more labour saving capital.

Probably it would start with middle class families choosing to keep a partner at home when children are young. But even such a small withdrawal would tighten the labour market in the skills removed and so push up wages.

As people withdraw from the labour market, this will force wages up, and will also increase the share of returns to those still in the labour market.

With labour tight, and wages rising this will also encourage adoption of more efficient, labour saving technology. With Bowley's ratio holding, returns to both labour and capital will go up, while more and more of the actual work done by the machines.

The aim would be to create mass underemployment, or even unemployment, but not, as presently happens by accidentally creating unemployment at the bottom of society.

Instead, the aim would be to create voluntary underemployment at the top of society, as people choose to live more on their investment income and less on their wages. As this then forces wages up, the process will then work its way down to poorer people.

The aim is to create underemployment at the top end of society, so creating full employment throughout society. So increasing wages for all, so increasing returns to labour, so, via the Bowley ratio, forcing up returns to capital.

The aim would be to build up the v40 so that it would consist of shares in companies owning machines carrying out fruit-picking, hospital-cleaning and personal rapid transport. Meanwhile the people who used to be agricultural labourers, cleaners and taxi drivers would get more rewarding and better paid jobs with shorter hours. They would be helped by the income from their own v40's.

A good aim would be to get the v40 sufficiently large for everybody that dividend payments pay the equivalent of two working days per week of total living costs, while people still work three days a week for their remaining income.

On retirement, the additional drawdown of capital would provide for five working days per week of income.

A three day working week seems a sensible aim. There will always be a need for human beings to provide education, caring and entertainment. Three days a week would be sufficient to give structure and integration in society, but would leave ample time for family, friendship and leisure.

To many the above will seem ridiculously naïve, but the example of Norway given previously shows that the numbers can add up, and a three day week is feasible. As long as enough capital is available.

Futurologists' predictions have gone wrong because of the workings of the Bowley Ratio. Understanding how the Bowley Ratio works may allow the future to be changed.

4.9 Bowley Squared

Going back again to the base model shown in figure 1.3.5, this shows financial wealth W being held by households in the form of stocks and shares as claims on the real wealth K in the productive companies.

Figure 1.3.5 here

In one important way, this is very unrealistic.

I personally don't own any shares. In reality very few people own shares directly. In fact, aside from housing, most people do not own any capital directly.

Most peoples' wealth is in the form of bank deposits, pension funds, insurance policies, mutual funds, etc.

All of these investments form financial claims on companies within the financial sector.

The companies in the financial sector then own the claims on the real assets of the non-financial sector.

When it works correctly this is just a sensible way of dividing labour. Most people who have money to invest do not want to spend their spare time investigating possible investments. Also they would prefer to spread their investments across different companies to spread their risk.

It makes lots of sense to lend their money to professional experts who can save costs by analysing investments on the behalf of lots of different investors at the same time.

This then results in a model of the form shown in figure 4.9.1:

Figure 4.9.1 here

While it might seem very sensible to set up a specialist finance sector in this manner, from a control systems point of view this is something of a nightmare.

This repeats the feed back loop of the simple macro economic model a second time. Instead of one simple feedback loop capable of creating endogenous cyclical behaviour, you now have two feedback loops both capable of creating endogenous cyclical behaviour, and more importantly, capably of interacting with each other to give even bigger more complicated endogenous cycles.

The original macroeconomic model can be considered to be a very simple unstable model on the lines of the Soay sheep model discussed briefly in section 1.2.1 In this model the companies grow too rapidly for the base level of labour that can support them, in the same way that Soay sheep breed too quickly for the grass to support them. Introducing a financial sector, installs a second population on top of the first. It is similar to adding wolves to predate on the sheep of the first model.

I have not attempted to construct this model mathematically. The models discussed in section 4 above already have sufficient loose parameters and dynamic complexity to produce confusing patterns of behaviour. They really need pinning down with real data before being expanded to the model in figure 4.9.1.

But even without modelling, some of the behaviour is easy to predict. In fact we have returned back to something very similar to the original fox and rabbits Lotka-Volterra model discussed back in section 1.2.

In this case, the rabbits are the non-financial sector and the foxes are the financial sector. Typically a boom would start with a small financial sector and a growing productive sector. As the productive sector grows the financial sector grows more and more rapidly taking up an increasing proportion of the economy. Then the productive sector will start to decline slowly. A short, but significant time after that, the financial sector will show a sudden and much more rapid decline.

The operation of the two business sectors is analogous to the fluctuations of biomass in a Lotka-Volterra model. First biomass builds up in the rabbits then in the foxes, then it declines in the rabbits and then the foxes. Similarly capital should build up in the productive and then financial sectors, followed by declines, in turn for each sector.

So a prediction of this model is that over the next five to ten years, the proportional size of the financial sector in countries such as the USA and UK should decline back significantly towards proportional sizes seen in say the 1980s or early 90s.

One other outcome of this model is that the two sectors can follow their own paths to a significant extent. In such a model, the secondary feedback loop, that of the finance system can vary much more dramatically than the underlying population, see figure 1.2.1.1, showing the original Hudson Bay lynx and hare populations.

This makes control of such a dual speed economy very difficult when you are only using the single weapon of inflation targeting and interest rates.

While the underlying economy may respond reasonably to interest rates, the liquidity generated in this productive economy can generate much larger changes in liquidity in the finance sector, which are harder to control. Also the fluctuations in the financial sector will not be in the same time phase as the main economy.

To take an analogy this model can be likened to an air-conditioning system. The main economy can be imagined as a large office block somewhere in the temperate northern hemisphere. Depending on the time of year or time of day this main block will need a certain amount of heating or cooling.

The financial sector can be seen as similar to a large atrium on the south aspect of the building, full of hothouse flowers. The two buildings will be connected together, and will be roughly aligned through the seasons and days, but will vary greatly in the amount of cooling and heating needed. The atrium will need more heating in winter and more cooling in summer. This will depend on the amount and direction of sun and the external air temperature. On some spring and autumn days, the atrium might need cooling when the building needs heating or vice versa.

The Bowley squared model is a complex system and needs full understanding to control effectively. The topic of financial sector liquidity and how to control it is revisited in some depth in section 8.2.1 below.

Despite the complexity of the model in figure 4.9.1, it remains the case that control of such a system should be straightforward using standard controls systems feedback theory.

4.10 Siamese Bowley - Mutual Suicide Pacts

In the previous section one Bowley model was placed on top of another, in a way that was multiplicative.

An alternative model would be to put two Bowley models side by side and allow individuals in one half of the model to own capital in the other half of the model. This is illustrated in figure 4.10.1 below.

Figure 4.10.1 here

This gives an international model, with international trade.

The discussion that follows borrows heavily from the work of Michael Pettis [Pettis 2001], whose writing I have found highly illuminating, in contrast to much standard economic work on international economics and finance.

Pettis's work takes a financial framework for analysis, and concentrates heavily on flows and stocks of capital and debt. As such it fits well with the analytical models described in this paper. Pettis's work also fits closely with the known facts of repeated booms and busts triggered in poorer nations by investment booms and financial crises initiated by capital investment typically from London or New York; a process documented beautifully by Reinhart and Rogoff in 'This Time is Different' [Reinhart & Rogoff 2009].

One aside with regard to the use of the word capital, which in international economics is used in a markedly different way to that in normal macroeconomics, or the preceding sections of this paper.

In this paper capital can refer to K, the stock of physical assets that produce real wealth in the form of goods and services. It can also mean W (or Q), the stocks of paper financial assets that are held as claims on those productive physical assets, such as stocks, shares and company bonds.

In international finance a 'capital flow' is used to refer to a flow of money in return for a stream of paper financial assets; sometimes financial assets of companies, but these can also be assets such as government bonds.

So a capital inflow from Britain to Brazil would indicate purchase of Brazilian financial assets by institutions in Britain. The ownership of these financial assets would then give the right of the British owners to receive a stream of financial income based on the wealth produced by the underlying real physical capital.

In theory such a capital inflow should be used to invest in physical capital goods in the recipient country so allowing the country to become more productive and pay the interest on the loans.

Unfortunately it is all too common for the 'capital flow' to be used as payments for imports into the country receiving the 'capital flow', eg, Brazil paying for imports from the UK. When this is the case, the original meaning of the word 'capital' is lost altogether, and the 'capital inflow' is simply a way of describing lending money as a form of debt, often effectively unsecured.

And as can be seen from the analysis of Pettis or the research of Reinhardt and Rogoff, it is this quick and natural split of countries into creditors and debtors that is symptomatic of financial trade.

International finance can be very confusing, with a large number of variables, especially when currency flows and exchange rates are taken into account.

Much analysis of international finance concentrates on the role of currency, along with control of interest rates and the role of inflation.

Actually, history suggests that different currencies are in fact something of a red herring. To get the basic model for analysis you don't need currencies.

Throughout history there are many examples of international trade, and gross trade imbalances occurring when countries shared a common currency. Pettis gives the first such well documented example as that of different parts of the Roman empire in a speculative property boom in 33AD. In this case the metropolis of Rome was the debtor, while the grain producing provinces were the creditors.

History is replete with currency unions or fixed exchange rate pegs coming to grief through trade imbalances. Many of the imbalances of the depression, when the US was a creditor and most of the rest of the world were debtors, were exacerbated by the fixed exchange rates of the gold standard. Most of the countries involved in the Asian financial crises of the 1997 were on fixed pegs to the US dollar. Mexico was forced off its fixed exchange rate during the tequila crisis of 1994 And Argentina suffered severe economic problems until it abandoned it's currency board in 2002. At the time of writing Greece, Ireland, Portugal and Spain are suffering major structural problems while Germany and it's near neighbours simultaneously enjoy good growth. The common currency of the euro is currently magnifying trade problems, not reducing them.

Another factor that can be ignored in a base model is relative wealth. Although it is most common for the rich nation to be the creditor nation and the poorer nation to be the debtor nation it is sometimes the other way round. Ancient Rome provides one example, where the rich metropolis was in hock to the poor provinces. A much better example is the current one of the rich USA being a very substantial debtor balanced by a much poorer China as a very substantial creditor.

In fact, when looking at trade imbalances, it is my belief that it is debt, or more particularly, savings rates, that are key.

In Europe rich Germany has a high savings rate while Ireland and the Mediterranean countries have lower savings rates and higher debt. On a bigger scale poorer China has one of the highest savings rates ever seen, and America has moved, in less than a century, from the world's creditor to the world's debtor.

It is unfortunate that this is often seen in moralistic terms, especially by creditor nations. In fact, though cultural reasons are clearly important, savings rates are often driven by deeper fundamentals.

As Lewis [Lewis 1954] pointed out lucidly, newly industrialising countries tend to have high savings rates as the newly rich elite have access to cheap land and cheap labour, and have little else to do with their money but save it.

The US complains bitterly about China's 'currency manipulation' causing an imbalance of trade, but the US made the same complaints about France and Germany in the 50's and 60's, about Japan in the 70's and 80's, and about the Asian tigers in the 90's.

The common denominator here is the US; the exceptionalism of the US in this case is their ability to issue the world's reserve currency. As issuers of the reserve currency, the US is able to borrow at cheaper rates than other countries, so it is hardly surprising that they have become the world's biggest debtor. An identical process happened in the UK in the 19th century.

In fact there appears to be a cycle in reserve countries over the last half a millennium. Reserve currency status has been held in turn by Portugal, Spain, Holland, France, the UK and now the US, with each country holding the status for a roughly a century. In each case it appears that a country starts with a solid productive base that put it at the heart of trade. This trade and creditor role then allowed its currency to become dominant in trade. Reserve currency status then allowed cheap borrowing and increased debt. The increasing debt, allied with 'imperial over-reach' defending trade routes, then caused a crisis and loss of reserve status to the next upstart.

So going back to figure 4.10.1 below;

Figure 4.10.1 here

We have two countries, Chermany with a high savings rate, and Medimerica with a lower savings rate.

The two countries could start with the same population and the same amounts of capital K and wealth W per head, but the situation is naturally unstable.

Chermany, with its higher saving rate will consume less than Medimerica and will accumulate more capital. After the first iteration, Medimerica will have a little less capital, but will still have a thirst to consume rather than save.

In the short term the flows can be balanced by an unholy trade off. Chermany can supply funds; 'capital outflow' to Medimerica in return for financial assets belonging to Medimerica. Medimerica can then use this cash to buy imports from Chermany, mopping up the extra production that Chermany's high savers don't need.

Unfortunately, although this balances the flows in the short term, it results in a grave problem with stocks. Chermany keeps on building up capital that it doesn't need. Meanwhile Medimerica increases it's financial debt to Chermany while simultaneously running down it's badly needed capital to pay for imports from Chermany.

This system is inherently unstable and can only end in tears. Eventually there will come a point where Medimerica simply can not pay the interest on it's debt. It no longer has sufficient real capital to generate the real income to do so. At this point Medimerica has to default one way or another. This can be by straight repudiation of debt, or by devaluation and inflation to reduce the value of the debts.

For Chermany this then gives two problems. Firstly the loss of value of the foreign assets owned. Secondly, and more importantly, the loss of markets for the exported goods produced by the excess capital that has been built up.

This was most dramatically demonstrated in the run up to the depression of the 30's when almost the whole world used the gold standard.

During the 20's as the world's creditor, the USA (and also France) slowly built up their proportion of the world's gold reserves until Germany, the UK and other nations ran low on gold and were forced off the gold standard. They were also forced to partially default on their debts to the USA. The US was left with a large productive capacity and no buyers for its goods and also sank into depression. The US cried foul, but with a large portion of the world's gold in the US it was not clear what the Europeans were supposed to use to buy American goods.

This bilateral instability goes back to the two forms of economic suicide introduced previously.

One form of economic suicide is to run up too much debt as discussed in section 4.6 which eventually becomes unsustainable. Running up debt can be very appealing, as it allows consumption to run ahead of real growth, and also inflates the values of financial assets. Until the party ends and the hangover kicks in, this feels good for public and politicians alike.

The second form of economic suicide is to allow capital to build up too quickly as discussed in section 4.4 above. Again in the short term this feels good because the rapidly expanding capital base increases employment and wages. (It can also have the unfortunate side effect of increasing pride in supposed national industriousness and thrift.)

While it is possible to carry out each form of suicide independently, this is not so easy. In a single isolated economy the results of too much debt or too much manufacturing capacity are difficult to ignore. It is difficult to keep increasing debt in a home market beyond a certain point, and it is also difficult to build up capital and carry out a mercantilist export policy without people to export to.

It is much easier to carry this out as a form of mutual suicide pact where one country takes on the role of debtor and the other of creditor, as described in the model above. The debtor country is able to borrow more and more at easy rates, the creditor country is able to sell more and more of its exports. Unfortunately neither of these processes can go on forever.

In the thirties it was the debtor countries that first collapsed one by one.

In the Plaza accord of 1985 the debtor countries laid down the law with over-exporting Germany and Japan. Germany took heed and rebalanced its economy (at least until the launch of the euro). Japan continued to push export led growth and imploded in 1989; to date it has not recovered.

From 2006 onwards the American economy started to sputter, stalled by too much debt. In 2008 the American economy imploded in the credit crunch taking other debtor countries such as the UK and Spain with it. At the time of writing, the creditors, primarily China and Germany, have rebounded, but with a world full of excess industrial capacity it isn't clear who they are going to keep exporting too. In Europe the need for rebalancing is obvious, Ireland and the Mediterranean members of the EU are moving into outright depression and are likely to default. In the world as a whole it remains to be seen whether China can rebalance in time to prevent a Japan style bust.

The big problem for China is that easing back on its export machine will result in mass unemployment and serious political unrest. A possible solution is to move capital into the hands of the workers, as discussed in section 1.8 above, so that workers would have more to spend, and would not be reliant on wages alone. All in all it would make sense for the Chinese and Germans to consume more of the goods that they make.

As with the Bowley squared model in the previous section I have not attempted to create a mathematical model on the lines of figure 4.10.1 above. Again there are a lot of different variables and the base models need first to be benchmarked against real data.

Conceptually however, the models should be straightforward to build. Again, this sort of system is common in control systems engineering, and should be familiar to most office dwellers.

To take the example of air conditioning systems again, an analogous system is where two large air-conditioning units are installed on an open office floor, each with its own independent control loop, set to control at exactly the same temperature. Common sense suggests that two identical systems like this should move up and down together in tandem. However in this case, common sense is wrong.

Unfortunately, although the two units may be wired separately, the flows of air from one part of the building to another mean that the two units are actually influencing each other in what is called a 'coupled system'.

Such a system can very easily become unbalanced, for example if their settings are slightly different or if part of the office is in shade and the other is receiving sunlight.

In the second example the a/c unit in the shady part will provide a little cooling, while the a/c unit in the sunny part will provide a lot of cooling.

Unfortunately, the cold air can then flow from the sunny part of the office to the shady part, while the warmer air from the shady part can flow to the sunny part. In fact convection will make this inevitable.

When this happens the a/c unit in the shady part reduces its cooling, while the a/c unit in the sunny part ramps up its supply of cold air, and the two units end up in an ever-increasing battle to control the temperature. Ultimately, the a/c unit in the shady part may even convert to heating mode. This results in stratified air, bad draughts, general discomfort and very expensive utility bills.

In this case the two a/c units are coupled but end up working in anti-phase; working in opposite directions. This is a common outcome in this type of control system. The same can happen with national economies, though it doesn't have to be the case.

For example, where a large country has good economic links with a smaller country, the smaller tends to move into phase with the larger. This is true for example with Canada and the United States. Although Canada can be influenced by external events such as commodity prices, its economy usually moves closely with that of the US.

The same is true of the many smaller countries around Germany, not only does this include euro users such as Austria, Finland, the Netherlands and Belgium, it also includes others such as the Czech Republic, Denmark, Sweden and the Baltic states. Together, these countries form a linked bloc, with all countries moving closely in phase with Germany.

In contrast, due to their size and different economic fundamentals, Italy, Spain, Portugal, Ireland and Greece have moved into anti-phase with the problems discussed in the model above. France remains uneasily stuck between the two conditions.

The model described in this section is analogous to a competitive Lotka-Volterra model (in contrast to the predator-prey Lotka-Volterra models we have discussed previously).

A competitive L-V model consists of, for example, sheep and rabbits living side by side eating grass on the same island. Depending on the different growth rates and breeding rates animals in these situations can come to different equilibria.

If the animals are similar, say sheep and horses, an equilibrium can be reached with fixed proportions of the two groups of animals.

If the animals are different the equilibrium is unstable and moves to one extreme or the other. So with say sheep and rabbits, depending on the start point, one or other group will dominate and drive the other group to extinction. One group of animals will take over all the biomass, just as in international trade it is possible for one company to take over all the real capital.

Clearly the above model could be adapted in many ways, most obviously by introducing different currencies. Empirical data from the history of failed monetary unions and fixed currencies suggests that independent currencies have a significant effect, largely beneficial. If managed correctly devaluation generally allows beneficial adjustment.

Obviously to introduce currency in international trade models, it first needs to be introduced in domestic economies, this is discussed in brief in section 4.11 below.

4.11 Where Angels Fear to Tread - Governments & Money

I move into a discussion of the theory of money, and the role of governments, with some trepidation. Of all the areas of economics, this seems to be the one in which a religious belief in theory unfounded on empirical fact seems to be most widespread. And discussions in this sphere seem to take on the character of arguments between religious zealots.

Exceptionally, Perry Mehrling writes on this field with great clarity and insight [Mehrling 2000].

It is my belief that an understanding based on flows and stocks, as followed in the rest of this paper could be productive.

It would be possible first to start by looking at commodity money as an actual commodity in line with section 3 above.

Using a commodity, such as gold, in the real world is problematic, because, as Robert Triffin noted, the supply of gold is insufficient to allow expansion of the money supply to keep pace with the size of the economy.

To get around this problem all modern economies have moved to systems of fiat money, generally with inflation targeting or some other control system.

While I have many grave reservations regarding 'Modern Money' theory (see for example [Wray 1998]) I find their central insight of treating money as an artificially created commodity flow as appealing. Diagram 4.11.1 below shows a typical treatment.

Figure 4.11.1 here

The big problems with modern money theorists is their almost religious belief that governments can expand public debt without limit when the economy is below full output capacity. A brief review of [Bernholz 2003], [Reinhart & Rogoff 2009] or [Pettis 2001], shows that the empirical data demonstrates that this is emphatically not true.

As Perry Mehrling [Mehrling 2000] points out very lucidly, the problem with the approach of Wray and others is that the state's ability to pay coupons on government bonds ultimately depends on the states ability to raise taxes, and also on the good use that the state puts those taxes to. In the simplistic examples of Modern Money, a colonial governor in a undeveloped rural economy raises hut taxes to pay for new roads and schools, and this clearly results in substantial economic improvements. That this can be translated into a modern western economy is not obvious. In fact, in industrialised countries, much money raised, whether by taxation or borrowing from private markets, is not invested in infrastructure but instead passed straight through to consumption as transfers. In this light the relationship between government and the private economy would appear to resemble the relationship between a debtor nation and a creditor nation in the Siamese-Bowley models above.

The modern money theorists are surely correct in their belief that a significant amount of government debt is good for the economy as it provides a secure asset that gives needed liquidity for effective private markets. To believe that this debt can be expanded indefinitely is to undermine the most important value this debt has; that of security.

In similar vein I find much of Milton Friedman's monetary theory terrifyingly naïve. However I have found the blogging of 'kitchen-sink' monetarists such as Simon Ward [Ward] and John Hussman [Hussman] enormously insightful and surprisingly able in their predictive power. Friedman's theories, though simplistic were also of course based on flows, and assumed delays in action. So although his formulation was not dynamic, his underlying model, and the data it was based on was.

I am insufficiently skilled to be able to judge whether either or both of the modern money and the monetarist approaches can be synthesised effectively into the modelling framework described into this paper. But I believe it may be an approach worth pursuing.

Another problem with monetary theory is that 'money' can be artificially created by at least two dynamic feedback mechanisms.

The first is the loop of fractional reserve banking that can allow a large multiplier of debt to be created for each sum of reserves pushed into the economy by the reserve bank.

A second multiplier is the endogenous creation of liquidity within the finance system this was seen in the models in section 4, and is discussed at length in section 8.2.1 of this paper.

Taking all the above together, this then ends up with a basic model of the financial system that works something like the diagram below.

Figure 4.11.2 here

This simple model, includes at least two amplification loops and two feedback loops with positive feedback. If housing were included in the diagram, with the leverage of mortgages, there would be more feedback and amplification.

With my control engineer's hat on, the only thing I can say about this as a control system is that if I was trying to design an effective control system, it definitely wouldn't look like the diagram above.

It is about as sensible as trying to control a steam engine with a system made out of cheap rusty shower mixer valves and some lengths of garden hose.

In democratic countries, central bankers are expected to control the whole of the country effectively by controlling the variables on the left hand side. Whatever they are paid, it is not enough.

4.12 Why Money Trickles Up

Before finishing this section on modelling, and moving on to a discussion of background theory, I would first like to revisit the premise of this paper.

At this point I am forced to confess to having committed a major offence that I have accused others of.

I used the phrase 'Why Money Trickles Up' as the title for this paper to give an emotional impact; the title should really have read 'Why Wealth Trickles Up' or perhaps 'Why Income Trickles Up'. I have only discussed monetary theory as a passing aside.

I believe however that I have given an authoritative explanation of both how and why wealth trickles up from the poor to the rich, as well as a detailed description of the mechanisms.

In brief, macroeconomic factors including interest rates, saving/consumption rates and debt define the Bowley ratio; the proportions of wealth returned as wages and profits.

The Bowley ratio then defines the parameters of the General Lotka-Volterra distribution that defines the distribution of wealth between individuals.

This distribution of wealth then defines the majority of the shape of the distribution of income.

That is why money trickles up.

Part B - Some Theory

5. Theory Introduction

Section A introduced a range of possible models to look at some of the basic interactions of economics. Though they may have had inspiration from other sources, the models are my own work.

In many ways the models are naïve and simplistic. Time will tell whether they prove useful or not. If the models survive unchanged I will be pleased, but also surprised. If the models are trashed and replaced I will be disappointed, but not particularly surprised. The accuracy of the models is beside the point.

The point of the models is that by using a set of tools selected from other areas of science in combination with ideas primarily from classical economics and finance, it is possible to create simple effective models that address basic, fundamental regularities in economics. This is the main point of the models. If the approaches of the models above are taken further, but the models themselves are superseded, then I will have achieved the main aim of this paper.

The scientific tools come primarily from physics, biology and pure mathematics. For almost all economists these tools; ideas such as chaotic mathematics, statistical physics, and entropy will be unfamiliar to the point of being quite alien. Even for most physicists, ideas such as the GLV and maximum entropy production will be unfamiliar, and I believe these will be of interest to many working in the field of complex systems whether this includes economics or not.

As to the economics, of course almost all scientists will be ignorant of the basics of economics. Sadly, with vary rare exceptions, even most physicists, mathematicians and modellers researching in economics seem to take a perverse delight in not knowing anything at all about basic economists.

This attitude seems to be something along the lines of "we know all about steel plate, diesel engines, turbo-chargers, power steering, inertial guidance systems, etc – why on earth should we spend our time learning about sailing boats?" However; although sailors take a lot of time and effort tacking backwards and forwards without getting anywhere particularly fast, some of their knowledge is quite useful; for example, where the shoals and reefs are, how to use a compass and sextant, why you should carry a fog-horn, not to mention lifeboats and life-jackets. And why it is a good idea to know how to swim.

In fact many of the economic ideas in this paper will be unfamiliar to many economists. The economic ideas come largely from finance, economic historians and classical and other heterodox economics; including, somewhat to my own surprise, Marxian economics. All of these ideas are outside the canon of mainstream neo-classical economics and so are not just ignored but are politely rubbished, in the case of economic history and finance, or very impolitely rubbished in the case heterodox economics. None of these ideas are included in undergraduate economics courses other than at the most maverick of universities.

As this paper is largely based on non-standard economics, I have gone to some efforts, not just to explain this background, but also to justify it to sceptical economists steeped in marginality and utility theory.

This is firstly to explain unfamiliar ideas to both economists and non-economic scientists. Also, for the economists, it is to explain how many other things, such as liquidity and dynamic scarcity, explain large apparent diversions from the idea of intrinsic value which is inherent in classical economics but absent in neo-classical economics. Once these diversions are understood and correctly modelled, classical economics becomes a much more powerful theoretical method than neoclassicism.

The economic historians such as Reinhart & Rogoff, Shiller, Smithers, Harrison, Napier and Bernholz have the advantage of the long sweep of history to prevent them from accepting high-faluting theory that disagrees with reality. This research shows clear patterns in economics, such as strong cyclical and mean reversion behaviour, that clearly supports Austrian, Minskian and similar views. This clearly supports the theory of intrinsic value, and discredits orthodox economics.

Similarly the inclusion of ideas from finance was not particularly surprising, people working in finance do not have the option of embracing intellectually beautiful ideas that don't describe reality; at least if they wish to remain working in finance. They are obliged to adopt rules of thumb that work. Some of the more thoughtful financiers, people such as Pettis, Shiller, Smithers, Cooper, Pepper & Oliver have then made insightful attempts to explain why these rules of thumb work in practice.

In the field of market-microstructure in particular these approaches have been researched systematically and are both close to regularisation, and are also close to melding with the work of the more insightful financial econophysicists, despite the fact that the econophysicists have approached these problems from a completely different direction.

Like econophysics, market-microstructure is highly mathematised, and very difficult to comprehend on a first reading. Perhaps because of this combination of complex mathematics and inscrutability, most curiously, market-microstructure appears to have been accepted as mainstream economics. This suggests most mainstream economists have never read any market-microstructure, as its rejection of marginality is, though very discreet, absolute.

Which brings us to heterodox economics. Firstly the parallels between market-microstructure and post-Keynesian pricing seem, to this author, both obvious, and of considerable practical importance. Though I stand to be corrected, this parallel does not appear to have been noted previously, presumably because post-Keynesians don't read market-microstructure papers and vice-versa.

The main reason for adopting classical economics was almost accidental. I had previously rejected the dabblings of both Foley and Wright into Marxian economics as misguided foolishness. I was wrong, they were right. My first reason for rejecting Marxian economics was because the labour theory of value is so obviously wrong-headed, the second was because I had believed that Marxian economics had been systematically disproved by neoclassical economics. More reading of economics quickly proved the second assumption to be false, Sraffa was the victor of the Cambridge capital controversies.

The labour theory of value is indeed nonsense. However the concept of absolute value is not nonsense, it is in fact very powerful. The concept of 'negentropy' as value, as articulated by Ayres & Nair [Ayres & Nair 1984] for example, is not just basic common sense; it works as a theoretical approach, as evidenced by the models in part A. Once the labour theory of value is replaced by a "negentropy theory of value", not only does classical economics make perfect sense, it also allows economics to become a self-consistent theory that is an obvious subset of the natural sciences. A very large, very interesting and very important subset; but a subset nonetheless.

In contrast, the fundamental innovation of neoclassical economics; that value is not inherent, but is set in the collective sub-conscious of buyers and sellers has proved to be a spectacular non-achiever.

This assumption also has the worrying theoretical feel that one somehow has to believe in fairies; that the value of a brick or a ham sandwich can dramatically change overnight just because a lot of people believe its value should change.

That is not to say that I have an intrinsic problem with believing in fairies. When studying quantum mechanics or information theory, I find the explanations seem to depend on a worrying existence of an intelligent external observer. Given the assumed existence of quantum mechanics and systems described by information prior to humanity's descent from the trees, I find this worrying.

However I feel obliged to accept both quantum mechanics and information theory because the maths works well, unbelievably well, in describing the characteristics of real world systems.

In contrast, neoclassical economics, despite 140 years of theoretical effort has singularly failed to achieve a single macroeconomic model of the slightest usefulness. Neoclassical theory failed spectacularly to predict the credit crunch of 2008; as it failed to predict the crash in Argentina in 2002 before that, or the failure of LTCM (despite the Nobels) in 1998, the multiple crashes in Asia in 1997, of Mexico in 1994, of the collapse of the European monetary system in 1992, or the collapse of Japan into deflation in the early 1990's.

At the time of writing it is clear that the central banks of the USA, the Eurozone, Japan, the UK, Switzerland, Sweden and others are all following their own significantly different policies, based primarily on experience and intuition. This is because they have no meaningful macroeconomic models. The ones they did have in 2008 have been quietly abandoned, and they are now largely flying by the seat of their pants with a finger in the air to check the weather conditions. Such is the legacy of a century and a half of neoclassical economics.

It is the belief of the author, that the movement instigated by neo-classical economics to subjective value, remains the biggest and most damaging wrong turn ever made in the history of the sciences.

The teaching of chaos, statistical mechanics and entropy is famously difficult. The concepts of liquidity and market microstructure are similarly opaque when first encountered.

Despite this, once the ideas are grasped they are actually quite simple and can become easily understood and then become very powerful tools to understand problems. I have neither the teaching skills nor the space in a paper of this length to do justice in explaining these ideas. What I have attempted to do in Part B is to give a basic feel for the ideas, with very simplistic models and almost no mathematics. I have than also pointed to other authors, authors more skilled than myself, who can give greater depth and clarity than I can.

Finally in section 13 I have included a reading list to point the way forwards into these subjects for mathematicians, economists and other scientists.

In the sections that follow I have included some lengthy quotes from some authors, primarily Duncan Foley, Steve Keen and Ian Wright. This is mainly because they explain some of the points I wish to make very eloquently. In most cases I have then attempted to explain the ideas in alternative ways in my own words. Some readers may not find the extracts easy to follow on first reading. If this is the case I suggest that readers skim these extracts and read my own words, then reread the extracts. It is hoped that the two different descriptions will help illuminate the underlying theories.

It goes without saying that the basic ideas in part B are not my own. The ideas of mathematical chaos, statistical mechanics and basic entropy are centuries old, as are the ideas of classical economics.

Other concepts such as maximum entropy production, market-microstructure, liquidity and post-Keynesian pricing theory are relatively recent; recent enough to be largely unknown in wider physics and economics circles.

My own limited input includes, firstly, occasionally suggesting possible practical examples and uses that emerge from the theory – the ideas are speculative, and whether they actually prove to be useful remains to be seen. The intention of these proposals is to encourage a new way of tackling problems in economics and finance.

More importantly, I believe I have pulled together an apparent rag-bag of ideas, from seemingly unconnected fields, that may allow a systematic approach to be put together that gives economics a strong, coherent, mathematically rigorous basis that transcends the petty boundaries of the many current competing economic models.

Part B.I – Mathematics

6. Dynamics

6.1 Drive My Car

Before moving into the ideas of non-linear dynamics and chaotic mathematics I would like to briefly start with a discussion of the difference between statics and dynamics.

Imagine that you own a car, or better a pick-up truck, a small vehicle with an open space at the back for carrying loads.

For the moment we will discuss what happens when the truck is parked, this is the case were the mathematics of statics is relevant.

If the truck is unloaded it will be high up on its springs, with a big space between the top of the back wheels and the top of the wheel-arch on the body. This is a particular static equilibrium, the force of the gravity and the force of the spring come to a balance at a particular point.

If you then put a dozen bags of cement in the back of the pick up truck, the truck will move down on its springs and the body will move closer to the wheels. This is a new static equilibrium at a different point where the new greater weight due to gravity balances with a new bigger force from the more compressed spring.

Now the truck will also have dampers; shock-absorbers fitted. In a normal pick-up truck these dampers will be quite beefy, and will slow down the movement from one static equilibrium point to another. These dampers provide a frictional force, and from the point of view of the static equilibria beloved of economists, they are very inefficient. They physically prevent rapid movement from one static equilibrium to another. From this line of thinking it would be better to reduce the size of the dampers or just remove the dampers altogether. Then, following a point change in the weight, the truck would move to its new equilibrium much faster.

Using this line of thinking, a neoclassical economist could also point out that, once you started driving you won't be changing the load anyway so you don't need to worry about the dampers as you won't be moving away from whichever static equilibrium you started at.

More thoughtful people will realise that this is not a sensible line of argument. A moving truck is in a dynamic situation. When you set off driving you will need to turn corners and you will sometimes hit bumps in the road, this will set of bouncing in the truck, and you need dampers to slow the up and down movements of the truck. Obviously if you drive down a dirt road, with a lot of bumps, you will need dampers or the truck will bounce about all over the place.

What very few people realise, even very thoughtful people, is that dynamic systems are much more difficult to control than that.

If you take the dampers off a car, and then you drive the car very carefully, down an absolutely flat, absolutely straight road (an airport runway say), within a few tens of seconds the car will start bucking like a bronco and will be almost undriveable. It doesn't matter how carefully you drive the car, the car will rapidly move into a strongly vibrating mode.

The problem is that as soon as you start driving the car, you introduce extra time based equations into the system of mathematics that describes the car. This new system of mathematics, the dynamic model, is completely different to the static solution. It is not an extension of the static model, it is not a modification of the static model. It is a different system with different solutions.

For a car without dampers the solution is similar to the Lotka-Volterra model seen in figures 1.2.1.2 and 1.2.1.3 in section 1.2 above. This solution is naturally unstable and rotates around a central point indefinitely. Even if you deliberately start the car off with conditions at the central point (which would be the solution to the static system), the car's movements will quickly spiral out to the circle of dynamic points. That is because this circle is the solution to the dynamic equations. The central point is not a solution to the dynamic system, so the car cannot stay at this point. The car will have a natural 'resonant frequency' and will move into this form of vibration. Like the Lotka-Volterra system, this vibrational mode is the equilibrium solution for this physical model. In this case the equilibrium is dynamic, it has constantly variable parameters.

If you put the dampers back on the car, then the central point is a solution to the dynamic system, the behaviour of the car then becomes similar to that seen in figures 1.2.1.4 and 1.2.1.5 in section 1.2 above, or to that seen in some of the commodity models of section 3 and the macroeconomic models in section 4. Even if the car hits a bump and starts bouncing, its movements will be damped and will quickly move back to the stable point. That is why cars have dampers, they automatically and very simply change an unstable dynamic equilibrium into a stable dynamic equilibrium.

In a static framework dampers are inefficient, they prevent rapid movement to a new equilibrium. In a dynamic framework, dampers are essential, they move the system from an ever-changing cyclical dynamic equilibrium close to the static dynamic equilibrium.

Similar problems are found in many other systems, a famous example is the Tacoma Narrows suspension bridge ("Gallopin' Gertie") in the United States that was destroyed by the wind (for a little entertainment do an internet search for videos of 'Tacoma Narrows'). Common sense suggests that the wind should not be strong enough to destroy a bridge built of steel. But the wind blew around the suspension cables and induced vibrations in the cables at their resonant frequencies. These then induced vibrations in the bridge at its natural frequency, which eventually built up enough to destroy the whole bridge.

Nowadays suspension bridges are normally built with dampers installed on the cables to prevent vibrations building up, as well as vanes to prevent alternate vortex shedding (similar vanes can usually be seen on tall steel chimneys).

More recently a similar problem occurred with the Millennium footbridge near St Paul's in London. This time the vibrations were induced in the bridge by pedestrians. In this case the pedestrians started movements in the bridge at the natural frequency of the bridge. The movements of the bridge then forced the pedestrians to walk at this natural frequency, so a feedback process built up that caused large movements in the bridge. The bridge had to be closed the day it opened, and stayed closed for some months until dampers could be installed.

Another very elegant example of how dynamic systems can behave in unexpected ways is the example of traffic flows. A video of a beautiful example of a system moving into a stable but chaotic zone of behaviour is given at [New Scientist 2008]. Here a number of drivers were asked to drive in a circle at a constant 30km/h. They signally failed to achieve these very simple instructions An alternative system quickly set itself up with a clear and stable wave pattern of blocked vehicles moving around the system at a steady speed. This system of flows being blocked and forcing rhythmical patterns of fast and slow is exactly analogous to the flows of goods, and changes in prices in economic systems.

For 140 years economists have treated economics as a static system. A Walrasian auctioneer compares all bids and offers in the market and then closes out all purchases and sales at a market clearing price. To compare two different economic points economists use 'comparative statics'. They look at one static point, say 'stationary truck unloaded', and then look at another point, 'stationary truck loaded', and then calculate the locus of movement from one point to the other.

From this view economists conclude that economic systems will quickly and naturally come to an equilibrium, they also conclude that frictional forces are bad and prevent rapid movement to the equilibrium.

In recent years economists have started using what they call 'dynamic' models. With the notable exception of the Goodwin models, these are lots of small stationary comparative static analyses carried out one after the other. This might be better described as 'high-frequency statics', and are equivalent to loading and unloading the truck rapidly with lots of small bags of cement.

Even the Goodwin model is highly confused, attempting to model growth process, presumably long term exponentials, via the Lotka-Volterra model, which although it shows short term growth and decline, is most certainly a long-term stable model, not a growth model.

Certainly none of the 'dynamic' models proposed in recent years have made it into the mainstream textbooks, for the simple reason that the models don't work and don't effectively model anything. To take the two mainstream economic texts cited in this paper, Mankiw [Mankiw 2004] has a dozen or so time based graphs, but all show actual data, not theoretical modelling. There are lots of theoretical graphs in Mankiw, but all are static or comparative static; almost all of them being variations of price versus quantity. Similarly Miles & Scott [Miles & Scott 2002], a much better book, has many dozens of time based data graphs but only one theoretical time based graph; their figure 7.2. There is no discussion of dynamic equilibrium in Miles & Scott, all theory is discussed in a comparative static framework.

A century and a half of neoclassicism has prevented economists getting in the car, turning on the ignition and releasing the handbrake.

Economics is a dynamic system.

Whether it is a trader selling shares on a stock market or shopper buying groceries in a supermarket, traditional auctions are notable by their absence. Prices are never formally closed, prices are settled dynamically in real time. They are set by price setters; market-makers or books in stock-markets, by suppliers in retail markets.

These prices are set by people who look at the prices of competitors, the rate of purchase of goods, the inventory of goods in the shops, the prices of raw materials, etc.

The values of all these items are historic, they are functions of past time.

With a shop, the competitor's prices may have been collected the previous day. For a stock trader the competitor's prices may only be seconds old. But with high-frequency trading, seconds old is definitely pre-historic.

So the most important variable in the functions that are used for setting prices is that of time. Price setting is a dynamic process, with a lot more equations than a static process.

These dynamic systems give feedback loops and often give unstable equilibrium solutions just as with biological Lotka-Volterra systems and car suspension systems.

This is painfully obvious to see in the cyclical behaviour of stock-markets, house prices, commodity prices, currency fluctuations, etc. These fluctuations are inherent in economics. Because economies are dynamic systems. The fluctuations of stock-markets, house prices, commodity prices are a result of natural dynamic equilibria.

Neoclassical economics states that the fluctuations shouldn't exist, and if they do it is a result of frictional inefficiencies. As a result the policy recommendations of neoclassical economists make the fluctuations in dynamic economies worse.

If neoclassical economists genuinely believe that comparative statics is a sensible way to analyse and manage dynamic systems like economies, they should prove it by taking the shock-absorbers off their cars.

6.2 Counting the Bodies - Mathematics and Equilibrium

In his book, Debunking Economics [Keen 2004], Steve Keen puts his finger on the problem at the heart of economics. Economists are using the wrong sort of mathematics when they attempt to solve their problems:

Economics remains perhaps the only area of applied mathematics that still believes in Laplace's dictum that, with an accurate enough model of the universe and accurate enough measurement today, the future course of the universe could be predicted.

For mathematicians, that dictum was dashed in 1899 by Poincaré's proof of the existence of chaos. Poincaré showed that not only was it impossible to derive a formula which could predict *the future course of a dynamical model with three or more elements to it, but even any numerical approximation of this system would rapidly lose accuracy….*

The more appropriate starting point for mathematical models of the economy are dynamic equations, in which the relationships between variables cannot be reduced to straight lines. These are known as nonlinear differential equations. The vast majority of these cannot be solved, and once three or more such equations interact, they are impossible to solve.

Table 1 summarises the situation. Economic theory attempts to analyse the economy using techniques appropriate to the upper left-hand part of Table 1 (with boldface text), when in fact the appropriate methods are those in the lower right-hand part (with cells shaded gray).

Table 1 The solvability of mathematical models (adapted from Constanza 1993)

Equations	**Linear**			**Non-linear**		
	One Equation	Several Equations	Many Equations	One Equation	Several Equations	Many Equations
Algebraic	**Trivial**	**Easy**	**Possible**	Very difficult	Very difficult	Impossible
Ordinary Differential	Easy	Difficult	Essentially Impossible	Very difficult	Impossible	
Partial Differential	Difficult	Essentially Impossible	Impossible			

Or alternatively, as Wright puts it:

The state-space of a system is the set of all possible configurations of the DOF [degrees of Freedom]. A particular configuration is a 'point' in state space. In general we find that many neat systems, if they enter equilibrium, tend toward a point or trajectory in state-space. A canonical example is a set of weighing scales. Place some weights on each arm and the scales will tend toward an equilibrium point in which the internal forces balance and the system is at rest. This is a simple kind of deterministic equilibrium, in which the equilibrium configuration is a subset of state-space. The classical mechanics concept of equilibrium was a founding metaphor of the 19th Century marginal revolution in economics (e.g., see Mirowski (1989)). And it appears in a more developed form in 20th Century neoclassical general equilibrium models (e.g., Debreu (1959)).

But most messy systems, if they enter equilibrium, do not tend toward a subset of state-space.

[Wright 2009]

And, of course, economics is not a neat system; economics is a messy system, economics is a multibody system.

Foley gives this background in more detail:

The concept of equilibrium states has played a decisive role in the development of quantitative sciences. The study of mechanical equilibrium, conceived as a balancing of forces in a static system, clarified the fundamental notions of force and mass in the course of the 17th century development of Newtonian physics. The 19th century saw the emergence of characteristically statistical descriptions and theories of mass phenomena (see Stephen Stigler, 1986; Theodore Porter, 1986) which migrated from the social sciences to physics, where they blossomed into the *marvellously successful and equally marvellously puzzling methods of statistical mechanics (see Lawrence Sklar, 1993). These statistical theories eschew the goal of describing in detail the situation of all the subsystems that constitute a large system with many degrees of freedom in favor of drawing strong conclusions about the observable macro behavior of the system based on statistical considerations. As Edwin T. Jaynes (1978), following the approach of J. Willard Gibbs, realized, statistical equilibrium in all its various applications occurs when the appropriately defined entropy of the system is maximized subject to the appropriate constraints. The entropy is a strictly concave function of the probability distributions describing the system, and the constraints are typically linear or convex functions, so that this maximization implicitly calculates shadow prices (Lagrange multipliers) for each of the constraints, which are uniform over the subsystems and characterize its important properties in equilibrium.*

One might have expected that these statistical methods would be a natural basis for the attempt to put social theory, and particularly economic theory, on firm mathematical and quantitative foundations. It is a commonplace of social and economic methodology to point out that human behavior, no matter how idiosyncratic and unpredictable it is in individual human beings, is subject to statistical regularity and predictability in the aggregate. The Maxwell-Boltzmann-Gibbs methods of statistical mechanics, furthermore, are based on the calculation of dual variables that have the dimension of prices, and effectively view the establishment of physical equilibrium as a kind of economizing process. Thus it would not have been surprising had economic theory developed a statistical concept of equilibrium.

By a curious turn of the history of thought, however, economic theory, despite an almost obsessive fixation on physical models and analogies (see Philip Mirowski, 1989), gave birth to an idiosyncratic conception of equilibrium fashioned more on the mechanical analogy, in the work of Leon Walras, Vilfredo Pareto, Irving Fisher, and Francis Y. Edgeworth (to name a few of the more important figures). In Walras' equilibrium each subsystem (firm or household) deterministically maximizes profit or utility facing uniform prices "cried out" by an "auctioneer". The auctioneer experiments until she has determined an equilibrium price system at which the offers to sell and buy each good in each market are exactly balanced. Because this theory assumes as an axiom that no transactions take place until the equilibrium prices are determined, households with the same preferences and endowment will always receive the same bundle of consumption goods in the equilibrium: horizontal equity (or equal treatment) is guaranteed by this a priori assumption. The Walrasian conception of equilibrium is in sharp contrast to the statistical thermodynamic conception in which the equilibrium energy distribution of subsystems (say, molecules) is achieved by their exchange of energy as they interact during the transient approach to equilibrium. In a thermodynamic context we would be astonished to find that two molecules that started in the same energy state generally end up in the same energy state.

Apparently physicists tried to alert Walras to the peculiar nature of the conception of equilibrium he was proposing, but without success, either because Walras did not understand the statistical point of view very well, or because he considered it and rejected it on other grounds. J. W. Gibbs served as Irving Fisher's thesis adviser at Yale apparently without raising questions about the non-statistical conception of the equilibrium systems Fisher was studying. Francis Edgeworth distrusted Walras' conception of the auctioneer enough to propose an abstract combinatorial model of exchange, based on the idea of recontracting among coalitions of traders (which has developed into the modern theory of the core). The recontracting feature of Edgeworth's theory, however, implies equal treatment of agents with the same preferences and endowments, thus reproducing the key elements of Walras' system.

One aim of Walras' and Edgeworth's theories was to explain the emergence of coherent market price systems from the decentralized interaction of atomistic traders. Unfortunately, both Walras and Edgeworth resort to strong and unrealistic assumptions to address this issue: Walras *invented a fictional information centralizing auctioneer, and Edgeworth posited costless recontracting among agents. The statistical approach offers an elegant alternative in this respect: market prices can be regarded as the shadow prices or Lagrange multipliers arising inherently from entropy maximization. In this view the system constraints (market clearing conditions) give rise to global prices just as the constraints of volume and energy in a physical system give rise to the emergent properties of pressure and temperature in a confined gas. The atomistic agents in a market "feel" the effects of these global constraints combinatorially as the relative difficulty of changing their holdings of goods, just as individual molecules "feel" the global constraints on energy and volume in terms of the likelihood of reaching any given energy state.*

[Foley 1996b]

Few physicists read economics books.

Even the physicists who are profoundly interested in economics, and produce papers on economics, rarely read economics books.

The main reason; for the scientifically trained, is the extraordinarily unscientific approach that they have. Statements such as 'Assume a demand curve……', 'assume a budget line………', etc simply inculcate an overriding feeling of 'why?'. Where on earth do these assumptions come from, and why should they be assumed.

For more intrepid physicists who persevere, it comes as something of shock to discover that utility theory was directly copied from the field theory of physics in the 1870's, and copied with gross errors. More extraordinarily, having absorbed field theory and adopted it as the core of economics, economics has studiously ignored the majority of mathematics developed since the 1870's (game theory being a notable exception) even though this mathematics would be much more appropriate for the analysis of economics.

In this regard economics resembles a tenacious terrier, unable to eat the plates of meat set down in front of it, due to its inability to let go of the very well chewed bone it has firmly gripped in its teeth.

The full horror of this calamity is recounted at length, in very entertaining detail, in Mirowski's book 'More Heat than Light'; a book that, contrary to its title, many economists might find enlightening reading. [Mirowski 1989]

The central point of Mirowski's book is that utility was copied from field theory, but in doing so economists threw away the basic conservation principles that give field theory any meaning. If fields are not conservative, then there is little point in drawing curves and lines to visualise them. Without conservation laws, two different paths between the same two points will give different values, and so the curves and lines do not have values that can be meaningfully represented; neither graphically nor mathematically.

The second problem with field theory as a basis for economics, is that it is simply, and absolutely, not appropriate for multibody systems.

In their different ways; gravity, electromagnetism, relativity and quantum mechanics are all varieties of field theory. But in the application of their mathematics interactions are limited to two bodies, so for example an electric current can be seen as a unified flow of the separate electrons, moving at the same speed in the same direction.

Newton's theory of gravity was the first and the classic description of field theory, and with two bodies; the sun and a single planet for example, Newton's theories work perfectly.

But even with a very simple multibody planetary system, Newton's theories break down, and fail to explain behaviour exactly. The errors are small, but the errors are there.

As soon as you get to three bodies; for example the sun, earth and moon, it becomes impossible to find exact solutions for the motions of the bodies. Even in a three body system the motions of the bodies become chaotic and unpredictable at a detailed level.

In 1890 Poincaré demonstrated that it is actually impossible to solve the equations for a three body system in a simple field system, so even a system as simple as the sun, moon and earth is chaotic, and can not be accurately predicted over the long term.

This, and a full history of analysing the motions of the planets is written up in the very enjoyable book by [Peterson 1993], Poincaré's work is discussed in chapter seven.

It is important to note that this chaotic motion is noticeable in objects as large as planets. This is not simply the chaos of quantum effects or the stochasticity found in Black-Scholes. This is 'deterministic chaos' or usually 'chaos theory'. The chaos is present even in problems that can be described in exact mathematics and are completely free from random exogenous or microscopic behaviour. The original Lotka-Volterra model is just such a mathematical system. In practice the meeting of foxes and rabbits will have a stochastic element, but the system at a macroscopic scale is described very well by deterministic equations. In deterministic chaos, the behaviour of the system can change dramatically according to very small changes in initial conditions, as described in the analogy of butterflies causing tornados a continent away.

However it is of course obvious, that although the positions of the earth, sun and moon can not be predicted exactly, they can be predicted to a very high degree of accuracy, and that their paths follow strongly constrained bands.

This is a different type of equilibrium, a constrained chaotic equilibrium, that never stabilises at a fixed point, and so never becomes a static equilibrium.

The Lotka-Volterra equilibriums (but not the General Lotka-Volterra's) fall into this class of equilibrium.

So in a simple eco-system, the number of rabbits and foxes can vary significantly, but a peak in the population of either will be followed by a trough; and the long term average values of both populations will be very stable.

In economics, Minskian, Austrian and Goodwin type systems fall into these categories, and the commodity and macroeconomic models discussed in sections 3 and 4 above attempt to model such systems.

Such systems can show different behaviour depending on their underlying characteristics. The systems can be very stable staying close to the long-term averages, they can oscillate strongly, or they can grow explosively to infinite positive or negative values.

And of course, real economies clearly follow the same patterns empirically. Business cycles have been evident and documented for at least two centuries. The periodicity may have changed as economies have changed, but the fluctuations remain. These can be short term cycles of building up and drawing down inventories, they can be the 15-20 year land cycles documented by Harrison [Harrison 2005], they can be the decadal mean-reversions of stock prices documented by Smithers and Shiller [Smithers 2009], they can also be the once per lifetime financial crises such as the great crash or the credit crunch caused by the retirement of all the people who remember the reasons why strict controls were imposed on the financial system after the last such crisis [Napier 2007].

And in the great crashes the system moves out of periodicity into explosive behaviour.

Fortunately in the last half-century or so there has been a great deal of progress in analysing such systems in the field known as 'nonlinear dynamics' and there are many standard ways of solving such problems.

In fact the Lotka-Volterra system is one of the simplest such systems and strictly is not necessarily non-linear, though in my models a little non-linearity has been introduced.

There are two big reasons, and one small one, why economics needs to use the mathematics of non-linear dynamics.

The first reason is the inclusion of time as a variable.

In comparative statics prices change with supply, and prices change with demand. Equilibrium is reached when the prices match each other and supply equals demand. The mathematical derivatives for the equilibrium relate the prices and the quantities.

In the real world prices cannot change instantaneously, the main derivatives of prices are with respect to time. The economy is constantly moving with a continuous series of trades, the economy rarely formally 'clears' prices. This is true even for goods such as cheap manufactures that show strong price stability, this is equivalent to a car moving smoothly down a motorway at constant velocity, it is not equivalent to a parked car. If you put a brick under the wheel of a parked car, a new equilibrium point will be reached in a couple of seconds, if you drive over a brick while doing 70mph, it might take a little longer for a new equilibrium to be reached.

In real economies the most important derivatives are the time derivatives, and the mathematical framework for economics must be cast in these derivatives.

Adding in the time derivatives allows extra degrees of freedom and complexity, and normally moves the real equilibrium away from the static equilibrium, it also allows oscillating and explosive solutions that do not have a short-term or any equilibrium respectively. The analogy between stock-market crashes and normal (eg car) crashes is a mathematically exact one. Comparative statics states that a temporary liquidity crisis should not bring an economy to its knees, in the same way that putting a brick under the wheel of a parked car should not destroy the car. However if the car is doing 70mph, it is quite likely that the car will end up wrapped around a lamp-post. Similarly a liquidity crisis in a debt-laden economy can turn into a general solvency crisis.

The most obvious way that time is important to the economy is with the delay of installation of capital in capital intensive sector and also with housing and office building. But time delays can be much shorter and still have strong effects, the research of Milton Friedman showed that monetary effects had delays of six months or more. Inventory stocking cycles operate on similar timescales. In financial markets time delays allow momentum effects on the scale of seconds.

The second big reason that economics needs non-linear dynamics is that the variables in economics have two-way effects (and as discussed above, the effects are fed back with time delays).

These mutual feedback loops are legion. For example:

Increasing prices of company shares creates new apparent wealth - new apparent wealth allows people to invest in companies, so pushing up share prices.

Increasing wealth in the productive sector allows more consumption – more consumption allows increased investment in the productive sector.

Increasing debt allows more liquidity and rising asset prices – rising asset prices gives more apparent capital against which more debt can be secured.

A decrease in saving propensity gives a boost to consumption and the productive sector – more earnings from the productive sector allows a decrease in saving propensity.

In all these cases, and many, many, more, economics has mutually reinforcing feedback loops. And in all these cases the feedback can reverse and work in the opposite direction.

In all these conditions you have coupled systems with feedback, where:

$$dx/dt = f(x,y) \qquad \text{and also} \qquad dy/dt = f(x,y)$$

In these systems y gives feedback to x, and x gives feedback to y. Even with linear systems this can give periodic and explosive behaviour.

All of these are analogous to the lynx and hares in the original model discussed in section 1.2 the populations of both can expand or contract over long periods before an external limit changes the direction of growth.

The imposition of limits brings us to the third reason for using non-linear dynamics. Some functions in economics are non-linear.

The most obvious ones are when you have genuine scarcity such as a fixed supply of labour or urban land suitable for house-building. Minerals such as gold, copper, platinum or oil also have scarcity, at least in the short-term, as installing capital is expensive and takes time. In finance, access to credit and other financing can be limited beyond a certain point and can lead to highly non-linear functions.

A very good text explaining these approaches, with lots of practical examples, is 'Nonlinear dynamics and Chaos: with Applications to Physics, Biology, Chemistry and Engineering', by Strogatz [Strogatz 2000], a good alternative is Hirsch, Smale & Devaney [Hirsch et al 2003]. Prior to either of these books, chapter eight of Keen gives a very good brief introduction to chaotic systems, Ruhla also gives an excellent introduction with a little more maths [Keen 2004, Ruhla 1992].

Although the approach may seem very new to most economists, actually the techniques are extensions of techniques familiar from basic economics. Most non-linear systems are not directly solvable, so mathematicians often resort to graphical representation in 'phase space' to resolve the problems. This ends up with intersecting lines and curves not dissimilar (and a bit more fun) than the diagrams found in comparative statics. Jacobian matrices, for example, appear a third of the way through Strogatz.

Although dynamic systems can be very complex and are often mathematically insoluble, there are standard approaches to analysing these systems, and it is usually possible to produce important mathematical conclusions out of such analysis. It is usually possible to identify the controlling variables and the different zones of stability and instability.

Indeed one of the interesting things about complex systems is that while they can be very difficult to analyse and describe, they are usually very easy to control. Usually it is just a question of installing suitable damping or time delays in the system. In engineering such systems are commonly encountered within control systems where problems of feedback can be highly deleterious.

On the plus side, control system engineering, and systems dynamics, have investigated the problems of such systems in detail, and when the underlying characteristics of the system are understood, relatively minor changes in the system can result in dramatic changes to the stability of the system. See for example Control Systems Engineering [Nise 2000].

In the following two sections, and also in section 9.2 I take a qualitative look at house prices and share trading and ideas of how the natural cycles in these markets could be damped out. Section 9.2 is somewhat out of order in the paper, this is because it is necessary to introduce some ideas of market microstructure first.

The ideas in these sections are pretty much common sense on the issue of housing, the ideas regarding share trading are much more speculative and contentious.

The main point of the discussion is to make it clear that, counter-intuitively, just as with shock-absorbers in cars, introducing damping can create a better system.

6.3 Chaos in Practice – Housing in the UK

It is a common aphorism of economics that it is a difficult science to progress, as it is not possible to carry out suitable experiments. This is tosh.

Experiments are regularly carried out in economics, though usually by accident. The problem is that economists ignore the results, even when the damage to the public is substantial.

The example of housing provides one of the clearest and most important experiments ever carried out in economics in the UK.

Figure 6.3.1 here

Figure 6.3.1 above shows the prices of housing in the UK from 1953 to 2010, divided by the average wage, prepared using data from the Nationwide Building Society and the UK Office of National Statistics. The high house prices immediately following the Second World War were a consequence of substantial loss of housing during the war and a suspension of house construction for the six-year duration of the war.

During the 1950s and 60s access to mortgages in the UK was tightly regulated and controlled by government micro-management of financial institutions, with direct lending ceilings imposed on banks and building societies; resulting in strict rules on eligibility, deposit sizes, etc.

During this period house prices showed remarkable stability at a cost of roughly 3.0 to 3.5 times average salary. It is very important to note that, despite the strong state controls on access to housing finance, the 50's and 60's were a time of substantial private house building in the UK, as the post war generation, including large sections of the working class, fled their city terraces for suburban semis. Despite the restrictions imposed by the state, even at these regulated 'low' prices, demand created lots of supply.

As can be seen in figure 6.3.2 below UK private house building reached a prolonged peak in the mid 1960s.

Figure 6.3.2
[ONS 2004]

Access to mortgages was liberalised in 1971 under the policy of 'Competition and Credit Control', which, despite its title, pretty much abandoned credit control; in line with neoclassical theory. This resulted in the 'Barber boom', starkly clear in figure 6.3.1, stimulated by the resulting rise in liquidity, and the first, of many, UK house price bubbles.

From the 1970's onwards, the UK housing market has been characterised by vicious cyclic booms and busts, with a very clear reversion to the pre-Barber long-term trend or 3 to 3.5 at the bottoms of the cycles.

These cycles are identical in form to the ones discussed in the commodity models in section 3 and the macroeconomic models in section 4. Compare figure 6.3.1 (or 6.3.3 below for the US) with the outputs in figures 3.3.2 and 4.3.3 in previous sections. These are exactly the outputs you would expect from a non-linear differential system that is showing quasi-periodic cyclical stability. In fact, if you look at the pre-1971 section it is possible to see the same cyclical fluctuations, just that the amplitude of the cycles is very much smaller.

It is important to note that at the bottom of both the actual housing data, and the commodities models, prices reach their 'real', 'fundamental', Sraffian values. At these prices the value of housing represents the cost of the inputs. The same can be seen even more clearly in data from the United States (this time deflated for cpi); see figure 6.3.3 below.

Figure 6.3.3 here
[Shiller 2010]

Supply is capable of balancing demand at these Sraffian prices. Any increase above these prices is pure speculation and rent-taking.

Indeed the persistence of these cycles is deep within the economy of the UK. In his book 'Boom, Bust, House Prices, Banking and the Depression of 2010' [Harrison 2005] (first published in 2005) Fred Harrison not only confirms how trivially easy economic forecasting is if you are willing to believe in fundamentals and cyclical behaviour, but also shows that the cycles in the UK go back to at least the middle of the eighteenth century.

As an experiment, you could scarcely ask for clearer data output. The basic system dynamics are substantially and dramatically changed following a point change in policy. Not only that, but this experiment has controls, Germany and Switzerland for example, have retained strict controls on mortgages for house purchases and don't suffer from strong cyclical booms and busts in house prices.

The consequences of this experiment are of some considerable importance to the welfare of all people living in the UK.

Figure 6.3.4 below has the average value of house prices included for the two periods.

Figure 6.3.4 here

On the scales used, average house prices from 1955 to 1970 were 3.3 times average salary.

During the last thirty years, from 1971 to 2009, average house prices in the UK have cost an average of 4.0 times average salary.

In the latest boom, prices have gone to even higher levels, though a meaningful average can't be given until the cycle has bottomed.

The net result of the liberalisation of credit in 1971 was the increase in average cost of housing for all Britons by roughly 23%. In the last cycle, from 1996 to 2010, prices were fully 40% higher than the '55-'70 baseline rate.

This represents a very significant reduction in welfare for residents of the UK. It also has many secondary negative effects. Many more poorer people are unable to afford housing, and are forced to rely on social housing and subsidies paid from taxation. This then helps to create ghettos of poorer people, which exacerbate employment and crime problems, which again requires more social spending and higher taxation.

Even for the well-off that can still afford to buy houses, on average they must spend more money on housing, reducing that available for saving, pensions, or simply enjoying life.

The beneficiaries here are the financial companies that issue the mortgages, or rather the investors and savers with these companies. Once again, exactly like the commodity cycles in section 3, We have a case of unjustified rent-taking on a massive scale. Given that private sector rents are substantially set by house prices, some of the rent-taking is literal. Taken as a whole, this represents a large transfer of wealth from the poor and middle income individuals to the rich.

Housing suffers from the same problem as capital-intensive commodities, as modelled in section 3 above. Construction of housing takes a finite time, and so house prices can go up significantly before market mechanisms have time to work. Unfortunately, housing also has the same problems of endogenous liquidity creation that is seen in the macroeconomic model. As house prices go up, people feel richer, and also as with shares 'momentum' kicks in, and house prices, and the economy as a whole keeps rising, until finally house prices become unaffordable for new entrants in the market, and the bubble bursts. As a capital-intensive industry, housing is naturally cyclical.

Although this conclusion is based on casual observation, it seems that housing seems to be much more dangerous to the overall economy than other asset classes. Booms in commodities and shares seem to be survivable when they turn into busts. Normally such collapses are followed by recessions and rebalancing for a couple of years, and then the economy picks up again. Housing crashes seem often seem to morph into financial crises, threatening the stability of the whole economy, and recovery from such crises normally takes much longer. It seems likely that this is because housing is the only highly-leveraged asset generally available to the public.

This again shows that the contrast between the comparative statics of neoclassical economics, and the real world of dynamic differential equations is stark.

With comparative statics it is easy to 'prove' that credit controls and other government interventions 'must' increase the price of goods, and so reduce the welfare of the public. So neoclassical economists always push for removal of such controls.

In the real world, where speculative cycles can be endogenously created within the economic system; credit controls and other 'interferences' in the market work beneficially by 'damping' the cyclical behaviour. It may be counterintuitive, but in the right circumstances, applying controls and apparent 'costs' to the market actually reduces the price of goods. And reduces them substantially. In the area of UK housing, the experimental data shows that the reduction would be over 20% if strict credit controls were reimposed tomorrow as they were in the '50s and '60s.

It is essential to understand that the logic of this argument is supported by the experimental data of figures 6.3.1 and 6.3.3. It also happens to be supported by the mathematical models, if you understand the right maths, but that is a secondary issue. The experimental data is clear; credit controls reduce the cost of houses, by very helpfully damping, and largely removing, the cyclical nature of house price movements.

If you reject this experimental data, and hold on to a theory that states, purely on theoretical logical grounds, that removing credit controls must make house prices cheaper, then you are not following science. You are following a religious dogma.

Again neoclassical economists, by failing to understand basic dynamic systems, accidentally support massive rent-taking by insisting on deregulation of markets in search of nebulous market efficiencies.

The 'Barber Boom' of the early 1970's ended with a spectacular crash and the 'secondary banking crisis' in which the Bank of England had to launch the 'lifeboat' to rescue thirty or so banks in the UK's very own dry run of the credit crunch. Despite this early warning, deregulation was not rolled back, but instead was systematically pursued in all areas of UK finance and economics. The results can be seen in figure 6.3.1, recurring housing bubbles in UK housing of increasing size and ferocity.

The strength of this religious dogma is quite profound. Since 1971 the UK has had ten chancellors and eight prime ministers, all advised by what must be many hundreds of the most intelligent economists that work in the UK. Despite this the 'reforms' of 1971 have never been questioned, never mind reversed. The citizens of the UK are consequently still obliged to spend their lives paying off their expensive mortgages. The worst economic experiment carried out in the UK in modern times continues.

The damage that this dogma has done to Britain is writ large in figure 6.3.4. From the early 1970s onwards, the liberalisation of credit has increased house prices in the UK by 23%. Another more subtle problem can be seen in figure 6.3.2. Private sector house-building continued at a roughly constant rate from the 1960s to the present. The liberalisation of finance failed spectacularly in encouraging new house-building; presumably because its main effect was to make houses more expensive.

What did change in the 1970s was the collapse of the provision of social housing. From the mid-1970s onwards the government reduced funding for social housing, primarily because, from the 1970s onwards, the UK has had ongoing severe budget problems. This was due to a dramatic increase in the need for welfare payments compared to the 1950s and 60s. The need for welfare payments was needed to cope with the dramatic rise in unemployed and the poorly paid in the 1970s, a problem that has never gone away. The blame for the steep rise in the poor in the 1970s has been blamed variously on oil price shocks, de-industrialisation, union power, foreign competition, etc. While all of these factors may have had contributions, it is the belief of the author that the main factor was the ongoing deregulation starting in the Barber era. This increased overall debt levels and changed the Bowley ratio and so the GLV distribution. This not only created the poor, but forced higher taxes on the rich.

It is perhaps time to end this experiment. Unfortunately the political drive for deregulation is powerful.

The biggest problem, at least in Anglo-Saxon countries, is that many people believe that housing is a good long term investment.

Going back to figure 6.3.1 or 6.3.3 for the UK and US it is clear that the 'investment' value of housing is a chimera. Over the long term, growth in the value of houses is derisory and barely keeps up with the growth in earnings.

Stock market growth is typically 5% higher than this.

Smithers discusses the dual properties of housing as both a form of consumption and investment in Wall Street Revalued p 107-108 [Smithers 2009]. The fact that housing is fundamentally consumption is demonstrated by the continuous reversion to a fixed proportion of wages. Equally this demonstrates that, for all the apparent growth in the booms, housing is a lousy investment, which over the full business cycle only manages to match the increase in wages.

Figures 6.3.1 and 6.3.3 show clearly that in the long-term housing is a proportion of wages, and behaves as consumption. Governments should treat it as such, and actively prevent houses being treated as investments, and most certainly should prevent them being treated as speculative investments.

Despite this the booms are usually longer than the crashes, and inflation often masks real falls in house prices. Both of these effects may explain the visceral attachment of the public, and worse politicians, to housing as investments. Historically, politicians have invented many ways of subsidising housing purchase; so assisting bubbles to form, and so unintentionally, and perversely, making housing more unaffordable. In the recent credit crunch the US did this so effectively as to put the financial system of the whole world at risk of collapse.

Politicians are a very big part of this problem. They seem profoundly addicted to housing booms. Encouraging home ownership is always popular, though if people don't have the wealth or income to maintain the homes they purchase, home ownership alone doesn't solve any problems. More worryingly politicians seem to enjoy the public's enjoyment of rising house prices. Very few politicians seem to be able to comprehend that house prices cannot rise above gdp growth rates over the long term, neither do they seem to appreciate that long-term rising house prices necessarily produces high, and ultimately unaffordable house prices.

This is puzzling. Whether you are a dyed in the wool socialist or a radical free marketeer, it should surely be the aim of any politician to ensure decent affordable housing for all.

In addition to the problem of the housing cycle causing over priced-houses, there are other very major issues. Firstly the diversion of resources to the housing sector that would be better used elsewhere, secondly and more importantly, as Harrison has shown, the cycles in housing appear to be the main driver of the cycles of boom and bust in economy as a whole.

One of the central themes of this paper is that governments should assist in the transfer of capital to poorer people. But housing is not productive capital, and it is the wrong target for such transfers.

Of course, housing can be a very good short-term investment if you get your timing right. Anybody who bought in the UK in 1970, 1978, 1983 or 1996 will almost certainly make a substantial unearned profit when they sell.

But this of course is simply speculation, and speculation in it's non-healthy form. This represents a transfer of wealth to the well informed, and usually already wealthy. This is wealth that is removed from the hands of ordinary people.

And this gives another big problem with allowing cyclical behaviour in economic systems. Most people buy without addressing the timing of booms and busts. If you are lucky and buy at the bottom you win, if you are unlucky and buy at the top you lose. As such allowing this cyclical behaviour in the housing market allows massive inter-generational transfers of wealth on a completely arbitrary basis.

Looking both at the UK data and the US data in figures 6.3.1 and 6.3.3, a very worrying development is that in both countries the size of the booms is steadily rising, though the falls back to normal are the same. From a controls point of view this is very worrying, it suggests that the cycles could be even more dramatic and dangerous in the future – as if the last two years were not traumatic enough.

Faced with a dynamic, cyclical system, standard control systems knowledge can be used to control the system. There are two ways to remove cycling (what engineers call 'hunting') in a control system.

One is to use deliberate counter-cyclical feedback; most central banks try to do this using interest rates to control the economy as a whole. As central bankers are only too aware, this is not an easy way to control anything. A good example of such a feedback loop is a domestic shower system. A combination of a difficult to use mixer valve, and the delay between making the change at the tap and feeling the change in the water temperature often results in alternating flows of water that is too hot or too cold .

Wherever possible, a much better solution is to use damping of the cycle. When done successfully this can result in a dramatic drop in oscillations with fairly minor, adjustments to the system. This is like the example of using shock absorbers with a car's wheels to prevent the car vibrating wildly on its springs every time it hits a bump.

The strict credit controls used in the UK prior to 1971 provided just such an effective damping system. If all else fails it is imperative that such controls are reintroduced in the UK.

However it may be possible that less draconian measures may be just as effective.

As a rule of thumb, to be effective, damping measures need to have a time span of a similar order to that of the natural cycle time of the system, as a minimum they should be of a length of half a cycle or so. For the UK Harrison [Harrison 2005] shows strong evidence for a fifteen to twenty year cycle for house prices. Sensibly, damping measures need to be of the order of ten years or so.

Looking closely at the US data in figure 6.3.3; there is the same flat trend as the UK at the bottoms of the cycles; showing the same reversion to real, non-speculative, prices. It is also clear that the booms are a relatively new phenomenon.

A subtly different experiment has been carried out in the US. The change in behaviour of the housing market appears to be correlated with the rise in non-standard mortgage products. Historically the US has used fixed-rate mortgages, only moving to adjustable rate mortgages comparatively recently. In the UK adjustable, or short term fixed mortgages have been the norm for many years, and it is very difficult to get fixed rate mortgages of more than five years.

The finance industry does not like fixed-rate mortgages. It leaves the issuers holding interest rate and inflation risk. Moving to adjustable rates gives the appearance of moving the risk to individual mortgage holders. This in itself is a practice to be questioned in a democratic society. Why sophisticated finance companies should be allowed to offload complex financial risk onto individuals with little mathematical, let alone financial, training is not clear.

In reality, offloading risk in systemic fashion like this simply creates systemic risk. As has been made abundantly clear in recent years; ultimately the only realistic holder of systemic risk is the taxpayer. Allowing financial companies to issue variable rate mortgages is to give the financial companies government subsidised one-way bets.

Figure 6.3.5 below gives a comparison of mortgage types issued in various different countries in Europe.

6.3.5 here

[Hess & Holzhausen 2008]

The mainly variable countries are Greece, Spain, Ireland, Luxembourg, Portugal, Finland and the UK. This pretty much speaks for itself.

The solution to this is trivially straightforward. All loans that are secured against domestic property should be limited to a ten-year minimum and a thirty year maximum. They should also be fixed rate, or, as a minimum, be a fixed percentage above rpi or cpi, throughout the period of the mortgage. This would move interest rate risk back on to the shoulders of the finance industry. Where it belongs.

Variable rate mortgages should be strictly illegal in any self-respecting democracy.

There are other sensible mechanisms to reduce the use of houses as investments, especially as speculative investments. The most obvious one is to have a capital gains tax that is more punitive than that for other investments. The tax should be charged on all houses, including first homes, without exception. Sensibly this would be a tapered tax; starting at say 20% for the first year, then drop by two percentage points per year, so reaching zero after ten years.

A much better approach would be to have a sales tax on all houses. This should be applied to the seller of all houses, whether they have increased or decreased in value. Again, sensibly, the tax should be tapered over the years.

A tapered capital-gains tax or house sales tax, with a ten-year taper should bring in the damping of the sort required to deal with a 15 to 20 year endogenous property cycle. People buying houses to live in would not be punished, speculators would be.

In addition annual property taxes, or land taxes, should be charged on the value of houses or on the value of the underlying land, rather than on the occupants, as many local taxes are.

Another sensible policy would be to have compulsory mortgage indemnity guarantee (MIG). House purchasers would be obliged to take out insurance to cover full potential losses against potential negative equity, ie the difference between mortgage loan value and likely sale value of house. Such insurance would be cheap if the purchaser had a large deposit and prices were below the long-term trend. The insurance would be very expensive if the deposit was small and it was the height of a boom. As such, compulsory MIG should act in a strongly counter-cyclical manner. (For an off topic discussion of a different sort of deposit protection, refer also to the endnote 6.3.1 below.)

Many countries enforce minimum deposit requirements [Hess & Holzhausen 2008]. This seems a very sensible policy, as those with small deposits are far more likely to default, see for example figure 6.3.6 below.

6.3.6 here

[FT/S&P 2010]

It can be seen that arrears rates increase dramatically as deposit sizes reduce. As with variable rate mortgages, when governments allow financial institutions to offer low deposit rates; that is highly leveraged asset purchases, they allow financial institutions to offload their risk onto the state.

There is a more sophisticated and better way of addressing this particular risk problem. Rather than prescribe laws on deposits, a more effective law would define a maximum limit of say 80% of the sale value of a house that could be repaid to pay off debt secured on the property.

So if a homeowner was foreclosed on, and their property was sold off, a minimum of 20% of the sale proceeds would go to the homeowner, and the other 80% would be shared by all the creditors who have loans secured on the property. This would have a number of advantages. It would have the same effect as a minimum deposit requirement of 20%. Banks would generally be reluctant to supply a mortgage of greater than 80% of the value of the house. It would also make it much more difficult to evade the minimum deposit rules by taking out secondary loans secured on the house.

More subtly it would also act in a counter-cyclical manner. When house prices were at historical lows, banks might be willing to lend 90% mortgages, confident that house price were likely to rise. Conversely, when house prices were significantly above their long-term averages banks would require larger and larger deposits due to their fears that house prices might drop in the future. Similarly they would be very reluctant to allow mortgage equity withdrawal.

In addition to the passive management techniques discussed above, there is also a strong case for active counter-cyclical monitoring and management of the economy by central banks and other monetary authorities.

Despite protestations to the opposite, housing bubbles are very easy to spot.

The first obvious measure is that shown in figures 6.3.1 and 6.3.3 for the US and UK. The ratio of house prices to median wages shows very strong patterns of reversion to mean.

Similar patterns are also seen in ratios of housing costs to rental costs. When house prices are correctly valued, housing costs (mortgage payments, etc) are close to rents on equivalent properties [FT 2010].

If either of these ratios increases significantly above the long-term trend then you are moving into a housing bubble.

At this point the central bank should intervene to prick the bubble as early as possible. This could be by increasing the sales tax or capital gains tax on houses, increasing deposit and MIG requirements or by imposing a tax on mortgage debt.

Finally, if none of the above work effectively to damp markets then the necessary solution is to simply bring back the same credit controls that the UK had prior to 1971.

It would also be wise to impose similar controls on commercial property, especially office accommodation, which also seems to be subject to dramatic fluctuations with the business cycle.

Of course, many economists, banks, building societies, estate agents, and most politicians will believe, and argue vociferously, that bringing in control measures such as those above will slow the economy and make homeownership available only to the few.

These people are wrong. The economic theories are wrong.

Experimental data confirms that these theories are wrong.

When listening to these people it is important to bear in mind that it was the very same economists, financiers and real estate professionals that created the recent housing booms, and the consequent crashes in the US, UK, Ireland and Spain.

Both housing and commercial building are very important as candidates for effective damping for two very big reasons. Firstly as leveraged assets the busts following the booms can be very financially damaging. Secondly housing and commercial construction have very big impacts on employment in the construction industry and so have large effects on the economy as a whole.

[6.3.1 An Aside on Deposit Insurance

Talking about deposit insurance, but wandering completely off-topic; it has puzzled me as to why compulsory default insurance is not instituted for bank deposits.

This would not be intended as a realistic way of insuring the deposits, but as a way of introducing market pricing into the risk of government bank deposit insurance. If done correctly this would also reduce the moral hazard element of public assurance of bank deposits.

Realistically, in a democratic capitalist society, a government run central bank will always need to be the lender of last resort and will need to guarantee the deposits of members of the general public to a basic level.

However, such guarantees remove all risk for all but the richest members of the public. It encourages them to move their deposits to the highest interest payers without any need to worry about whether the bank is well run or in danger of collapse.

This then encourages all banks, even the well run, to compete on interest paid while ignoring the risk taken. Indeed the well run banks are forced to match the foolishness of their badly run competitors if they wish to stay in business.

A way to resolve this is to insist that all deposit-taking banks apply compulsory deposit insurance on their deposits. The insurance would be strictly in the form of a percentage charged on the deposits, and this would be displayed in parallel to the interest rate paid by the bank.

It would be illegal for a particular bank to offer its own insurance on its own accounts, and it would be compulsory for banks to offer all alternative insurance from all alternative deposit taking banks.

Bank customers would be able to swap their insurance simply and electronically at any time they wished, from a visible list of alternatives available via the account.

All deposit taking banks would be obliged to offer a price for insurance for all their competitors. They may wish to price their insurance at a high level, but they would be obliged to price, and would be obliged to take on the insurance at the price offered.

In the event of a bank failing, the insuring banks would be obliged to pay the deposits of the insured depositors from their own bank's funds (to avoid spreading systemic risk, reinsurance of this risk would be prohibited; banks would be obliged to carry a portion of funds against these risks on their balance sheets).

The central bank would remain the ultimate insurer of the deposits but would only step in if there was a pattern of systemic risk, and even then only after bank shareholders and all bondholders were wiped out. In the event of a single bank failure due to poor management, the other banks, the insurers, would carry the costs by themselves.

Further rules would apply even in the event of systemic failure. Government deposit guarantee would apply up to a maximum limit (say £100,000), but this maximum guarantee would apply across all deposits for a single person, no matter how many accounts failed at any number of banks. The maximum paid out would be £100,000 even if the person invested £10k in each of 20 different accounts, all of which failed simultaneously. Similarly the government deposit guarantee would only cover £100,000 maximum over any 10-year rolling period.

Individual bank customers would only be able to waive the compulsory bank insurance where they could demonstrate that they already had £100,000 deposited in insured accounts.

Although the above may sound complex, it would be trivial to put in place in a modern electronic retail banking system.

The net effect of this would be to create a market in retail bank deposit insurance. While the Bank of England may have been surprised by the collapse of Northern Rock, Bradford & Bingley and HBOS; the author was not. The rumours of all these impending bank failures were wandering around internet forums from early 2007 onwards. Banking insiders knew that the funding models for these banks were unsustainable and dangerous.

Forcing banks to insure each other's deposits would force banks to price the risk on badly run banks like Northern Rock at higher rates than better run banks such as HSBC and Barclays. By pricing this risk strictly as a percentage rate, the general public would gain direct visibility of the default risk.

Under this regime, a well-run bank might still pay lower interest rates, but would be compensated with even lower insurance rates. This should make the net interest rate; interest less insurance, of the low risk bank better than that of the risky bank. Competition would no longer be on interest rates alone.

With the best will in the world, such a system would not be capable of insuring all deposits in the event of a systemic bubble. But that is not the point.

The point is; that by introducing effective market based pricing of risk, the general public and the banks would be penalised for indulging in the risk-taking that encourages bubbles in the first place.

Additionally, the general rates of insurance should act as both an early warning system for the monetary authorities and even as a counter-cyclical assistance in popping bubbles in the first place.

In normal times, insurance rates for all but the most foolish of banks should be ridiculously low. In the event of the economy moving into bubble conditions, insurance rates would start to creep up on the riskiest banks. This would then start to pass on the infection, via the insurance, to other banks, but at a much earlier stage than normally happens when entering a financial bubble. Faced with the obligation of holding more reserves on their balance sheets to cover the deposit failure of others, all banks would be obliged to cut back on credit in general. All banks would be affected, but with the strongest effects on the worst run and most highly leveraged banks.

Monitoring of individual and overall insurance rates would give the central banks live data on the perceived risks of the banks in their charge, as well as the financial system as a whole.]

6.4 Low Frequency / Tobin Trading

THE spectacular collapse of so many big financial firms during the crisis of 2008 has provided new evidence for the belief that stockmarket capitalism is dangerously short-termist……… Shareholders can no longer with a straight face cite the efficient-market hypothesis as evidence that rising share prices are always evidence of better prospects, rather than of an unsustainable bubble.

If the stockmarket can get wildly out of whack in the short run, companies and investors that base their decisions solely on passing movements in share prices should not be surprised if they pay a penalty over the long term. But what can be done to encourage a longer-term perspective?……

In the early 1980s shares traded on the New York Stock Exchange changed hands every three years on average. Nowadays the average tenure is down to about ten months. That helps to explain the growing concern about short-termism. Last year a task force of doughty American investors (Warren Buffett, Felix Rohatyn and Pete Peterson, among others) convened by the Aspen Institute, a think-tank, published a report called "Overcoming Short-Termism". It advocated various measures to encourage investors to hold shares for longer, including withholding voting rights from new shareholders for a year. [Economist 2010a]

Warren Buffet is of course a value investor, the sort of investor who intuitively understands the workings of the companies models in section 2 of this paper. The sort of investor that the efficient market hypothesis states cannot exist. Value investors also intuitively understand that the short-term liquidity and momentum effects seen in the commodity and macroeconomic models in sections 3 and 4 not only make value investing difficult, but also add no value to the process of creating wealth that capitalism aspires to.

The proposals of the Aspen Institute were pretty much stillborn for a number of reasons. Firstly, because orthodox economics assumes, erroneously, that any cost imposed on market transactions must increase costs to the consumer. Secondly because such a tax would destroy a substantial part of the finance industry, which makes the majority of its profits by charging rents on the very volatility they create in the first place. And thirdly, and more reasonably, if such a tax were imposed in one country, trading would simply move to an alternative jurisdiction.

To understand just how short-term the finance industry has become, it is worth noting that stock-trading is now dominated by 'high-frequency trading' (HFT). In the major stock-markets supercomputers trade billions of dollars of trades in seconds using automated algorithms. Individual bids and offers may be held open for fractions of a second. High frequency trading systems are now being co-located within stock-exchange buildings as the speed of light now means that companies trading from a few blocks away are at a significant disadvantage.

To anybody who has actually worked in a real company, the idea that the real market value of a normal company can change from millisecond to millisecond is bizarre; it is palpable nonsense. A full discussion of high-frequency trading is postponed to section 9.2 below.

It is my belief that Buffet, Shiller, Smithers et al are correct, and that the unnecessary volatility is induced endogenously in share markets, causing excessive movements away from real value on timescales from seconds to decades.

It is my belief that the decadal movements are caused by liquidity at a macroeconomic scale, a problem that will need tackling at a macroeconomic level – this is discussed in detail in section 8.2.1 below.

Other timescales are much shorter and give the appearance of being quasi-periodic momentum effects. Although the evidence is controversial, typical time-scales for the periodicity appear to be on the order of fifty and two hundred trading days, with other shorter time scales also present.

A system is proposed below that would dampen the fluctuations on these timescales.

The solution proposed is a private-sector approach, independent of government. Following the same logic as housing in the previous section, it is proposed to introduce damping with losses imposed on early retrading on the lines of those proposed by Buffet et al. This would be done by introducing a new class of shares, or special investment certificates, in the companies. These shares would have different rules as to their trading. The issuing of such shares would be voluntary, at the choice of the companies involved.

In the same way as housing, damping would be imposed with a haircut of say 10% imposed on anybody who sold a share within the specified time period. The haircut would be paid back to the company in which the share is held at the time of sale, as such it would be effectively a 'negative dividend' on the share, paid by the owner to the company. The haircut would automatically be deducted from the sale proceeds. In extremis the haircut would be imposed for a period of say three years.

However unlike housing it is not proposed that the haircut on all shares be imposed for the full term of three years. This would present great problems for pricing of the shares. If a large purchase was made of a company's shares this would kill the market in that company's shares for years at a time, which would make price discovery for the company almost impossible.

Instead it is proposed that all shares that have been sold are marked as 'locked'. This would be in contrast to all the remaining shares that would be 'unlocked'.

Every trading day a random selection would be made across all the currently 'locked' shares and 1% of all the currently locked shares would be unlocked. The owners of these newly unlocked shares would then be able to sell the shares immediately without penalty.

Assuming 250 days of trading per year, then this release of 1% of shares per trading day would give a half-life for locked shares of roughly six months.

This means that if every single share was bought on day one, and no further trading took place, roughly half the shares would be unlocked after six months, more than 70% would be unlocked by the end of the first year, over 90% would be unlocked by the end of the second year and almost 98% would be unlocked by the end of year three. At this point, after three years, any remaining locked shares would be automatically unlocked.

This system would be a compromise between ensuring a haircut on fast resellers, while ensuring that shares were continually made available to the market for further trading. For an individual purchaser who bought a block purchase, their haircut on day one, if they resold all their shares would be 10%, if they sold all shares after a year the haircut would be slightly below 3%, after two years it would be 1%. After three years the haircut would be zero.

In these circumstances purchasing shares for value investment would have very little risk as in such a circumstance the period would be expected to be a minimum of a few years. Speculative investment would be risky, and effectively pointless.

Even better for value investors, it should be noted that the losses taken by the early sellers accrue to the company in which the shares are held, and so ultimately to the other shareholders. The losses of the speculators are transferred directly to the value investors.

All of this could be simply organised electronically through the same systems that currently manage dividend payments.

Interestingly, although such a system may seem complex, it may actually be one that would be driven to adoption by the market. For well managed companies, issuing such shares would give direct benefits to value investors, but much more importantly issuing such shares would in its own right be a very powerful signalling mechanism to the market. It would be very foolish for a company that is manipulating a short-term rise in its share price to issue such shares, the subsequent burning of locked-in investors would cause significant reputational loss. On the other hand, for well-run companies with long-term investment horizons, issuing such shares would be a way of signalling the long-term commitment of the management. This would particularly be the case if managers share options were restricted to these shares. Eventually, failing to issues such shares might become a good indication of a poorly managed company.

Such a shareholding pattern might form a useful compromise between the pattern of 'Anglo-Saxon' free trading of shares and the 'European' model of very long-term share-holding with very low levels of open trading.

6.5 Ending the Chaos

A third example of controlling chaotic financial systems is discussed in section 9.2, this ordering is necessary as it needs to follow discussions on market microstructure.

In economics there has been a traditional split between the laissez-faire who wish to minimise perceived barriers to trading, and the dirigiste who wish to regulate trade to minimise perceived speculation and profiteering. Both viewpoints are based on a static assumption of economic activity. The examples above assume a dynamic system, and so introduce time-based restrictions on regulation. This is designed to eliminate short-term speculation while encouraging long-term value investment.

It is the belief of the author that the controls proposed for housing in section 6.3 are practical. Those in sections 6.4 and 9.2 for share trading are much more speculative.

The point however is that changing dynamic, chaotic, systems to remove endogenous oscillations, is of profound importance, and usually very easy if the system is understood.

The oscillations result in mispricing and misallocation of capital and are enormously wasteful.

In general control of such systems is straightforward. One way to control is to use external feedback loops. Inflation targeting with interest rates is a classic example of this. This is generally fraught with danger, if the feedback control is not set up correctly, it is very common for such feedback loops to exaggerate cyclical behaviour rather than reduce it.

It is nearly always better to introduce a damping mechanism into a naturally oscillating system. If the damping is of the order of the systems natural oscillations, then the system should move to stability very rapidly.

My own personal experience as a commissioning engineer has shown the truth of this. It is eye-opening to see a system that is 'hunting'; moving rapidly backwards and forwards erratically, suddenly flatline as the time delay on a feedback loop is gently increased.

Rather than follow the seat of the pants methods of sections 6.3, 6.4 and 9.2, a better method is to analyse the data of asset price changes and then build models using non-linear dynamics and control theory similar lines to those in part A of this paper. Then the models and data can be analysed, and the natural frequencies of the systems can be identified. Finally the control variables can be identified and modified to allow the system to be moved to a stable equilibrium point.

Standard control theory books such as Nise give systematic ways to analyse and control dynamic systems, chapter six of Nise, on stability, is of particular interest [Nise 2000].

With chaos we have looked at dynamic systems with up to a dozen or so parameters, or 'degrees of freedom'. Now I would like to move on to the problems of what happens when you have much larger numbers of degrees of freedom. This leads us into the fields of entropy.

7. Entropy

7.1 Many Body Mathematics

At a theoretical level, Poincaré's conclusions have permeated higher economics, though at the cost of some pain. In a piece of tragi-comedy, Poincaré's work appeared shortly after the marginalists had transformed economics by putting their version of field theory into the very foundations of economics. In the 1980s after much deep intellectual work theoretical economists 'proved' that the Walrasian system could not produce stable equilibria, so reproducing Poincaré's conclusions some eight decades after the original, without a hint of irony, let alone embarrassment. As Foley describes it:

There is no doubt, however, that the outcome of these investigations have been surprises that raise unexpected and disturbing questions about the general validity of the Walrasian approach.

The initial attack on infinite commodity spaces involved the development of specific models examining economic growth, international trade, and public finance problems over time. In these models the equations of supply and demand give rise to difference or differential equations, whose solution paths represent the equilibrium allocations and prices of the model. The simplest behavior of these solutions occurs when they converge asymptotically to a steady-state in which the levels or ratios of the relevant variables remain unchanged forever. This type of stability is called saddle-point stability in mathematical jargon. In infinite horizon models which exhibit saddle point stability most of the key results of the finite-commodity economy carry over. The equilibrium paths are locally unique, so that comparative statics (which now becomes comparative dynamics, the comparison of equilibrium paths) methodology still works. Furthermore, in models with some infinitely lived agents, the first welfare theorem will hold as well. The difficulty with this line of work was that the hypothesis of saddle-point stability was not in general a consequence of the basic assumptions of the model together with the Walrasian requirements of market clearing, that is the equality of supplies and demands in each period. Researchers had to add hypotheses to assure saddle-point stability. The careful workers introduced such hypotheses into their models of technology and preferences at the price of reducing the generality and persuasiveness of their conclusions. Less careful workers simply assumed the saddlepoint property, at the risk of making erroneous statements, or confined their analysis to saddle-point paths, at the risk of reaching unjustified conclusions within their own models.

A more sophisticated attack by mathematically trained theorists on this problem (see William Baumol and Jess Benhabib (1989)) revealed the surprising fact that the equilibrium paths of even very standard economic models were much richer than the saddle-point literature had suggested. Equilibria might not approach a steady-state, but could end in limit cycles, in which variables endlessly repeated cyclical movements, or even in chaotic paths of a highly irregular kind, confined to a local region of the price allocation space. The assumptions necessary to rule out these complex solutions were very strong. Thus the saddlepoint literature has limited general validity, and the problem of generalizing the finite-commodity space Walrasian results remains unresolved.

[Foley 1990]

The method of economics remains comparative statics. To study a phenomenon, the economist proposes a model, in which certain variables are taken to be exogenous, or unexplained, and other endogenous variables are taken to be determined by equilibrium conditions. The method of explanation requires that the specification of the exogenous variables determine the endogenous variables in some sense, so that the effect of changes in exogenous variables on the endogenous variables can be traced unambiguously…

…In fact Walras' conception of equilibrium, even in the finite commodity space case, is not very satisfactory in this regard, because, except in the case where all the agents can be regarded as a single consumer (the representative agent case), competitive equilibrium is not unique. There may be several different price systems at which supply and demand are equal. (A related serious problem is that no natural and robust concept of stability of equilibrium can be developed within the Walrasian model, because it lacks a clearly articulated dynamics.) High theory in the '60s and '70s was able (through the work of Gerard Debreu) to show that generically equilibria are locally unique. Thus the comparative static use of the theory rested on the methodological assumption that after a change in exogenous variables the economy would follow the equilibrium state it initially occupied to a new configuration of prices…

…I would like to underline the fundamental significance of this technical problem. If the model is not determinate in some sense, either it must be abandoned, or the comparative statics methodology must be revised.

[Foley 1990]

The reason for the surprise at the complex equilibrium paths remains unclear.

Having failed to produce a mathematical system for dealing with multibody problems, economics then took an unfortunate route to solve the problem. By going to a single consumer model with just one 'representative agent', economics made the maths solvable by returning the model to a two-body system. This makes the sums easier, but dramatically decreases the believability of the model.

As the number of bodies, or variables, increases, solution of such systems becomes more and more intractable. The problems become insoluble in detail. Once the numbers of independent bodies move into double figures the maths of field theory becomes useless by itself. A good example is the asteroid belt in the solar system, where trajectories of the asteroids can only be predicted in the short term, and individual asteroids can be ejected from the asteroid belt on an apparently random basis. Indeed as the numbers of bodies increase the description is no longer at the level of an individual body but instead becomes that of a probability distribution.

And this is where the beauty and power of statistical mechanics steps in.

Faced with the same problems a century and a half ago, physics borrowed statistical ideas from the social sciences and took a different route that proved much more fruitful. Effectively, physics took large numbers of identical 'representative agents' but abandoned looking at individual interactions and simply looked at probabilities of outcomes. This process became very effective, and became known as statistical mechanics.

Statistical mechanics is an approximation method for describing systems characterised by deterministic chaos, see for example [Gould & Tobochnik 2010 section 1.7]. Although it is an approximation, it is capable of very accurate predictions of macroscopic properties. Counter-intuitively, with statistical mechanics, the more bodies, the more accurate the predictions.

The contrast between physics and economics here is stark. Alongside Ludwig Boltzmann, the work in this field was pioneered by James Clerk Maxwell.

In physics Maxwell was 'Mr field theory'. He started with the same Newtonian field theories that were adopted by the neoclassicals. He expanded them rigorously to cover the whole of optics, electricity and magnetism. This remains the crowning achievement of field theory, in the second great unification in physics, second only to the work of Newton. As a sideline he also analysed chaotic control systems and so produced the first effective governor systems for steam engines.

When he started looking at the many-body systems of energy in gases he promptly junked his field theory knowledge and built on the infant science of statistical analysis pioneered by Quetelet and Buckle in the social sciences. By bringing a much greater level of mathematical sophistication and inventing statistical mechanics; Maxwell, along with Boltzmann, was able to explain the microscopic behaviour of molecules in a gas, link the microscopic to the macroscopic and explain the microscopic origins of pressure, entropy and the gas laws.

In contrast economists have been attempting to apply field theory to many body systems for 140 years without success.

If we go back to the table from Keen/Constanza:

Table 1 The solvability of mathematical models (adapted from Constanza 1993)

Equations	**Linear**			**Non-linear**		
	One Equation	Several Equations	Many Equations	One Equation	Several Equations	Many Equations
Algebraic	**Trivial**	**Easy**	**Possible**	Very difficult	Very difficult	Impossible
Ordinary Differential	Easy	Difficult	Essentially Impossible	Very difficult	Impossible	
Partial Differential	Difficult	Essentially Impossible	Impossible			

The statements that many equation systems are impossible to solve are strictly correct. However, when you get to a many body system with thousands or more of independent variables, you can look at the statistics and probabilities of events happening, and things actually become easier again.

It then turns out that some outcomes are so probable that they become inevitable. As a consequence of this, highly predictable system variables arise straight out of pure statistical considerations. In these circumstances, underlying microscopic drivers of behaviour become almost irrelevant, they are drowned out by the statistical effects. Counter-intuitively, in a many body situation, the statistical properties outweigh the underlying interactions, and often produce unexpected results, results that go against obvious common sense.

The most important thing about this statistical mechanical approach is that a new sort of equilibrium is formed. Equilibria that are very stable. In these equilibria individual agents can change their values very significantly, but the overall distributions of values are very stable.

From a mathematical point of view, statistical mechanics also has another big advantage; the maths of statistical mechanics is better behaved than the mathematical agglomeration of utility/field theory:

Statistical equilibrium is much better behaved mathematically than Walrasian equilibrium. Statistical equilibrium exists and is unique for arbitrary finite offer sets without restrictions of concavity. The logarithm of the economy wide partition function …… is a concave potential for the statistical demand functions, which as a result have a negative definite Jacobian.

From an economic point of view the statistical market equilibrium differs from Walrasian equilibrium in two important respects. First, it does not exhibit horizontal equality, since two agents of the same type will in general end up at different points in their offer sets, representing different final consumption bundles. Thus the statistical market process induces some inequality in the final allocation of the economy that was not present in the original states of the agents. This market induced inequality is a consequence of agents' trading at different, disequilibrium, prices. Second, the statistical equilibrium in general leaves some mutually advantageous trades unconsummated. The market moves the economy toward Pareto-efficiency, but does not fully achieve it. Thus certain pervasive phenomena in real markets, such as unemployment of productive factors like labor and excess productive capacity, which are inconsistent with Walrasian equilibrium, are consistent with statistical equilibrium.

[Foley 1996b]

To take an example of the power of statistical mechanical drivers, the income data from the UK shown in figure 1.1.1 shows strong equilibrium properties. This data set runs from 1992 to 2002 with the shape of the distribution almost constant throughout this period. The actual UK economy changed through very different phases during this period, including a major recession at the beginning of the 90's; yet the shape of the distribution is almost constant.

This approach also explains the fascination that statistical physicists have with wealth and income models. Although the mathematical theory of income and wealth distribution is a quiet backwater in economics, this area has attracted physicists and statistical mathematicians and engineers in significant numbers since at least the work of Champernowne.

The reason is simple; to a statistical physicist, economics is obviously a multi-body phenomenon. It is messy. There are millions of agents in a typical economy, and their behaviour is not coordinated at a high level. In such a system, as physicists intuitively understand, statistics must take over from microscopic drivers, and entropy raises its head.

Apart from income distribution, the other area of economics in which physicists have taken a large interest is in that of finance.

The earliest work on random walks was that done by the mathematician Bachelier on stock prices. Bachelier's work predates Einstein's own random walk model of Brownian motion.

For half a century Bachelier's work was largely forgotten. The use of random walks in finance was rekindled and ultimately led to the option pricing formulae of Merton, Black and Scholes.

Unfortunately the random walk process has been removed from it's many body background, and individual prices are treated as moving randomly isolated by themselves. But Black-Scholes is simply the diffusion equation, and things don't diffuse with random jumps in a vacuum. The random movements of dust particles undergoing Brownian motion are caused by interactions with air molecules. Black-Scholes is used in economics without looking at the overall picture of all price movements

Although it is rarely considered as such, Black–Scholes is a many body mathematical approach. Necessarily, the random movements in prices effectively assume multiple random interactions; in the real world random buys and sells by investors. If this was analysed properly, analysis should be taken across all the different stock prices changing at same time. In an investment world with no new money supplied, the purchase of one stock must be balanced by the sale of another.

In a simplistic case then a conservation law would hold if money supplied to the stock market was constant. This would give an overall distribution of price changes different to a B-S application to a single stock. Without such assumptions, B-S applied to a single stock allows for infinite growth in individual stock prices, an impossible assumption without supply of unlimited liquidity.

Clearly a more sophisticated multi-stock model would need to take into account increases in money supply, exogenous and endogenous, as well as movements of investment between different asset classes.

Michael Stutzer has started some useful research in this direction [Stutzer 2000] using maximum entropy approaches.

Despite its unrealistic use on isolated stocks, Black-Scholes has been enormously successful; possibly the only piece of theoretical economics to be used on a daily basis to successfully calculate the prices of anything.

Bizarrely, the success of B-S and the apparent randomness of stock market data has been used to support the theories that stock markets are efficient and fully incorporate all knowledge about stocks.

That these beliefs continue to be widely held is puzzling. That B-S does not work fully is well known. Mandlebrot first discovered that price movement distributions had fat tails in the early sixties, which clearly disprove the efficient market hypothesis. The EMH needs a log-normal distribution. Smithers gives a wealth of data that debunks the efficient market hypothesis [Smithers 2009].

Given the theoretical origins of Black-Scholes, to simultaneously believe in the validity of Black-Scholes, and also believe in the Efficient Market Hypothesis is a bit like accepting that the earth goes round the sun, while still maintaining that it is flat.

Although it remains isolated in finance, and used incorrectly, statistical mechanics, in the form of Black-Scholes is the most successful piece of theoretical mathematics in economics. In the next section the concepts of statistical mechanics and entropy are discussed briefly, but hopefully in a way that gives a little clarity as to why and how statistical mechanics and entropy can give a more useful approach to the whole of economics.

7.2 Statistical Mechanics and Entropy

A long quote from Wright to begin with:

Farjoun and Machover (1989), in their path-breaking work on Political Economy, 'Laws of Chaos', make a simple but important methodological point. They observe that an economy is a dynamic system composed of millions of people in which 'the actions of any two firms or consumers are in general almost independent of each other, although each depends to a very considerable extent on the sum total of the actions of all the rest' (Farjoun and Machover (1989), p.39); in other words, a market economy has a huge number of degrees of freedom (DOF) with weak micro-level coordination. They argue that the appropriate equilibrium concept for such a system is a statistical equilibrium in which the macro-level regularities take the form of probability distributions. Let's explore their thesis for a moment.

The economy of the United States has a civilian labor force of approximately 155 million individuals. The kinds of economic activities performed by these individuals spans the whole range of human experience and subsumes a great variety of tasks, skills, situations, enjoyments and motives. An enormous variety of both mundane and novel decision-making contexts are routinely presented to the individuals that constitute the economy. The space of possible configurations of this system is of course astronomically large.

Local economic decisions are globally coordinated primarily through the 'invisible hand' of supply and demand dynamics in markets distributed in time and space. The economy gropes this way and that, from one configuration to another, generally in a 'bottom-up' manner, adapting continually to new economic circumstances. The existence of this type of emergent coordination does not significantly reduce the DOF since there is no top-down plan or 'Walrasian auctioneer' to synchronize the local behavior.

Systems that have a huge number of DOF and weak micro-level coordination ('messy' systems) behave very differently to systems with a small number of DOF and strong micro-level coordination ('neat' systems). This is reflected in the different kinds of equilibrium they can exhibit.

The state-space of a system is the set of all possible configurations of the DOF. A particular configuration is a 'point' in state space. In general we find that many neat systems, if they enter equilibrium, tend toward a point or trajectory in state-space. A canonical example is a set of weighing scales. Place some weights on each arm and the scales will tend toward an equilibrium point in which the internal forces balance and the system is at rest. This is a simple kind of deterministic equilibrium, in which the equilibrium configuration is a subset of state-space. The classical mechanics concept of equilibrium was a founding metaphor of the 19th Century marginal revolution in economics (e.g., see Mirowski (1989)). And it appears in a more developed form in 20th Century neoclassical general equilibrium models (e.g., Debreu (1959)).

But most messy systems, if they enter equilibrium, do not tend toward a subset of state-space. So in the physical sciences the tools of statistical, not classical, mechanics are used to study messy systems. A canonical example is an ideal gas in a container. The internal forces never balance. Instead, at the micro-level, there is ceaseless motion and change, a process that effectively samples the whole state-space in a random fashion. Yet at the macro-level a certain kind of regularity does emerge. The probability that a randomly selected gas particle will have a certain energy is constant over time (in this case, the probability distribution is Boltzmann-Gibbs). In this simple kind of statistical equilibrium the equilibrium configuration is not a 'point' or subset of state-space but a probability distribution over an aggregate transform of the state-space (in this case, the number of atoms with a given energy level).

Since an economy is more like a messy than a neat system we should expect any empirical regularities to be better captured by the concept of a statistical, rather than a deterministic, equilibrium. Essentially this is Farjoun and Machover's point.

The importance of statistical equilibrium in economics has been emphasized by other authors, notably Steindl (1965), and more recently Aoki (1996, 2002) and Foley (1994).2 Nonetheless, thinking that the relation between micro and macro in statistical mechanics is related to the analogous problem in economics remains the 'less trodden path'. One reason, perhaps, is that it calls into question the need for explicit microfoundations.

A counter-intuitive property of statistical mechanics is that macro-level regularities are in an important sense relatively independent of the precise mechanisms that govern the micro-level interactions. So the adoption of macro-level statistical equilibrium as an explanatory principle has a concomitant implication for micro-foundations. For example, classical statistical mechanics represents the molecules of a gas as idealized, perfectly elastic billiard balls, which is a gross oversimplification of a molecule's structure and how it interacts with other molecules. Yet statistical mechanics can deduce empirically valid macro-phenomena. Khinchin (1949), who pioneered the development of mathematical foundations for the field, writes:

> *Those general laws of mechanics which are used in statistical mechanics are necessary for any motions of material particles, no matter what are the forces causing such motions. It is a complete abstraction from the nature of these forces, that gives to statistical mechanics its specific features and contributes to its deductions all the necessary flexibility. ... the specific character of the systems studied in statistical mechanics consists mainly in the enormous number of degrees of freedom which these systems possess.* Methodologically this means that the standpoint of statistical mechanics is determined not by the mechanical nature, but by the particle structure of matter. *It almost seems as if*

> *the purpose of statistical mechanics is to observe how far reaching are the deductions made on the basis of the atomic structure of matter, irrespective of the nature of these atoms and the laws of their interaction. (Eng. trans. Dover, 1949, pp. 8–9, emphasis added).*

So, analogously, the method by which individuals choose (the 'mechanical' nature of individuals) is not as important as the fact that a huge number of individuals are choosing with respect to each other but are weakly coordinated (the 'particle' nature of individuals). The approach of implicit microfoundations adopts this methodological 'rule of thumb'.

Given the aim is to determine 'how far reaching are the deductions made on the basis' of the particle nature of individuals while abstracting from the mechanics of individual rationality, it makes sense, at least initially, to 'bend the stick' as far as possible in the direction of implicit microfoundations. But how do we abstract from the 'mechanics' of individual rationality and represent individuals as 'particles'? Sometimes it is possible to predict choice behavior in controlled experimental settings or in situations where conventions or rules play an important role. But in general the everyday creativity of market participants who aim to satisfy their goals in open-ended and mutually constructed economic situations is unpredictable.

For example, Aoki (2002) writes,

> *'Even if agents inter-temporally maximize their respective objective functions, their environments or constraints all differ and are always subject to idiosyncratic shocks. Our alternative approach emphasizes that an outcome of interactions of a large number of agents facing such incessant idiosyncratic shocks cannot be described by a response of the representative agent and calls for a model of stochastic processes'.*

The unpredictability of choice behavior suggests representing the choice mechanism as a random process. So the implicit approach represents economic agents not as 'white box' sources of predictable optimizing behavior but instead as 'black box' sources of unpredictable noise; that is, they are particles that choose in a random manner subject to objective constraints (e.g., a budget constraint). The single representative agent with well-defined choice behavior has been replaced by a huge number of heterogeneous agents with random choice behavior. This is the simplest possible starting point for implicit microfoundations and provides a null hypothesis against which claims of the importance of explicit microfoundations can be measured. For example, as a starting point, randomness can be modeled as selection from a uniform distribution, in accordance with Bernoulli's Principle of Insufficient Reason that states that in the absence of knowledge to the contrary assume all outcomes are equally likely. The aim is 'to explain more by saying less', or at least start by saying less and see how far that takes us (c.f. Farmer et al. (2005)).

The principle that many market outcomes are determined more by the objective social structure than the particulars of individual rationality is not new. For example, Gode and Sunder (1993) show that the results of an economics experiment are broadly similar when classroom students are replaced with 'zero-intelligence', random agents; Farmer et al. (2005) show that the assumption of 'zero-intelligence' agents can explain many of the statistical features of double-auction trading data from the London Stock Exchange; and Wright (2008) shows that 'zero-intelligence' agents in a simple commodity economy can instantiate supply and demand dynamics that approach efficient allocation of resources and equilibrium prices (see also Cottrell et al. (2009)).

A natural objection at this point is the observation that economic agents do not act according to random rules. They often think very carefully before acting. Surely it is necessary, therefore, to model individual rationality, even when considering macro-level phenomena? But the objection *elides the distinction between epistemology and ontology, a picture with reality. A 'black box' probabilistic model of individual agency does not imply that choice mechanisms are in fact random, only that, when placed in the range of situations routinely presented by a dynamic, large-scale economy, they are operationally equivalent, at the aggregate level, to an ensemble of random process. So the precise detail of the choice mechanism is not a decisive factor in the determination of macro-level outcomes.*

Randomness in a theory can be viewed as an unmodeled residual, like assuming a constant in physical theories (e.g. the constant of gravitation). Residuals should eventually be eliminated and replaced by a more encompassing theory (e.g. a theory that explains the value of the gravitational constant). But the 'rule of thumb' of implicit microfoundations says something different: eliminating randomness won't necessarily yield a better explanatory or predictive theory since the randomness represents an essential property of 'messy' systems. We should expect rapidly diminishing explanatory returns from increasingly explicit microfoundations.

[Wright 2009]

And a shorter one from Von Neumann, reported by Claude Shannon:

"My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage."

Claude Shannon [Tribus & McIrvine 1971].

I find the last quote very reassuring. In my own opinion, though less well known than Einstein, Von Neumann ranked close to Einstein in terms of genius. Like Einstein, he didn't merely bring in a single profound new idea; but seemed to change radically, for the better, any field that he investigated.

Despite this, it appears he found entropy as philosophically puzzling as most other people who encounter it do.

Entropy is a famously abstract concept, bordering on the mystical. I believe that the main reason for this is that entropy does not have a straightforward analogue in day to day human experience, so it is simply very difficult to relate to.

I do not wish to write a book on entropy and statistical mechanics, and the following section is intended only as a brief introduction. Fortunately there are two very well written introductory, non-mathematical books on entropy and statistical mechanics; one by Atkins [Atkins 1994] and the other by Ben-Naim [Ben-Naim 2007] to which the reader can go for more illumination.

There is one key fact about entropy that this section will attempt to illuminate, a key fact which goes against the whole practice of economic theory from the days of the physiocrats right up to the present day.

The key fact is that statistical equilibrium is more powerful than any local equilibrium. And so local equilibria are a very poor guide to overall equilibria. The statistical equilibrium will normally be in a different place to the local equilibrium, and the system will come to rest at the statistical equilibrium, not the local equilibrium. In general, low-level information is close to irrelevant as a guide to macroscopic outputs.

If economics is to make theoretical progress, the process of extrapolating from the bottom up must be abandoned. More importantly, in many cases, economic 'common sense' must also be abandoned.

Einstein famously stated that 'God does not play dice', with regard to quantum mechanics, and was forced to back track. In economics God runs a casino.

Going back to the Von Neumann quote, another important thing to note is that entropy in its statistical form was effectively 'discovered' twice.

In his ground breaking work on information theory, Shannon rediscovered the mathematics that Boltzmann had discovered 80 years previously while trying to explain the macroscopic entropy of heat flow.

This is not to devalue Shannon's work, which if anything is more generally applicable than that of Boltzmann and Gibbs.

The two introductory texts on entropy mentioned above follow these two different approaches at looking at entropy.

The heat entropy approach is explained beautifully in the book 'The Second Law' by Atkins [Atkins 1997]. In this, entropy is explained through the traditional concept of disorder, or more accurately 'dispersion'. The more dispersed something is, the higher its entropy, and the less its value. In particular, the statistical concentration followed by dispersion of heat in heat engines giving rise to useful power.

'Entropy Demystified' is another very good book, by Ben-Naim [Ben-Naim 2007]. This follows the information path of counting systems statistically.

There has been some considerable debate as to whether the two approaches of heat entropy and information entropy are isomorphic or merely analogous. This is a debate I do not wish to enter.

Certainly, from a human cognition point of view, neither approach is fully satisfactory. The information approach is more obvious mathematically, but like quantum mechanics, somehow seems to imply the necessary presence of an outside observer.

Dewar's work, discussed in section 7.3 below may shed some light on this discussion.

Both Atkins's and Ben-Naim's books are short and well written, and I commend them both.

A later book by Ben-Naim points out the basic fact that through an accident of history, the sign of entropy (like that of the electron) is intuitively wrong. In most of the things that humans count, more is better, while with entropy less is normally better. In later sections I follow Schrödinger in using the concept of 'negentropy' (negative-entropy) to get round this problem.

The important point of the energy/information debate is that the same fundamental mathematical models fall out in two fields that appear to be widely different, one explaining how steam engines work, the other explaining how much information can be squeezed down a telegraph line.

What the two approaches have in common is an abandonment of detailed analysis of the system and replacing it with the concept of counting very carefully all of the possible available states of a system.

It turns out that this simplistic approach is both very powerful and very generally applicable.

Very briefly the concepts that are explained at length in the two books above are as follows.

Entropy is a measure that counts all the possible statistical states that a system can occupy.

When this counting is carried out, it is normal to find that a subset of these possibilities is much more probable than all the other possible states. Because of this, this subset of possible states dominates the behaviour of the system; almost absolutely.

To get a feel for how this works, it is worth first looking at a similar concept that is often used, in fact rather over-used, in economics. That is the central limit theorem.

The CLT states that if values are randomly selected from different underlying distributions and added together, the resultant distribution will be a normal distribution. If the underlying values are multiplied together, then you get a log-normal distribution.

In both cases the resultant distributions are independent of the underlying distributions.

This is a simple result of statistics. Take, for example, two underlying distributions which are both uniform distributions. Each of the underlying distributions is much more skewed, with more extreme values, than a normal distribution. A naïve investigator might assume that the result of adding samples from two uniform distributions would be another uniform distribution. However this is not true, because of the likely sampling.

For example, while it is quite possible that you will get a high value when sampling one of the underlying distributions, it is quite unlikely that you will simultaneously get two high values from the two distributions, or two low values when sampling from both distributions.

So the resultant distribution, when the underlying samples are added, is bunched towards the centre, and if enough samples are taken, you get a normal distribution.

Similar arguments produce a log-normal distribution when the underlying samples are multiplied together.

If a researcher was unaware of the CLT, they might assume that different underlying distributions would produce different resultant output distributions, and that by studying the underlying distributions carefully they would be able to predict the resulting output distribution.

Knowledge of statistics in this case solves a lot of unnecessary work. It doesn’t matter what the underlying distributions are, if you take enough samples, statistics gives you a normal distribution if you add the samples, and a log normal distribution if you multiply the samples.

In these circumstances, the underlying distributions are irrelevant.

Another trivial example is that of flipping coins.

If you take two coins and toss them randomly, the chance of getting all heads is one quarter. If you use three coins the probability of getting all heads 1/8 if you use ten coins the chance of getting all heads is 1/1024.

All sequences are equally likely, but in each case only one out of all possible sequences is all heads.

In contrast the number of sequences that is close to 50:50 heads and tails gets proportionally larger and larger as the number of coins gets larger.

More importantly the average variation from the mean becomes smaller and smaller. This can be seen clearly in figure 7.2.1. below.

Figure 7.2.1 here

So one narrow band of similar distribution results becomes so likely it becomes inevitable, others become negligible

Where this gets much more interesting, and much more powerful is when external constraints, or boundary conditions are introduced.

We have already seen one example of this. The log-normal distribution can be considered as a normal distribution where there is a boundary condition set at zero. This is why it is used (erroneously) as the assumed base distribution for Black-Scholes theory. The price of shares is assumed to be able to increase infinitely but cannot go below zero. In the absence of more detailed knowledge of the underlying distribution, the log-normal was the sensible choice to use in the earliest models. Following Mandlebrot, the log-normal needs replacing with an alternative distribution. There are of course many other distributions that fit the characteristics of not having negative values, many of which (including the GLV) have the required fat tails that the log-normal lacks.

In statistical physics, perhaps the most well known example of the operation of an external constraint is that of a conservation principle. For example, under the external constraint of conservation of energy, distributions form a standard shape known as the Maxwell-Boltzmann distribution, typically given in the form:

$$F(x) = xe^{-x} \qquad (7.3a)$$

This is a special case of the gamma distribution, and gives a shape that can be closely modelled by a log-normal distribution [Willis 2005].

For example all the molecules of air in a room have kinetic energy. Ignoring heat losses and gains through the walls of the room, the total kinetic energy of all the molecules is conserved. That is, if one molecule gains a unit of energy, another molecule must lose an equivalent amount.

In theory it is possible that all the molecules could have exactly the same amount of energy (a uniform distribution), but there is only one way of creating this distribution, so it is very unlikely that this is in fact how the energy will be shared. This state has only one configuration.

A second possibility is to give all the energy to one molecule, with all other molecules having zero energy. This can happen in N different ways, where N is the number of molecules, so this distribution is much more likely than the previous uniform distribution. In fact it is N times more likely. This state has N configurations. But the difference between this and the first option would be enormous, there would be many moles of air in a room, so N would be much greater than 10^{23}. However this second distribution is still a very unlikely distribution.

A third option would be to give to give two thirds of the energy to one molecule and the other third to a second molecule, with all other molecules having zero energy. This distribution could be formed in N(N-1) ways. So this distribution would be (N-1) times more likely than the second option above, and N(N-1) times more likely than the uniform distribution.

Clearly as energy is shared out in different ways between all N different molecules the number of possible distributions becomes enormous, with some distributions being much more likely than others. Fortunately it is relatively easy to show mathematically that the most likely distribution in this case is of the form:

$$F(x) = x^2 e^{-x^2} \qquad (7.3b)$$

(The power of two arises because kinetic energy is proportional to mv^2.)

It is always possible that the distribution could take a form that doesn’t fit the above function, for example, in theory it is possible that a single molecule could have all the energy. However the probability of the distribution being of the above form in (7.3a) is so high that you would have to wait for time periods of the order of the universe to observe a noticeable deviation from the form above.

This result is a maximum entropy equilibrium.

Counting of all the possible states indicates that this distribution is the most likely, and so, by definition, has the maximum entropy. Moving away from this equilibrium would require the expenditure of energy (or information) such as the use of ‘Maxwell’s demon’, or for that matter ‘Walras’s auctioneer’.

This maximum entropy solution relies only on the statistical analysis. It does not depend on the underlying interactions between the atoms or molecules in the gas.

For instance, it is possible to compare say a bottle of a noble gas such as neon with a bottle of water vapour.

Neon is a noble gas with all its electron shells full. As a result it does not form chemical reactions, and when two neon atoms collide their local interaction should be very close to a perfect inelastic collision. The results of this collision can be accurately predicted and are highly likely to be ‘unequal’ with a high probability of energy being transferred from one atom to another.

Water molecules are at the other end of the scale. Two molecules of water can form temporary hydrogen bonds when they collide; they also have many options for temporarily storing energy in rotational and vibrational modes. In general, collisions between two water molecules are likely to be more ‘equal’ with both molecules of water likely to emerge from a collision with similar amounts of energy.

However, no matter how different the behaviour of the atoms / molecules at a local level, the resulting distribution of velocities will be a Maxwell-Boltzmann distribution for both the neon and the water vapour, as long as the water vapour is above boiling temperature.

It doesn't matter how many years you spend studying the interactions of water molecules, and their energies following collisions, you will never be able to extrapolate up to the overall energy distribution.

Consequently a maximum entropy equilibrium can be very different from a market clearing equilibrium. This is the basic problem with using a marginal approach; the probability of reaching a marginal solution is vanishingly small, almost infinitely small. Within economics, almost uniquely, Foley has made substantial progress in moving from a Walrasian approach of pricing to a more sensible maximum entropy approach. This is discussed in the following papers, [Foley 1996b, 1999, 2002], while a good example of the failure of market clearing is given in 'Statistical equilibrium in a simple labor market' [Foley 1996a]

In the general framework of maximum entropy in economics, supply and demand are just forces driving in directions, just as electrical fields drive directions in physics. However entropy can overpower these forces.

When looking at such models, a subtle point is that you don't need complete randomness to create a maximum entropy output, only an element of randomness. There has been a history in econophysics of creating 'pure' exchange models. In most of these models, hypothetically, a beggar could meet Bill Gates in the street, and walk away a billion dollars richer. Although intellectually pleasing, such models are clearly highly unrealistic. In the models of income and companies discussed above in this paper only a small amount of randomness was introduced. But even this small amount was sufficient to destabilise the system away from an intuitively logical Pareto type outcome to one based on maximum entropy.

Where microscopic effects do remain important is in the ranking of individuals in the distribution, whether looking at people with different basic abilities and savings preferences, or companies with different capital efficiencies; the ranking of the individual or company is given by the ranking of abilities. However the rewards are defined by the shape of the outcome distribution. The output distribution is defined by entropy, not by the underlying input distributions. So the rewards are not 'fair'.

7.3 Maximum Entropy Production

There is another very substantial, and very interesting, difference between the thermodynamic systems discussed in the section on entropy above, and the various models discussed in this paper.

All the discussion so far on entropy has been about what physicists call equilibrium thermodynamic models. In these models the system has been allowed to evolve until there are no temperature differentials or net energy flows across the system. Everything has stabilised with uniform macro level variables.

It should be noted that this is very different to the traditional equilibrium mathematics used in economics, which is entirely static.

In the thermodynamic equilibrium models of physics, individual molecules are still swapping energy and changing their places in the distribution of energies – however the shape of the distribution is stable.

Historically, these equilibrium thermodynamic models are well understood and can be described exactly mathematically, with entropy values directly calculable.

The models in this paper consist of sources of wealth generation in companies and sinks of consumption at households, with a continuous flow from one to the other. In this they resemble models that have continuous flows of heat in and out of the system and that have different temperatures in different parts of the model.

Such models are described by physicists as 'out of equilibrium thermodynamics systems', or simply non-equilibrium systems; though it is the belief of the author that this nomenclature may need to be revisited.

Traditionally, such systems have been very difficult to describe mathematically, however recent work by Lorenz, Paltridge, Ackland & Gallagher, and others in the field of planetary ecology and, also that of Dewar, Levy, Solomon and others in the field of theoretical physics appear to have changed things substantially.

In the 1970s Garth Paltridge produced papers looking at the absorption of sunlight by the earth and the re-radiation of heat into space. Paltridge's model is profoundly simplistic. He split the earth into just ten cells by latitude and set up basic energy balance flows between the cells and attempted to produce a simple system of formulae to give an overall balance. In so doing he 'accidentally' rediscovered the basic formulae for entropy first discovered by Carnot two hundred years previously. This is recounted, entertainingly, in chapter three of 'Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond' [Kleidon & Lorenz 2005]. Despite the very rudimentary nature of the model, the model was able to give surprisingly accurate predictions of the temperature and cloud cover at different latitudes of the earth, this can be seen in figure 7.3.1 below.

Figure 7.3.1 here

[Ozawa 2003]

This is typical of the power of entropy. All the detail of evaporation rates, wind speeds, precipitation, etc were irrelevant and unnecessary for production of the model. A simple application of entropy was sufficient.

What was new, and ground breaking, with regard to the model is that this was a successful analysis of an 'out of equilibrium' thermodynamic model.

At one end of the model is the Sun at 5800 degrees Kelvin, at the other end is deep space at 3K, with the earth in the middle.

In such a system, entropy is not maximised, but is being produced continuously; as heat flows continuously from hot to cold.

What Paltridge and Lorenz discovered was that the earth appeared to act in a 'deliberate' manner, by adjusting the temperatures across the globe, to a give a maximum possible rate of entropy production.

Although it is early days, this principle of 'Maximum Entropy Production' or 'MEP' appears to be widely applicable, and also appears to make many previously insoluble systems much more tractable.

Analysis by other authors suggests that the same Maximum Entropy Production principle is true for the re-radiation of heat from Mars and Titan. It also appears that the use of MEP may be applicable to many other systems such as convection in the earth's mantle, and turbulent systems. Ozawa et al give an excellent review of the history and uses of MEP, while the book edited by Kleidon & Lorenz gives much more detail [Ozawa et al 2003, Kleidon & Lorenz 2005].

In Paltridge's model, earth becomes what is known as a 'dissipative structure'. Dissipative structures include things such as planets, and life forms. Dissipative structures are counter-intuitive from a normal equilibrium thermodynamic point of view. Dissipative structures are highly concentrated, highly organised, and so have very low entropy. From the point of view of ordinary equilibrium thermodynamics, they shouldn't exist.

However from an MEP point of view, dissipative structures do make sense.

To take a simple example of a dissipative structure, consider the convection cells (Bénard cells) that can appear in a pan of water that is being heated at the bottom.

Figure 7.3.2 here

[Georgia Tech 2010]

Figure 7.3.3 here

[Eyrian 2007]

The pan has a high temperature at the bottom and a low temperature above it (assume no heat flow through the walls).

Conduction is not a particularly effective method of heat transfer in water. So if conduction was the only mechanism for transferring heat through the water then the heat flow, and so the rate of entropy production would be constrained.

However, heating the water decreases its density so allowing the hot water to float to the top and release heat to the atmosphere. Meanwhile colder water sinks from the top to the bottom to replace the heated water.

In theory the water could circulate chaotically, or it could form one large loop. In practice, at heating rates low enough not to create bubbles of gas, the water 'self-organises' into hexagonal cells. These cells are low entropy, complex, 'dissipative structures'. However their existence allows a higher rate of entropy production, transferring heat rapidly from hot to cold. This allows the total entropy of the system, that is the heating source, pan and atmosphere, to be increased, despite the local drop in entropy associated with the creation of the hexagonal dissipative structures.

This is no minor parlour trick. The continents of the earth are pushed around on the global equivalent of Bénard cells, forming mountain ranges and oceans as they do so. Our geography is an accidental high entropy output caused by the need to move heat formed by radioactive decay in the core of the planet to the earth's surface.

As previously discussed, the earth's atmosphere operates as a dissipative structure moving hot equatorial air to the poles. The circulation of the oceans carries out exactly the same functions.

Interestingly, it also appears that the existence of plants changes the earths albedo in ways that also maximises entropy production. Animals, then appear as efficient redistributors and processors of vegetable matter.

When looked at in this manner almost everything on planet earth becomes a dissipative structure. This includes of course human society, and indeed, human economic systems. This is of considerable importance, and is returned to in section 8.1 below.

(As a brief aside, it should be noted that the discussions here relate to maximum entropy production. This is a different theoretical approach to that of Prirogine who has discussed dissipative structures under a minimum entropy production principle. While Prirogine's ideas appear valid in a certain number of examples with strongly defined constraints, the minimum entropy production approach has failed to find widespread application.)

In chapter 9 of 'The Second Law', Atkins gives a brief but very well written review of dissipative structures, using as one example the creation of a simple fox-rabbit ecology and introducing the Lotka-Volterra dynamics. This brings us full circle to where we began.

In parallel with the above work in the field of ecology; Levy, Solomon and various co-workers have carried out pioneering theoretical work looking at the dynamics of the Generalised Lotka-Volterra distribution and how it works mathematically.

In their mathematical analysis of the GLV, Levy and Solomon show that the entropy of multiple Boltzmann distributions gives the power law tails found in the GLV distribution [Levy & Solomon 1996].

In contrast the Maxwell-Boltzmann distribution of a normal thermodynamic equilibrium comes from an additive process. This is a direct conservation law, in such a system the addition and subtraction are direct and total energy is conserved absolutely. This results in a distribution with an exponential tail.

The GLV comes from a multiplicative process. And multiplicative process cannot be directly conservative. The GLV process does however remain conservative in total, at least in the long term; the process of this conservation is discussed further below.

Because of its multiplicative nature, the output of the GLV includes a power law tail.

This can be seen as analogous to the central limit theorem.

Under the CLT an additive process gives a normal distribution, a multiplicative process gives a log-normal distribution, with an exponential tail.

Under an additive, maximum entropy process, the output is a Maxwell-Boltzmann distribution, with an exponential tail. Under a multiplicative, maximum entropy production process, the product is a GLV distribution, with a power tail. The mathematics of this is quite robust and works under lots of different models as long as they meet some basic requirements, again more below.

'One sees therefore that a power law is as natural and robust for a stochastic multiplicative process as the Boltzmann law is for an equilibrium statistical mechanics system. Far from being an exception and requiring fine tuning or sophisticated self-organising mechanisms, this is the default.' [Levy & Solomon 1996]

As such, the GLV distribution might better be considered to be a 'log-Maxwell-Boltzmann' distribution and the Lotka-Volterra seen as a special, non-equilibrium version of this log-Maxwell-Boltzmann.

Within the fields of ecology, these ideas have been taken forward in some very interesting work by Ackland & Gallagher [Ackland & Gallagher 2004] on the modelling of ecosystems. This modelling shows that, by using simple GLV models, and some very basic assumptions it is possible to produce full food webs with all the complexity of a real ecosystem. This model allows and includes for constant evolution and transformations of predators and prey within the system. Despite this the overall parameters of the food web become highly stable in things such as numbers of predators, prey, varieties of species, etc.

It is particularly interesting that a large array of different species, different types of dissipative structures, appears so as to maximise the total biomass flow.

"We monitored this during our simulations and found a remarkable result—the total flow of resource (and hence total biomass) increases with time reaching a plateau after many thousands of steps—the steady-state linkstrength ensemble distribution appears to be the one which maximizes the use of resource. This type of optimisation is consistent with what has been observed in other ecological models. If the model is recast in terms of flow and dissipation, the maximization principle is equivalent to maximum entropy production: the mathematical equivalent of "entropy production" is just the total death rate, and hence the flow out." [Ackland & Gallagher 2004]

It is the belief of the author that the economies of the world are acting in exactly the same manner. An economy is an MEP dissipative structure, and when it is at equilibrium it is maximising the rate of entropy production.

In the natural world an ecosystem develops to a complex but stable equilibrium of different groups of animals, plants, herbivores, carnivores, etc each adapted to its niche.

Similarly, in an economy, a complex ecosystem evolves which splits into extractive industries, manufacturing, services, finance etc.

In both systems the apparently stable system involves constant microscopic competition, evolution and change.

Clearly, this maximum entropy production approach to economics links through to evolutionary economics and theories of the sources of endogenous growth.

It should be noted that this is not just an analogy. In entropy production terms, the human economic system is simply a complicated and interesting sub-section of the MEP function of the earth as a whole.

Returning to the mathematics, Dewar [Dewar 2005] has produced a seminal paper that derives maximum entropy production from the first principles of information theory and simple maximum entropy considerations.

This derivation of a Maximum Entropy Production (MEP) approach appears to be applicable to non-equilibrium systems in general.

Instead of looking at the counting of all possible statistical states, and finding the most probable, Dewar looks at the counting of all possible paths through a flow system, and finds that these can be counted using the same maximum entropy approach used by Boltzmann, Gibbs, etc.

Dewar does this by maximising the path information entropy, following the ideas of Shannon and Jaynes. This follows from Shannon's interpretations of information entropy and Jaynes generalisation of the maximum entropy approach as a general recipe for statistical inference.

In Dewar's words:

"Jaynes saw the Gibbs algorithm as a completely general recipe for statistical inference in the face of insufficient information (MAXENT), with useful applications throughout science, not just in statistical mechanics. Viewed as such, it is a recipe of the greatest rationality because it makes the least-biased assignment of probabilities, i.e., the one that incorporates only the available information (imposed constraints). To make any other assignment than the MAXENT distribution would be unwarranted because that would presume extra information one simply does not have, leading to biased conclusions.

But if MAXENT is essentially an algorithm of statistical inference (albeit the most honest one), what guarantee is there that it should actually work as a description of Nature? The answer lies in the fact that we are only concerned with describing the reproducible phenomena of Nature.

Suppose certain external constraints act on a system. Examples include the solar radiation input at the top of Earth's atmosphere, the temperature gradient imposed across a Bénard convection cell, the velocity gradient imposed across a sheared fluid layer, or the flux of snow onto a mountain slope. ***If, every time these constraints are imposed, the same macroscopic behaviour is reproduced (atmospheric circulation, heat flow, shear turbulence, avalanche dynamics), then it must be the case that knowledge of those constraints (together with other relevant information such as conservation laws) is sufficient for theoretical prediction of the macroscopic result. All other information must be irrelevant for that purpose. It cannot be necessary to know the myriad of microscopic details that were not under experimental control and would not be the same under successive repetitions of the experiment (Jaynes 1985b). We can only imagine with horror the length of scientific papers that would be required for others to reproduce our results if this were not the case.***

MAXENT acknowledges this fact by discarding the irrelevant information at the outset. By maximising the Shannon information entropy (i.e., missing information) with respect to pi subject only to the imposed constraints, MAXENT ensures that only the information relevant to macroscopic prediction is encoded in the distribution pi. Therefore, if we have correctly identified all the relevant constraints (and other prior information), then macroscopic predictions calculated as expectation values over the MAXENT distribution will match the experimental results reproduced under those constraints.

But of course that last if is crucial. In any given application of MAXENT there is no a priori guarantee that we have incorporated all the relevant constraints. But if we have not done so, then MAXENT will signal the fact a posteriori through a disagreement between predicted and observed behaviours, the nature of the disagreement indicating the nature of the missing *constraints (e.g., new physics). MAXENT's failures are more informative than its successes. This is the logic of science.'*[Dewar 2005]

The bold emphasis is my own. What holds true for 'atmospheric circulation, heat flow, shear turbulence, avalanche dynamics' also holds true for such regularities as wealth and income distributions, distributions of company sizes and the ratio of returns on labour and capital.

Because these regularities are found across multiple different economies, following Jaynes' logic, their causes can be determined using a max-entropy approach along with appropriate constraints and conservation laws.

'We can only imagine with horror the length of scientific papers that would be required for others to reproduce our results if this were not the case…' As a word 'horror' accurately captures the emotional reaction when an individual with a passing understanding of the power of entropy becomes acquainted with the amount of time and energy that highly intelligent economic theoreticians have invested in attempting to produce macroeconomic models from observed (and even worse, supposed) microeconomic behaviour.

I have not yet seen any theoretical work formally linking the work of Dewar to that of Levy & Solomon, however I am firmly convinced that they are isomorphous; that Levy & Solomon's mathematical derivations of the GLV should also be reproducible via working from Dewar's principals of path entropy.

It is my belief that Levy, Solomon and Dewar have produced some very important and very general principles. I believe that the max entropy production model, and GLV distributions will be found to give general and stable descriptions of many complex systems that have hitherto been seen as insoluble.

What Dewar, Levy and Solomon's systems consist of are three critical elements; a source, a sink, and some sort of self-limiting behaviour.

This model is potentially very powerful, as this simple model is typical of many complex systems. The sources and sinks are typically energy, but can also be population, or the wealth created in an economic system, or many other things.

The reason such systems are very common is because most other systems are inherently dull, at least in the longer term.

Without the source, the system quickly disappears.

Without the sink the system will quickly explode and disappear.

Without the self-balancing mechanism the system will either explode or disappear depending on the direction of the imbalance.

The self-balancing mechanism is the key to the long-term preservation of the process, and this reintroduces the conservation principle.

In a classical 'static' thermodynamic equilibrium conservation is absolute.

In a Dewar, Levy, Solomon type 'dynamic thermodynamic equilibrium', conservation is approximate and long term. Input and output can differ over the short term, but are brought back into balance automatically in the long term. Indeed such systems can wander backwards and forwards in a Lotka-Volterra type manner at a macroscopic level, while maintaining GLV type equilibrium at a microscopic level.

In section 7.4 below, I discuss the statistical mechanics of this further, as I believe there may be a shortcut way of unifying and simplifying the approaches of Dewar, Levy and Solomon by recasting the flow model into an equivalent exchange model.

In economics the source is production, the sink is consumption. Going way back to section 1.2 and section 1.6 there was a discussion of the different ways of producing power laws. These different methods were combinations of two exponential processes, multiplicative process, and self-organised criticality (SOC). As discussed in section 1.6 it is the belief of the author that the first two processes, double exponentials and multiplicative processes are in fact different ways of describing the same process. In the GLV this becomes obvious if you look at the difference equation (1.3o), which can be seen as either a way of multiplying the variables (a multiplicative process) or a way of modelling two different growth rates (double exponential process). However a single GLV can have many different possible equilibria. Dewar ties this together, and shows that dynamic systems tend to a single point of maximum entropy production point, a single dynamic equilibrium, at the limit of stability, at the point of self organised criticality.

This appears to be typical of many systems, and may explain the fact that many power law distributions have values between two and three even though they arise from substantially different underlying models (see Newman table 1 for example [Newman 2005]).

Indeed Dewar points out that many very chaotic systems; systems close to 'self organised criticality' such as earthquakes, avalanches, forest fires and the archetypal sandpiles, can be characterised by slow steady underlying growth rates (eg tectonic plate movement for earthquakes, tree growth rate for forest fires). He also explains that such systems can be included in the Maximum Entropy Production modelling approach, even though such systems are traditionally characterised as being very far from equilibrium.

Financial markets, especially asset markets, also show many of the characteristics of such SOC systems with steady growth intermittently interrupted with dramatic crashes.

This analogy may shed some light on the role of debt in finance discussed in section 4.6 above. An example, for those that can remember them, is the traditional old-fashioned egg-timers. When well-built, these represented a very well behaved sandpile. In a high quality egg-timer, the sand is very fine, with equal sized smooth grains, the sand is dry and friction is very low. In such an egg-timer the sandpile has a near constant, flattish, inverted conical shape, and close observation shows that the avalanches are small but near-continuous. With a 'normal' sandpile the sand behaves much more erratically. With a little 'stickiness', caused by damp or a wide distribution of grain sizes, the pile can build up significantly into steeper and steeper hills as grains are added at the top. Eventually a dramatic collapse occurs which changes the steep hill into a much shallower one, then the process restarts.

In human managed forests this lesson has been learned, though at a cost. In the middle of the last century forest managers attempted to fight forest fires by removing undergrowth and ignition sources. This appeared to work in the short run, but eventually this simply led to much larger, and more devastating and dangerous fires. In recent decades foresters now often manage nature reserves by deliberately starting fires on a frequent basis. This results in a steady stream of much smaller fires.

It is the belief of the author that increasing debt, and liquidity, in an economy above a certain point is actually counter productive in that it moves the economy closer to an unstable point, the point of SOC, approaching the scale-free system in which large fluctuations become much more likely.

It also suggests that there will always be strong pressure from financiers and politicians to move towards increasing debt. They are pushed in this direction by the forces of entropy. But any marginal increases in efficiency are outweighed by the increase in the instability of the economy. The comparison with forest management is apposite. Allowing a forest to grow freely, and removing sources of small fires, will, in the short term, and indeed on an ongoing average basis, marginally increase the total amount of wood growing per acre. But this is of precious little reassurance when you find your village surrounded on all sides by fire, of afterwards when you discover there isn't a tree standing for twenty miles in any direction.

In this light, Cooper's suggestion that central banks should aim for a pattern of small business cycles is eminently sensible. Simply reducing leverage and excess liquidity may be a better approach, if done correctly it should move the economy out of a cyclical mode altogether.

For these reasons, I believe that the nomenclature of such systems needs to be reviewed. In many cases I believe that many complex systems that are currently described as 'out of equilibrium' should be described as being in 'dynamic thermodynamic equilibrium' or 'MEP equilibrium'. This form of equilibrium is reached when the system has reached the point of maximum entropy production and continues indefinitely in that state.

7.4 The Statistical Mechanics of Flow Systems

In the following section I would like to briefly bring together some ideas on the statistical mechanics of power laws, from various sources cited in this paper, and also discuss their relevance to dynamic equilibrium both in economics and in flow systems in general. This section is aimed at statistical physicists, mathematicians and theoretical economists, and assumes that readers have read Glazer & Wark or the equivalent as a minimum. It is also highly speculative. It will not be easy to follow for many readers, who may wish to skip to the economics of section 8.

In this section I would like to make some suggestions as to possible ways forward for a statistical analysis of the flow systems described by Levy, Solomon and Dewar.

I would like to do this by attempting to reduce these models to equivalent exchange models.

I have previously been somewhat scathing of exchange models, primarily because they do not provide models that realistically capture the processes of real economic systems. For these reasons I have built the models in part A following the flow pattern of the GLV of Levy and Solomon. However for a core production of the statistical mechanics I believe appropriately designed exchange models may be useful proxies for flow models.

Very many exchange models have been produced by econophysicists, with many different underlying mechanisms, see section 1.1 above. In a very perceptive paper; 'The Rich Are Different!: Pareto Law from asymmetric interactions in asset exchange models' [Sinha 2005] Sitabhra Sinha points out that these models share a very basic pattern. When these models have a symmetric pattern of exchange they produce a traditional Maxwell-Boltzmann distribution. When the exchange mechanism is made to be asymmetric, then a power law is produced. Indeed; in one case an asymmetric mechanism was deliberately introduced to assist the poor, but instead produced a power law tail; so giving the opposite result of that intended.

I believe it is a similar simple asymmetry that drives the multiplicative flow models of Levy, Solomon and Dewar.

If we go back to the base equation (1.3o) for a single agent in the economic models from section 1.3:

$$\mathrm{w}_{i,t+1} = \mathrm{w}_{i,t} + \mathrm{e} + \mathrm{w}_{i,t}\mathrm{r} - \mathrm{w}_{i,t}\Omega \qquad (1.3o)$$

I would firstly like to generalise this to the following:

$$w_{i,t+1} = w_{i,t} + e_{i,t} - \tau + w_{i,t}r - w_{i,t}\Omega \qquad (7.4a)$$

The first change above is to allow a distribution of different possible earnings incomes $e_{i,t}$; this was actually the case in the later income models in section 1.3 where the uniform distribution was replaced with a normal distribution, though both distributions were defined to be exogenous.

The second change is to introduce a term τ. This was first discussed in passing in section 1.9.1 and represents what I called compulsory consumption, or what economists normally call non-discretionary spending. This is assumed (in my discussions) to be a base constant value that includes for basic housing, as well as minimum requirements for food, clothing, heating etc. All other spending is assumed to be discretionary, and proportional to wealth and so included in Ω.

If we now do a summation of equation (7.4a) across all individuals we get:

$$\sum \mathrm{w}_{i,t+1} = \sum \mathrm{w}_{i,t} + \sum \mathrm{e}_{i,t} - \sum \tau + \sum \mathrm{w}_{i,t}\mathrm{r} - \sum \mathrm{w}_{i,t}\Omega \qquad (7.4b)$$

Let us then assume that the dynamic flow model is at a dynamic equilibrium, ie that it is neither growing nor shrinking through time, though it is still flowing. At this equilibrium the total wealth is constant between times steps, so the term on the left hand side is equal in value to the first term on the right hand side. This gives:

$$0 = \sum \mathrm{e}_{i,t} - \sum \tau + \sum \mathrm{w}_{i,t}\mathrm{r} - \sum \mathrm{w}_{i,t}\Omega \qquad (7.4c)$$

The obvious way to balance this economic flow system is as an accounting identity as follows:

$$\sum \mathrm{e}_{i,t} + \sum \mathrm{w}_{i,t}\mathrm{r} = \sum \tau + \sum \mathrm{w}_{i,t}\Omega \qquad (7.4d)$$

This balances the total incomes on the left and the total consumption on the right. And indeed this would be the natural way to balance any similar physical flow system model, because this is the way to balance the flows in and out of the system.

However, from a point of view of statistical analysis, I believe it would be more fruitful to show a different balance as:

$$\sum e_{i,t} - \sum \tau = \sum w_{i,t}\Omega - \sum w_{i,t} r \qquad \text{or:}$$

$$\sum (e_{i,t} - \tau) = \sum w_{i,t}(\Omega - r) \qquad (7.4e)$$

This gives additive (but flowing) things on the left hand side of the exchange system and multiplicative (flowing) things on the right hand side of the exchange system.

Given that τ, r and Ω are all constants it also reduces a somewhat complex flow system to an exchange system with only two variables, the earnings, $e_{i,t,}$, on the left hand side and the wealth, $w_{i,t,}$, on the right hand side.

This, I believe, is close to the base model that Sinha was describing; an asymmetric exchange model.

In equation (7.4e) the left hand side additive flows must balance with the right hand side multiplicative flows. The balance of flows is between net earnings; that is earnings minus base living costs, and net consumption; which is discretionary consumption less unearned income.

In a normal exchange model both sides of equation (7.4e) would be additive, and indeed identical.

I believe this model, with only two variables and lots of boundary conditions, may be simple enough to be tractable to a traditional statistical mechanical analysis on the lines of Dewar, or indeed Champernowne.

Before moving into further discussion I would first like to follow the maths through a little more. I would like to do two things. Firstly I would like to neglect τ for the moment; we will come back to τ later. Secondly I would like to divide by Ω. That then gives us the following:

$$\sum \left(\frac{e_{i,t}}{\Omega}\right) = \sum w_{i,t}\left(1 - \frac{r}{\Omega}\right) \qquad (7.4f)$$

This brings us back to some old friends. The term (1-r/Ω) gave us our definition of α, the exponent of the powertail, included in equation (4.5q). Equation (7.4f) itself is just a restatement of Bowley's law as defined in section 4.5 of this paper. These relations imply that the suggested approach in this section may have promise.

A second observation, which may be completely wrong, is that equation (7.4e) has the feel of simple differential equation, with wealth on one side, and earnings, the time derivative of wealth, on the other. Instinctively the solution of this would be of exponential form.

Given that the solution of a symmetric exchange is Maxwell-Boltzmann with an exponential tail, then a solution of (7.4e) could reasonably be expected to be a Maxwell-Boltzmann with an exponential-exponential, or a power law tail, as per Reed and Hughes or Baek, Bernhardsson and Minnhagen or others [Reed & Hughes 2002, Baek et al 2011].

An alternative approach is to look at equation (7.4e) from a maximum entropy, statistical mechanical point of view, where you need to maximise the entropy over two different distributions.

On the left hand side, you have a traditional additive term that should produce a standard Maxwell-Boltzmann distribution of earnings. On the right-hand side you also have a distribution to maximise, however in this case the distribution is multiplicative, and so the ladder of energy levels are proportionately distributed. So the resultant Maxwell-Boltzmann is exponential-exponential, or power law. This seems very close to the original model built by Champernowne, and rediscovered by Levy & Solomon [Simkin & Roychowdhury 2006].

It may be possible to maximise each of these entropies independently, however it seems likely that the distributions on each side will affect each other.

At this point it is worth looking at the left hand side in more detail, as this may answer a quandary discussed back in section 1.9.2, though it raises as many questions as it answers. In this section it was noted that returns from waged employment appear to follow an offset Maxwell-Boltzmann distribution, or an 'additive GLV distribution'.

Looking at equation (7.4e) the answer to why earnings are distributed as a Maxwell-Boltzmann becomes, in one sense, trivial.

The distribution is a Maxwell-Boltzmann because that is the maximum entropy solution for the distribution of earnings. For a statistical mechanic that is good enough.

Indeed, statistical mechanics would predict a Maxwell-Boltzmann distribution of earnings even when all the individuals had identical skills.

However two questions are raised immediately; why is it offset? And what is the actual mechanism for creating the distribution?

The first question is one for which the answer is not at first obvious. Intuitively, the maximum entropy distribution would extend to zero, because, given a fixed total amount of incomes, this would also allow the maximum values of earnings in the tail to increase, and so give a wider total spread, which would have a higher overall maximum entropy.

However, although the model above attempts to reduce the system to an exchange model, it must be remembered that it is a flow system that is being analysed. I believe that Dewar is absolutely correct that these systems must be modelled by maximising the entropy flow, not just by maximising the entropy.

So, with two distributions, one on each side of the exchange, the simplistic (traditional) solution would be to maximise the entropy of the two distributions; that is to multiply the two different partition functions and maximise the single resultant function. However, both distributions are modelling distributions of flow. As well as maximising the entropy embodied in the two distributions, there is a simultaneous need to maximise the entropy embodied in the size of the flows. Hopefully this will be a straightforward trade off between the three (four?) different entropies being enumerated. Intuitively, given this extra contribution to total entropy from the flow, an offset Boltzmann distribution may achieve extra entropy flow to compensate for its narrower spread and the lower entropy in its distribution.

Going back to the concept of dissipative structures and negentropy generators, a narrower Boltzmann distribution could be seen as a dissipative structure with lower entropy, but which is capable of allowing larger entropy flows through the system. Ultimately, if it allowed very high entropy flows the earnings distribution might even collapse into a very low entropy uniform distribution, or, as is often seen in both real world monopolies and many econophysics models, all wealth and income would go to one individual.

With a dissipative structure approach, presumably there is a negentropy flow associated with 'maintaining' the dissipative structure in its low entropy form; Maxwell's demon is continually at work narrowing the spread of the distribution. However if this negentropy flow is smaller than the entropy flow through the system, enabled by the dissipative structure, then the flow system as a whole, including the dissipative structure can be stable and long-lived.

As long as a factory is making money, it is worth diverting part of the profits to maintain it. If a proposed new factory is predicted to be profitable in the long-term, it is worth borrowing money to build it.

As a first approximation, it might be possible to simply maximise the product of the entropies of the two distributions multiplied by the flow that results from the macrostates.

The second question; of the mechanism for creating income distributions, is also problematic.

For the right hand side of equation (7.4e) the mechanism of wealth condensation producing a feed-back loop for increasing wealth via returns on assets discussed in this paper seems, to me at least, very plausible.

The self-organisation of salaries into a Maxwell-Boltzmann distribution is a harder process to visualise; people do not randomly exchange jobs and salaries with each other.

The first problem is letting go of the fundamental economic belief that people are fairly rewarded for their employment. We have already seen in section 4 that this is not normally true at the aggregate level. I do not believe it is true at an individual level either. I have worked as a risk manager in the water and nuclear industries, doing roughly similar jobs at roughly similar salaries (we will come back to this). However, water is cheap, electricity is expensive, especially when compared to the amount of capital installed, so the amount of value I gave in the nuclear job was many orders higher than that I gave in the water jobs.

I was in fact paid a bit better in the nuclear job, but nowhere near enough to compensate for the extra value created. Similarly, as a risk manager, the wealth I created was many factors higher than that created by a security guard or a cleaner, but I was not paid many times the rate of these people, though I was certainly paid more.

But it is a well-known economic puzzle that some industries, such as the oil industry, pay better than others even for secretaries, cleaners and security guards, where the jobs are identical. An entropic, Maxwell-Boltzmann, distribution of wages, varying by the wealth of the industry might explain this puzzle. Similarly, at the high end, it may explain the persistent high pay and bonuses of executives and even mid ranking staff, in financial industries that have very high cash flows but low profits.

In fact when employers take on new employees they don't do a detailed analysis of the individual's probable contribution of wealth to the company. They decide if the employee is needed, they look at the market rates for the skills required and they pay the going rate. Certainly overall wage levels are checked carefully against total revenues, and deadwood is chopped back wherever possible. But wages are set externally in the market, not internally by potential wealth creation.

Note also, that in a stable economy, the total sum Σe of earnings available will be fixed, giving the boundary condition necessary for a Maxwell-Boltzmann distribution to develop.

Given that wages are set in the market, a maximum entropy distribution becomes more possible. As long as there is a minimum amount of stochastic churn in the market, with competition and movement up and down a ladder of earnings levels, then creation of a Maxwell-Boltzmann distribution becomes possible.

Moving to a different issue, an element that is missing from this model, and indeed all my models, is that of unemployment. Wright's models are superior in this regard, and may shed light on this dynamic.

Equation (4.7e), and a Maxwell-Boltzmann distribution, especially an offset one, would seem to imply that all would have jobs and earnings.

I can see two possible causes for persistent mass unemployment. A first explanation is given by reintroducing τ, the compulsory consumption or non-discretionary spending. It is possible that when the values of $e_{i,t}$ at the low end of the distribution becomes less than the value of τ individuals are removed from the distribution altogether.

A second source of persistent unemployment could come from a combination of the maximum entropy flow, dissipative structure model combined with differing actual skill levels. With differing skill levels greater flows of entropy might be achieved by diverting all earnings to highly skilled individuals with no flows to the low skilled. Although the distribution would have lower entropy, total entropy flows might be higher.

At this point I would like to return to the issue of equity, which has been a central theme of this paper. Equation (4.7e) implies that a group of identical individuals will be forced into an unequal distribution of earnings incomes. In practice, with non-identical individuals the individuals will be ranked into the Maxwell-Boltzmann distribution by their abilities.

Following this the individuals with the highest earnings will then be distributed into the highest income groups of the GLV distributions as we saw in section 4. Even ignoring unemployment effects, the whole system becomes deeply iniquitous.

Finally, and much more speculatively, I would like to consider what might happen when equation (7.4e) does not balance.

$$\sum(e_{i,t} - \tau) = \sum w_{i,t}(\Omega - r) \qquad (7.4e)$$

I think that equation (7.4e) will balance in many situations of flow systems; most physical and biological systems will come to a dynamic equilibrium when the flows in and out of the system are equivalent. This will define a pair of distributions and an entropy flow that will have a combined system maximum entropy production.

However for most economic systems the above is not true. Once a market system is installed in a country, the economy starts growing and is characterised by long-term persistent levels of growth. The growth level is so persistent that this can also be characterised as being stable, in that the parameters of the system; gdp growth rate, interest rates, stock-market growth rates, etc, are very stable over decades or even centuries. This was discussed in section 4.5.

For newly industrialising economies this is characterised as having high levels of gdp growth up to 10% per annum, with associated high interest rates and stock-market rates. Ω is typically low, around 0.5.

For mature economies, gdp growth and interest rates are typically 2-4% and Ω is typically 0.7.

In these cases Ω can be seen as the external variable.

Given this external value of Ω, it could then be possible that there is a set level of gdp growth, interest rates and stock-market returns that gives a maximum entropy production output for the sum of the terms represented by equation (7.4e).

If this was the case then the persistence of endogenous growth would have an explanation.

Even more speculatively, let us reintroduce τ to the discussion. τ will be defined somewhere endogenously within the system. It will basically be defined in terms of the proportion of average wage level required to provide basic housing, food, heating, etc. In a developing society it will probably be defined largely by the subsistence wage level needed to provide basic food and shelter. In an advanced economy it will be defined by basic housing rental costs and ultimately the costs of scarce land. This might explain the very similar rates of growth seen in industrialising economies. It could also explain the higher long term growth rates in the US, with its plentiful land compared to the lower rate for the UK, were land has been scarce for centuries.

If τ can be defined endogenously within the system, then Ω should be definable endogenously in terms of τ. People will need to save enough during their working lives to pay for their annual τ during their retirement.

In theory, then the whole system becomes an endogenous equilibrium, with the only real exogenous factor being scarce land prices in advanced economies.

So, after a lot of background, we have moved back to the economics.

Part B.II – Economic Foundations

8. Value

8.1 The Source of Value

The source of value is humanly useful negative entropy, or simply 'negentropy'.

This of course raises the question of what is negentropy.

Erwin Schrödinger first introduced the concept of negentropy, in his 1944 popular-science book 'What is life?' [Schrödinger 1944] and it was used in the discussion of living systems. Schrödinger explained his use of this phrase:

'…if I had been catering for them [physicists] alone I should have let the discussion turn on free energy instead. It is the more familiar notion in this context. But this highly technical term seemed linguistically too near to energy for making the average reader alive to the contrast between the two things…" [Schrödinger 1944]

I am going to leave the definition of negentropy deliberately vague for two reasons.

The first is that the exact definition of concepts such as negentropy and free energy are difficult and can vary by situation and definition of systems.

The second, more honestly, is that I remain unclear in my own views as to the detailed definition, and consequent measurement of negentropy within economic systems, when working from a physical bottom up point of view.

I am however convinced that this lack of clarity is of no great consequence.

Over the last two hundred years physicists, chemists and engineers have proposed different quantities such as entropy, enthalpy, free energy, Landau potential, etc to deal with entropic calculations in different systems. This has been done primarily to make the sums add up in a meaningful way subject to different restraints. Maximum entropy production is a very new set of models, a new way of adding up entropy, which has yet to be made systematic in the life sciences where it originated, never mind in economics.

But in the short term, this is of academic interest only. We don't need to invent a new entropy concept for economics. Human beings intuitively understand this particular negentropy, they call it 'value' and it is measured in non-SI units such as dollars, euros, pounds or yen.

It doesn't actually matter that much, whether we call it negentropy or free energy, it is what people think of value, and costs £ or $ or euro to get some of it. It accumulates during the production process and disappears in the consumption process.

Most importantly it is objective; while people may have different utility values, the value of a good or service has an intrinsic value. Although they may have disagreed, and indeed been wrong, on the ultimate source of value, the classical economists, from Quesnay to Marx were correct in believing that value was a real, meaningful intrinsic quantity.

There are of course a number of natural objections to an intrinsic concept of value, these are discussed in section 8.2 below.

For a feel of what 'negentropy' is we can go back to the concept of entropy being, in very general terms, a measure of dispersion. In general terms the more dispersed something is, the less useful it is, the more concentrated it is, the more useful it is. More concentration means more negentropy, means more value.

So for example, very rough estimates suggest there are about 15,000 tonnes of gold in the world's oceans, but the concentrations are so low that extracting it would not be economic, it is too dispersed; its entropy is too high, its negentropy is too low.

Wealth creation is usually a process of concentration, whether this be discovery of a concentrated ore in a gold mine or oil in an oil well, the concentration of baked beans into a can in a concentration of cans in a supermarket, art works in a gallery or a concentration of people in to a factory, a physical market or a city. The example of a traditional physical market is particularly apposite; the whole point of markets is to concentrate the goods from the surrounding area, so allowing goods to be exchanged. A traditional market is a creator of negentropy, a creator of value, even though the physical goods are unchanged in the process.

Classical economics produced an effective method for valuing geographically dispersed negentropy by introducing the concept of marginality. This is very useful for the pricing of genuinely scarce resources such as specific minerals, agricultural land, housing in cities, etc. Unfortunately this mathematical trick has been extended, most unwisely across the whole of economics.

From an information entropy point of view, negentropy is also increased by increasing the uniqueness of an object. Whether you turn a piece of gold into a piece of jewellery, some steel into machine tools, or raw ingredients into a restaurant meal, you are increasing the complexity, the concentration of information, and so the negentropy.

Another way of creating negentropy is to create artificial scarcity as a way of decreasing dispersion.

This is found in almost all luxury goods, whether they are unique works of art, haute couture dresses, first edition books, penny blacks, beanie babies, etc. In this case the artificial scarcity is maintained by use of copyrights and patents, which allow the price of the goods to be raised above the value of their inputs.

Money of course forms a very special form of a good that has its artificial scarcity carefully controlled. The paper currency itself is controlled through criminal legislation against counterfeiting; money creation in more general terms is controlled by banking and other legislation and the monopolistic actions of central bank policy.

Although marginality has been very useful in pricing simple economic goods that have scarcity, there are other, and I believe much more effective ways of measuring value in such systems. Through the entropy of information, entropy of mixing, etc, science has an extensive mathematical toolbox for dealing with the sort of negentropy found in economics. And the way to deal with it is in a statistical-mechanical way. It is the belief of the author that pricing in this manner will provide much more general ways of pricing, and that marginality will drop out as a special simplified case. Using entropic systems should also remove many of the theoretical problems associated with imaginary Walrasian auctioneers and other such difficulties. Foley has made very significant inroads into this way of carrying out pricing [Foley 1996a, 1996b, 1999, 2002].

All of the above form the first main class of humanly useful negentropy. These are economic products; goods and services generated through concentration and specialisation. These are the stores of negentropy, of economic value.

The specialisation for creation of tools, as against pure decoration, leads on to the second main form of negentropy.

The creation or extraction of all the above products is mitigated through the action of 'negentropic machines' or in the language of maximum entropy production, 'dissipative structures'.

Dissipative structures come in many forms and can be looked at different levels. They include, beasts of burden, trucks, tractors, computers, farms, factories, mines, power stations, markets, supermarkets, stock markets, cities, national economies and the world as a whole. The most important dissipative structures in economics are human beings.

Dissipative structures are 'negentropy machines'. They do two things simultaneously; firstly they produce outputs of high negentropy goods. Secondly they simultaneously produce a much larger output of high entropy waste products, normally mostly in the form of high entropy, low temperature heat. Basic physics demands that the high entropy waste stream must be larger in quantity than the high value negentropy stream. The value of dissipative structures in economic terms lies in their ability to produce large amounts of products with high negentropy values.

The ultimate source of the negentropy can come from a variety of sources, the most important ones for human economies at the moment are the sun, fossil fuels, and human ingenuity, though there are plenty of other minor sources.

The sun provides negentropy for the essential input of food for human beings, it also provides the evaporation and wind to provide the rain for the crops.

The other main negentropy sources are coal and natural gas for electricity production, and oil products for transport.

Prior to the industrial revolution almost all negentropy came ultimately from the sun, providing food for human beings and draft animals, and wood for heating.

As the industrial revolution progressed, energy negentropy from plants and animals was displaced by fossil fuels. That is to say, the physical labour of draft animals and human beings was slowly replaced by the mechanical labour of machines powered by coal, oil and gas.

In more recent times, the information revolution means that computers are increasingly providing the information negentropy that used to be provided from the human brain.

The interchangeability of negentropy sources means that on one point Marx's theories were fundamentally wrong; labour is not the only source of value. In this, Smith was correct, both draft animals and machines can provide value. The physiocrats were also correct in their belief that land can provide value, in its role in capturing the sun's rays.

Where Smith, Marx, Ricardo, Sraffa etc were correct was in their belief that value is an inherent quantity, embodied in the goods and services in the economy.

However, as was found above, Marx was accidentally more than half right, as Bowley's law shows that the negentropy from human beings is roughly twice as important as the negentropy from all other sources put together.

Although, at a local level, sources of negentropy are interchangeable and substitutable, labour does have one very important property that makes it different from all other sources of negentropy. All non-human forms of negentropy can be, and normally are, owned in the form of capital. In the absence of slavery, humans can not be owned.

Where negentropy sources have useful value, normally they are owned. Ownership of a mine or oil well gives ownership of the oil, coal or uranium as power sources, or ownership of the concentrated ores for raw materials.

Ownership of land gives rights to the sun and rain falling on it that allows growing of high negentropy food.

More subtly, the ownership of land in the centre of a city gives the right to use a location that has high negentropy, in that it is a location where it is possible to meet many people and do business with them.

In this regard, it should be noted that adding of negentropy value is not restricted to traditional manufacturing production. Retailers add value by bringing goods to people, by 'concentrating' them in their shops. Travel agents aggregate many different holidays together to allow quick and easy selection of the best value. The financial industry concentrates knowledge of many different investments to find the best ones for their clients.

While the above arguments are clear qualitatively; quantitative calculation of values of negentropy from first principles are problematic.

There has been a recent history of attempts to calculate economic entropy in this manner, mostly dating back to the work of Georgescu-Roegen [Georgescu-Roegen 1971], these attempts are common in the energy economics and environmental economics fields, and are very problematic.

I would like to make it clear that the negentropy being discussed in this paper is definitely not 'emergy'; embodied energy, or 'exergy' or other similar concepts originating from these sources.

I believe that these particular concepts are only useful in narrow well-defined areas of application such as in the energy industry. For a general application in economics they fail to take into account three important issues; the role of locational or concentration entropy, the role of information entropy, and the concept of 'humanly useful' in entropy.

The role of concentration negentropy, the opposite of dispersive entropy or the entropy of mixing, has already been discussed above. Calculation of this for things such as land prices in city centres, and the existence of markets is of large importance. The engineering like approaches of emergy or exergy fail to capture this important source of wealth. Indeed it is the belief of the author that many historic attempts to map thermodynamics to economics have failed as they have concentrated on finding analogies for pressure and temperature, etc. The key parallel is that of chemical potential.

Information entropy has always been important to human economies once they moved beyond subsistence to agricultural markets. Writing was invented in Babylonia as a way of recording storage and sales of crops. Numbers and calculation were invented for the same reasons. Since the oil shocks of the 1970s information negentropy has become one of the main sources of economic growth. For the last thirty years Western Europe has enjoyed substantial growth and significant increases in material wealth despite having an almost constant rate of energy usage. In Europe, information negentropy is the primary source of new wealth.

In theory, pure information entropy is directly measurable in bits and bytes. But the actual negentropy embodied in the display on a computer screen is very different from the simple information displayed on the screen.

Calculating the effects of information entropy is not straightforward, to take a recent trivial example, there has been a revolution in the United States in the extraction of natural gas. Innovative rock fracturing techniques (new information) allows extraction of substantial amounts of cheap gas from shale and other 'tight gas' sources. The information is easily passed from company to company and so the addition of a very small amount of information has resulted in a very significant amount of new useful energy negentropy. How this would be calculated is not straightforwardly obvious (though measuring the drop in natural gas prices is easy enough).

The third problem is the concept of 'humanly useful' negentropy. It is possible to run cars on petrol (gasoline), or alternatively on compressed natural gas, and store similar amounts of calorific value in similar sized cars. However, as a liquid, petrol is far easier to store and handle, so it is more 'humanly useful', and so has a higher negentropy value to a human being.

Due to accidents of genetic history, some things are much more 'humanly useful' than others. I don't believe that an emergy approach can usefully calculate the difference in values between petrol and cng. I am very confident that an emergy approach will struggle to calculate the intrinsic value of an edition of 'Playboy' from first principles

This may all seem very negative and suggest that absolute value is not calculable. Actually, whether negentropy is calculable or not is not important. If it is functioning correctly, a very big if, the market automatically calculates these values for us, and prices the negentropy in $, euro, £ etc.

As Sraffa and von Neumann showed, the long-term prices of goods should reflect the value of the inputs. For most manufactures and services this is easy to observe and prices stabilise quickly, prices are set by the value of the inputs.

Given the day-to-day fluctuations of the prices of things such as food, petrol, shares and houses, many readers may disagree with the concept of intrinsic value. I hope to deal with these objections in the next section.

8.2 On the Conservation of Value

Trivially, value is not conserved.

If I drop a Ming vase on the floor, crash my car, or my country goes to war, then wealth is arbitrarily destroyed.

Similarly, if I bake a cake or build some shelves then wealth is created.

However, I normally buy both my cakes and my shelves, from somebody who has produced them.

If I am sensible, I insure both my car and my vases, so that this possible accidental destruction of wealth is transferred to deliberate consumption; in the form of regular payments of premiums on insurance policies, along with the consumption of the cake, and in the fullness of time the wearing out of the shelves.

I also vote for people who I believe will avoid wars, and democracies very rarely go to war with each other – and most economists would accept that the normal rules of economics don't apply in wartime.

At all times wealth is being continuously destroyed, via eating, wearing out clothes, heating homes, crashing cars, etc. At the same time it is continuously being created on farms, from mines, in factories, offices, etc.

Between the very deliberate acts of production and consumption, people do their utmost to conserve whatever wealth they have.

That they often fail to achieve these aims with financial investments is related to the inherent instability discussed in the macroeconomic and commodity models above.

It is my personal belief that there is a very strong argument for saying that wealth in all its forms is close to being a conserved substance between the acts of production and consumption.

This is supported by the fact that Lotka-Volterra and GLV models in this paper work effectively as models; and that they produce outcomes such as income distributions with power law tails, company size distributions with power law tails, and splits in capital / labour returns that match Bowley's law. These models would not work without the conservation of intrinsic wealth. This in itself strongly suggests that wealth or value is an approximately conserved quantity through the market system.

There are many reasons that people might have for not believing that value is intrinsic, and can appear to be set arbitrarily. The most obvious reason for this is because the prices of things such as petrol (gasoline), houses, artworks, computers, and share prices can vary rapidly according to time and place.

I believe there are five main reasons for these fluctuations in prices, being:

1. locational scarcity
2. artificial scarcity
3. technology change
4. dynamic scarcity
5. liquidity

The reason for variety in the price of 'identical' houses is locational scarcity. The fact that land in the centre of London is more expensive than land in the mountains of Wales has been dealt with in both classical and neo-classical economics using the concepts of marginality.

Artificial scarcity, the reason that diamonds, artworks, vintage cars, beanie babies and money have stores of value that are manifestly different to there production costs is due to the artificial limiting of these items.

Both locational and artificial scarcity were discussed briefly above in section 8.1. While marginality is an effective tool for analysing value in these areas, it is the belief of the author that the dispersional and information properties of entropy will enable a better way of explaining and calculating such values.

Technology change is easily dealt with. All the previous discussions in this paper have been based on economies without technological change, and so values of goods remain constant, ignoring inflation effects, as new capital is created by temporary shortfalls in supply.

Clearly in a real, modern economy, the rapid progress in IT and other high tech industries can result in rapidly dropping prices. This itself is a consequence of the maximum entropy production principle continually working to improve the efficiency of the dissipative structures.

Dynamic scarcity is most obvious with commodities as modelled in section 3. This scarcity, affects things such as oil, metals and agricultural products. The capital intensity and long timescales needed for installation of the capital for commodity production can result in dramatic changes in prices, though generally they take the form of short term spikes in long term stable base prices. The same bubble mechanism is responsible for the dramatic changes in house prices over time.

Liquidity is a much more difficult, and interesting topic. Liquidity is a measure of how easy or difficult it is to buy and sell things. It has already been shown in the macroeconomic model above that liquidity can be artificially generated in a financial system simply by the known short-termism of markets combined with standard financial pricing procedures.

Liquidity has been the subject of much interesting research in recent years. This research suggests that liquidity could be of key importance in the apparent failure of markets to price assets correctly, and in the failure of financial markets in general. It does not appear that this research has so far made much impact in the fields of economics, finance or, with rare exceptions, in econophysics, which I believe is unfortunate.

I therefore propose to give a brief review of some recent research on liquidity and discuss aspects which relate to my own models, and also which are of more general importance.

8.2.1 Liquidity

"Liquidity is not a virtue in and of itself unless it produces a benefit to the real economy."

Yves Smith [Smith 2010]

"But there is one feature in particular which deserves our attention. It might have been supposed that competition between expert professionals, possessing judgment and knowledge beyond that of the average private investor, would correct the vagaries of the ignorant individual left to himself. It happens, however, that the energies and skill of the professional investor and speculator are mainly occupied otherwise. For most of these persons are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield of an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public. They are concerned, not with what an investment is really worth to a man who buys it "for keeps", but with what the market will value it at, under the influence of mass psychology, three months or a year hence. Moreover, this behaviour is not the outcome of a wrong-headed propensity. It is an inevitable result of an investment market organised along the lines described. For it is not sensible to pay 25 for an investment of which

you believe the prospective yield to justify a value of 30, if you also believe that the market will value it at 20 three months hence.

Thus the professional investor is forced to concern himself with the anticipation of impending changes, in the news or in the atmosphere, of the kind by which experience shows that the mass psychology of the market is most influenced. This is the inevitable result of investment markets organised with a view to so-called "liquidity". Of the maxims of orthodox finance none, surely, is more anti-social than the fetish of liquidity, the doctrine that it is a positive virtue on the part of investment institutions to concentrate their resources upon the holding of "liquid" securities. It forgets that there is no such thing as liquidity of investment for the community as a whole. The social object of skilled investment should be to defeat the dark forces of time and ignorance which envelop our future. The actual, private object of the most skilled investment to-day is "to beat the gun", as the Americans so well express it, to outwit the crowd, and to pass the bad, or depreciating, half-crown to the other fellow.

This battle of wits to anticipate the basis of conventional valuation a few months hence, rather than the prospective yield of an investment over a long term of years, does not even require gulls amongst the public to feed the maws of the professional; - it can be played by professionals amongst themselves. Nor is it necessary that anyone should keep his simple faith in the conventional basis of valuation having any genuine long-term validity. For it is, so to speak, a game of Snap, of Old Maid, of Musical Chairs - a pastime in which he is victor who says Snap *neither too soon nor too late, who passes the Old Maid to his neighbour before the game is over, who secures a chair for himself when the music stops. These games can be played with zest and enjoyment, though all the players know that it is the Old Maid which is circulating, or that when the music stops some of the players will find themselves unseated."*

JM Keynes [Keynes 1936]

The following is a brief review of current research and emerging ideas within the field of liquidity. This section is something of a diversion, but research in this area is proceeding rapidly, and important new conclusions have been reached in recent years. It is my belief that these conclusions are important for finance and economics in general, and econophysics in particular, but they don't appear to have become widely know.

This section is somewhat technical, and assumes a basic knowledge of finance, for example through reading a standard text such as [Brealey et al 2008].

Where it is important for my own modelling is that my commodity and macroeconomic models predict endogenous creation and destruction of liquidity. If such models are to be successfully built and calibrated; understanding and meaningful measurement of liquidity will be an essential ingredient.

The discussion is largely confined to liquidity within stock markets and its effects on the pricing and trading of stocks and shares.

More sophisticated readers will also be amused at a discussion that is largely based on the marginalist approaches used in the CAPM and related models. Approaches that are otherwise treated with some derision in the rest of the paper. Like many other aspects of economics; I believe that recasting asset pricing models into a dynamic, chaotic framework will give significant advantages. For the moment almost all the research on liquidity, other than that carried out by econophysicists such as Bouchaud, Potters, Mézard, Wyart, French, Farmer, and others, has been carried out against the traditional models of Debreu, Arrow, et al, and I am obliged to follow this in my review.

As a concept, liquidity rivals entropy in it's opacity. Both the definition and measurement of liquidity presents problems. Historically stock market liquidity has been defined as the ability to trade large quantities of shares quickly, at low cost and with minimal price impact. Unfortunately this actually describes a range of desirable outcomes rather than an underlying concept or property.

Similarly, measurements of liquidity may focus on trading quantity, trading speed, trading cost, volume of trade, etc. Historically it has not been clear whether these different measures were in fact measuring the same thing or not.

In the last decade a large number of papers have been produced giving comparisons of measurements of liquidity and illiquidity, see for example: [Chordia et al 2000, Porter 2008, Korajczyk & Sadka 2006, Goyenko et al 2009]. Many different variables have been used to measure liquidity including trading volume, frequency of shares traded, bid-ask spreads, order imbalances, amongst many, many others. The variety of measures used reflects the difficulty of pinning down exactly what liquidity is. As well as individual measures, composite measures have been created in an attempt to capture the multiple dimensions of liquidity. Indeed there seems to be something of a cottage industry in the creation of new measures of liquidity.

In the more recent papers such as those above, it appears that more sophisticated measures of the different dimensions of liquidity do in fact correlate closely. It also appears that annual and monthly, long time scale data, correlates well with daily data [Goyenko et al 2009]. These results appear to hold true for both stock markets as a whole and individual company shares.

The research above suggests that the different measures of liquidity are in fact measuring the same underlying property, however the exact definition of this underlying property remains elusive.

It appears that including liquidity risk as a factor may explain a number of prominent 'market failures'. The following are given as examples:

Historically, domestic closed end funds have traded at a discount to the underlying shares, while international closed end funds have traded at a premium. These results can be explained by the greater liquidity of the domestic shares vis-à-vis the funds, and the less liquid foreign stock markets compared to the US fund share market [Amihud et al 2005 – 3.4.5].

Similarly, in most countries, where companies have two classes of shares for nationals and foreigners, the national owned shares trade at a lower price than foreign owned shares. In China the reverse is true. This appears to be a consequence of the high level of liquidity in the Chinese domestic stock market [Chen & Swan 2008], while in most countries the domestic market is less liquid than international markets.

Similar arguments can be used to explain the discounts on restricted stocks [Amihud et al 2005] as well as the differences between prices of treasury notes and treasury bills [Amihud et al 2005 – 3.3.1] and also of treasury notes versus corporate bonds; where the price difference can not be accounted for by default risk alone [Amihud et al 2005 – 3.3.2].

Chordia et al, have demonstrated that liquidity problems can explain the post earnings drift that follows unexpectedly high or low earnings announcements [Chordia et al 2009]. While Korajczyk and Sadka show that liquidity can explain up to half the benefits of momentum strategy anomalies documented by Jegadeesh & Titman [Korajczyk & Sadka 2006].

To date I haven't seen a paper discussing the anomalies of dual listed companies such as Royal Dutch Shell, however I confidently expect liquidity to explain the long-term diversion of such share prices.

While all the above are interesting, probably the most important result of recent research into liquidity, is that liquidity, or more correctly, liquidity risk appears to be a major component of asset pricing.

Amihud et al, give a full review of these results, which demonstrate that a liquidity augmented Capital Asset Pricing Model (CAPM) gives much better results than a traditional CAPM [Amihud et al 2005 – 3.2.3]. Other work supporting this view has been carried out by Acharya & Pederson and Pastor & Stambaugh using single measures of liquidity [Acharya & Pedersen 2005, Pastor & Stambaugh 2003], Goyenko et al, Korajczyk and Sadka [Goyenko et al 2009], Liu [Liu 2006 & 2009] and Lee [Lee 2005].

Given the poor historical performance of the CAPM, the Fama-French three factor model has often been used as an alternative. This uses firm size and book-to-market ratio in addition to a market index. The book to market ratio is fundamentally equivalent to the ratio of K to W in the modelling of part A; where K is the real capital, the book value, and W is the market capitalisation.

Results from the research above strongly suggest that a single liquidity measure can replace both firm size and book to market ratio and give improved results. This suggests that both firm size and book to market ratio may be surrogate measures for liquidity risk.

As discussed in section 2.1 above regarding the companies model, Fama and French's own work indicated that as well as the factors of risk, firm size and book to market ratio, a fourth momentum factor needs to be included to fully explain share price movements. If the research in liquidity stands up to further investigation, it suggests that share price movements can be explained by just risk, liquidity and momentum.

Further to that, and in line with the workings of the macroeconomic models of section 4 above, the work of Korajczyk and Sadka [Korajczyk & Sadka 2005] suggests that provision of liquidity also reinforces momentum strategies. This suggests that short term momentum pricing is not 'behavioural' or even plain stupidity, but is 'rational' behaviour for participants, until the market finally reaches a position far out of equilibrium, and endogenous liquidity creation is stopped.

Taken together it appears that a new 'three factor' asset pricing model involving the market beta, liquidity risk, and momentum may be superior to both the CAPM and the Fama-French 'three factor' model.

This then becomes much more significant at the level of the whole stockmarket, especially in the light of the extensive work by Shiller and Smithers regarding the long-term valuation of stock markets. This work is very well summed up in 'Wall Street Revalued' [Smithers 2009].

The central thesis of this work is straightforward. Shiller and Smithers find that stock market prices do not follow random walks, but are in fact mean reverting over decadal timescales. Two measures in particular are able to capture the over or under valuation of the stock market, the two measures that do this are CAPE and Tobin's q.

Tobin's q is of course the same thing as the book-to-market ratio, the same value used at company level in the research of French & Fama and various other researchers in liquidity. q is just the ratio of K to W.

At a whole stock market level, both company risk factors and company size are averaged out, leaving only book to market value as a meaningful indicator.

It appears that by measuring the value of Tobin's q, researchers such as Shiller and Smithers have simply been measuring the liquidity of the whole stock market, with Tobin's q acting as a close proxy measure for liquidity.

On the other hand the 'CAPE' is the 'cyclically adjusted price to earnings ratio', which is simply the price to earnings ratio adjusted to a long time period; normally ten years. The CAPE also provides a very good measure of over/under valuation, and consequently correlates very closely with Tobin's q.

Working backwards, the logical conclusion is that the over- or under-valuation of the stock market, defined by long term earnings and prices, is simply a measure of the overall liquidity in the stock market, and that deviations away from the long term average are almost wholly due to liquidity.

The anecdotal evidence that equity prices are linked to liquidity is certainly plausible. The dramatic fall in share prices during the 2008 Credit Crunch and the subsequent rebound following the introduction of quantitative easing and other fiscal loosening are strongly suggestive of a direct link between liquidity in the economy as a whole and equity prices.

To date there appears to have been relatively little research in this area, which is unfortunate considering its potential importance.

Pepper & Oliver [Pepper & Oliver 2006] have produced an extensive study of this issue. Their work is very persuasive, and an excellent discussion of how liquidity works in practice, but the attempts to link share price levels to monetary data, while compelling, are not conclusive. This reflects the problems of finding trustworthy monetary data, a problem that the new approach using liquidity measures may alleviate.

More recently, Chordia et al, Jones, and Pastor & Stambaugh, have used different measures of market liquidity and have all noted correlations of liquidity to market movements; particularly sharp declines in liquidity associated with declining markets. [Chordia et al 2001a, Jones 2002, Pastor & Stambaugh 2003].

Liu has carried out a longer and more detailed analysis and concludes that there is evidence for mild changes in liquidity corresponding to market movements, and that this is consistent with the argument that liquidity is a state variable important for asset pricing [Liu 2006, 2009].

Chordia et al have carried out an empirical analysis of the relations between liquidity in the stock, bond and money markets, and suggest important links between liquidity, volatility, and monetary policy [Chordia et al 2005].

Important work in this area has also been carried out by econophysicists such as Farmer, Bouchaud and Wyart, this is discussed further in section 9.1 below.

While it is early days, it appears that not only is liquidity of fundamental importance in the pricing of stocks and other financial assets, it appears that it may in fact be a fundamental state variable of the stock market, and one that is straightforward to measure on a timely basis. If this is true, then there are some big implications for both finance and economics.

Historically, attempts to measure liquidity at a national level have focused on measurements of money supply. Most notably in the UK in the early 1980's monetary policy was used in an attempt to control the economy. The policy was quickly discredited, primarily due to the difficulties of collecting timely and accurate monetary data, and also due to the ease with which the sources of such data could be manipulated by financial institutions, see Pepper & Oliver for more details.

In marked contrast, some of the liquidity measures used in more recent liquidity research, for example those of [Chordia et al, 2002 & 2005], are easily calculated on a daily basis from stock market information. It would be trivially easy for indices and sub-indices of liquidity to be set up that could be observed and used by both the financial markets and economic actors.

The research on liquidity suggests two implications for finance that are both quite profound, the first area relates to the pricing models based on Black-Scholes, the second to the pricing of shares under the CAPM.

Almost all modern option pricing theory is based on the Black-Scholes model, or other closely-related models. Black-Scholes has been one of the most important mathematical contributions to economics or finance, and certainly the only one to have come into widespread day to day use within the financial industry.

However, one of the core assumptions of B-S is that options on shares, as well as the underlying shares themselves, can be bought and sold easily in highly liquid markets.

The recent body of work studying liquidity of financial assets suggests that this assumption is profoundly flawed. It seems likely that prices of both options and underlying assets will be affected significantly by liquidity. It also seems likely that the effects might not be the same for the option and the underlying. Consequently this would suggest that B-S models would, as a minimum, need modifying to take into account the effects of liquidity.

That liquidity should be a concern for quantitative finance in general seems obvious; Long Term Capital Management (LTCM) was brought to earth largely through trading in products that became illiquid overnight, and illiquidity was a major factor in the collapse in asset prices that took place during the credit crunch.

Clearly the effect of liquidity on asset prices appears to be an area ripe for more quantitative analysis. The possibility of a relationship between liquidity and volatility seems particularly interesting. Other than the work of Chordia et al [Chordia et al 2005] discussed above, there appears to be little published research in this area.

If it is true that liquidity is an easily measurable state variable of shares, and that also there are mathematical relationships between liquidity and volatility (which seems plausible), then it may be that measurement of liquidity might be able to give good timely measures for current volatility that can be used directly in Black-Scholes models; rather than the current practice of imputing from historical volatility.

A second significant area of interest for the application of liquidity in finance is to asset pricing models. The research to date suggest that liquidity can replace both the size and book to value elements in the Fama-French three factor model, leaving only risk and liquidity, along with momentum, as the determinants of equity prices. Or to put it another way, liquidity risk appears to be the main missing risk element of the various CAPM models. This knowledge gives the intriguing possibility that it should be possible to fully hedge an asset portfolio, and, more questionably, that this might even lead to self-stabilising markets in asset prices.

As discussed above, some of the liquidity measures are easily calculated on a daily basis from stock market information.

It would be trivially easy to set up a standard 'liquidity index', similar to the VIX index for volatility, and encourage trading of futures in the index and so allow a deep market to form in this liquidity index.

Investors would then be able to go long on shares, or stock market indices, and simultaneously short the liquidity index to protect against a reduction or collapse in liquidity. If the recent work on liquidity is correct, this should give almost full protection on an asset portfolio of investments.

Interestingly, this should act in a strongly counter-cyclical manner. Given the mean reversion properties of the market as per Shiller and Smithers, liquidity protection of this type should be cheap at historical liquidity lows, but increasingly expensive as liquidity bubbles formed; if, of course, it was correctly priced.

If such hedging functioned correctly, the cost of protecting against excessive liquidity would itself prevent excessive overpricing of assets and would automatically withdraw liquidity from the market as prices became frothy.

As well as having an overall liquidity index, there would also be scope for sub-indices tracking individual sectors. Indeed it may make sense to re-sort companies from traditional 'industry' sectors into groupings that share a common pattern of historical liquidity and volatility behaviour.

Clearly correct pricing, and the formation of a sufficiently deep market to cover even a portion of the stocks traded might be problematic. There are also clear possibilities of counter-party default dangers of the sort that afflicted AIG following their substantial underpricing of CDS risk.

If liquidity risk is the main missing factor in the CAPM model, and also it proves possible to enumerate and hedge against this risk, then by analysing the resultant data, it may be also be possible to analyse and quantify the remaining residual risks in the pricing of assets.

In an ideal world, under these circumstances, it seems possible that momentum trading would become difficult and short-term speculation might be a difficult and profitless activity. This could lead to financial investment becoming a predictable and rather dull area of both business and economics. Common sense, and the weight of history, does suggest that this is more likely to be a possibility rather than a probability.

However, if deep and efficient markets in liquidity futures did form, then speculative interest would allow liquidity index pricing to change in response to external factors such as government policy, oil shocks and other exogenous events. This leads to the possibility that liquidity measures could also be very useful for macroeconomic control.

Having liquidity indices of this form could assist governments in targeting liquidity in stock markets, and in the economy in general. This might answer the problem of the poor quality and timeliness of traditional monetary data.

Casual observation suggests that there is poor short-term correlation between the supply of liquidity to financial markets and the health of the economy as a whole. In the United States for example, in 2005 and 2006 the stock market was booming, with very high liquidity, even though the economy as a whole was struggling (as expected in the Bowley squared predator-prey model of section 4.9).

In such circumstances, central bankers face acute problems. With the single tool of interest rates, governments are in a cleft stick. This was admitted to recently by Kate Barker, an ex-member of the UK monetary policy committee [Guardian 2010]. In 2005 the UK appeared to be in both a housing bubble and a stock market bubble, but the general economy was sluggish, and inflation was historically very low, with the threat of deflation in the wings. Raising interest rates to calm down the housing and financial markets risked initiating a recession, possibly moving into outright deflation. However, failing to raise interest rates caused an ongoing bubble to continue its expansion, which had very unfortunate consequences including the collapse of Northern Rock, and the bailing out of Bradford & Bingley, Royal Bank of Scotland and other institutions.

As discussed above, in the past, attempts have been made to control the economy through the control of the money supply, or as Cooper correctly describes it by controlling the supply of debt. Historically these attempts have not worked well, partly because the money supply is difficult to quantify and measure reliably.

I believe a second problem is that there are two sources of liquidity. The money supply and debt is one of them, but the endogenous creation of liquidity within the pricing system is another, and in my view this is the prime source and the larger source. So, certainly increasing the money supply and debt can increase liquidity in the stock market. But increases in stock valuations also create their own liquidity, and also provide apparent extra wealth against which new debt can be secured. These two sources of extra liquidity feed on each other in a most unhealthy way.

I believe targeting a liquidity measure in stock markets may be more effective than monetary targeting, as a liquidity measure is measuring the output, the residual, of the liquidity creation process. A certain amount of debt and new money supply is needed in an economy. If insufficient is supplied, then the stockmarket declines, if too much is provided the stockmarket booms, the stockmarket is normally a good weather vane for liquidity in the economy as a whole.

An important caveat here is the role of housing, which as discussed above in section 6.3 is more important than even the stockmarket as a driver of booms and busts.

Controlling liquidity and money supply for an economy will only be effective if the housing market is stabilised. Absent an effective measure of liquidity in the housing market, then other damping measures and long term indicators need to be used such as historical ratios of house prices to wages and ratios of mortgage payments to rents.

The macroeconomic models in section 4 above suggest that liquidity can be formed endogenously, in exactly the way proposed by Minsky. This suggests that, just as central banks are expected to control changes in the money supply caused by fractional reserve banking, it seems appropriate that they also be obliged to control money supply growth caused by Minskian asset price bubbles.

The recent research in liquidity, and the models in this paper, suggest that liquidity needs to be targeted separately, in addition to the inflation targeting of the overall economy. The ease and timeliness with which liquidity can be calculated, and compared to historical liquidity levels, suggests that this would be relatively straightforward to do.

For instance it might be possible to use active management of the bond market as has recently been done under 'quantitative easing', on a regular basis to increase and decrease the liquidity of financial markets generally. So in 2005-06 it might have been sensible to actively embark on 'quantitative tightening' to restrain the financial markets, while simultaneously lowering interest rates to assist the larger non-financial economy.

This takes us back to the building-atrium air-conditioning model discussed in the 'Bowley squared' model in section 4.9 above. The figure is shown again below:

Figure 4.9.1 here

On these lines there was an interesting recent proposal by Martin Weale of the National Institute for Economic and Social Research [Telegraph 2010a] to introduce a specific tax on debt. If this tax were differentiated for housing borrowing, financial borrowing and non-financial borrowing (industry, services and other non-financial borrowing), in theory it might be possible to kill bubbles in the housing or stock markets while maintaining economic growth. Whether this would be practical remains to be seen, it is likely for example that there would be significant problems preventing companies gaming such a system.

8.2.2 On the Price of Shares

In his excellent book 'The Origins of Financial Crises' George Cooper [Cooper 2008] points out that one of the clever sleights of hand of neoclassical theory is to demonstrate how supply and demand works well for simple commodities, and then blithely assume that this pricing system is equally valid for financial 'commodities' such as share prices.

As Cooper points out this is clearly wrong as the whole point of financial assets is that they have genuine scarcity or 'artificial scarcity' value and, very specifically, supply can not be ramped up to meet demand. People invest in gold because most of the world's existing gold deposits have been found and are owned. Similarly people invest in shares under the expectation that companies will not arbitrarily keep issuing more shares to other investors.

I would like to discuss this idea in practical terms and look at why share prices are so different to commodity prices, and discuss one possible way of looking at what is the source of company value, and what is the 'price' of a stock.

The prices of most goods and services, especially those that do not depend on scare mineral inputs are characterised by long-term stable prices. Though, as has been shown in the commodities model, even the prices of some basic commodities can vary widely through the results of delays in installing capital.

The valuation of a company can take these 'simple' commodities as a starting point. Most companies take in one sort of commodity from their suppliers and produce a more sophisticated commodity, which they then sell on to their customers.

Speaking as a humble engineer the 'value' of a company is how many useful things it produces every month.

However even an engineer is forced to admit that, for the owners, the meaningful value is the difference in output of the manufactured items and the various inputs; raw materials, labour, rent, etc. The market capitalisation is based on the profit stream, as discussed in section 2.

Assume that the outputs are more or less homogenous with a single price level in the market. Assume also that there is one main input responsible for the majority of costs, normally one of the following: a single raw material input, energy, rent, capital or labour.

If the difference between inputs and outputs is 10%, as the preceding models have assumed, then a 5% change in the cost of the main input, or the price of the main output price will result in the company's value changing by roughly 50%.

A moment's thought shows that the value of the company becomes a derivative price based on (at least) two underlying prices. And this derivative has very substantial leverage on the underlying prices of the commodity inputs and outputs.

This is further complicated by market interest rates. Let us assume, firstly that the input and output prices are absolutely stable, and that therefore the business has an absolutely stable dividends stream. Then to price the company's shares, this stream must then be compared to the risk free market rate.

So the price of the company's shares will go up as the risk free rate goes down, and down as the risk free rate goes up. As such the company's share price will vary in the manner of a bond, or more accurately, a perpetuity.

So the market 'value' of the company is in fact the price of an artificial perpetuity based on a derivative of two or more underlying commodity prices.

On top of this variability needs to be added the effects of things such as liquidity and momentum as discussed in the previous sections.

Looking at stocks and shares in this way, even such a simplistic model shows that the 'price' of a company is related to the prices of normal commodities in a very complex way. This gives an insight into why share prices are so volatile.

It certainly makes it clear that supply and demand cannot operate in a normal manner on company share prices, the price is not simply set externally by the utilities and preferences of buyers and sellers of shares.

9. Supply and Demand

9.1 Pricing

An interesting puzzle in the history of economic thought is why the mathematization of economic theory in the 1940s and 50s took place through the formalization of the static Walrasian model, rather than through the study of infinite horizon production based models that arise from the Classical view. This puzzle is particularly intriguing because the best mathematician who ever worked on economic problems, John von Neumann, introduced the key mathematical tools in a study of such a Classical, infinite-horizon, production-based model before Arrow and Debreu used the same tools (mostly topological) to formalize Walras.

[Foley 1990]

The idea that production rather than exchange is the source of value is contentious within mainstream economics, though why this is so is puzzling. Both the theoretical history and empirical data support this central view of production.

The theoretical debate goes back at least to the work of Sraffa and the Cambridge Capital Controversies. The conclusions of the debate have been discretely forgotten; Sraffa's work demonstrated that the production function approach of marginalism was not appropriate, and that pricing of produced commodities through the long period classical approach was the appropriate way forward.

Sraffa and others demonstrated that pricing of capital can not be carried out using marginality. The original work of von Neumann was also classically based, and also showed that a coherent system of prices can be built using the approaches of classical economics. The work of Sraffa and von Neumann has since been systematically synthesized by Kurz & Salvadori [Kurz & Salvadori 1995] to give a modern classical framework.

Meanwhile Arrow, Debreu and others took von Neumann's insights and battered them into a neo-classical framework; back into the realms of field theory.

With regard to the clash between classical and neo-classical approaches, the work of Burgstaller [Burgstaller 1994] is particularly intriguing, in that he proposes that both the neo-classical and classical approaches can be presented as subsets of a unified approach.

In particular he shows that the neo-classical approach is appropriate when no labour is involved, as for example in a pure exchange process, while the introduction of labour results in the necessity of a classical approach. (It is the opinion of the author that the neo-classical approach is only appropriate when no value is added, whether by labour or machines. This remains however only an opinion.) Burgstaller's work suggests that the neo-classical approach is only suitable for processes such as the purchase of raw materials, or interestingly in the exchange of financial products.

In this light, marginalism would at first appear to be very useful in defining the mechanics of the purchase and sale of financial assets. With financial assets, owners have strong preferences for ownership, based on different preferences for risk, liquidity, etc. At a particular point in time, they will also have set initial endowments.

Following an exogenous event, such as an unexpected change in dividends, interest rates, etc, market participants will presumably want to rebalance their endowments to bring them into line with their preferences.

Unfortunately, the financial field of market microstructure, with its wealth of data, has long moved on from the simple cartoons of static supply and demand curves.

Research in market microstructure has shown that the determinants of prices are stocks (inventories not shares), information, liquidity, etc, while marginality has been quietly sidelined. This is primarily caused by the problems of matching supply and demand over a time basis. When time is taken into account, marginality is replaced with a focus on inventories of financial assets owned, and the information encoded in order flow. Or, as has been pointed out previously, comparative statics cannot model effectively in a dynamic environment.

These conclusions on sources of costs are based on substantial quantitative research, supported by some very interesting theoretical work. This work is well reviewed in papers by Stoll, Madhavan and Biais et al, Stoll is a particularly good introduction [Stoll 2003, Madhavan 2000, Biais et al 2005].

Lyons discusses this with great clarity in 'The Microstructure Approach to Exchange Rates' [Lyons 2001]. In sections 6.3 to 6.5 Lyons captures the difference between the 'Tastes & Technology' approach of traditional economics and the 'Information & Institutions' approach of market-microstructure.

What Lyons is too polite to point out is that the utility approach of 'tastes & technology' rests on hypothetical foundations invented in the late 19th century, while the 'information & institutions' foundations are based on theoretical models proposed to fit large scale data sets through the finish of the 20th century and the start of the 21st.

As Lyons notes: "The microstructure approach also includes utility maximization, but as we saw in chapter 4, utility is specified very simply, typically in term of terminal nominal wealth."

Market microstructure has analysed two main forms of markets, those composed of continuous double auctions and those made by market makers. The second is of particular interest.

Market makers buy and sell shares or other financial assets in financial markets. Financial markets involve buying and selling things in a dynamic time frame. There is no guarantee that somebody will want to buy something at exactly the same time that someone else will want to sell something. Market makers keep markets working by 'providing liquidity' and ensuring that there is always somebody who is willing to buy and sell shares at any particular time.

Market makers make markets by acting as intermediaries and do not normally hold on to shares on a long-term basis. They make their living by maintaining a small margin between the prices at which they buy and sell. Market makers are normally obliged by market rules to post prices at all times, and are obliged to fulfil purchases and sales at their advertised prices. They normally have to do this while in competition with other market makers. The speed of trading means that markets never formally 'clear' and market makers are often working 'blind' with little information other than the recent trading history of themselves and their competitors, and the knowledge of the level or inventory of assets that they currently have on their books.

Market makers make money by having a margin between the prices at which they sell and buy, this is known as the 'bid-ask-spread' or simply spread.

Market microstructure empirical research, experiments and theory have left the models of supply and demand behind; primarily because there is no evidence to suggest that market makers use marginality in pricing, and significant evidence that other factors are used in their pricing strategies.

Research suggests that the bid-ask spread is made up of five main components, these are discussed briefly below, for a more detailed review see Stoll, Madhavan or Lyons.

The first type of cost is administrative or 'handling' costs and other overheads. These reflect the costs of renting offices, paying wages, running systems etc. For modern electronic share-dealing these costs are generally very small, though the arms war of high-frequency and algorithmic trading, which demands both expensive technology and highly numerate employees may be pushing these costs back up. For non-standardised 'over-the-counter (OTC) products, these costs can also be higher.

Another cost may be caused by non-competitive practices, such as industry standards on tick sizes or standardised bid-ask spreads.

A third source of cost is related to the cost of holding unwanted inventory. Market makers are like bookies at horse races. Bookies probably know the horses and jockeys far better than the punters, but they don't make their money by betting on the horses. They make their money be balancing the supply and demand of the various punters, and making sure they take a small margin in the middle. It is dangerous for them to take a lot of bets on one horse, even if they think the horse will probably lose, because if it does win they will be wiped out. If they do get a lot of bets on one horse, even if they think the horse is lame, they will increase the price of that horse (by decreasing its odds) and decrease the price of the other horses (by increasing the odds) until they bring their positions back in to line and ensure that they will make a small profit whichever horse wins. In the same way market makers also generally know their markets much better than their customers. But they do not normally wish to hold large positions in a single stock, because if the price of that stock should collapse unexpectedly then they could go bankrupt overnight. Because of this managing and hedging inventory can be a significant source of costs.

This leads on naturally to the fourth source of cost, the cost of 'adverse information'. However well the market-maker knows his markets, he will never know them as well as 'informed traders', that is people who are closely linked or even working for the company whose shares are being traded, and so will have knowledge of good or bad news about the company before the market-maker. These 'informed traders' are able to make money out of the market-maker, and for the market-makers to stay in business, they must collectively recoup this money from the 'uninformed traders', they do this by having an appropriate extra margin in their bid-ask spreads.

A final source of costs is what is known as the 'free option' cost. In a well administered market, providers of liquidity are forced to hold their quotes open for a fixed minimum period. Priority rules then ensure that orders are closed out in a fair manner normally based firstly on price priority, then on time priority, where prices are equal. These rules force market-makers to compete with each other and so protect the ordinary share-trading public.

One problem with this is that it forces the market-maker to hold his price for a fixed time period; in this time the market price may move, giving an advantage to a well informed customer who can make money out of this 'free option'. To protect themselves, market makers add a small extra margin into the bid-ask spread.

As well as the work of pioneers in finance and economics covered in these papers, this area has also recently been extensively researched by others from the field of econophysics such as Farmer et al, Wyart et al and Bouchaud et al [Farmer et al 2005, Wyart et al 2008, Bouchaud et al 2009].

The convergence of the work of economists and physicists in this area is interesting both in its own right, and also more remarkably as it demonstrates that physicists and economists are in fact capable of both reading and citing each others recent research work.

Taken together, the fields of market microstructure and econophysics seems close to providing full models for financial market functions that combine good theoretical underpinnings with good fits to actual data.

There also appears to be strong areas of similarity between the research that has been carried out in the area of market microstructure and that of post-Keynesian pricing theory. To the best of my knowledge these parallels do not appear to have been investigated.

Post-Keynesian pricing theory is primarily empirical, and its empirical basis is of a depth and surety rarely found in economics. In 'Post-Keynesian Price Theory' [Lee 1999] Frederic Lee gives an excellent review of how far disconnected from reality is the marginal approach to the pricing of manufactures. Despite the book's title, 80% of the book provides an excellent review of extensive historical research showing how businesses actually carry out pricing policy.

The results of the research show that, in the real world of business, marginality is non-existent. In particular, most businesses have their maximum profitability at maximum output. Diminishing returns simply don't appear in real world manufacturing, this has been clear for decades, see for example [Eiteman & Guthrie 1952]. In almost all production processes costs decrease with production right up to maximum output, and extra capacity, in the form of new factories, can be added easily and speedily. Under these conditions; of decreasing returns to scale, marginality is irrelevant as it simply cannot work.

In the real world almost all companies carry out their pricing using some variation of an average cost and 'mark-up' basis, with standard additional costs being added to the prices of the inputs.

It is important to note that, as with market makers; manufacturers and retailers also price their goods in advance of sale when supplying to the public. They also often do so on long-term defined contracts when supplying to other companies. It is also notable from the post-Keynesian research that manufacturers and retailers focus strongly on inventory levels and the prices of their competitors for their decisions on prices and production quantities.

An interesting piece on pricing in industry by Langlois [Langlois 1989] looks at pricing in the automobile industry. Particularly interesting in this research is the prime role manufacturers give to the monitoring of inventories in pricing goods and controlling output.

All this is immediately obvious to anybody who has actually worked in a factory environment, including of course pin factories.

The existence of mark-up pricing and controls based on inventory levels, along with the absence of diminishing returns, is strongly supportive of the classical economists' point of view.

One of the main conclusions of all the research into the real world of business is the absolute irrelevancy of marginal pricing outside the areas it was originally used by the classical economists, areas such as land or mineral extraction.

The parallels between post-Keynesian pricing theory and market microstructure theory are clear.

Companies are obliged to behave as market makers.

Complex market makers; but market makers none the less.

For a company, their 'mark-up' is directly analogous to the 'bid-ask spread' of the financial market-maker; though the weightings in the spread are a little different.

An easy example to follow is that of a retailer. A shop buys goods from manufacturers and sells the same goods on to the general public. So in this case the main inputs and outputs are identical; in the same manner as a financial market maker. While overheads for a financial market-maker are very small, for the retailer they are much larger, and need to pay for the remaining inputs of staff wages, distribution costs, rental of shop space, services, advertising, etc. They also need to include for payment of profit on capital and interest on debt. But just like stock markets, prices never formally 'clear', and pricing is based on information from competitors, rates of sales turnover, and levels of inventories of goods held. Purchases of new goods are strongly influenced by inventories of goods within the supply chain. Prices are raised when turnover is high; at Christmas for example, and are dropped in the January sales to get rid of excess inventory.

Manufacturers, or providers of services, follow exactly the same logic, but now the stocks bought and the stocks sold are of different goods, and the 'bid-ask spread' is even larger and now includes the costs of the value adding processes used in production.

It appears that the substantial body of post-Keynesian empirical work could benefit strongly from looking at analytical ideas from market-microstructure and econophysics research.

Indeed the processes of market-making and market microstructure approaches in general appear to be ubiquitous and universally applicable in its role of price formation in economics as well as finance. Perry Mehrling provides a very thoughtful analysis of the US banking system using market microstructure approaches, while Lyons does the same for currency trading [Mehrling 2010, Lyons 2001].

The processes described by market microstructure concentrate on order flow and spread. They arise from markets in which prices are dynamic and not formally settled, where prices ultimately are linked to long-term values, but public information on those values is usually not complete.

In this price discovery process information is found, and long-term prices are defined on different levels. Long-term prices will ultimately link to fundamental values, but as has been shown above, 'correct prices' will also vary with the point that has been reached in different cycles, on levels of liquidity and debt in the economy, on levels of government activity in the markets, on relative levels of trade and capital flows between different countries and levels of inventories of financial assets in the portfolios of different investors. As described in section 8.2.2, in such complex systems, the link to 'fundamental' values is weak and time dependent.

Market microstructure describes the mechanisms that allow buyers and sellers to discover these 'correct' values.

It is not the balance of buyers and sellers that define these values.

As ever Foley hits the nail on the head:

I believe that the informational view of prices brings modern economics closer to the Classical economists than to Walras. The Classical economists argued that costs of production are the fundamental determinant of prices. Costs of production are the relevant transversality condition for durable and reproducible commodities. Thus forward-looking speculators will price current commodities on the basis of their estimates of long run costs. The new information that disturbs asset prices is, in this way of thinking, primarily information about long-run costs of production.

[Foley 1990]

….it makes more sense to interpret the commodity bundles of agents as stocks, such as stocks of consumer durables (the food in the refrigerator, for example). The availability of well-organized markets permits agents to keep close to their desired stocks at equilibrium prices at all times. Since agents are human beings who get hungry, wear out clothes, and in general deplete stocks, it is necessary for them to make transactions more or less continuously to keep close to their desired stocks (selling their labor-power, paying their rent, buying food, and so forth). These transactions, which generate national income, are not in this way of thinking the result of irreversible movements from far-from-Pareto endowments to a Pareto allocation, but the result of agents' constant effort to maintain their desired stocks given equilibrium prices. Something like Hicks' Sunday night, in which the economy and its agents are suddenly moved to a point far from the Pareto set, occurs only rarely as the result of external shocks to the system.

If we regard actual data on economic transactions as arising in this way, conventional specifications of demand functions in which flow transactions are a function of market prices and incomes are inappropriate. The prices at which transactions in a close-to-Pareto allocation economy take place are in fact equilibrium prices, which we can thus observe directly. The quantities transacted, however, depend on the dynamics of consumption and depreciation of stocks, which require specific modeling. The assumption that agents generally remain close to desired stocks, and that the economy can as a result be analyzed with the concept of reversible transformations, is a strong abstraction. For example, an agent who loses her job typically feels that she has been forcibly (irreversibly) moved to a lower utility level. Real economies experience shocks (wars, revolutions, depressions, and technological innovations, for example) that intuitively seem to be best understood as irreversible transformations. The gradual processes of economic growth and development move agents to higher indifference surfaces, *but on a time scale much longer than that of the establishment of market prices. We would like to emphasize the notion that the method of reversible transformations is best adapted to analyzing ongoing economies operating more or less normally.*

[Smith & Foley 2008]

And, of course, moving to a focus on stocks means moving to a world of dynamic equilibrium, of Lotka-Volterra models, predators and prey and maximum entropy production.

The work of Sraffa, von Neumann, Kurz & Salvadori, Burgstaller, etc give a very good starting point for the calculation of long term prices in such a world. Unfortunately the approach used by these authors remains one based on static processes and single period analyses. Recasting this work into a dynamic approach should be straightforward. A sensible way forward would seem to be by using the market microstructure, market maker / post-Keynesian approach to attack the single-commodity, multiple-commodity, joint-production, etc, problems. If a simulation approach was used, rather than an algebraic approach, this might also reduce the ratio of headaches to results.

Unfortunately the non-existence of diminishing returns, and the work of Sraffa et al leave a problem as to what exactly does form the limit on the volume of goods produced. An obvious limit is scarcity, the restraints on growth provided by a limited planet. I would like next to explore in detail just how much scarcity there actually is in the world at the start of the 21st century.

9.2 An Aside on Continuous Double Auctions

In previous sections I have been scathing about the fashion for high-frequency trading. In an act of some foolishness I would like to look at this in more detail. I do this with some trepidation, moving into an area where debate is vociferous and my knowledge is limited. However, despite my inexperience, from my naïve viewpoint it appears that the structure of financial markets often seems perverse and appears to be incentivised against easy price discovery and the simple execution of large trades.

This discussion is also in the wrong section, and logically fits with liquidity or the control of dynamic systems, however for reasons of intelligibility it was necessary to leave this discussion until after the discussion of the role of market microstructure.

Finally, the debate in this section is somewhat technical, and not core to the paper. It is simply an example of how using a controls system mindset might allow more efficient markets to be constructed. I suggest that those who are not interested in these issues skip this section. For those who are interested, but are unfamiliar with market microstructure, reading the excellent paper by Stoll [Stoll 2003] should give sufficient background to follow this section.

As discussed previously, stock trading is now dominated by 'high-frequency trading'. On the major western stock markets the majority of trading is done by high-frequency algorithmic trading. In these stock markets supercomputers trade billions of dollars of trades in seconds using automated algorithms. Individual bids and offers may be held open for fractions of a second.

This is done in the sacred name of 'liquidity', which is assumed to be always a good thing.

The current data suggests that high-frequency traders largely provide their liquidity to well-traded shares in preference to infrequently traded ones. They also prefer doing so at times of low volatility to high volatility. By definition this is opposite to the requirements of effective liquidity supply, and the reverse of a couple of centuries of defining the role of liquidity suppliers.

The quote from Keynes at the beginning of section 8.2.1 gives his views of the benefits of liquidity, and it appears reasonable to assume his opinion of high-frequency trading would not have been positive.

More recently other experienced financiers have shared similar views [Noser 2010], and at least one commodity trade body has denounced 'parasitic' traders [FT 2011b].

That my concerns are more widely held is supported by the recent decision of Credit Suisse to start a 'light-pool' for institutional investors. This is deliberately aimed at large volume traders and 'opportunistic traders' will be specifically denied access to the system [FT 2011a].

My own fundamental problems with high-frequency trading are three fold.

Firstly it is trivially obvious that the value of companies does not change from microsecond to microsecond. In fact research suggests that publicly announced information has negligible effect on trading, see for example [Joulin 2008, Ranaldo 2008, Bouchaud et al 2009]. In fact information largely comes from large trades by institutional traders, and as Bouchaud et al make clear, the savagery of the market means that such large trades now need to be broken up into small trades and fed into the markets in a piecemeal fashion, sometimes in periods as long as months, to prevent adverse price movements.

This brings me to my second fundamental problem with hft. In a dynamic, chaotic system, reducing the time constant of trades, allowing trades to be faster and faster, increases the speed and volatility of short-term momentum processes.

To go back to the idea of a traditional market, if I was a customer trying to buy or sell oranges from or to a stall-holder, I would naturally prefer to see all the stall-holders displaying their prices while I get the opportunity to walk around and chose the best price. If each stall holder just flashed a quote for one second and told me to take it or leave it, things would be much more difficult for me.

Finally, and leading on from the above, there is very little evidence that high-frequency trading does in fact provide liquidity. The paper by Bouchaud et al is magisterial in its depth, and the main conclusions are that, although a lot of shares are traded, revealed market liquidity is very low. Like the orange sellers in my example, the short time of quotes makes it very difficult for buyers and sellers to move large volumes without changing the prices.

In their role as liquidity providers, high-frequency traders have taken over the role of market-makers as being traders who do not buy shares to hold in their own right, but simply buy and sell to others and make a profit on this trading.

Unfortunately the traditional duty of market-makers to ensure an orderly market, and not to favour themselves over their clients, seems to have been lost in the cracks somewhere.

As Noser points out, there are well-established rules for order book precedence in market-making and there is no obvious reason why high-frequency traders should be exempted from these rules.

As a minimum high-frequency trading needs reforming, with a return to the rules traditionally imposed on market makers, including a minimum required time for a quote to be offered of say five seconds, along with reinstatement of the normal price and time rules for filling orders. (Traditionally market order books are filled first by precedence of price, and then by time of arrival of the quote.)

This would allow competition to revert to that of price and spread, rather than speed. The resultant recreation of meaningful bid-ask spreads, though possibly larger would be much better at providing signalling of liquidity requirements, which is of course the whole point of market-making in the first place. The increase in price transparency should far outweigh the cost of the free options offered.

Looking more broadly, speed of trading, and narrowness of spread are not the only benefits required from a liquid market. As is seen in Bouchaud et al's paper, high speed does not guarantee the ability to trade a large volume. Similarly, a narrow spread does not mean good value if the upper and lower bands of the spread move against you rapidly as soon as you start trading.

In fact, a good liquid market has a combination of three dimensions, the ability to trade large volumes, at good prices, at high speeds. The way markets are structured allows high-frequency trading to prioritise the advantages of speed at the expense of price and volume.

Supporters of high speed trading show reduced spreads as the main benefit of their technologies, with the implicit assumption that this has clearly reduced costs for all market participants. But the reduced spreads have been accompanied by increased volumes of trading. It is the belief of the author that the increased speed of trading, and the faster reaction of markets to order flow mean that short term momentum effects have been increased, so obliging all traders to balance their portfolios more frequently.

It is trivially obvious that if spreads are halved, but traders are forced to trade three times more frequently, then overall trading costs have been increased by 50%. If the majority of gains are going to the algorithmic traders, then costs to normal traders have been increased even further. And here 'normal traders' ultimately means the general public as savers, and genuine capitalists raising money to invest in productive capacity.

One possible way to manage this is to change the trading rules so that they also reward providers of volume, longer quotes, and so good stable pricing.

The big advantage in offering larger volume quotes is clearly that more trading can be done faster, and at lower cost. The existence of over-the-counter 'upstairs' markets suggests that institutional investors often want to sell and buy large quantities at the same time, however the ad-hoc nature of upstairs markets can make such exchanges slow and expensive, indeed 'dark-pools' appear to be part of an ongoing process to formalise this upstairs market. Whether 'light-pools' form an extra step in this process remains to be seen.

The big disadvantage of trading large volumes is that it gives a large information signal and cause large movements if only one side of such a potential trade advertises their potential trade.

Similarly, if more bidders provided longer quotes this would give more quotes available, more price transparency and greater competition. Unfortunately, as discussed above, a long-life quote gives a 'free-option' to traders who can predict the direction the market is going to move. This therefore encourages short quotes, which in a circular reinforcement, encourages rapid movements.

It is possible that Credit Suisse, or other organisers of 'light pools' may be able to increase the effectiveness and liquidity of their trading platforms if they used rules along the lines of the following for filling orders against the limit order book.

1. All quotes to be quoted with both a size and a 'valid-to' time as well as a price. The quote would stand at least to the valid-to time. The valid-to time could be extended, or be rolling from the present time, but the quote could not be cancelled before the valid-to time, and a rolling quote would only be convertible into a valid-to quote of the same length.
2. Impose a minimum valid-to time of a few seconds.
3. Fill orders firstly according to price.
4. Where offers have the same price the offer with the furthest 'valid-to' time is selected first.
5. Where offers have the same prices and 'valid-to' time, the offer with the largest volume is selected first.

All incoming orders would follow the same rules, any that crossed the existing order book would be settled immediately, any that don't cross would be obliged to remain on the book until at least the end of the minimum 'valid-to' time.

This would be a 'no-time-wasters' market. It is possible that all quotes submitted would be for the minimum valid-to time, with small quotes competing on price only. However it is the belief of the author that such a market would encourage competition first on length of quote and then on volume.

The minimum time period would form an initial 'level playing field' and would discourage opportunistic bids. Given an existing price level, a new quote on the market that wanted to ensure a sale could simply quote a better price. Alternatively they could put in a quote at the same price but with a later 'valid-to' time. If the extension of time was relatively short, this second course would probably be cheaper than quoting a better price, especially if the market was stable. So at first the market should get a greater amount of quotes going further into the future.

With more bids on each side of the limit book, dealers that had large positions to move would then be able to compete on volume. If they did this alone in the current hft market it would be suicidal, but with more 'revealed liquidity' on each side of the book, the proportion of new information revealed would be smaller.

This process should allow more visibility and stability in pricing and so better price discovery. This could then feed back into more competition on quote duration and volume. Ultimately, if this system did work it would have more quotes, more volume and more revealed liquidity than other markets, and ultimately, smaller spreads.

The whole point of the proposed system above is to make traders behave more like fruit stall holders, or better, shop-keepers; to incentivise them to advertise their prices for longer periods and greater amounts of goods, so allowing better competition.

Counter-intuitively under such a system much greater liquidity, and better overall price value may also be achieved by limiting the intervals in prices at which shares can be traded and also by limiting the frequencies at which 'valid to' times can finish, say every 2 seconds. Infinite granuality would be reserved just for volume. This would be a reversal of recent history in the management of stock-exchanges. This would prevent price competition at very small fractional levels of price and time, and so encourage more competition on quote time length and volume.

A second area in which the current structure of markets seems sadly lacking is at the opening and closing of sessions. Currently this is commonly done by complex bidding procedures and crossing algorithms to dictate median prices. The suspicion that these procedures don't work; that the median prices are not in fact the market prices, is reinforced by the fact that the majority of trading in equity markets takes place in the first and last hour or so of the trading day. Figure 9.2.1 below gives the price (thick line), and volume (smaller grey shading towards bottom) for shares in HSBC, a large UK bank. Although the scale is a bit small, it can be seen from the volume that the majority of the trading takes place at the start and end of the trading sessions. This is typical of share trading patterns.

Figure 9.2.1 here

[FT.com Markets/data]

The problem here is that as the market opens, liquidity goes from zero to near infinite instantaneously. Conversely, at the close of the market, liquidity goes from infinite to zero instantaneously.

It is well known that increasing liquidity decreases spreads, so conversely, deliberately decreasing liquidity should increase spreads. This suggests an alternative to crossing procedures.

Opening a market could be managed by steadily increasing the liquidity over the first half hour. This could be done easily by opening the market with a very large minimum trade size, in the UK market this would be a minimum multiple of the normal market size 'NMS'. With this large minimum trade size, bids and offers would be a long way apart, and it is very unlikely that any trading would take place. Over the first half hour the minimum bid size would then slowly be moved from a large multiple of NMS to the normal minimum quote size. At some point during this process the bid and offer prices would come close enough for trading to start. This starting point would then be exactly the correct market price. A similar process could be used in reverse for closing markets.

Following the ideas above, it might be better to use the length of time that a quote is held open as the way of manipulating liquidity. At the opening of the market, minimum quote length would be in the order of minutes, and would then be steadily shortened. This would have the same effect of bringing the bid and ask prices together slowly, while having the advantage of not discriminating against small traders.

In fact, although this process would be very useful for restarting a stopped market, it wouldn't generally be necessary. Some commodities markets have already solved this problem. For example the oil futures market run by ICE has trading hours between 01:00am and 11:00pm (UK time). Again the figure below gives price (thick line), and volume (smaller grey shading towards bottom).

Figure 9.2.2 here
[FT.com Markets/data]

Although this might raise fears of traders being forced to work anti-social hours, actually the reverse is true. Trading through the night is low, and then trading and liquidity both rise to a morning peak, followed by a larger afternoon peak before dropping off again. Clearly this has settled to a standard pattern where people who have large trades wait for the liquidity peak to build before they move in to trade.

It would certainly be feasible to do the same for the major stock-exchanges, if only for the larger shares such as those in the FTSE100 index.

All the above are the suggestions of an amateur game theorist. Within economics in recent years there has been an explosion of literature on game theory and auction theory, but this seems to have had little practical input to the trading of financial assets in general and market microstructure in particular. The systematic application of game theory to continuous double auction markets would appear to be a very productive potential future field.

9.3 Supply - On the Scarcity of Scarcity, or the Production of Machines by Means of Machines

"Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses." [Robbins 1932, p. 15].

I write this section with some trepidation, given the beliefs, held by a significant number of intelligent people, that the world is simultaneously on the edge of a dramatic ecological crisis and about to run out of many critical resources, most notably oil.

I had considered a third alternative title of 'The Confessions of a Cornucopian'. Because, after twenty years working daily in the environmental industry, and having read widely on engineering, technology and economics, I am philosophically a strongly committed cornucopian.

I personally agree with the binary economists and Amartya Sen that there are more than enough commodities in the world for everybody; just that most people do not have the money to buy them.

From the point of view of the analysis of economics, it is most unfortunate that heterodox economics appears to have been significantly infected by this Malthusian virus of scarcity. As can be seen from the discussions above, the environmental and ecological scientists have the right mathematical tools for creating a radical and effective new economics, both specifically in the form of the Lotka-Volterra and maximum entropy production models, and more generally in their understanding of complex inter-related evolving systems.

Ultimately this maths finds it's roots in the work of Malthus and Sismondi, unfortunately the environmental and energy economist movements seem to have also inherited Malthus' pessimism lock, stock and barrel.

A significant source of the problem appears to be the work of Georgescu-Roegen [Georgescu-Roegen 1971], the first person to successfully introduce the concept of entropy into economics. I believe that Georgescu-Roegen's work was of ground-breaking importance, and profoundly insightful. However some of his conclusions, though meaningful at the time of writing, have proved to be wrong in hindsight. This is not his fault. In the middle of the 20th century there was no sign of declining human fertility, green revolutions or cheap photovoltaic cells.

Georgescu-Roegen's work can be seen as similar to that of Lord Kelvin. Kelvin's scientific genius is not in question, but few people today believe that the earth is in imminent danger of 'heat death'.

In almost all recent work in environmental and energy economics the supposed 'restraints' imposed by entropy, first proposed by Georgescu-Roegen have been treated as fundamental truths. Unfortunately these precepts are trivially mistaken.

The paper of Ayres & Nair [Ayres & Nair 1984] provides a typical example. I have found this paper profoundly useful in guiding my ideas as to how to link the mathematics of economics to the real world of science. The present paper would never have been written without the assistance of Ayres & Nair. But the atmosphere of doom and gloom runs deeply through the paper, from beginning to end, with the clear prediction, on the last page, made in 1984: *"What are the prospects of avoiding a resource-depletion catastrophe? It will not be avoided without a major effort, we believe."* And much more in the same vein.

Despite these predictions, the economies of the West have plodded on at their long term 2-4% annual growth rates, the developing world has managed at least double this, and China has lifted a billion people out of poverty in the greatest single advancement of welfare that humanity has ever seen.

Oddly, although neoclassical economists are generally very optimistic with regard to the market providing for peoples needs, and are often philosophically opposed to the environmental movement; neo-classical economics shares with the greens a bizarre fixation with the concept of scarcity. As typical examples the second page of my edition of Mankiw [Mankiw 2004] states that *'Economics is the science of how society manages its scarce resources…….'*, while on page 4 of 'Macroeconomics', Miles & Scott give the definition of the whole of economics as *'Economics is the study of the allocation of scarce resources'.*

Robbins is generally credited as being the originator of this meme; he is quoted above at the beginning of this section. Prior to Robbins the study of economics was generally defined as the study of the distribution of wealth, as for example by Ricardo at the very beginning of this paper.

The conversion to a definition of economics based on scarcity represented the absolute victory of marginalism over common sense. The definition using scarcity seeks to define the whole of a scientific field in terms of one cheap mathematical trick. It is as if; exactly as if, 100 years ago the field of physics had been defined as the study of conservative fields.

In the following sections I would like to briefly discuss these apparent constraints of scarcity.

<u>Population</u>

The world's population is rising, and because of the relative youth of most people in developing countries, it will continue to rise for some time.

However the decline in fertility in recent years has been dramatic. China dropped below replacement rate years ago, along with most of the rest of East Asia. India's fertility rate is dropping dramatically and will soon be below replacement rate. High fertility is now confined to a small number of countries in Africa and the Middle East, and is dropping quickly in most of these places.

The pattern of fertility drops is strong, and is clearly linked to women's education levels and general economic wealth, both of which continue to improve at a rapid rate in all countries except for the few that are at war or are failed states.

Many examples can be seen at the excellent site 'gapminder' [gapminder].

The current median predictions, from the UN in 2008, for the future world population are nine billion people in 2050, which is expected to be close to the peak [UN 2008].

For those who are horrified by these numbers, a little context might be useful.

If there were 10 billion people alive in 2050 and every single one of them lived in the USA, the population density of the USA would be 712 people/km^2, that lies somewhere between the present day population densities of the Island of Jersey at 789 people/km2 and the Palestinian Territories at 667 people/km^2 [UN 2004].

Oddly enough, the island of Jersey has a long and successful history of selling itself as a bucolic rural tourist destination and a quiet millionaires playground. Meanwhile the Israelis appear to have signally failed to realise that the Palestinian Territories are overpopulated, and have installed an extra half a million people there as illegal settlers in the last forty years.

Energy

Very roughly, world energy production can be split as follows, one third oil, one quarter coal, one quarter natural gas, while the rest is made up of nuclear, biomass, hydro and other renewables.

Nearly all of the oil is used for transport. Most of the coal, gas and other is used for producing electricity for industrial and domestic use; other than a minority of the natural gas which is used directly for heating homes in the Northern hemisphere.

All of this usage can be replaced easily and rapidly from other sources, at only moderate extra cost.

The amount of solar radiation received on the earth is many orders higher than the amount of energy used by human beings, roughly one hour's sunlight hitting the earth would supply humanity's energy needs for a year. The black dots on the map below show how little area would be needed to supply all the world's energy needs.

Figure 9.3.1 here

[Loster 2006].

For electrical generation, and so also space heating, solar power can be used directly, or with storage for use at night. Current storage options include hydroelectric, already used in Northern Europe with storage in Norway, hot oil in CSP plants, grid scale sodium/sulphur batteries, or just old fashioned domestic electrical storage heaters with bricks in them.

The cost of photovoltaic solar energy has already reached grid parity in Italy, where the sun is plentiful and electricity is expensive [NEA 2010].

The recent experience of both Germany and Spain have shown that when there is an economic incentive, solar power can be rapidly installed in industrial quantities.

As solar has reached grid parity in Italy, and will shortly do so in Spain, Australia and the South West of the USA, installation will proceed rapidly unless it is forestalled by something cheaper such as shale gas or an innovative form of nuclear power.

Given both that solar has already reached grid parity, and will inevitably continue to get cheaper, and coal power will inevitably get more expensive; this means that CO_2 emissions will also peak in the very near future and will decline rapidly thereafter.

The recent revolution in shale or 'tight' gas, illustrates the pitfalls of the Malthusian tradition of underestimating the combined powers of economic incentives and human ingenuity. As recently as five years ago it was believed that natural gas supplies in the USA had peaked, and so very substantial money was invested in natural gas import facilities in the USA. With the new techniques for extracting shale gas, the estimated gas reserves of the US increased by a third from 1998 to 2008 [EIA], and have increased substantially in the last two years as more unconventional gas reserves have been made accessible.

Exploration for tight gas supplies in the rest of the world has barely started, and there is good reason to believe that reserves similar to those found in the US will be found in similar rock formations worldwide.

The one third of the world's energy supply, mostly oil that is used for transport cannot be replaced directly by solar, but many other alternatives exist.

With the boom in shale gas; compressed natural gas for transport use is the trivial short-term solution. In countries such as Argentina and Turkey, natural gas is already widely used for transport, and if the shale gas revolution continues at its current speed, this short-term change over appears inevitable worldwide.

Ethanol, specifically cellulosic ethanol, is another alternative. Brazil already supplies more than half its needs for car fuel from ethanol. Brazil alone could supply enough ethanol to replace all current world oil needs, using just a quarter of its land, supplying cellulosic ethanol from sugar cane [Biopact 2007].

In the meantime, battery technologies are improving rapidly, and cars are slowly becoming more and more hybridised. This will increasingly allow grid electricity (from solar if needs be) to supply power for short distance commuting transport – which of course forms the majority of usage.

Already half the two-wheeled motorised vehicles produced in China are electrically powered. In China electric bicycles are already outnumbering petrol-powered motorcycles. Around the world hybrids are replacing straight diesel engines for buses, refuse lorries and parcel delivery vans. This is being driven by economics, in any application that involves rapid stop-start cycles hybrids are already competitive with traditional drive chains.

The other longer-term reason for expecting oil demand to drop is the installation of personal transport systems in urban areas. This is no longer a high tech dream, but a reality with the Ultra system working at London Heathrow [Ultra].

For techie people it is perhaps worth briefly looking at EROEI; the energy returned on energy invested. The EROEI of photovoltaic solar has come close to that of natural gas, which is why it has reached grid-parity in Italy. The EROEI of solar is continuing to drop at a steady rate. The only reason solar might not expand dramatically is because shale gas has dropped the EROEI of natural gas dramatically overnight. Similarly the EROEI of sugar cane ethanol is already equal to that of gasoline, and is steadily decreasing. It is only tariff barriers in the US and EU that are preventing its widespread adoption. Given excess land for both solar and growing sugar producing crops, analysis of EROEI shows that there is no energy crisis. Indeed EROEI is a poor measure of efficiency. If engineers used 'free-energy returned on free-energy invested' or 'negentropy returned on negentropy returned' then their measures would be much better from a scientific and engineering point of view. However if they did this it would of course render EROEI pointless, as it is in fact just an engineers ham-fisted way of calculating market values.

Food

In recent years, changing appetites in Asia, combined with oil price movements pushing up the price of natural gas, and so fertiliser prices, have created spikes in food prices that have panicked people into believing the world faces food shortages.

As discussed above, the world's population is surprisingly small compared to the amount of land available. Until very recently both China and India were self-sufficient in food and fed large populations (nearly half the world) on comparatively small land areas. Compared to China and India; Russia, Ukraine, the USA, Canada, Brazil and most of Africa are empty. All of these places have substantial potential for extra food production.

The table below is FAO data, pulled from a good background article by the Economist [Economist 2010b].

Figure 9.3.2 here

[Economist 2010b]

To put things in perspective it is worth looking back at the European common market between 1986-1989. This is in the period after Spain and Portugal joined, but before East Germany joined.

In this period, the EEC had a population density of 150 people/km^2. This population fed itself, and lived on a high protein, high meat, high dairy, highly unhealthy, westernised diet. Not only did Europe produce enough food to feed its population in this way, it also; under the Common Agricultural Policy, built up large mountains of surplus food and subsidised food exports to other countries, so destroying agricultural economies across the third world.

In comparison, the future population density of the whole world assuming 10 billion population, and excluding the land in Antarctica, would be 74 people/km^2, or less than half of the EEC during 1986-1989.

But clearly not all the world's land is fertile. So for the sake of argument, we can also exclude Australia, which is almost all desert, and Russia, which is mostly tundra. The European part of Russia can approximate for the Himalayas and central Asian deserts. Similarly we can exclude Canada, largely tundra, and also to compensate for the Rockies and deserts of the USA and Mexico. Brazil can be excluded for its rainforest and to compensate for the Andes. Finally we can exclude half of Africa to compensate for the Sahara, Kalahari and Namib deserts. (Agronomists can be forgiven a wry smile, as we have now excluded most of the world's major bread baskets.)

If you exclude all this land area, and also assume a world population of 10 billion, you get a world population density of 130 people/km^2, still less than the EEC at the height of its butter mountain madness.

In the meantime the rapid expansion in the supply of natural gas from shale means that natural gas prices have de-linked from oil prices and so we are guaranteed cheap fertiliser for the indefinite future.

Other Natural Resources

As discussed above, energy and agricultural land are available in excess.

For the construction of engineering equipment, buildings, cars, etc; steel, aluminium, and silica are also all available in excess. As the oil price has increased, so plastics are already being manufactured from sugar cane.

With a single possible exception, all other raw materials are fungible, if something becomes scarce, its use can easily be replaced with something else.

The only near convincing case for an essential resource being consumed, than cannot be replaced, is that of phosphorous, an essential part of fertiliser for agriculture. Oddly enough, the last time I designed a wastewater phosphorous removal plant, the waste iron phosphate was sent to land fill for dumping. Phosphorous is still so cheap and plentiful that it has no commercial value for recycling. In these circumstances, worrying about it running out seems a little premature.

Capital

As discussed above, with the possible exception of phosphorous, there are no meaningful restraints on the supply of raw materials, energy and food for provision of goods and services to human beings. Supply of raw materials is unlimited in meaningful terms for all reasonable human needs.

With regard to capital, with raw materials available in excess, the only natural limit is human ingenuity, at least in the short term, and there does not seem to be any great constraint on human ingenuity at the present time.

In manufacturing industry, the current level of automation can be extraordinary. In state of the art factories human beings are largely confined to a supervisory and intermittent maintenance role. In western countries, to a great extent, the production of machines is already carried out by machines, and increasingly this is moving beyond the scope of the factory floor.

A first recent examples of how machines are able to provide additional humanly useful value is the rise of fruit picking machines. These are fully automatic, capable of travelling over rough ground, identifying fruit, checking their ripeness and removing them from trees or vines without damaging them. The complexity involved in these processes is enormous, which is why this task has remained a labour intensive process, at least until now [Economist 2009]. A second example is an automated hospital, where a fleet of robots will automatically perform duties such as removing clinical waste, delivering food, cleaning operating theatres and dispensing drugs [BBC 2010b].

These two examples, along with personal rapid transit systems are interesting in that fruit picking, cleaning and taxi driving remain three of the last significant redoubts remaining for the employment of unskilled labour. The IT revolution has already taken over swathes of semi-skilled labour; the clerks that used to dominate offices have largely disappeared, sacrificed to spreadsheets. And IT is slowly but surely eating into skilled and managerial work.

Casual observation confirms that supplying new capital is trivially easy. Whether it is the supply of new manufacturing capacity in Japan in the nineties, housing in the USA or Spain in the noughties, or more recently solar panels in Spain and Germany, history has shown that provision of new capital in large quantities has never been a problem.

The problem has always been how to ensure good use is made of the capital once it has been provided.

The work of Sraffa, von Neumann, Kurz & Salvadori, Burgstaller and others leaves a major quandary; an inverted Malthusian quandary. Whether you use the mathematical approaches of these academics, or the commonsense engineering approach in this section, it is clear that there simply are no external constraints on the wealth that individual human beings can own.

In an exact reversal of Malthus's fears, we are left with an (intellectual) problem, the population of the world is stabilising rapidly, the production of commodities by commodities should be an exponential growth.

The problem is the one economic input that is truly scarce: labour.

The reasons for the apparent scarcity are simple. The scarcity is a consequence of the financial system and Bowley's law. Capital, and so wealth can not be accumulated because the amount of capital that can be built up is dependent on the amount of labour available.

If there is no shortage of supply, what of demand?

9.4 Demand

On the supply side, things are relatively straightforward. Costs are defined by 'negentropy', or intrinsic value, and, bubbles aside, prices adjust to these costs in the long term. With a few minor exceptions, and the two major exceptions of labour and land in cities, marginality is of little relevance.

On the demand side things are a bit more complex, and a lot more fraught.

I would like to start this particular section of discussion by stating categorically that, unlike many physicists, I firmly believe in the concept of utility.

That ownership of a second car, for example, gives less benefit than the ownership of a first car seems obvious and plausible, and indeed important.

I have worked for many years in the water industry, where measurement of utility (or rather disutility) has become important.

Water is expensive to transfer over long distances, so water companies are natural monopolies, this, along with the non-substitutability water, and the potential for trace contamination of supply, means that water companies are normally subject to strict regulatory control.

In practice the base cost of treating water can be very low, but the expense can vary enormously according to required service levels. These service requirements include things like interruptions to supply, (harmless) discolouration or odour, pollution of watercourses, etc. All of these are externalities and/or low-probability high-risk events that are not easily priced by means of the market. Eliminating all possible service failures would be enormously expensive, and there is no easy way of using supply and demand to fix the levels customers believe appropriate for rare events.

To get round these problems, it has become commonplace in the water industry to use surveys of customers which use pairwise comparison exercises. These allow customers to choose which of two outcomes are worse or better, repeatedly over large numbers of different outcomes. By using some comparators that are directly costable, it is possible to measure and build up disutility curves for customers.

The results are interesting; the curves can be highly non-linear. For example nearly all customers are highly tolerant of a short break in water supply, say up to eight hours, seeing this as a much lesser problem than, for example, a spillage that kills hundreds of fish.

However most customers are highly intolerant of a water supply outage of more than say twelve hours. Under these circumstances sympathy for aquatic life evaporates.

Another notable feature of these surveys is that, despite outliers, the disutility curves are very similar for most people.

Whether you follow the psychological, hierarchy of needs theories of Maslow [Maslow 1954], or the marketing methods of multinationals, this is hardly a startling observation.

People are very similar in their basic needs and desires, which for economic goods is a simple hierarchy of needs to define.

A basic list starts with food, ranking through needs for shelter, transport, leisure/entertainment, health care, education and retirement security.

Utility curves do change according to sex, parenthood and age, but the basic requirements of food on the table and a roof over the head are fundamental.

That the list changes dramatically with wealth is well known. In developing countries people start buying bicycles en masse at one wealth level, motorbikes at another higher level and cars at another level higher than that.

These markets are predictable, opening up at threshold levels of average income. That is why Coca Cola, a very cheap 'luxury', is marketed in almost every country in the world, but Ferrari, at the time of writing, appears to have only two dealerships in the whole of sub-Saharan Africa, unsurprisingly, in Cape Town and Johannesburg.

Rich people may appear to have very different utilities to poor people, but that is only because they are rich, not because they are different. And, as has been noted above, it is the GLV distribution that defines the split of people into rich and poor, into 'workers' and 'capitalists'.

So utilitarian desires are fixed by wealth, which, as we have seen, are fixed by entropy. Attempting to use utility as a foundation for the whole of economics is to put the cart firmly before the horse.

A further problem with utility theory is it's absolute approach to relative value. As stated above, I can see the common sense in the fact that for a single person, ownership of a second car clearly has less utility than ownership of a first car. But this is only true if I am obliged to retain ownership of both cars.

If I am allowed to sell either of the cars, then the 'utility' of the second car is simply it's market value less the transaction costs (my time and trouble of selling it).

Following the discussion of the sections 8.1 on value above, the value of other cars on the market will be the fundamental value of the long term costs of producing them, so the 'utility' of my second car will simply be its fundamental value (less transaction costs).

Personally, I am not much of a fan of Picasso; Picasso paintings have negligible 'utility' for me. But if somebody offered to sell me one for a thousand pounds (and I was sure of provenance and ownership) I would certainly buy it.

And quickly resell it. Living in the UK I prefer to receive wealth in the form of pounds and pence, but will happily accept alternate stores of value if I stand to gain from the deal. Although the utility of a Picasso painting to me is low, its value is actually set in the market by entropic measures discussed in section 8.1 above.

In fact human beings' desires for economic goods fit into two categories. With one very big exception almost all goods are satiable. Realistically most people only want one bicycle, one car, one house, one set of furniture to put in the house etc.

Even large houses are a disadvantage in countries that don't have a supply of cheap labour to clean the rooms and maintain the garden. In the UK it is striking that the majority of owners of large houses; 'stately homes' have been obliged to let the public visit them and assist with payment of the upkeep. In high-income countries second homes are largely confined to people who are required to live in cities for work reasons.

This holds true even for smaller low value items, as Steve Keen has pointed out;

"Two bananas per day may well be preferred to one; three per day is probably pushing the envelope for most humans; and you would have to be a monkey to, for example, prefer twenty bananas in a day to nineteen. Most humans would kill rather than consent to eating a twentieth banana in a day. Thus, when we consider consumption as a function of time, anyone who behaves in a fashion which economists call rational–always preferring more bananas per unit of time to less–is clearly insane" [Keen 2004].

While some people own collections of things, these fall into two main categories. Either they are the low cost collections of hobbyists, and so count as leisure activity. Or, they are investments.

Investments are of course the exception to the rule of satiability. Unlike things that give actual utility, human beings seem to have a near insatiable desire to accumulate stores of wealth; 'potential utility' or better 'potential negentropy', whether this is in the form of property, shares, artworks, prestige cars or just 'money in the bank'.

Taken together, this suggests that there are straightforward ways of using concepts from physics to model utility and the resulting distributions of goods between individual human beings.

Statistical physics has standard methods for dealing with localised fields en masse. It gives each agent a 'potential energy well' which can be filled in a quantum mechanical manner. Such potential wells normally have defined levels at which the levels can be filled. So for molecules of gas, translation energy levels can be filled at a certain temperature, rotational energy levels can be filled at another higher temperature and vibrational levels can be filled at a third level. In this way the energy 'needs' of the molecule can be filled at different temperatures. Human beings could be modelled in a similar way, with bicycle needs filled at a certain level of GDP, motorcycle needs filled at another level of GDP, and motor car needs filled at a third level of GDP.

Statistical mechanics has well-understood and consistent mathematics for dealing with such problems, and the probability of the levels being filled is defined by different statistical rules according to the type of 'good' that is filling the potential well.

So, for example, for modelling investment wealth ('money in the bank') classical statistical mechanics would be appropriate. For most other goods that are wanted in limited quantities, the correct statistical mechanical approach will be some variation or modification of Bose-Einstein or Fermi-Dirac statistics.

Such an approach could be very powerful, instead of using a single representative agent, as current macro-economic models do, it would be possible to use a large number of identical representative agents and calculate the macroeconomic parameters from the statistical mechanical properties.

In economics, some interesting work along these lines has already been carried out by Foley [Foley 1996a, 1996b, 1999, 2002].

Foley's work is also important as it gets to the heart of the problems of using utility and marginality alone in a multi body system. Even where there are genuinely scarce resources such as labour or special minerals, in all but the simplest cases, the market will never clear at the marginal price. Statistical mechanics will force it's own equilibrium. However, as Foley has demonstrated, the statistical mechanical approach is more powerful, with functions that behave sensibly and can give meaningful equilibria.

Part B.III – The Logic of Science

10. The Social Architecture of Capitalism

All the above however does still leave the central question of supply and demand unanswered.

Given the reasonable assumption that basic human desires are fundamentally the same, it is clear that in most parts of the world, even the most basic needs for food, clothing, water and shelter are not fulfilled.

In most of the rich world demand for good health, education, housing and most importantly secure and decent retirement income is not satisfied for the majority of people.

If supply is unlimited in the basic physical sense, and demand is far from being satiated even in the rich world, then a basic question needs to be answered.

What exactly is it that controls the balance of supply and demand, or more importantly, why are the basic needs of human beings, for decent housing, education, health, pensions and leisure not provided by the capital available, or the capital that could very easily be made available?

The reason for this substantial market failure is the structure of economics and finance, or to again borrow Ian Wright's phrase, it is a consequence of the 'Social Architecture of Capitalism'.

In Wright's paper of this name he put together the first ever, coherent, effective, meaningful model of an economic system based on capital, a model that can be applied to feudal land-owning, Victorian owner managers, or with minor modifications modern disintermediated capitalism.

Wright's models are much less 'knowing' than my own, with no financial sector, and no preordained mathematical basis. With this simplistic approach Wright shows just how powerful a statistical mechanical approach is. The behaviour of a normal economy 'emerges' naturally out of the model without any assistance from the model builder.

Although the make up of Wright's models is very different to that of my own, the interesting point regarding the comparison of Wright's models with my own is not the differences but the similarities. And this is entirely a consequence of statistical mechanics.

Both Wright and myself make some very basic and similar assumptions, these are briefly:

- Economies are multibody, chaotic, stochastic, statistical mechanical systems
- Wealth is produced in companies
- Wealth is conserved in exchange
- Wealth is destroyed in consumption
- Returns on capital are proportional to wealth owned

Once these assumptions are made, and these assumptions are trivially obvious assumptions to anybody who has a passing knowledge of statistical mechanics, and has worked for a living in a factory, then it doesn't really matter how dodgy your models are. The 'Social Architecture of Capitalism' drops out in short order; replete with gross inequalities of wealth, bubbles, crashes, inflation, recessions and persistent unemployment.

And as discussed in length in section 4 above, you simply don't need utility, consumption or production functions and the rest of the marginalist paraphernalia to explain all this. Neither Wright's nor my models even need economic growth.

The maths, and indeed the gut feel of statistical mechanics, can be initially quite daunting. However what the models of Wright and myself demonstrate is that this approach makes life much easier. Many millions of complex local interactions get washed out in the sweep of entropy.

This modelling approach is very powerful, and offers an effective way of building comprehensive economic models along the lines below.

The big problem with modelling any multi-body, thermodynamic system; which includes economic systems, is the large number of parameters. Care is needed in identifying and reducing the active variables initially so that the most general and basic model can be built first, before then being expanded.

The role of economic growth is a good example. Almost all macroeconomic models include economic growth as a variable. Yet casual observation of the depression era, or the last two decades in Japan, demonstrate trivially that capitalism can survive indefinitely with its structures operating intact in a long-term zero growth environment.

Growth is clearly not a primary variable, and should not be included in base models. Its inclusion at this level merely causes confusion with too many variables.

The first part of an approach is to deliberately limit the modelling to the stable macro-economic zone of the economy from the point of view of a high level Lotka-Volterra systems. That is to say, for this level of modelling, the macroeconomic Minskian/Austrian cycles are deliberately ignored. The maths is deliberately constrained so that the economy is deliberately 'damped' to a stable dynamic equilibrium.

At this point the economy will be in a state of maximum entropy production (MEP), and so it is in a 'stable dynamic thermodynamic equilibrium', at this point microeconomic conditions such as income distribution, company size, unemployment and debt can be investigated in detail. The effect of traditional underlying economic relations can be investigated. Also the economy can be moved through different loci of stable points, such as those that are defined by the Bowley equation (4.6g). The models would be expanded slowly to include other factors such as housing and other compulsory spending, as well as corporate and household saving and debt. Eventually such models would include government taxation and spending, currency, imports, exports and exchange rates. However they would still be held at stable dynamic equilibria. At this stage growth would be ignored.

Under the above circumstances, total consumption is equal to total production, and both are unchanging.

The total demand is set by the balance between labour and capital at the maximised MEP, given by the appropriate version of Bowley's law.

Total wealth and income is dictated by the amount of productive capital installed, which will depend on the equilibria above.

Wealth and consumption for each individual is set by their place in the GLV, their earning ability, their compulsory spending on housing and other goods, the consumption and saving preferences etc. This then generates different classes of owners and workers in society and a final equilibrium solution. This creates a total quantifiable aggregate level of demand. This equilibrium also defines the prices of the different types of capital; primarily companies and housing.

In such a system supply balances to equal the above demand, where supply costs of commodities are calculated by aggregating costs of inputs on a cost plus basis, as per Sraffa, (un-reconstituted) von Neumann, market-microstructure and post-Keynesian analyses, but done on a dynamic equilibrium basis.

At this level, price distortions due to capital hoarding, and the delay of installation of capital would be prevented.

In this system supply and demand are both constrained by maximum entropy production and Bowley's law, so issues such as increasing returns on capital are not problematic.

This first level of modelling allows many underlying interactions to be quantified and analysed in detail, and correlated to real world data. This first approach would primarily define microeconomic interactions, though it would also give insight into macroeconomic models.

Once such models have been built satisfactorily, then the models can be relaxed and the damping progressively removed. In a normal economy, inherent instabilities exist due primarily to the basic process by which capital is priced. This creates endogenous cyclical behaviour within economies, so you then move to the states that may be characterised as 'quasi-periodic quasi-stable dynamic thermodynamic equilibrium'. Under these conditions Minskian and Austrian theory; variations in capital, debt and liquidity, along with relevant behavioural economics can be analysed.

There are many secondary sources of instability that can destabilise economic systems, which may either exacerbate the fundamental instability of capital pricing, or create their own cyclical patterns.

Such destabilisers include capital cycles in commodities, housing and commercial property as well as corporate behaviour such as 'capital hoarding', both of which have been modelled in a simplistic manner in part A. It would also be appropriate to look at the effects of savings and investments at this level.

Other destabilisers include household, corporate and government debt, fractional reserve banking and feedback from monetary authorities.

Investigation at this level, allowing models to evolve dynamically around their points of stability, would allow detailed analysis of changing macroeconomic variables.

Finally, when such models are well understood, longer-term trends can be modelled; trends such as population growth, economic growth, technology change, productivity growth, cultural change, institutional change, etc. This leads into fields such as evolutionary economics, institutional economics and growth theory.

11. The Logic of Science

In their abstract to 'Worrying trends in econophysics' Mauro Gallegati, Steve Keen, Thomas Lux and Paul Ormerod wrote:

'Our concerns are fourfold. First, a lack of awareness of work that has been done within economics itself. Second, resistance to more rigorous and robust statistical methodology. Third, the belief that universal empirical regularities can be found in many areas of economic activity. Fourth, the theoretical models which are being used to explain empirical

phenomena. The latter point is of particular concern. Essentially, the models are based upon models of statistical physics in which energy is conserved in exchange processes. There are examples in economics where the principle of conservation may be a reasonable approximation to reality, such as primitive hunter–gatherer societies. But in the industrialised capitalist economies, income is most definitely not conserved. The process of production and not exchange is responsible for this. Models which focus purely on exchange and not on production cannot by definition offer a realistic description of the generation of income in the capitalist, industrialised economies.' [Gallegati et al 2006]

I am slightly embarrassed to admit that, due to both time constraints and limited experience in econometrics, the present paper remains significantly remiss with regard to the second criticism above.

But, then again, to rephrase Ernest Rutherford; if you need to use statistics to prove your theory, you ought to have thought of a better theory.

In the event of some party choosing to award me remuneration for my ongoing research I would hope to remedy these shortcomings in future papers.

However I believe the present paper has come a long way in answering the other criticisms.

In particular, I believe criticisms one and four have been fully addressed in this paper.

I believe however that the authors' third criticism is fundamentally flawed.

It is the nature of science that a field can appear complex and difficult to make any sense of until a significant insight can bring sudden clarity. It has taken time for physicists to bring this clarity to economics, but to physicists, the multi-body nature of economic and financial systems meant that the belief that universal empirical regularities would be explained was only a matter of time. It is this insight that drove Champernowne half a century ago. It is this insight that resulted in Wright, Souma & Nirei and myself independently producing similar models near simultaneously.

It is a canard among economists that physicists have moved into economics and finance due to the lack of job opportunities in mainstream physics. This may be the case for quants in the city, but it is not the case for econophysics researchers. For the research oriented physicist the attraction is a mathematical field that has not been effectively analysed, but that clearly has parallels with other fields that have been regularised. Finding wide open research areas like this in the mainstream sciences or engineering is difficult. Economics offers the low-hanging fruit of major new research findings, that is if you can truly describe a field full of watermelons as 'low-hanging'.

Indeed the '*universal empirical regularities'* pooh-poohed by Gallegati et al where always there.

Wealth distributions, company size distributions, and the split of the returns from labour and capital are all long-standing 'anomalies' within economics. Economists such as Schumpeter and Gabaix have noted these regularities [Gabaix 2009]. Why almost all other economists, even heterodox economists such as Gallegati et al, have shown such disinterest in investigating these recurrent and profound features of economics has always been puzzling to physicists.

Economics has systematically treated such persistent 'anomalies' as anomalies, ignoring raw data while retreating into the comforts of intellectual hypothesis, whether this be neoclassical, Keynesian, Marxian, behavioural or other. Even in areas such as post-Keynesian or behavioural economics, were data collection has become something of a fetish, the flavour of the data collection still gives the feel of data being ransacked to prove the previously held opinions of the researcher, rather than the data being looked at and analysed as it is found.

This behaviour is the behaviour that has kept economics as a branch, to be generous, of political philosophy. It is this behaviour, understood intuitively by the general public, and explicitly by natural scientists, that is responsible for the very low regard that both have for economics as a science. To be brutally honest, given the fixed holding of ideologically motivated positions against the evidence of recurrent arbitrary destruction of wealth in bubbles, widespread poverty and persistent unemployment, economics as currently practiced should be considered a branch of religious philosophy, fitting somewhere between fundamentalism and cargo cults.

At least all the mainstream religions include compassion and charity as compulsory elements. The main difference between economics and religion being that, in the majority of countries, members of the public are allowed to voluntarily remove themselves from the experiments of zealots.

It is precisely by investigating 'anomalous' but persistent data outputs that the natural sciences have progressed.

By definition, if data output is persistent, it is not 'anomalous'.

If data output is persistent, it is normal.

It may be 'anomalous' with regard to current theory. But that simply makes the current theory by definition 'anomalous', not the data.

In these circumstances the theory must be abandoned, not the data.

Einstein, for example, is usually characterised as a theoretical physicist. But his biggest single insight (amongst many) was to treat the experimental fact of the constancy of the speed of light as a given. From this he abandoned 'common sense' and simply worked out the mathematical consequences of this fact. Thus was relativity born.

Economists seem to prefer the route of Einstein's peers, forever producing more complex theories to substantiate the existence of a hypothetical aether.

Science can not be built simply on common sense, intuition and intellectual rigour.

Science must start with the observed facts if it is to make progress.

This, at a much deeper level than that intended by Jaynes, is the logic of science.

For any multi-body system, entropy has to be the guiding force, it has taken time for physicists and mathematicians to get to the root of this, mainly because the entropy was dynamic path entropy rather than static state entropy, but the driving power of entropy in economics is immediately obvious to anybody who has a passing understanding of entropy.

Economics is a specialised study of entropy. It is a branch of thermodynamics, a branch of physics.

Like information theory, in fact even more so than information theory; economics is a very complex, interesting and important subject in its own right. But nonetheless it is a subset of thermodynamics.

It is the application of dynamics and statistical mechanics to political economy.

It is econodynamics.

11.1 Afterword

It was noted in the introduction that this paper was researched and written in a little over a year, without financial support or academic supervision.

Foolishly, I have gone against a basic conclusion of this paper, and spent a significant portion of my own capital in producing it.

If you have found the paper of interest or value, any donation to defray the costs of writing it, no matter how small, would be gratefully received.

Those who wish to make a donation can do so by clicking on the Paypal link below:

click here to make donation

(Paypal accept all major credit cards, you do not need to have a Paypal account.)

Part C - Appendices

12. History and Acknowledgements

Between 1980 and 1982 I was taught A-level physics by Malcolm Ruckledge using the innovative Nuffield Foundation Physics course. This was a powerful combination of an outstanding teacher with outstanding material. The section on statistical mechanics was particularly well written and taught, and gave me an early and profound intuitive insight into the power and simplicity of entropy. I suspect this paper would not have been written without this insight.

Sometime in my first year studying physics at the University of Manchester, in 1992/3, while looking at a picture of the Maxwell-Boltzmann distribution of molecular velocities on a blackboard, it occurred to me that wealth in a society was shared out in a similar manner; a lot of people with a little wealth and a few with a lot of wealth. It further occurred to me that the underlying systems, involving a lot of freely interacting particles/individuals, where fundamentally similar. At the time I imagined this was a unique and very clever insight, however it turned out that a lot of other physicists and mathematicians have had similar insights, some preceding mine by many decades.

After this, nothing very much happened for a decade or so, though the idea refused to go away, and being by nature an engineer at heart, I thought a lot about how income and wealth inequality might be tackled as well as to why it exists.

In 2003 I had a letter published in the New Scientist, this is reproduced at the end of this section. This encouraged me to take my ideas more seriously, and while working abroad in 1995 I had the opportunity to write down my ideas at that stage into a fairly amateurish paper.

On returning to the UK I circulated the paper around various individuals I thought might be interested. The paper was greeted on a spectrum that largely went from disinterest through to derision.

One exception was Michael Stutzer, who suggested I forward it to Duncan Foley, with whom I had a brief but very rewarding correspondence. I remain very thankful to both these individuals and especially to Duncan Foley for encouraging my work even when it was at this very early and amateurish stage.

After this nothing very much happened again for some years, as I lacked the skills, in both economics and mathematics to take the work forward. I did however read a paper by Ayres & Nair 'Thermodynamics and economics' which I found very useful in linking the concept of entropy to the economic concept of value.

This changed in August 2000 when, via the New Scientist, I discovered the work of Bouchaud & Mézard and other researchers, primarily physicists but also some heterodox economists, working

in the new field of econophysics. The majority of the work was in the field of asset pricing in finance, but there was also a parallel stream looking at wealth and income distributions.

Over the next few years I attended a number of econophysics and related conferences where I learned a lot more about both the maths and economics from the other participants.

During this period I was given support and guidance, from Steve Keen, Thomas Lux and others, but most particularly from Juergen Mimkes, for which I would like to give thanks. Thomas Lux gave me some very useful insight into the real meaning of value and wealth that helped to generate the ideas in this paper. Steve Keen gave interesting discussions on economics and also pointed me in the direction of James Galbraith who was also very supportive.

As stated in the introduction I met Wataru Souma at the Econophysics of Wealth Distributions conference at Kolkata in 2005. I almost certainly attended his lecture on his paper 'Universal Structure Of The Personal Income Distribution'. I found Souma & Nirei's model complex and difficult to follow, and did not knowingly use it further.

Judging from the pile of papers that I rediscovered it in; it appears that I read Ian Wright's 'The Social Architecture of Capitalism' some time shortly after the Kolkata conference. I remember reading this paper quite clearly, as the style of the paper was unusual. The paper is very strongly a modelling paper, with very little formal mathematical content. This resulted in my finding it very difficult to make much sense of, and in fact I didn't understand the paper until some years later. I also, at the time, found the Marxian approach very naïve and off-putting, particularly in the insistence on the use of the labour theory of value. This seemed to me plainly wrong; so at this stage I dumped this paper in the 'irrelevant' pile and forgot about it. That was a big mistake.

In 2006 it was suggested to me that the general Lotka-Volterra distribution might make a good fit to some high quality income data I had acquired from the UK Statistical Office. It turned out that the data did fit the GLV exceptionally well; better than alternative distributions.

As a scientist, this dictated that building models along the lines of the GLV would be the most sensible way forward.

By this stage, my knowledge of economics had expanded a little, and I was somewhat dismayed by the naivety and complexity of the approaches taken to economics by most physicists. It seemed to me that power law distributions, and gross inequality, had a universality through geography and more importantly history (cf the paper regarding inequality in ancient Egypt [Abul-Magd 2002]), and that they appeared to be valid in any society where wealth, including land, was traded.

This could be contrasted with, for example, community owned land systems in Africa, which though associated with general poverty appeared to be characterised by low levels of inequality. In my view any model for wealth distributions should be able to accommodate payments to capital in the broadest sense, whether this be via dividends, interest rates, or rent on land and property.

With this in mind I attempted to fit, in the simplest way possible, basic economic concepts to two different generating equations that I was aware were capable of producing GLV distributions. These two systems were the exchange system of Slanina and the GLV system of Levy & Solomon. I wrote a note and circulated it to a number of academics in early 2006, I have reproduced the note in full below in section 12.1.

Unfortunately, none of the academics proved interested in my proposals. Also unfortunately, I did not send the note to Wright or Souma & Nirei, as it had been some time since I read their papers, and I didn't consciously connect them to this present work.

I lacked the mathematical or programming skills to take this forward, so once again, nothing much happened for a few years.

In 2009, in the middle of the post-credit-crunch recession, I took the opportunity to start an MSc in Finance at Aston University. Due to some very unfortunate circumstances I was unable to complete the course.

However in the two terms I attended the course I acquired a lot of useful knowledge regarding basic finance and economics. I would also like to give thanks to Patricia Chelley-Steely for giving me important insights into the role of market-microstructure in general and liquidity in particular.

I was also able to gain invaluable assistance from George Vogiatzis and Maria Chli with regard to producing simulations of my models proposed in 2006. The exchange model proved difficult to construct. However, in March 2010 Maria and George produced the first Matlab model for me based on the GLV model in the second part of my 2006 note. Somewhat to my surprise, this produced a perfect GLV distribution on its first run, though no power law.

It turned out that, to generate the power law, the profit ratio had to be increased substantially from the initial 5% proposed to somewhere near 50%. A little investigation revealed that the returns to capital where indeed on this scale, and so this was realistic.

At this point George and Maria politely, but firmly, suggested that I conquer my technophobia and learn to program in Matlab myself. I followed their advice and discovered that it is a lot easier than other programming languages I had encountered. From the first programme, I produced all the other programmes in this paper in short order, with almost all programming work being done in May 2010. I remain deeply indebted to Maria and George for their initial assistance and support with this work.

The income model followed naturally from the wealth model. The companies model followed naturally from the wealth and income models. The commodity model followed naturally from the companies model.

During the modelling process I was rereading Steve Keen's 'Debunking Economics' and had also read some of the Goodwin modelling work while investigating the ratio of returns to labour and capital. It seemed to me that by combining the wealth, company and commodity models it would be possible to generate a much simpler but effective Goodwin style macroeconomic model. This proved to be the case, with a resultant simple base model that appeared to produce Minskian/Austrian cycles endogenously.

At some point after the modelling was largely complete, while rereading a large volume of papers I had collected over the years, I reread Wright's 'The Social Architecture of Capitalism'. For the second time I found it difficult to follow, and found the labour theory of value difficult to accept. However something in the paper was nagging at me. I reread the paper for a second time, more carefully; and slowly realised that, though coming from a completely different angle, Wright had built a model that was both making the same base assumptions as my own, and producing many of the same outputs. Indeed, in many ways Wright's models produced better results than my own.

Given the very different ways that Wright and myself produced our models, I believe that my approach was not influenced by Wright. My original proposals of 2006 were deliberately, mathematically based on the GLV, and were also focused on a financial sector with returns paid on capital. Wright's models are significantly different to my own, most notably in not involving a financial sector. Also, unlike the present paper Wright takes a 'black box' and 'zero intelligence' approach to modelling which eschews formal fitting of the models to mathematical equations.

Despite this belief, I am obliged to accept that I may have been influenced subconsciously by Wright's work.

Much later in the writing of this paper, close to it's completion, I reread the work of Souma & Nirei. Again I found the complexity of the mathematical approach of Souma & Nirei very difficult to follow, and I believe this complexity is unnecessary, and that my own approach is more useful as a basis for analysing economics. However the parallels between their work and my own are significant. Most notably Souma & Nirei use consumption as a dissipative part of their model in a way that is almost identical to my own models.

They also use capital as a main source of new wealth in their model, which is analogous to my own, though less strongly than with consumption. Souma & Nirei use capital growth as the main form of supplying new wealth to their model. They justify this by using supporting data from the Japanese economy. While this may have seemed sensible at the time, given the collapse of the Japanese stock-market and property prices over the last two decades, this now looks less sensible. Although I believe that capital growth can form a part of wealth generation, on a long-term cyclical basis this is likely to be very small. I believe that my simple model of returns to capital in the form of interest, dividends and rent is a better basis for future economic modelling.

As with Wright, I do not believe I was influenced directly by Souma & Nirei. My first model in 2006 was a simple exchange model, quite different to that of Souma & Nirei, while I generated the second model by simply substituting what to me were the most obvious and simple economic terms into Levy & Solomon's generating equation. Indeed my original model was a little over-complex and significantly different to that of Souma & Nirei.

However, even more so than Wright's work, the parallels between the models of Souma & Nirei and my own are striking. And the possibility that I was subconsciously assisted by their work seems significant.

I would like to state in the strongest terms that I believe that the work of Wright, Souma & Nirei is of considerable importance. These three academics have been able to bridge the gap between the physics and the economics in a way that no other academics have been able to. Also they all carried out this work prior to my own.

Where my own work differs to that of the gentlemen above is that it has a clear mathematical basis, unlike that of Wright, and that the mathematical basis is dramatically simpler than that of Souma & Nirei.

It is my hope that Wright, Souma, Nirei and myself can share the credit for finally bringing an effective mathematical and modelling approach to the understanding of economics.

Figure 12.1 here

12.1 Proposed Models 2006

Pair exchange process, after Slanina;

$$W_{i,t+1} = W_{i,t} + \beta_{ij} – \beta_{ji} – p_i + r * W_{i,t}$$

$$W_{j,t+1} = W_{j,t} + \beta_{ji} – \beta_{ij} – p_j + r * W_{j,t}$$

β_{ij} would be a good or service received,
β_{ji} would be money exchanged for the good or service,
(or vice versa) you could make this more 'economist friendly' by using:
β_{gs} for a good or service received,
β_m for money exchanged for the good or service,

typically β_{ij} would be a factor smaller than $W_{j,t}$ in size

$\Delta\beta = \beta_{ij} – \beta_{ji}$
is a small random difference in wealth due to the exchange not being exactly equal, typically $\Delta\beta$ would be a few percent of β_{ij} (economists would argue that $\Delta\beta$ would be equal to zero at equilibrium, I believe this is not the case, however it is much easier just to argue that there will be small random differences in the wealth exchange, which is a very plausible assumption) I see the $\Delta\beta$'s as the main stochastic driver in this model.

p_i

is the profit taken by a third party. If I buy a car directly off you, then p_i equals zero, but if I buy a car off you via e-bay, a small percentage of β_{ij}; p_i and/or p_j is taken by e-bay. (In e-bay's case, the seller is charged, so $p_i = 0$). Ignoring the example of e-bay, I would initially model this by assuming that all p_i's are a fixed small percentage of the exchange. So:

$$p_i = \beta_{gs} * p_{rate}$$

r
is the interest rate (factored down to a weekly or daily rate, whatever Δt is) Annual real interest rates (after inflation) are very stable, varying between 0.5 and 4% (annual) over long time periods. I would also initially model this as a small fixed percentage. (To get a working model with equations that balance it may be necessary to have a fixed relationship between and p_{rate} and r ;
p_{rate} = const * r)

I do not see any reason to make the r 's a distribution set. Most peoples investments are stable, poor peoples especially so. Rich people will only hold a portion of investments in riskier, more variable funds. I would only really see a need to introduce a distribution set if it was the only way we could generate the necessary curve.

So in this model the change in wealth comes from a small random element from the exchange, a small element taken in profit, and a small gain of interest which, crucially, is proportional to current wealth.

From a max entropy type approach I would then add the following two conditions:

$$\Sigma W_{i,t} = \Sigma W_{i,t+1}$$

ie, all wealth is conserved (ie. there is no economic growth or recession).

And:

$\Sigma\ p_i = \Sigma\ r * W_{i,t}$

ie, all profit is recycled as interest on peoples wealth.

In this model the stochastic variability comes from the wealth exchanges; the Δβ's. This combined with the assumption of conservation of wealth would provide a boltzmann type distribution if profits p_i and interest r were equal to zero.

I believe the extra terms of profit and interest will be a circular reinforcing mechanism that should produce the power tail.

If you can solve, this or something similar, hopefully you will get a wealth distribution that is a GLV with alpha = 1.5

<u>GLV type process;</u>

$W_{i,t+1} = W_{i,t} + Inc_i * \Delta t - p_{Inc} - Con_i * \Delta t - p_{Con} + r * W_{i,t}$

Inc_i
is waged income; income from employment. Realistically I would expect this to be a stable distribution, very much on the lines of Juergen's arguments. (http://arxiv.org/abs/cond-mat/0204234)

p_{Inc}
is the small profit taken by the employing organisation. Modelled as previous model.

Con_i
is consumption, which includes food, clothes, new cars, petrol, rent#, mortgage payments#, holidays, etc. (# not completely sure about these two). Consumption is the big variable, and is where I would expect the stochastic element to come in strongly.

p_{Con}
is the small profit taken by the shop, landlord, building society, etc.

r
As previous model.

Again, from a max entropy type approach, I would then add the following two conditions:

$\Sigma\ W_{i,t} = \Sigma\ W_{i,t+1}$

again, all wealth is conserved.

And:

$\Sigma\ (p_{Inc} + p_{Con}) = \Sigma\ r * W_{i,t}$

Again; all profit is recycled as interest.

From this equation you can derive something like:

Total Income = I_j

I_j = [Inc_i + (r * $W_{i,t}$ / Δt)] = [wages + interest, etc]

I_j = [Con_i + ([($W_{i,t+1}$ - $W_{i,t}$) + (p_{Inc} + p_{Con})] / Δt)]

If you can solve this or something similar, hopefully you will get an income distribution that is a GLV with alpha = 4 to 5.

13. Further Reading

One major aim of this paper has been to introduce existing concepts of mathematics and economics to audiences that may not be familiar with them. Primarily this means introducing the mathematics of chaos and statistical mechanics to economists, and introducing some basic economic and finance theory to mathematicians, scientists and engineers. However, the majority of the economics and finance referred to in this paper is heterodox, and so will also be new to most economists.

Figure 13.1 below shows a suggested route through the more central works referred to in this paper. The top section discusses statistical mechanics, the second section chaos and the bottom half economics and finance.

Figure 13.1 here

*/# - alternative / additional texts available

B, K & M – Brearley, Myers & Allen

G & W – Glazer & Wark

K & L – Kleidon & Lorenz

M & S – Miles & Scott

P & O – Pepper & Oliver

R & R – Reinhart & Rogoff

The diagram above is for assistance and is not intended to be prescriptive. The arrows simply indicate that, for example, the review by Ozawa will be easier to follow if Atkins and Ruhla has been read beforehand. If you have a strong mathematical bent and significant knowledge of finance then by all means start with Bouchaud et al.

To get a strong feel for how statistical mechanics works, both Atkins and Ben-Naim are essential reading, both use the minimum of mathematics and superb writing to explain difficult concepts very lucidly. Atkins follows a traditional energy approach, while Ben-Naim follows an information approach. I strongly recommend that anyone new to statistical mechanics read both books [Atkins 1994, Ben-Naim 2007].

‘The Physics of Chance’ by Charles Ruhla [Ruhla 1992], is also a very good book, well written with clear explanations, it builds from the foundations of probability into the basic ideas of both statistical mechanics and chaotic systems, and forms a natural bridge between Atkins/Ben-Naim and more formal textbooks.

Following this, Glazer & Wark is a well written basic statistical mechanics text book with a more mathematical treatment [Glazer & Wark 2001]. Gould & Tobochnik is an alternative, though it also covers standard thermal physics material; for statistical mechanics start at chapter three [Gould & Tobochnik 2010], Engel & Reid is a similar alternative, start at chapter 12, [Engel & Reid 2006], both are less easy to follow than Glazer & Wark.

For a discussion of the origin of power law tails, the paper by Newman is excellent, though I also recommend reading Mitzenmacher and Simkin & Roychowdhury [Newman 2005, Mitzenmacher 2004, Simkin & Roychowdhury 2006].

Unfortunately, the jump from standard statistical mechanics to the General Lotka Volterra work of Levy & Solomon is significant. The GLV approach is new and I don’t know of any good book discussing the GLV. It is for this reason that I have attempted to explain the GLV in some detail in section 1.2 of this paper. I have included Solomon’s own review of the GLV in the proposed reading scheme, but it is highly mathematical [Solomon 2000].

Following from Atkins and entropy in general, the paper by Ozawa et al gives an excellent review of the research and theory of maximum entropy production. This is expanded on with a set of very interesting papers in Kleidon & Lorenz [Ozawa et al 2003, Kleidon & Lorenz 2005]. The paper by Dewar [Dewar 2005] is of particular importance, and, in my opinion, links directly to the work of Levy & Solomon.

The website of Kumar [Kumar 2006] gives a brief but good introduction to plain Lotka-Volterra systems, and so an introduction to chaotic systems in general. Chapter eight of Keen gives a very good brief introduction to chaotic systems, Ruhla also gives an excellent introduction with a little more maths.

‘Nonlinear Dynamics and Chaos’ by Strogatz is an extraordinarily well written book, giving a full understanding of highly complex systems, including the mathematics, while using lots of clear examples and clear writing to keep things easy to follow. An alternative work by Hirsch, Smale & Devaney is also very good [Strogatz 2000, Hirsch et al 2003].

Following Strogatz or Hirsch, the works by either Britton or van den Berg move into the mathematics of more complex biological systems, where the Lotka-Volterra forms one of the simplest models [van den Berg 2010, Britton 2003,]. It is my belief that either of these books will prove a treasure-trove for people trying to find suitable models for economic and financial systems.

On similar lines, especially for financial regulators, Nise or similar standard control engineering texts show how straightforward it is to analyse and control complex dynamic systems [Nise 2000].

With regard to economics books, the most important thing is what not to read.

Almost all standard economics textbooks are pure neoclassicism with a few scraps of Bowdlerised Keynes thrown in. Unfortunately, despite being very wrong, neoclassicism is intellectually coherent and can be interesting to study, in the same way that for example ancient Latin or Greek is. It is however still wrong.

To understand the historical reasons why it is wrong read Mirowski, which is highly entertaining, but not necessary to learn about real economics [Mirowski 1989].

To understand the theoretical reasons why neoclassical economics is wrong I suggest reading Cassidy, Cooper and most importantly Steve Keen's 'Debunking Economics'.

John Cassidy's 'How Markets Fail' [Cassidy 2009] is ostensibly about the recent credit crunch. However the first two-thirds of the book gives a superb potted history of economic theory and how it measures up to reality. He includes heterodox economists such as Hayek and Minsky, as well as monetarism, behaviourism and game theory, along with neoclassical economics. The result is an outstanding review of economic history without any mathematics.

George Cooper [Cooper 2008] follows on from Cassidy with a more detailed look at finance, in an equally well written, non-mathematical book.

For a more mathematical, and very sharp analysis of the state of economics then you need 'Debunking Economics' by Steve Keen [Keen 2004].

If you only read one book out of those listed in this section, make sure that it is Debunking Economics (If you only read two books, make sure they are Keen and Ruhla). Keen explains in detail the main faults of neoclassical economics, and why the theories in the textbooks are simply wrong. He also discusses how economics needs to be changed, most notably by introducing proper dynamic modelling. He also reviews the various alternate strands of heterodox economics.

In parallel to the theoretical background of Keen, I would recommend the books by Smithers, Harrison, Reinhart & Rogoff, Bernholz and Lee [Smithers 2009, Harrison 2005, Reinhart & Rogoff 2009, Bernholz 2003, Lee 1999]. These books deal with share prices, house prices, financial crises, inflation and pricing respectively. Each is written with a long historical viewpoint and very full data. They give a clear feel for how real economies actually work, and the first three in particular make the dynamic, cyclical nature of economics clear.

The most important of these books is 'Post Keynesian Price Theory' by Lee which shows in careful detail how pricing is actually carried out in non-financial markets.

Finally, having been fore-armed with the theoretical background of Keen, and the real data of the six writers above I would recommend Miles & Scott as a standard macroeconomic text and Bodie, Kane & Marcus as a standard finance text [Miles & Scott 2002, Bodie et al 2009]. Miles & Scott use neoclassical techniques throughout, but are unusually honest with their questioning of the validity of assumptions. Their book is very good on giving underlying data on economics, and is a very good guide to the jargon and thinking in mainstream economics. Bodie, Kane & Marcus is similarly well written and is well supported with data, just keep in mind Cassidy, Cooper and Smithers' demolitions of rational markets in mind as you read it.

In international economics, Pettis has produced a profoundly insightful work that builds highly plausible theory to explain the history presented by Reinhart & Rogoff [Pettis 2001]. Mehrling [Mehrling 2000] gives a similarly insightful discussion of monetary economics.

For domestic financial markets Pepper & Oliver provide a short and highly, readable account of how liquidity and central banks affects markets from a practitioners point of view. The review by

Amihud et al 2005 gives much more background on the recent mathematical research in liquidity.

For the pricing and trading of financial assets in general; the field of market-microstructure is essential. Unfortunately there is not yet a good introductory book to cover this emerging and mathematical field. A very good introduction is given in a paper by Stoll, while an alternative discussion is given in Madhavan [Stoll 2003, Madhavan 2000]. The book by Lyons deals with market-microstructure in foreign exchange markets; this is in contrast to most market-microstructure work which is with equities. Despite this, Lyons is a well written work which deals very well with the basics of market-microstructure theory.

Finally, the work of Bouchaud, Farmer, Wyart and others in the econophysics community are bringing together detailed data analysis with theoretical work from market-microstructure and econophysics. [Farmer et al 2005, Wyart et al 2008, Bouchaud et al 2009].

14. Programmes

The programmes used for most of the modelling are included below. The income and company models were modelled in Matlab, the commodity and macroeconomic models were modelled in Excel.

If the Matlab models are pasted directly into the Matlab program editor they should run straight away. Minor modifications are needed to some of the programs to model different scenarios, the modifications required are indicated in the commented sections of the programmes. (nb. I am not by nature a programmer. The one thing I have learnt from my brief experience with Matlab is that whatever way you just did something, there was a better way. I ask for forbearance with my amateurish programming.)

The Excel files need to be pasted into a text editor such as Notepad, then imported into Excel. They then need further columns of formulae to be copied over, and graphs to be produced from the data. Finally, different data needs to be pasted in for each separate model. This is explained in full detail for each of the models.

14.1 Model 1A (Matlab)

```
rand('state',0);
randn('state',0);


profit_rate = 0.5;

number_runs = 10000;

agents = 10000;


halfway_wealth = zeros (1,agents);
consumption = zeros (1,agents);
```

```
waged_income = zeros (1,agents);
investment_income = zeros (1,agents);
total_income = zeros (1,agents);
total_waged_income = zeros (1,agents);
total_investment_income = zeros (1,agents);

profit = zeros (1,agents);

average_final_wealth = 1000;
initial_wealth = 1000 * ones (1,agents);
final_wealth = 1000 * ones (1,agents);


production = 200 * (ones(1,agents));

consumption_rate = 0.3 * (ones(1,agents));


for p = 1:number_runs

    profit = zeros (1,agents);

    total_profit = 0;
    total_wealth = 0;

    initial_wealth = final_wealth;


       for j = 1:agents

          consumption_rate(j) = 0.3 * ( 1 + 0.3*randn );

       end    %j

    consumption = initial_wealth .* consumption_rate;

    waged_income = (1 - profit_rate) * production;

    initial_wealth = initial_wealth + waged_income - consumption;

    profit = profit_rate * (production); % + consumption);

    total_wealth = sum (initial_wealth);
    total_profit = sum (profit);

    investment_income = (initial_wealth * total_profit) / total_wealth;

    final_wealth = initial_wealth + investment_income;

    average_final_wealth = (sum (final_wealth)) / agents;

    % halfway check results
    if p < (number_runs / 2)
    halfway_wealth = final_wealth;
    end  %p

    %income gathering - last 1000 runs
    if p > (number_runs - 1000)

        total_income = total_income + waged_income + investment_income;

    total_waged_income = total_waged_income + waged_income;
    total_investment_income = total_investment_income + investment_income;

    end     %p

    average_total_income = (sum(total_income))/agents;

end     %p
```

```
% deciles
deciles = ones (1,(agents/10));

% earnings deciles
production = sort (production);
decile_production = zeros ((agents/10),10);
decile_production(:) = production;
production_deciles = deciles * decile_production;
production_decile_ratio = production_deciles(10)/production_deciles(1);

% wealth deciles
final_wealth = sort (final_wealth);
decile_final_wealth = zeros ((agents/10),10);
decile_final_wealth(:) = final_wealth;
final_wealth_deciles = deciles * decile_final_wealth;
wealth_decile_ratio = final_wealth_deciles(10)/final_wealth_deciles(1);

% income deciles
total_income = sort (total_income);
decile_total_income = zeros ((agents/10),10);
decile_total_income(:) = total_income;
total_income_deciles = deciles * decile_total_income;
income_decile_ratio = total_income_deciles(10)/total_income_deciles(1);


% gini coefficients

index = zeros(1,agents);
for i = 1:agents

    index(i)=i;

end     %i

gini_earnings =((2*sum(production .* index))/(agents*sum(production))) -
((agents+1)/agents);
gini_wealth =((2*sum(final_wealth .* index))/(agents*sum(final_wealth))) -
((agents+1)/agents);
gini_income =((2*sum(total_income .* index))/(agents*sum(total_income))) -
((agents+1)/agents);


% relative poverty levels

poverty_number_wealth = 0;
poverty_ratio_wealth = 0;
poverty_number_income = 0;
poverty_ratio_income = 0;

for i = 1:agents

    if final_wealth(i) < average_final_wealth /2
        poverty_number_wealth = poverty_number_wealth + 1;
    end

     if total_income(i) < average_total_income /2
        poverty_number_income = poverty_number_income + 1;
    end

end     %i

poverty_ratio_wealth = poverty_number_wealth/agents;
poverty_ratio_income = poverty_number_income/agents;


%vertical display data

display_wealth = final_wealth';
```

```
display_income = total_income';
display_waged_income = total_waged_income';
display_investment_income = total_investment_income';
display_halfway_wealth = halfway_wealth';
display_consumption_rate = consumption_rate';
display_production = production';
```

14.2 Model 1B (Matlab)

```
rand('state',0);
randn('state',0);

profit_rate = 0.5;

number_runs = 10000;

agents = 10000;

halfway_wealth = zeros (1,agents);
consumption = zeros (1,agents);

waged_income = zeros (1,agents);
investment_income = zeros (1,agents);
total_income = zeros (1,agents);
total_waged_income = zeros (1,agents);
total_investment_income = zeros (1,agents);

profit = zeros (1,agents);

average_final_wealth = 1000;
initial_wealth = 1000 * ones (1,agents);
final_wealth = 1000 * ones (1,agents);

production = 200 * (ones(1,agents) + 0.2 * randn (1,agents));

consumption_rate = 0.2 * (ones(1,agents));

for p = 1:number_runs

    profit = zeros (1,agents);

    total_profit = 0;
    total_wealth = 0;

    initial_wealth = final_wealth;

    consumption = initial_wealth .* consumption_rate;

    waged_income = (1 - profit_rate) * production;

    initial_wealth = initial_wealth + waged_income - consumption;

    profit = profit_rate * (production);

    total_wealth = sum (initial_wealth);
    total_profit = sum (profit);

    investment_income = (initial_wealth * total_profit) / total_wealth;

    final_wealth = initial_wealth + investment_income;

    average_final_wealth = (sum (final_wealth)) / agents;

    % halfway check results
    if p < (number_runs / 2)
    halfway_wealth = final_wealth;
```

```
    end  %p

    %income gathering - last 1000 runs
    if p > (number_runs - 1000)
    total_income = total_income + waged_income + investment_income;

    total_waged_income = total_waged_income + waged_income;
    total_investment_income = total_investment_income + investment_income;

    end     %p

    average_total_income = (sum(total_income))/agents;

end     %p

% deciles
deciles = ones (1,(agents/10));

% earnings deciles
production = sort (production);
decile_production = zeros ((agents/10),10);
decile_production(:) = production;
production_deciles = deciles * decile_production;
production_decile_ratio = production_deciles(10)/production_deciles(1);

% wealth deciles
final_wealth = sort (final_wealth);
decile_final_wealth = zeros ((agents/10),10);
decile_final_wealth(:) = final_wealth;
final_wealth_deciles = deciles * decile_final_wealth;
wealth_decile_ratio = final_wealth_deciles(10)/final_wealth_deciles(1);

% income deciles
total_income = sort (total_income);
decile_total_income = zeros ((agents/10),10);
decile_total_income(:) = total_income;
total_income_deciles = deciles * decile_total_income;
income_decile_ratio = total_income_deciles(10)/total_income_deciles(1);

% gini coefficients

index = zeros(1,agents);
for i = 1:agents

    index(i)=i;

end     %i

gini_earnings =((2*sum(production .* index))/(agents*sum(production))) -
((agents+1)/agents);
gini_wealth =((2*sum(final_wealth .* index))/(agents*sum(final_wealth))) -
((agents+1)/agents);
gini_income =((2*sum(total_income .* index))/(agents*sum(total_income))) -
((agents+1)/agents);

% relative poverty levels
poverty_number_wealth = 0;
poverty_ratio_wealth = 0;
poverty_number_income = 0;
poverty_ratio_income = 0;

for i = 1:agents

    if final_wealth(i) < average_final_wealth /2
        poverty_number_wealth = poverty_number_wealth + 1;
    end

     if total_income(i) < average_total_income /2
```

```
        poverty_number_income = poverty_number_income + 1;
    end

end     %i

poverty_ratio_wealth = poverty_number_wealth/agents;
poverty_ratio_income = poverty_number_income/agents;

%display vertical
display_wealth = final_wealth';
display_income = total_income';
display_waged_income = total_waged_income';
display_investment_income = total_investment_income';
display_halfway_wealth = halfway_wealth';
display_consumption_rate = consumption_rate';
display_production = production';
```

14.3 Model 1C (Matlab)

```
rand('state',0);
randn('state',0);


gini_vector = zeros (7,19);

wealth_vector = zeros (agents,19);
income_vector = zeros (agents,19);

for m = 1:19

profit_rate = m * 0.05;

%profit_rate = 0.5;

number_runs = 10000;

agents = 10000;

cross_check_randn = zeros (1,agents);

halfway_wealth = zeros (1,agents);
consumption = zeros (1,agents);

waged_income = zeros (1,agents);
investment_income = zeros (1,agents);
total_income = zeros (1,agents);
total_waged_income = zeros (1,agents);
total_investment_income = zeros (1,agents);

profit = zeros (1,agents);

average_final_wealth = 1000;
initial_wealth = 1000 * ones (1,agents);
final_wealth = 1000 * ones (1,agents);
rent = 0 * ones (1,agents);

production = 200 * (ones(1,agents));
consumption_rate = 0.2 * (ones(1,agents) + 0.1 * randn (1,agents));


for p = 1:number_runs

    profit = zeros (1,agents);

    total_profit = 0;
    total_wealth = 0;
```

```
    initial_wealth = final_wealth;

    consumption = initial_wealth .* consumption_rate;


    waged_income = (1 - profit_rate) * production;

    initial_wealth = initial_wealth + waged_income - consumption;

    profit = profit_rate * (production);

    total_wealth = sum (initial_wealth);
    total_profit = sum (profit);


    investment_income = (initial_wealth * total_profit) / total_wealth;

    final_wealth = initial_wealth + investment_income;

    average_final_wealth = (sum (final_wealth)) / agents;

    % halfway check results
    if p < (number_runs / 2)
    halfway_wealth = final_wealth;

    end  %if p

    %income gathering - last 1000 runs
    if p > (number_runs - 1000)
    total_income = total_income + waged_income + investment_income;

    total_waged_income = total_waged_income + waged_income;
    total_investment_income = total_investment_income + investment_income;

    end     %if p

    average_total_income = (sum(total_income))/agents;


end     %p


% deciles
deciles = ones (1,(agents/10));

% earnings deciles
production = sort (production);
decile_production = zeros ((agents/10),10);
decile_production(:) = production;
production_deciles = deciles * decile_production;
production_decile_ratio = production_deciles(10)/production_deciles(1);

% wealth deciles
final_wealth = sort (final_wealth);
decile_final_wealth = zeros ((agents/10),10);
decile_final_wealth(:) = final_wealth;
final_wealth_deciles = deciles * decile_final_wealth;
wealth_decile_ratio = final_wealth_deciles(10)/final_wealth_deciles(1);

% income deciles
total_income = sort (total_income);
decile_total_income = zeros ((agents/10),10);
decile_total_income(:) = total_income;
total_income_deciles = deciles * decile_total_income;
income_decile_ratio = total_income_deciles(10)/total_income_deciles(1);


% gini coefficients
```

```
index = zeros(1,agents);
for i = 1:agents

    index(i)=i;

end     %i

gini_earnings =((2*sum(production .* index))/(agents*sum(production))) -
((agents+1)/agents);
gini_wealth =((2*sum(final_wealth .* index))/(agents*sum(final_wealth))) -
((agents+1)/agents);
gini_income =((2*sum(total_income .* index))/(agents*sum(total_income))) -
((agents+1)/agents);


% relative poverty levels

poverty_number_wealth = 0;
poverty_ratio_wealth = 0;
poverty_number_income = 0;
poverty_ratio_income = 0;

for i = 1:agents

    if final_wealth(i) < average_final_wealth /2
        poverty_number_wealth = poverty_number_wealth + 1;
    end

     if total_income(i) < average_total_income /2
        poverty_number_income = poverty_number_income + 1;
    end

end     %i

poverty_ratio_wealth = poverty_number_wealth/agents;
poverty_ratio_income = poverty_number_income/agents;

%vertical displays
display_wealth = final_wealth';
display_income = total_income';
display_waged_income = total_waged_income';
display_investment_income = total_investment_income';
display_halfway_wealth = halfway_wealth';
display_consumption_rate = consumption_rate';
display_production = production';

gini_vector (1,m) = profit_rate;
gini_vector (2,m) = gini_wealth;
gini_vector (3,m) = gini_income;
gini_vector (4,m) = wealth_decile_ratio;
gini_vector (5,m) = income_decile_ratio;
gini_vector (6,m) = poverty_ratio_wealth;
gini_vector (7,m) = poverty_ratio_income;

for j = 1:agents

    wealth_vector (j,m) = display_wealth (j,1);
    income_vector (j,m) = display_income (j,1);

end     %j


end     %m
```

14.4 Model 1D (Matlab)

```
%    Note, different commented sections below need
%     to be uncommented to model maximum wealth.
%     compulsory saving and model 1E

rand('state',0);
randn('state',0);

profit_rate = 0.5;

number_runs = 10000;

agents = 10000;

maximum_wealth = 1500;

cross_check_randn = zeros (1,agents);

halfway_wealth = zeros (1,agents);
consumption = zeros (1,agents);

waged_income = zeros (1,agents);
investment_income = zeros (1,agents);
total_income = zeros (1,agents);
total_waged_income = zeros (1,agents);
total_investment_income = zeros (1,agents);

profit = zeros (1,agents);

average_final_wealth = 1000;
initial_wealth = 1000 * ones (1,agents);
final_wealth = 1000 * ones (1,agents);
rent = 0 * ones (1,agents);

production = 200 * (ones(1,agents) + 0.1 * randn (1,agents));

consumption_rate = 0.2 * (ones(1,agents) + 0.1 * randn (1,agents));
%     for model 1E change 0.2 to 0.3 in the equation above.

for p = 1:number_runs

    profit = zeros (1,agents);

    total_profit = 0;
    total_wealth = 0;

    initial_wealth = final_wealth;

    consumption = initial_wealth .* consumption_rate;

%     %  compulsory saving - START
%     %     uncomment this section for compulsory saving
%     for j = 1:agents
%
%     if  initial_wealth(j) < (0.9*average_final_wealth)
%         consumption(j) = 0.8*consumption(j);
%
%     end %if
%
%     end %for j
%     %  compulsory saving - END

    waged_income = (1 - profit_rate) * production;

    initial_wealth = initial_wealth + waged_income - consumption;

    profit = profit_rate * (production);

    total_wealth = sum (initial_wealth);
    total_profit = sum (profit);
```

```
    investment_income = (initial_wealth * total_profit) / total_wealth;

    final_wealth = initial_wealth + investment_income;

    average_final_wealth = (sum (final_wealth)) / agents;

    % halfway check results
    if p < (number_runs / 2)
    halfway_wealth = final_wealth;

    end  %if p

    %income gathering - last 1000 runs
    if p > (number_runs - 1000)
    total_income = total_income + waged_income + investment_income;

    total_waged_income = total_waged_income + waged_income;
    total_investment_income = total_investment_income + investment_income;

    end     %if p

    average_total_income = (sum(total_income))/agents;

%    %  maximum wealth barrier  - START
%    %   uncomment this section for maximum wealth barrier
%    %   also choose whether to enforce with decreased production or
%    %     increased consumption
%
%     for j = 1:agents
%
%
% %  %  uncomment for decreased production (and comment if below)
% %      if  final_wealth(j) > maximum_wealth
% %          production(j) = 0.95 * production(j);
%
%    %  uncomment for increased consumption (and comment if above)
%       if  final_wealth(j) > maximum_wealth
%           consumption_rate(j) = 1.05 * consumption_rate(j);
%
%     end   %if
%
%     end   %j
%    %  maximum wealth barrier  - END

end     %p

% deciles
deciles = ones (1,(agents/10));

% earnings deciles
production = sort (production);
decile_production = zeros ((agents/10),10);
decile_production(:) = production;
production_deciles = deciles * decile_production;
production_decile_ratio = production_deciles(10)/production_deciles(1);

% wealth deciles
final_wealth = sort (final_wealth);
decile_final_wealth = zeros ((agents/10),10);
decile_final_wealth(:) = final_wealth;
final_wealth_deciles = deciles * decile_final_wealth;
wealth_decile_ratio = final_wealth_deciles(10)/final_wealth_deciles(1);

% income deciles
total_income = sort (total_income);
decile_total_income = zeros ((agents/10),10);
decile_total_income(:) = total_income;
total_income_deciles = deciles * decile_total_income;
```

```
income_decile_ratio = total_income_deciles(10)/total_income_deciles(1);


% gini coefficients
index = zeros(1,agents);
for i = 1:agents

    index(i)=i;

end     %i

gini_earnings =((2*sum(production .* index))/(agents*sum(production))) -
((agents+1)/agents);
gini_wealth =((2*sum(final_wealth .* index))/(agents*sum(final_wealth))) -
((agents+1)/agents);
gini_income =((2*sum(total_income .* index))/(agents*sum(total_income))) -
((agents+1)/agents);

% relative poverty levels
poverty_number_wealth = 0;
poverty_ratio_wealth = 0;
poverty_number_income = 0;
poverty_ratio_income = 0;

for i = 1:agents

    if final_wealth(i) < average_final_wealth /2
        poverty_number_wealth = poverty_number_wealth + 1;
    end

     if total_income(i) < average_total_income /2
        poverty_number_income = poverty_number_income + 1;
    end

end     %i

poverty_ratio_wealth = poverty_number_wealth/agents;
poverty_ratio_income = poverty_number_income/agents;

%vertical displays
display_wealth = final_wealth';
display_income = total_income';
display_waged_income = total_waged_income';
display_investment_income = total_investment_income';
display_halfway_wealth = halfway_wealth';
display_consumption_rate = consumption_rate';
display_production = production';
```

14.5 Model 2A (Matlab)

```
rand('state',0);
randn('state',0);


number_runs = 10000;

companies = 10000;

total_capital = 10000000;

minimum_capital = 10;

initial_capital = (total_capital/companies)* ones(1,companies) ;

final_capital = initial_capital;

initial_market_cap = initial_capital;
```

```
upside_payout_factor = 1.0;
downside_payout_factor = 1.0;

production_rate = zeros(1,companies);

production = zeros(1,companies);

expected_returns = zeros(1,companies);

actual_returns = zeros(1,companies);

halfway_capital = zeros (1,companies);


for p = 1:number_runs

    initial_capital = final_capital;


    for k = 1:companies

        production_rate(k) = 0.1 * (1 + 0.2 * randn);

    end %end k


    production = initial_capital  .* (production_rate);  % production generated

    expected_returns = initial_market_cap * 0.1;


    for k = 1:companies

       if  production(k) > expected_returns(k)

           actual_returns(k) = (expected_returns(k) * upside_payout_factor) +
(production(k) * (1 - upside_payout_factor));

       else actual_returns(k) = (expected_returns(k) * downside_payout_factor) +
(production(k) * (1 - downside_payout_factor));

       end %if

    end %end k


    final_capital = initial_capital + production - actual_returns;

    initial_market_cap = actual_returns .* 10;

    total_final_capital = sum(final_capital);

    final_capital = (final_capital * total_capital)/total_final_capital;


    % halfway check results
    if p < (number_runs / 2)

        halfway_capital = final_capital;

    end  % end if p

end

display_capital = final_capital';
display_halfway_capital = halfway_capital';
```

14.6 Model 2B (Matlab)

```
rand('state',0);
randn('state',0);


number_runs = 10000; %100000

companies = 10000;

total_capital = 10000000;


initial_capital = (total_capital/companies)* ones(1,companies) ;

final_capital = initial_capital;

initial_market_cap = initial_capital;


upside_payout_factor = 0.9;
downside_payout_factor = 0.9;

production_rate = zeros(1,companies);

production = zeros(1,companies);

expected_returns = zeros(1,companies);

actual_returns = zeros(1,companies);

halfway_capital = zeros (1,companies);


for p = 1:number_runs

    initial_capital = final_capital;

    for k = 1:companies

        production_rate(k) = 0.1 * (1 + 0.2 * randn);

    end %end k


    production = initial_capital  .* (production_rate);  % production generated

    expected_returns = initial_market_cap * 0.1;


    for k = 1:companies


       if  production(k) > expected_returns(k)

           actual_returns(k) = (expected_returns(k) * upside_payout_factor) +
(production(k) * (1 - upside_payout_factor));

       else actual_returns(k) = (expected_returns(k) * downside_payout_factor) +
(production(k) * (1 - downside_payout_factor));

       end %if

    end %end k


    final_capital = initial_capital + production - actual_returns;

    initial_market_cap = actual_returns .* 10;
```

```
    total_final_capital = sum(final_capital);

    final_capital = (final_capital * total_capital)/total_final_capital;


    % halfway check results
    if p < (number_runs / 2)

        halfway_capital = final_capital;

    end  % end if p

end

display_capital = final_capital';
display_halfway_capital = halfway_capital';
```

14.7 Model 2C (Matlab)

```
rand('state',0);
randn('state',0);

number_runs = 10000;

companies = 10000;

total_capital = 10000000;

initial_capital = (total_capital/companies)* ones(1,companies) ;

final_capital = initial_capital;

initial_market_cap = initial_capital;

upside_payout_factor = 0.9;
downside_payout_factor = 0.5;

production_rate = zeros(1,companies);

production = zeros(1,companies);

expected_returns = zeros(1,companies);

actual_returns = zeros(1,companies);

halfway_capital = zeros (1,companies);


  for k = 1:companies

        production_rate(k) = 0.1 * (1 + 0.1 * randn);

  end %end k

production_rate = sort (production_rate, 'descend');


for p = 1:number_runs

    initial_capital = final_capital;

    production = initial_capital  .* (production_rate);  % production generated

    expected_returns = initial_market_cap * 0.1;
```

```
    for k = 1:companies

       if  production(k) > expected_returns(k)

           actual_returns(k) = (expected_returns(k) * upside_payout_factor) + 
(production(k) * (1 - upside_payout_factor));

       else actual_returns(k) = (expected_returns(k) * downside_payout_factor) + 
(production(k) * (1 - downside_payout_factor));

       end %if

    end %end k

    final_capital = initial_capital + production - actual_returns;

    initial_market_cap = actual_returns .* 10;

    total_final_capital = sum(final_capital);

    final_capital = (final_capital * total_capital)/total_final_capital;


    % halfway check results
    if p < (number_runs / 2)

        halfway_capital = final_capital;

    end  % end if p

end

display_capital = final_capital';
display_halfway_capital = halfway_capital';
display_initial_market_cap = initial_market_cap';
```

14.8 Model 3 - Commodity (Excel)

<u>Instructions</u>

Open a text editor programme - in windows you can go to ‘all programs’ / ‘accessories’ and open 'notepad'.

From the text below; under ‘Program’, select and copy all the text between the two rows of asterisks - but do not select the asterisks themselves.

Go to the text editor and paste all the text into the text editor.

The first line in the text editor should read: this writing should be in cell A1.

If you have pasted the asterisks into the text editor, delete them. If there is a space above the first line delete it.

Save the data as a plain text file in a location you will be able to find easily.

Open excel, open a new worksheet.

Go to 'Data' / 'Get External Data' / 'Import Text File'. Use the explorer window to find and open the text file you saved above.

Select ‘delimited’ and then ‘Next.

Select ‘Comma’, also unselect ‘Tab’, select ‘Next’.

Select ‘Finish’.

Put the data in the existing worksheet, in cell $A$1.

The phrase: ‘this writing should be in cell A1’ should be in cell A1. If it isn’t, select all the text and move it en masse so that the phrase is actually in cell A1.

Check all the formulae are all working as formulae. The process above should work, however sometimes the formulae still keep the apostrophe (‘) in front of the equal signs (=) from the CSV input. If there are any apostrophes in front of any equal signs, delete them before going on to the next step. (Note that if a formula is showing “#DIV/0!” it is working correctly as a formula; the next step below will provide the missing data to prevent the division by zero errors.)

Select all the data from cell K16 to K34 inclusive.

Copy this data over into cells L16 to HB 34; the easiest way to do this is by moving the cursor over the small black square at the bottom right of the selection, right-clicking on it and dragging across to column HB. If this is done correctly row 16 should automatically increment from 1 to 200 timesteps.

To create a graph, select the whole area from I16 to HB34, and then press the chart wizard button. Set up the graph, using x-y scatter, with data points connected by smoothed lines.

Once you have your graph you can format it, make a copy of it, and delete unwanted data series as required.

Now you can run different parameters in the model to see what happens. Enter the parameters in column J, between J3 and J8.

The parameters for the models in this paper are given in cells D1 to F11.

Program

```
******************************************************************************************
this writing should be in cell A1,,,values for different models below,,,,,,required values entered in column J below,
,,,Model 3A,Model 3B,Model 3C,
,,,0.1,0.1,0.1,,,interest_rate,0.1,
,,,0.2,0.2,0.2,,,production_rate,0.2,
,,,0.4,0.4,0.4,,,consumption rate,0.4,
,,,1,1,0.9,,,upside_payout_ratio,0.9,
,,,1,1,0.9,,,downside_payout_ratio,0.9,
,,,0,2,0,,,lag (max 10),0,
,,
,,,-9,-9,-9,,,m,-9,
,,,1000,1000,1000,,,c,1000,
,,
,,
,,,,,,,,initial,
,,,,,,,,values,
,,,,,,,timesteps,0,1,
,,,,,,,expected_returns,,=J34*$J$3,
,,,,,,,commodity payments variable component,,=$J$10*J22+$J$11,
,average,average,,,,,,commodity payments mimimum component,100,=$J$19,
,timesteps,timesteps,,,,,,commodity payments - actual,,=MAX(K18:K19),
,1 to 200,21 to 200*,,,,,,production_rate * capital,,=$J$4*J33,
,=MAX(K22:HB22),=AVERAGE(AE22:HB22),,,,,,actual production - smaller of 2 above,100,=MIN(K20:K21),
,=AVERAGE(K23:HB23),=AVERAGE(AE23:HB23),,,,,,prices,,=K20/K22,
,* allows equilbrium to form,,,,,,,capital_employed,,=K22/$J$4,
,,,,,,,production_revenue,,=K20-K22,
,,,,,,,downside returns,,= (K17 *$J$7) + (K25 * (1 -$J$7)),
,,,,,,,upside returns,,= (K17 * $J$6) + (K25 * (1 -$J$6)),
,,,,,,,returns_selector,,"=IF(K25<K17,1,0)",
,=AVERAGE(K29:HB29),=AVERAGE(AE29:HB29),,,,,,actual_returns,,=(K26*K28+K27*(1-K28)),
,,,,,,,,,=K20-K22-K29,
,,,,,,,capital procured in line above - do not enter values in the line above,
,,,,,,,capital_added,,"=OFFSET(K32,-2,($J$8*-1),1,1)",
,=AVERAGE(K33:HB33),=AVERAGE(AE33:HB33),,,,,,capital_available,500,=J33+K32,
,=AVERAGE(K34:HB34),=AVERAGE(AE34:HB34),,,,,,capital_wealth,500,=K29/$J$3,
******************************************************************************************
```

14.9 Model 4 - Macroeconomy (Excel)

Instructions

Open a text editor programme - in windows you can go to 'all programs' / 'accessories' and open 'notepad'.

From the text below; under 'Program', select and copy all the text between the two rows of asterisks - but do not select the asterisks themselves.

Go to the text editor and paste all the text into the text editor.

The first line in the text editor should read: this writing should be in cell A1.

If you have pasted the asterisks into the text editor, delete them. If there is a space above the first line delete it.

Save the data as a plain text file in a location you will be able to find easily.

Open excel, open a new worksheet.

Go to 'Data' / 'Get External Data' / 'Import Text File'. Use the explorer window to find and open the text file you saved above.

Select 'Delimited' and then 'Next.

Select 'Comma', also unselect 'Tab', select 'Next'.

Select 'Finish'.

Put the data in the existing worksheet, in cell $A$1.

The phrase: 'this writing should be in cell A1' should be in cell A1. If it isn't, select all the text and move it en masse so that the phrase is actually in cell A1.

Check all the formulae are all working as formulae. The process above should work, however sometimes the formulae still keep the apostrophe (') in front of the equal signs (=) from the CSV input. If there are any apostrophes in front of any equal signs, delete them before going on to the next step. (Note that if a formula is showing "#DIV/0!" it is working correctly as a formula; the next step below will provide the missing data to prevent the division by zero errors.)

Select all the data from cell N17 to O37 inclusive.

Copy this data over into cells P17 to HE37; the easiest way to do this is by moving the cursor over the small black square at the bottom right of the selection, right-clicking on it and dragging across to column HE. If this is done correctly row 17 should automatically increment from 1 to 200 timesteps.

To create a graph, select the whole area from L17 to HE37, and then press the chart wizard button. Set up the graph, using x-y scatter, with data points connected by smoothed lines.

Once you have your graph you can format it, make a copy of it, and delete unwanted data series as required.

Now you can run different parameters in the model to see what happens. Enter the parameters in column M, between M3 and M12, values for capital should be changed in M32 and M33.

The parameters for the models in this paper are given in cells F3 to I14.

To experiment with Bowley ratios and cash balances you will need to use Solver. You may need to install this if it isn't already installed. To check, make sure a cell (any cell) is selected on the spreadsheet. Go to the 'Tools' menu and look for 'Solver'. If Solver is on the list open it, if Solver is not available, go to 'Add-Inns', tick the box for 'Solver' and click OK. You will then be able to install Solver, you may need your original software discs.

Once you have Solver open you can target particular levels of cash wealth or Bowley ratio (earnings / total_returns).

To target a cash wealth, insert the value of your required cash wealth in cell G34, open Solver, set the target cell as H34 (cell H34 is a formula – do not enter any values in cell H34). Select 'Min' on 'Equal To:'. Under 'By Changing Cells:' select cell M32 – the Capital(K). Then select 'Solve'.

To target a Bowley ratio, insert the value of your required Bowley ratio in cell G37, open Solver, set the target cell as H37 (cell H37 is a formula – do not enter any values in cell H37). Select 'Min' on 'Equal To:'. Under 'By Changing Cells:' select cell M32 – the Capital(K). Then select 'Solve'.

Program

```
******************************************************************************************
this writing should be in cell A1,,,,values for different models below,,,,,,,,required values entered in column M below,
,,,,Model A,Model B,Model C,Model D,Model E,,,,,
,,,interest_rate,0.1,0.1,0.04,0.04,0.04,,,interest_rate,0.1,
,,,production_rate,0.2,0.2,0.2,0.2,0.4,,,production_rate,0.2,
,,,omega,0.4,0.4,0.4,0.5,0.5,,consumption rate,omega,0.4,
,,,upside_payout_ratio,1,1,1,0.7,0.8,,,upside_payout_ratio,1,
,,,downside_payout_ratio,1,1,1,0.7,0.8,,,downside_payout_ratio,1,
,,,lag,0,0,3,0,1,,,lag,0,(max 12)
,,,labour_required,1,1,1,1,1,,,labour_required,1,
,,,A,1,1,1,1,1,,,A,1,
,,,B,4,4,4,4,4,,,B,4,
,,,CC,100,100,100,100,100,,,C,100,
,,,capital (K),100,400,100,100,100,
,,,capital_wealth (W),100,100,100,300,100,
,
,,,,average,average,
,,,,timesteps,timesteps,,,,,,timesteps,0,1,2,
,,,,1 to 200,21 to 200*,,,,,,expected_returns,0,=M33*$M$3,=N33*$M$3,
,,,,=AVERAGE(N19:HE19),=AVERAGE(AH19:HE19),,,,wealth * omega,,goods_payments,0,=$M$5*M35,=$M$5*N35,
,,,,* allows equilbrium to form,,,,,production_rate * capital,,potential production,0,=$M$4*M32,=$M$4*N32,
,,,,,,,,,smaller of two above,,production,0,=MIN(N19:N20),=MIN(O19:O20),
,,,,,,,,,,,capital_employed,0,=N21/$M$4,=O21/$M$4,
,,,,=AVERAGE(N23:HE23),=AVERAGE(AH23:HE23),,,,quadratic,,earnings_income,0,=$M$9*$M$10 * (($M$11 * N22 - $M$12)*N22)/1000,=$M$9*$M$10 * (($M$11 * O22 - $M$12)*O22)/1000,
,,,,,,,,,,,production_revenue,0,=N19-N23,=O19-O23,
,,,,,,,,,,,downside returns,0,= (N18 * $M$7) + (N24 * (1 - $M$7)),= (O18 * $M$7) + (O24 * (1 - $M$7)),
,,,,,,,,,,,upside returns,0,= (N18 * $M$6) + (N24 * (1 - $M$6)),= (O18 * $M$6) + (O24 * (1 - $M$6)),
,,,,,,,,,,,returns_selector,0,"=IF(N24<N18,1,0)","=IF(O24<O18,1,0)",
,,,,=AVERAGE(N28:HE28),=AVERAGE(AH28:HE28),,,,,,,actual_returns,1,=(N25*N27+N26*(1-N27)),=(O25*O27+O26*(1-O27)),
0,0,0,0,0,0,0,0,0,0,0,0,0,=N19-N23-N28,=O19-O23-O28,
,,,,,,,,,,,capital procured in line above - do not enter values in the line above,
,,,,,,,,,,,capital_added,0,"=OFFSET(N31,-2,($M$8*-1),1,1)","=OFFSET(O31,-2,($M$8*-1),1,1)",
,,,,=AVERAGE(N32:HE32),=AVERAGE(AH32:HE32),,,,,,capital (K),100,=M32+N31,=N32+O31,
,,,,=AVERAGE(N33:HE33),=AVERAGE(AH33:HE33),,,,,,capital_wealth (W),100,=N28/$M$3,=O28/$M$3,
,,,,=AVERAGE(N34:HE34),=AVERAGE(AH34:HE34),0,=ABS(F34-G34),,,,cash_wealth,0,=M34-N19+N23+N28,=N34-O19+O23+O28,
,,,,=AVERAGE(N35:HE35),=AVERAGE(AH35:HE35),,,,,,total_wealth,100,=N33+N34,=O33+O34,
,,,,=AVERAGE(N36:HE36),=AVERAGE(AH36:HE36),,,,,,total_returns,0,=N23+N28,=O23+O28,
,,,,=AVERAGE(N37:HE37),=AVERAGE(AH37:HE37),0.7,=ABS(F37-G37),,,,earnings/total_returns,0,=N23/N36,=O23/O36,
,,,,,,set targets,minimise,
,,,,,,in cells,values above,
,,,,,,column G,in column H,
,,,,,,above,using solver.,
,,,,,,,do not enter,
,,,,,,,values in,
,,,,,,,column H,
******************************************************************************************
```

15. References

Abul-Magd 2002	Abul-Magd AY, 2002. Wealth distribution in an ancient Egyptian society. Phys. Rev. E 66, 057104.
Acharya & Pedersen 2005	Acharya V, Pedersen L, 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375-410.
Ackland & Gallagher 2004	Ackland GJ, Gallagher ID, 2004. Stabilization of large generalized Lotka-Volterra foodwebs by evolutionary feedback, Phys. Rev. Lett., 93, doi:10.1103/PhysRevLett.93.158701.
Amihud et al 2005	Amihud Y, Mendelson H, Pederson H, 2005. Liquidity and Asset Prices, Foundations and Trends in Finance 1:4.
Atkins 1994	Atkins PW, 1994. The Second Law: Energy, Chaos, and Form. New York, Scientific American Library.
Atkinson & Bourguignon 2000	Atkinson AB, Bourguignon F, (Eds.), 2000. The handbook of income distribution, V.1. Amsterdam and New York, Elsevier.
Ayres & Nair 1984	Ayres RU, Nair I, 1984. Thermodynamics and economics. Physics Today November: 62–71
Baek et el 2011	Baek SK, Berhardsson S, Minnhagen P, 2011. Zipf's law unzipped. New Journal of Physics 13 043004.
Bai et al 2006	Bai CB, Chang-Tai H, Qingyi Q, 2006. Returns to Capital in China. Brookings Papers in Economic Activity 2006(2).
Bandourian et al 2002	Bandourian R, McDonald JB, Turley RS, 2002. A Comparison of Parametric Models of Income Distribution across Countries and Over Time. Working Paper No. 305. Luxembourg: Luxembourg Income Study.
Baquero & Verbeek 2009	Baquero G, Verbeek M, 2009. Investing: Evidence from Hedge Fund Investors. Presented at 2009 Conference of the European Financial Management Association, http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2009-milan/EFMA2009_0646_fullpaper.pdf
Barbosa-Filho Taylor 2006	Barbosa-Filho NH, Taylor L, 2006. Distributive and demand cycles in the US economy – a structuralist Goodwin model. Metroeconomica, 57: 389-411.
BBC	BBC. GCSE Bitesize Science, Compete or Die, Predators and Prey. http://www.bbc.co.uk/schools/gcsebitesize/science/ocr_gateway/environment/2_compete_or_die2.shtml
BBC 2010a	16 May 2010. Household wealth grows five-fold in past 50 years. http://www.bbc.co.uk/news/10118346
BBC 2010b	BBC News, 18 June 2010. Forth Valley Royal Hospital to use robot 'workers', http://www.bbc.co.uk/news/10344849
Ben-Naim 2007	Ben-Naim A. Entropy Demystified, The second law of thermodynamics reduced to plain common sense.; World Scientific: Singapore, 2007.
Ben-Rephael et al 2008	Ben-Rephael A, Kadan O, Wohl A, 2008. The diminishing liquidity premium. WP presented to the NBER workshop on financial microstructures. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1099829
Bernholz 2003	Bernholz P, 2003. Monetary Regimes and Inflation: History, economic and political relationships. Cheltenham: Edward Elgar.
Biais et al 2005	Biais B, Glosten L, Spatt C, 2005. Market microstructure: A survey of microfoundations, empirical results, and policy implications. Journal of Financial Markets 8:217–264.
Biopact 2007	Biopact, 2007. Brazilian biofuels can meet world's total gasoline needs. http://news.mongabay.com/bioenergy/2007/05/brazilian-biofuels-can-meet-worlds.html
Bodie et al 2009	Bodie Z, Kane A, Marcus AJ, 2009. Investments. McGraw Hill, Boston.
Borges 2002	Borges EP, 2002. Empirical nonextensive laws for the geographical distribution of wealth. cond-mat/0205520.
Bouchaud & Mezard 2000	Bouchaud JP, Mezard M, 2000. Wealth Condensation in a Simple Model of Economy. Physica A282:536-545, cond-mat/0002374.
Bouchaud et al 2009	Bouchaud JP, Farmer JD, Lillo F, 2009. How markets slowly digest changes in supply and demand. In Handbook of Financial Markets: Dynamics and Evolution, ed. T Hens, KR Schenkhoppe, pp. 57–130. Amsterdam: Elsevier
Brav et al 2005	Brav A, Graham J, Harvey C, Michaely R, 2005. Payout policy in the 21st century. Journal of Financial Economics 77, 483–527.
Berg 2010	van den Berg H, 2010. Mathematical Models of Biological Systems. OUP, Oxford.
Brealey et al 2008	Brealey RA, Myers SC, Allen AJ, 2008. Principles of corporate finance. New York: McGraw-Hill.
Britton 2003	Britton NF, 2003. Essential Mathematical Biology. Springer Verlag.
Burgstaller 1994	Burgstaller A, 1994. Property and Prices. Cambridge: Cambridge University Press
Campbell 2003	Campbell JY, 2003. Consumption-Based Asset Pricing. In G. Constantinides, M. Harris, and R. Stulz (eds), Handbook of the Economics of Finance, North-Holland, Amsterdam, 803–887.

Cassidy 2009	Cassidy J, 2009. How Markets Fail. Allen Lane
Champernowne & Cowell 1998	Champernowne DG, Cowell F, 1998. Economic Inequality and Income Distribution. Cambridge, UK: Cambridge University Press.
Chatterjee et al 2007	Chatterjee A, Sinha S, Chakrabarti B, 2007. Economic inequality: Is it natural? Current Science 92 (10), 1383.
Chatterjee & Chakrabarti 2007	Chatterjee A, Chakrabarti BK, 2007. Kinetic exchange models for income and wealth distributions. The European Physical Journal B 60, 135–149.
Chen & Swan 2008	Chen Z, Swan P, 2008. Liquidity Asset Pricing Model in a Segmented Equity Market. Stock Market Liquidity, ed Lhabitant & Gregoriou, Chapter 23.
Chordia et al 2000	Chordia T, Roll R, Subrahmanyam A, 2000. Commonality in liquidity. Journal of Financial Economics 56, 3–28.
Chordia et al 2001a	Chordia T, Roll R, Subrahmanyam A, 2001. Market liqudity and trading activity. Journal of Finance 56, 501–530.
Chordia et al 2002	Chordia T, Roll R, Subrahmanyam A, 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.
Chordia et al 2005	Chordia T, Sarkar A, Subrahmanyam A, 2005. An empirical analysis of stock and bond market liquidity. Review of Financial Studies 18, 85–129.
Chordia et al 2009	Chordia T, Goyal A, Sadka G, Sadka R, Shivakumar L, 2009. Liquidity and the Post-Earnings-Announcement-Drift. Financial Analysts Journal, 65-4, 18-32.
Clementi Gallegati 2005a	Clementi F, Gallegati M, 2005. Income Inequality Dynamics: Evidence from a Pool of Major Industrialized Countries. Talk at the International Workshop on the Econophysics of Wealth Distributions, Kolkata. http://www.saha.ac.in/cmp/econophysics/abstracts.html
Clementi Gallegati 2005b	Clementi F, Gallegati M, 2005. Power law tails in the Italian personal income distribution. Physica A 350, 427–438.
Cooper 2008	Cooper G, 2008. The origin of financial crises: Central banks, credit bubbles and the efficient market fallacy. London: Harriman House.
Corbett & Jenkinson 1997	Corbett J, Jenkinson T, 1997. How is investment financed? A study of Germany, Japan, UK and US, Manchester School, 65 supplement, 69-93.
Dewar 2005	Dewar RC, 2005. Maximum entropy production and nonequilibrium statistical mechanics. In Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond, edited by A. Kleidon and R. D. Lorenz, Springer Verlag, Heidelberg, Germany.
Dial & Murphy 1995	Dial J, Murphy K, 1995. Incentives, downsizing, and value creation at General Dynamics. Journal of Financial Economics 37, 261–314.
Dragulescu & Yakovenko 2001	Dragulescu A, Yakovenko VM, 2001. Evidence for the exponential distribution of income in the USA. Eur. Phys. J. B 20:585–89
Dynan et al 2004	Dynan KE, Skinner J, Zeldes S, 2004. Do the Rich Save More? Journal of Political Economy, 112, 397-444.
Economist 2005	Economist, 14 May 2005. Sugar-coating the piggy bank. http://www.economist.com/node/3974797
Economist 2009	Economist, 10 December 2010. Fields of automation. http://www.economist.com/node/15048711
Economist 2010a	Economist, 20 February 2010. A different class. http://www.economist.com/node/15544310
Economist 2010b	Economist, 28 August 2010. The miracle of the cerrado. http://www.economist.com/node/16886442
Economist 2010c	The World in 2011, Economist publications.
EIA	US Energy Information Administration. U.S. Dry Natural Gas Proved Reserves. http://www.eia.doe.gov/dnav/ng/hist/rngr11nus_1a.htm
Eiteman & Guthrie 1952	Eiteman WJ, Guthrie GE, 1952. The shape of the average cost curve. American Economic Review 42: 832-838.
Engel & Reid 2006	Engel T, Reid P, 2006. Thermodynamics, Statistical Thermodynamics, and Kinetics. San Francisco, CA: Pearson-Benjamin-Cummings.
EWCO 2010	EWCO, 2010. Working poor in Europe – Norway http://www.eurofound.europa.eu/ewco/studies/tn0910026s/no0910029q.htm
Eyrian 2007	Eyrian, 2007. Convection Cells, Wikipedia http://commons.wikimedia.org/wiki/File:ConvectionCells.svg
Fama & French 1992	Fama EF, French KR, 1992. The cross section of expected stock returns. Journal of Finance 47, 427–465.
Farmer et al 2005	Farmer D, Patelli P, Zovko I, 2005. The predictive power of zero intelligence in financial markets. Proceedings of the National Academy of Sciences of the United States of America, 02(6):2254{2259, 2005.
Ferrero 2010	Ferrero JC, 2010. The individual income distribution in Argentina in the period 2000-2009, A unique source of non stationary data. Arxiv preprint; http://arxiv.org/ftp/arxiv/papers/1006/1006.2057.pdf

Foley 1990	Foley DK, 1990. Recent developments in economic theory. Social Research, vol. 57, no. 3, 666–87
Foley 1996a	Foley DK, Statistical equilibrium in a simple labor market. Metroeconomica, 47(2), 125–147.
Foley 1996b	Foley DK, 1996. Statistical Equilibrium Models in Economics. Presented at Summer Meetings of Econometric Society. Available at: http://cedar.newschool.edu/ foleyd.
Foley 1999	Foley DK, 1999. Statistical Equilibrium in Economics: Method, Interpretation, and an Example. Paper presented at the XII Workshop on "General Equilibrium: Problems, Prospects and Alternatives", Certosa di Pontignano, Siena, Italy, http://homepage.newschool.edu/~foleyd/stateqnotes.pdf
Foley 2002	Foley DK, 2002. Maximum Entropy Exchange Equilibrium. New School for Social Research, New York. http://homepage.newschool.edu/~foleyd/maxentexeq.pdf
Friedman 1962	Friedman M, 1962. Capitalism and Freedom. Chicago: University of Chicago Press.
FT 2010	Financial Times, May 21, 2010. Can't spot an asset bubble? Blogs, Money Supply. http://blogs.ft.com/money-supply/2010/05/21/cant-spot-an-asset-bubble/
FT 2011a	Financial Times, Jan 12, 2011. Let Their Be Light…Pools. Blogs, Alphaville. http://ftalphaville.ft.com/blog/2011/01/12/456666/let-there-be-light-pools/
FT 2011b	Financial Times, Feb 8, 2011. Sugar Body Denounces 'Parasitic' Traders. http://www.ft.com/cms/s/0/d1b5e3ae-33c1-11e0-b1ed-00144feabdc0.html#axzz1HMTcGYtI
FT/S&P 2010	Financial Times, Nov 4, 2010. Mortgage arrears linked to low deposits. http://www.ft.com/cms/s/0/9f0ab4aa-e83c-11df-8995-00144feab49a.html#axzz1GCRHIJ00
Gabaix 2009	Gabaix X, 2009. Power Laws in Economics and Finance. Annual Reviews of Economics, 2009, 1, 255-93.
Gaffeo et al 2003	Gaffeo E, Gallegati M, Palestrini A, 2003. On the size distribution of firms: Additional evidence from G7 countries. Physica A, 324:117–123.
Gallegati et al 2006	Gallegatti M, Keen S, Lux T, Ormerod P, 2006. Worrying trends in econophysics. Physica A 370, 1–6.
gapminder	Rosling H. http://www.gapminder.org/
Georgescu-Roegen 1971	Georgescu-Roegen N, 1971. The Entropy Laws and Economic Progress. Cambridge, MA: Harvard University.
Georgia Tech 2010	Georgia Tech, 2010. ME6601: Introduction to Fluid Mechanics, http://www.catea.gatech.edu/grade/mecheng/mod8/mod8.html
Glazer & Wark 2001	Glazer M, Wark J, 2001. Statistical Mechanics: A Survival Guide. Oxford University Press.
Gollin 2002	Gollin D, 2002. Getting Income Shares Right. Journal of Political Economy 110(2):458-474.
Gould & Tobochnik 2010	Gould H, Tobochnik J, 2010. Statistical and Thermal Physics: With Computer Applications. Princeton University Press.
Goyenko et al 2009	Goyenko R, Holden C, Trzcinka C, 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92 2, 153-181.
Griffiths & Tenenbaum 2006	Griffiths TL, Tenenbaum JB, 2006. Optimal predictions in everyday cognition. Psychological Science, 17, 767–773.
Guardian 2010	Guardian Business, 24 May 2010. Top policymaker says Britain risks copying Japan's lost decade. http://www.guardian.co.uk/business/2010/may/24/kate-barker-bank-complacent-financial-crisis
Harrison 2005	Harrison F, 2005. Boom, Bust, House Prices, Banking and the Depression of 2010. Shepheard Walwyn, London.
Harvie 2000	Harvie D, 2000. Testing Goodwin: growth cycles in ten OECD countries. Cambridge Journal of Economics, 24 (2000), 349-376.
Hayek 1931	Hayek F, 1931. Prices and Production. London: Routledge & Kegan Paul.
Hess & Holzhausen 2008	Hess A, Holzhausen A, 2008. The structure of European mortgage markets. Special Focus: Economy & Markets 01/2008.https://www.allianz.com/static-resources/en/images/wp_europaeische_hypothekenmaerkte_eng.pdf
Hirsch et al 2003	Hirsch MW, Smale S, Devaney R, 2003. Differential Equations, Dynamical Systems, and an Introduction to Chaos. Academic Press, 2 edition.
Homer & Sylla 1996	Homer S & Sylla R, 1996. A history of interest rates. 3rd ed., rev. New Brunswick, N.J., Rutgers University Press.
Hussman	Hussman JP. Weekly Market Comment. http://www.hussmanfunds.com/weeklyMarketComment.html
Hussman 2011	Hussman JP, 4 April 2011. Will the Real Phillips Curve Please Stand Up? http://www.hussmanfunds.com/wmc/wmc110404.htm
Jones 2002	Jones C, 2002. A Century of Stock Market Liquidity and Trading Costs. Working Paper. Columbia University, New York, NY. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=313681
Joulin 2008	Joulin A, Lefevre A, Grunberg D, Bouchaud JP, 2008. Stock price jumps: news and volume

play a minor role. Technical report.
http://arxiv.org/PS_cache/arxiv/pdf/0803/0803.1769v1.pdf

Kaldor 1956	Kaldor N, 1956. Alternative theories of distribution. Review of Economic Studies 23: 94-100.
Keen 1995	Keen S, 1995. Finance and Economic Breakdown: Modeling Minsky. Journal of Post Keynesian Economics, vol.17, pp. 607-635.
Keen 2004	Keen S, 2004. Debunking Economics: The Naked Emperor of the Social Sciences. Zed Books.
Keynes 1936	Keynes JM, 1936. The General Theory of Employment, Interest and Money. London: Macmillan.
Keynes 1939	Keynes JM, 1939. Relative Movements of Real Wages and Output. Economic Journal 49, 34-51.
Kleiber & Kotz 2003	Kleiber C, Kotz S, 2003. Statistical Size Distributions in Economics and Actuarial Sciences. New York: Wiley
Kleidon & Lorenz 2005	Kleidon A, Lorenz RD, (Eds), 2005. Non-equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond. (Understanding Complex Systems) Springer, ISBN-10: 3540224955.
Korajczyk & Sadka 2006	Korajczyk R, Sadka R, 2006. Pricing the Commonality Across Alternative Measures of Liquidity. Working paper, Northwestern University. http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2007-Vienna/Papers/0034.pdf
Korajczyk & Sadka 2005	Korajczyk R, Sadka R, 2005. Are momentum profits robust to trading costs. Journal of Finance 59, 1039–1082.
Kumar 2006	Kumar A, 2006. Lotka Volterra Model. http://www.personal.psu.edu/auk183/LotkaVolterra/LotkaVolterra1.html
Kurz & Salvadori 1995	Kurz HD, Salvadori N, 1995. Theory of Production, A Long-Period Analysis. Cambridge University Press, Cambridge,
Kydland & Prescott 1990	Kydland FE, Prescott EC, 1990. Business Cycles: Real Facts and a Monetary Myth. Federal Reserve Bank of Minneapolis Quarterly Review 14 (Spring 1990), 3–18.
Langlois 1989	Langlois C, 1989. Markup pricing versus marginalism: a controversy revisited. Journal of Post KeynesianEconomics 12: 127-151.
Lee 1999	Lee FS, 1999. Post Keynesian Price Theory. Cambridge: Cambridge University Press.
Lee 2005	Lee KH, 2005. The World Price of Liquidity Risk. Ohio State University working paper. http://www.cob.ohio-state.edu/fin/dice/seminars/LCAPM_lee.pdf
Lettau & Ludvigson 2001	Lettau M, Ludvigson S, 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815-849.
Levy & Solomon 1996	Levy M, Solomon S, 1996. Power Laws are Logarithmic Boltzmann Laws. International Journal of Modern Physics C, 7 (4), 595—600.
Lewis 1954	Lewis WA, 1954. Economic Development with Unlimited Supplies of Labor. The Manchester School of Economic and Social Studies, 22(2): 139-191.
Loster 2006	Loster M, 2006. Total Primary Energy Supply - From Sunlight. http://www.ez2c.de/ml/solar_land_area/
Lotka 1925	Lotka AJ, 1925. Elements of physical biology. Baltimore: Williams & Wilkins Co
Liu 2006	Liu W, 2006. A liquidity-augmented capital asset pricing model. Journal of Financial Economics 82-3, 631-671.
Liu 2009	Liu W, 2009. Liquidity and Asset Pricing: Evidence from Daily Data over 1926 to 2005. Nottingham University Business School Research Paper No. 2009-03 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1345953
Lyons 2001	Lyons RK, 2001. The Microstructure Approach to Exchange Rates, MIT Press, (Cambridge, Massachusetts).
Madhavan 2000	Madhavan A, 2000. Market microstructure: A survey. Journal of Financial Markets 3, 205-258.
Mankiw 2004	Mankiw G, 2004. Principles of Economics (3rd edition). International student edition, Thomson South Western.
Maslow 1954	Maslow AH, 1954. Motivation and Personality. Harper and Row, New York.
Measuring Worth	Williamson SH, Lawrence H. Measuring Worth. http://www.measuringworth.com/index.php
Mehrling 2000	Mehrling P, 2000. What is Monetary Economics About? Barnard College, Columbia University mimeo. http://www.econ.barnard.columbia.edu/faculty/mehrling/papers/What%20is%20Monetary%20Economics%20About.pdf
Mehrling 2010	Mehrling P, 2010. Monetary Policy Implementation: A Microstructure Approach, David Laidler's contributions to Macroeconomics. Ed. Robert Leeson, pp 212-232. Palgrave Macmillan, 2010.
Miles & Scott 2002	Miles D, Scott A, 2002. Macroeconomics, Understanding the Wealth of Nations. John Wiley and Sons.

Mirowski 1989	Mirowski P, 1989. More Heat Than Light: Economics as Social Physics, New York, Cambridge University Press.
Mirowski 2010	Mirowski P, 2010. The Great Mortification: Economists' Responses to the Crisis of 2007–(and counting). http://www.iasc-culture.org/publications_article_2010_Summer_mirowski.php
Mitzenmacher 2004	Mitzenmacher M, 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. Internet Mathematics, 1(2):226–251.
Napier 2007	Napier R, 2007. Anatomy of the Bear. Harriman House Ltd.
NEA 2010	Nikkei Electronics Asia, April 2010. Photovoltaic Cells on the Verge of Explosive Growth. http://techon.nikkeibp.co.jp/article/HONSHI/20100326/181377/
Newman 2005	Newman MEJ, 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46:323–351.
New Scientist 2008	New Scientist, 04 March 2008. Shockwave traffic jam recreated for first time. http://www.newscientist.com/article/dn13402-shockwave-traffic-jam-recreated-for-first-time.html
Nirei & Souma 2007	Nirei M, Souma W, 2007. A two factor model of income distribution dynamics. Rev. Income Wealth 53:440–59
Nise 2000	Nise NS, 2000. Control Systems Engineering. 23rd ed. London: John Wiley.
Norton Ariely 201	Norton MI, Ariely D, 2010. Building a Better America – One Wealth Quintile at a Time. Forthcoming in Perspectives on Psychological Science http://ncsjp.org/news/wp-content/uploads/2010/09/Building-a-Better-America.pdf
Noser 2010	Noser G, 2010. Whose Stock Market is this Anyway. Institutional Investor. http://www.institutionalinvestor.com/Article.aspx?ArticleID=2652254
ONS 2003	Office for National Statistics, 2003. New Earnings Survey. Stationary Office Books, http://www.statistics.gov.uk/statbase/Product.asp?vlnk=5749&More=Y
ONS 2004	Office for National Statistics, 2004. Housebuilding completions: by sector: Social Trends 34. http://www.statistics.gov.uk/StatBase/ssdataset.asp?vlnk=7317&More=Y
Ozawa et al 2003	Ozawa H, Ohmura A, Lorenz RD, Pujol T, 2003. The second law of thermodynamics and the global climate system – A review of the maximum entropy production principle, Rev Geophys 41: 1018.
Pareto 1896	Pareto V, 1896. Cours d'economie politique. Reprinted as a volume of Oeuvres Completes (Droz. Geneva, 1896-1965)
Pastor & Stambaugh 2003	Pastor L, Stambaugh R, 2003. Liquidity risk and expected stock returns. Journal of Political Economy 113, 642–685.
Pepper & Oliver 2006	Pepper G, Oliver M, 2006. The Liquidity Theory of Asset Prices. Wiley Finance.
Peterson 1993	Peterson I, 1993. Newton's Clock. WH Freeman & Co Ltd.
Pettis 2001	Pettis M, 2001. The Volatility Machine: Emerging Economies and the Threat of Financial Collapse. Oxford University Press.
Porter 2008	Porter R, 2008. The Multiple Dimensions of Market-Wide Liquidity. Implications for Asset Pricing. Stock Market Liquidity, ed Lhabitant & Gregoriou, 2008, Chapter 21.
Ranaldo 2008	Ranaldo A, 2008. Intraday Market Dynamics Around Public Information Arrivals. Stock Market Liquidity, ed Lhabitant & Gregoriou, 2008, Chapter 11.
Reed & Hughes 2002	Reed WJ & Hughes BD, 2002. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. Phys. Rev. E 66, 067103.
Reinhart & Rogoff 2009	Reinhart CM, Rogoff K, 2009. This Time is Different: Eight Centuries of Financial Folly. Princeton: Princeton University Press.
Ricardo 1817	Ricardo D, 1817 (1951) . On the Principles of Political Economy and Taxation. The Works and Correspondence of David Ricardo, vol. 1, ed. P. Sraffa, Cambridge: Cambridge U. Pr.
Robbins 1932	Robbins LC, 1932. An Essay on the Nature and Significance of Economic Science. London: Macmillan.
Ruhla 1992	Ruhla C, 1992. The physics of chance: From Blaise Pascal to Niels Bohr. Oxford: Oxford University Press.
Schrödinger 1944	Schrödinger E, 1944. What is Life? Cambridge Univ. Press, Cambridge, 1944.
Shiller 2010	Shiller R, 2010. Robert Shilller Online Data, http://www.econ.yale.edu/~shiller/data.htm
Simkin & Roychowdhury 2006	Simkin MV, Roychowdhury VP, 2006. Re-inventing Willis. preprint, http://arxiv.org/abs/physics/0601192
Sinha 2005	Sinha S, 2005. The Rich Are Different!: Pareto Law from asymmetric interactions in asset exchange models. In: Econophysics of Wealth Distributions, 177-184 eds. A. Chatterjee, B. K. Chakrabarti and S. Yarlagadda, Springer Verlag, New York 2005.
Slanina 2004	Slanina F, 2004. Inelastically scattering particles and wealth distribution in an open economy, Phys. Rev. E 69, n 4, 46102-1-7.
Smith 2010	Smith Y, 4 Oct 2010. Is Liquidity All That It's Cracked Up to Be? http://www.nakedcapitalism.com/2010/10/summer-rerun-is-liquidity-all-that-its-cracked-up-to-

be.html

Smith & Foley 2008 Smith E, Foley DK, 2008. Classical thermodynamics and economic general equilibrium theory. Journal of Economic Dynamics and Control 32, 7–65.

Smithers 2009 Smithers A, 2009. Wall Street revalued: imperfect markets and inept central bankers. John Wiley & Sons

Solomon 2000 Solomon S, 2000. Generalized Lotka Volterra (GLV) models [of stock markets]. In G. Ballot and Weisbuch G., editors, Applications of Simulation to Social Sciences, pages 301–322. Hermes Science, Paris. http://arxiv.org/PS_cache/cond-mat/pdf/9901/9901250v1.pdf

Souma 2001 Souma W, 2001. Universal structure of the personal income distribution. Fractals, 9, 463–470.

Souma & Nirei 2005 Souma W, Nirei M, 2005. Empirical study and model of personal income. In Econophysics of Wealth Distributions, 2005 Chatterjee, Yarlagadda, and Chakrabarti, Springer-Verlag, 2005, pp. 34–42.

St Louis Fed 2004 Pakko M, 2004. Labor's Share, St Louis Fed, National Economic Trends, Aug 2004. http://research.stlouisfed.org/publications/net/20040801/cover.pdf

Stoll 2003 Stoll HR, 2003. Market microstructure. In Constantinides, G., Harris, M. Stulz, R. (Eds.), Handbook of the Economics of Finance. North-Holland, Amsterdam. http://202.194.27.133/jrtzx/uploadfile/pdf/books/handbook/15.pdf

Strogatz 2000 Strogatz SH, 2000. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering. Cambridge, MA: Westview/Perseus Books Group.

Stutzer 2000 Stutzer M, 2000. Simple entropic derivation of a generalised Black-Scholes option pricing model. Entropy, 2, pp. 70–77.

Subramanian 2008 Subramanian A. 2008. What Is China Doing to Its Workers? Peterson Institute for International Economics, Op-ed in the Business Standard, New Delhi. http://www.iie.com/publications/opeds/oped.cfm?ResearchID=885

Telegraph 2010a Conway E, 01 May 2010. Radical tax on debt put to parties. http://www.telegraph.co.uk/finance/economics/7663978/Radical-tax-on-debt-put-to-parties.html

Telegraph 2010b Spencer R, 2010. Welfare reform: What George Osborne and IDS can learn from socialist paradises. http://blogs.telegraph.co.uk/news/richardspencer/100050874/welfare-reform-what-george-osborne-and-ids-can-learn-from-socialist-paradises/

Tribus & McIrvine 1971 Tribus M & McIrvine EC, 1971. Energy and information. Scientific American 224(3), 179-88.

UN 2004 United Nations - World Population Prospects - The 2008 Revision, via wikipedia

UN 2008 United Nations - World Population Prospects - The 2008 Revision http://esa.un.org/unpd/wpp2008/pdf/WPP2008_Highlights.pdf

Ultra Ultra PRT, 2010. ULTra at London Heathrow Airport. http://www.ultraprt.com/heathrow/

Volterra 1926 Volterra V, 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. Mem. R. Accad. Naz. dei Lincei. Ser. VI, vol. 2.

Ward Ward S. Money Moves Markets. http://www.moneymovesmarkets.com/

Ward 2008 Ward S, 2008. A (very) long term look at US equities, Money moves markets. http://www.moneymovesmarkets.com/journal/2008/11/3/a-very-long-term-look-at-us-equities.html

Wilkinson & Pickett 2009 Wilkinson R, Pickett K, 2009. The Spirit Level: Why More Equal Societies Almost Always Do Better. Allen Lane

Willis & Mimkes 2005 Willis G, Mimkes J, 2005. Evidence for the Independence of Waged and Unwaged Income. Presented at the 'Econophysics of Wealth Distributions' International Workshop, Kolkata, India. March 2005. http://xxx.arxiv.cornell.edu/abs/cond-mat/0406694

Willis 2005 Willlis G, 2005. Relieving Poverty by Modifying Income and Wealth Distributions. In Econophysics of Wealth Distributions, Springer Verlag, ISBN 8847003296, Sept 2005.

Woo et al 2010 Woo B, Rademacher I, Meier J, 2010. Upside Down - The $400 Billion Federal Asset-Building Budget. The Annie E. Casey Foundation & CFED, http://www.aecf.org/~/media/Pubs/Initiatives/Family%20Economic%20Success/U/UpsideDownThe400BillionFederalAssetBuildingBudget/033%2010_UpsideDown_final.pdf

Wray 1998 Wray LR, 1998. Understanding Modern Money: The Key to Full Employment and Price Stability. Cheltenham: Edward Elgar.

Wright 2005 Wright I, 2005. The social architecture of capitalism. Physica A 346, 589–620.

Wright 2009 Wright I, 2009. Implicit microfoundations for macroeconomics. Economics (e-journal) 3, 2009-19,

Wyart et al 2008 Wyart M, Bouchaud JP, Kockelkoren J, Potters M, Vettorazzo M, 2008. Relation between bid-ask spread, impact and volatility in double auction markets. Technical report, 2006. http://arxiv.org/PS_cache/physics/pdf/0603/0603084v3.pdf

Young 2010 Young AT, 2010. One of the things we know that ain't so: is US labor's share relatively stable?

Journal of Macroeconomics (forthcoming). Vol 32, 1, March 2010, 90-102.

16. Figures

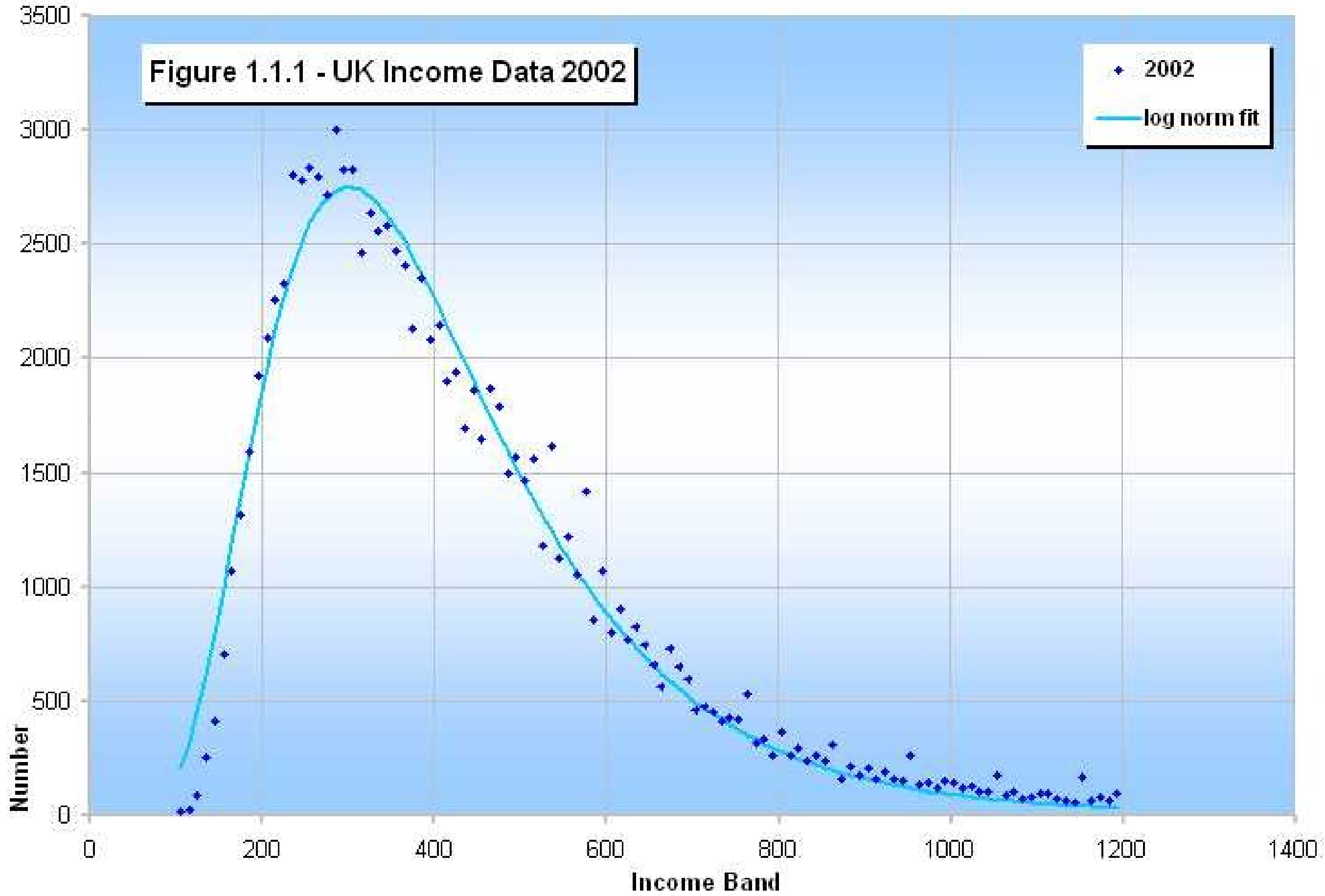


Figure 1.1.1 - UK Income Data 2002
2002
log norm fit
Number
Income Band
0
500
1000
1500
2000
2500
3000
3500
0
200
400
600
800
1000
1200
1400

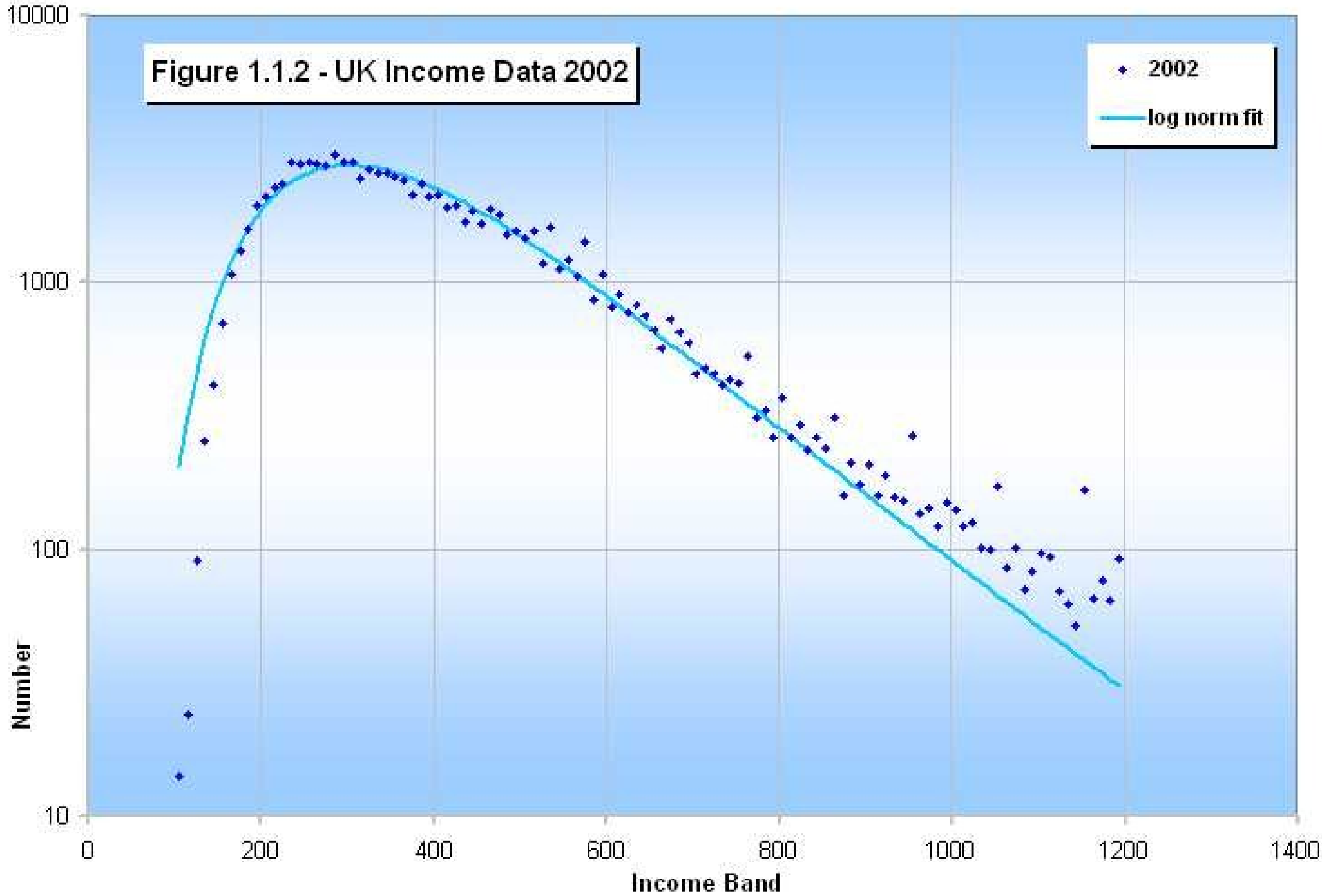


Figure 1.1.2 - UK Income Data 2002
2002
log norm fit
Number
Income Band
10
100
1000
10000
0
200
400
600
800
1000
1200
1400

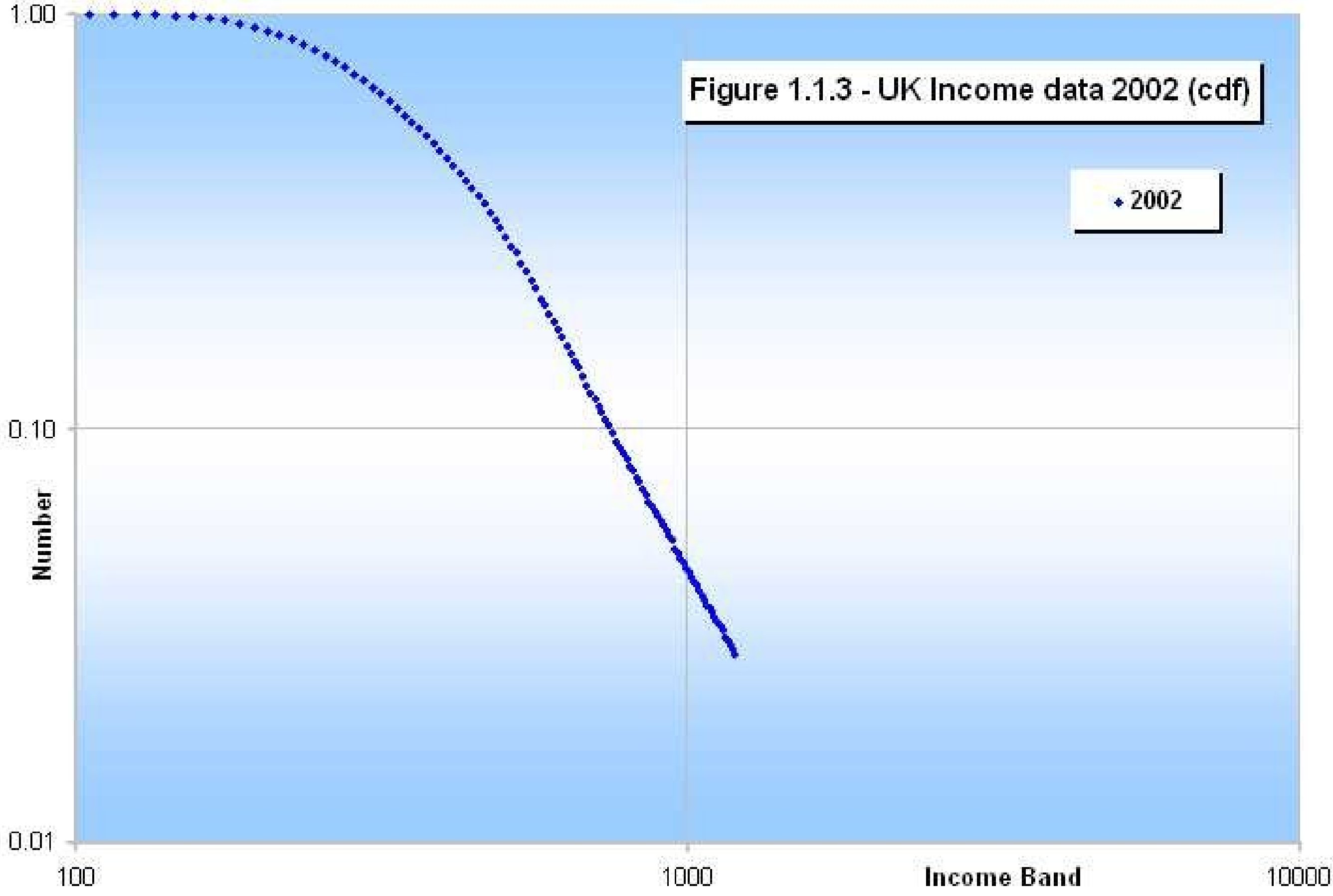


Figure 1.1.3 - UK Income data 2002 (cdf)
2002
1.00
0.10
0.01
Number
100
1000
10000
Income Band

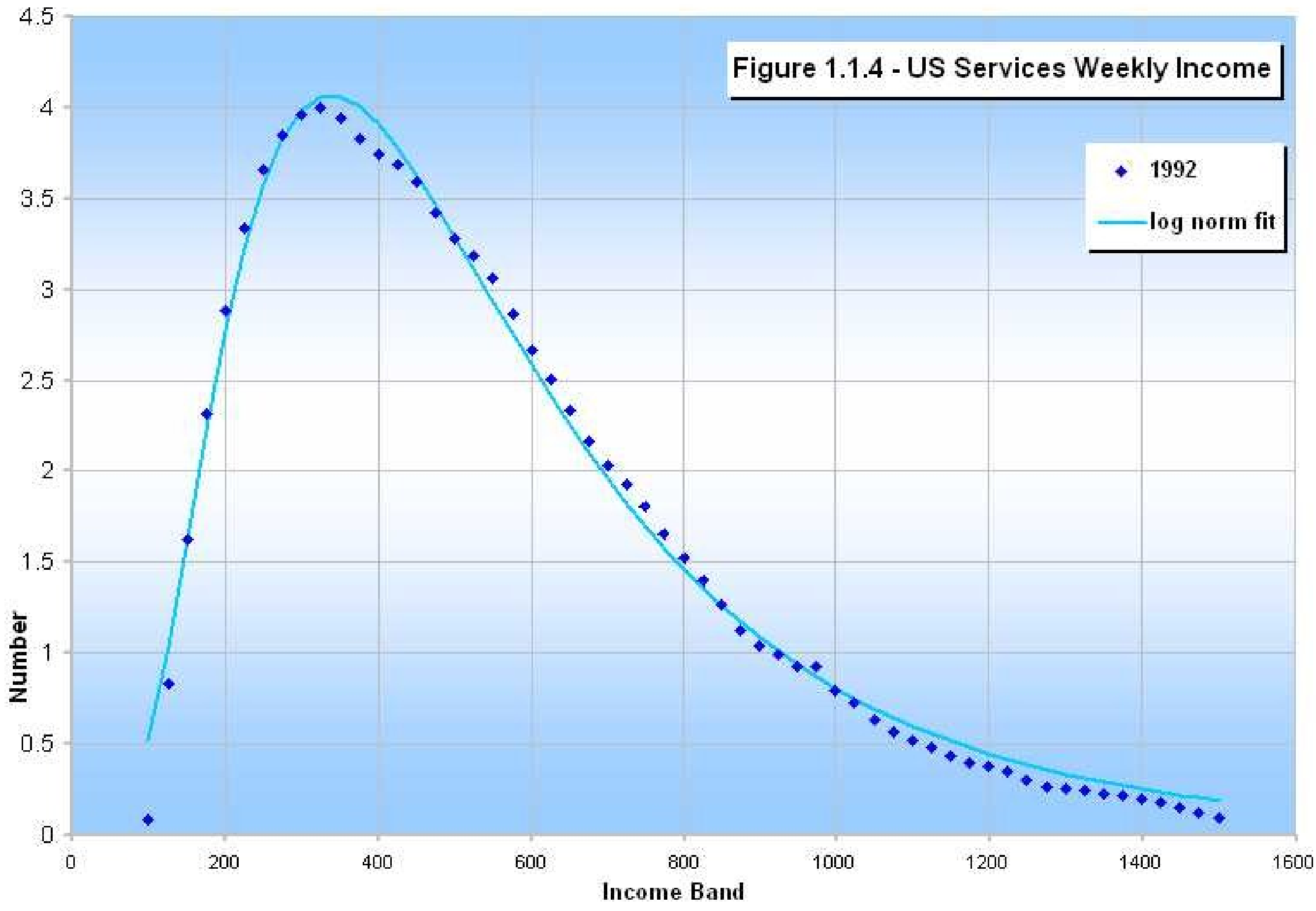


Figure 1.1.4 - US Services Weekly Income
1992
log norm fit
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Number
0
200
400
600
800
1000
1200
1400
1600
Income Band

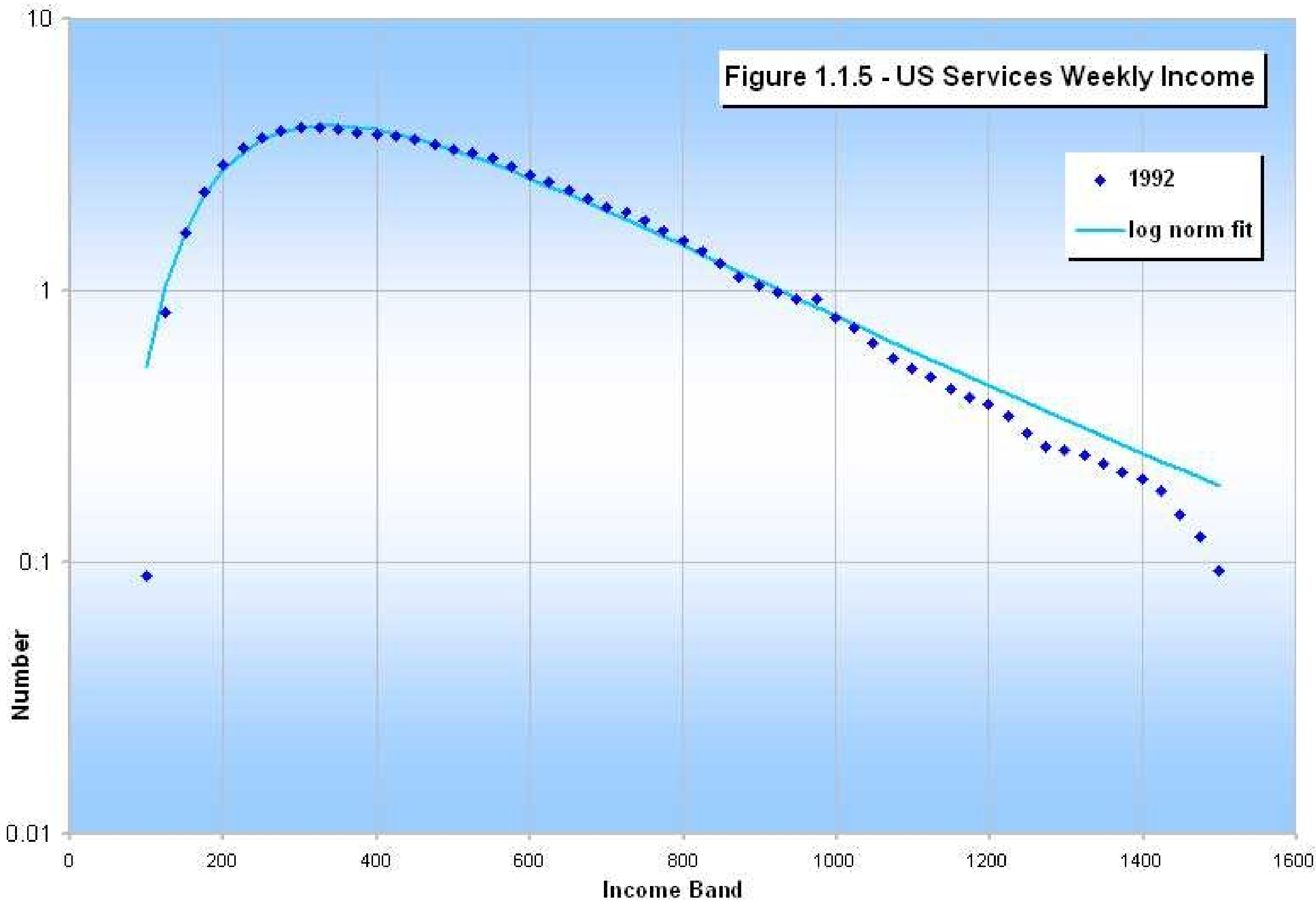


Figure 1.1.5 - US Services Weekly Income
1992
log norm fit
Number
Income Band
10
1
0.1
0.01
0
200
400
600
800
1000
1200
1400
1600

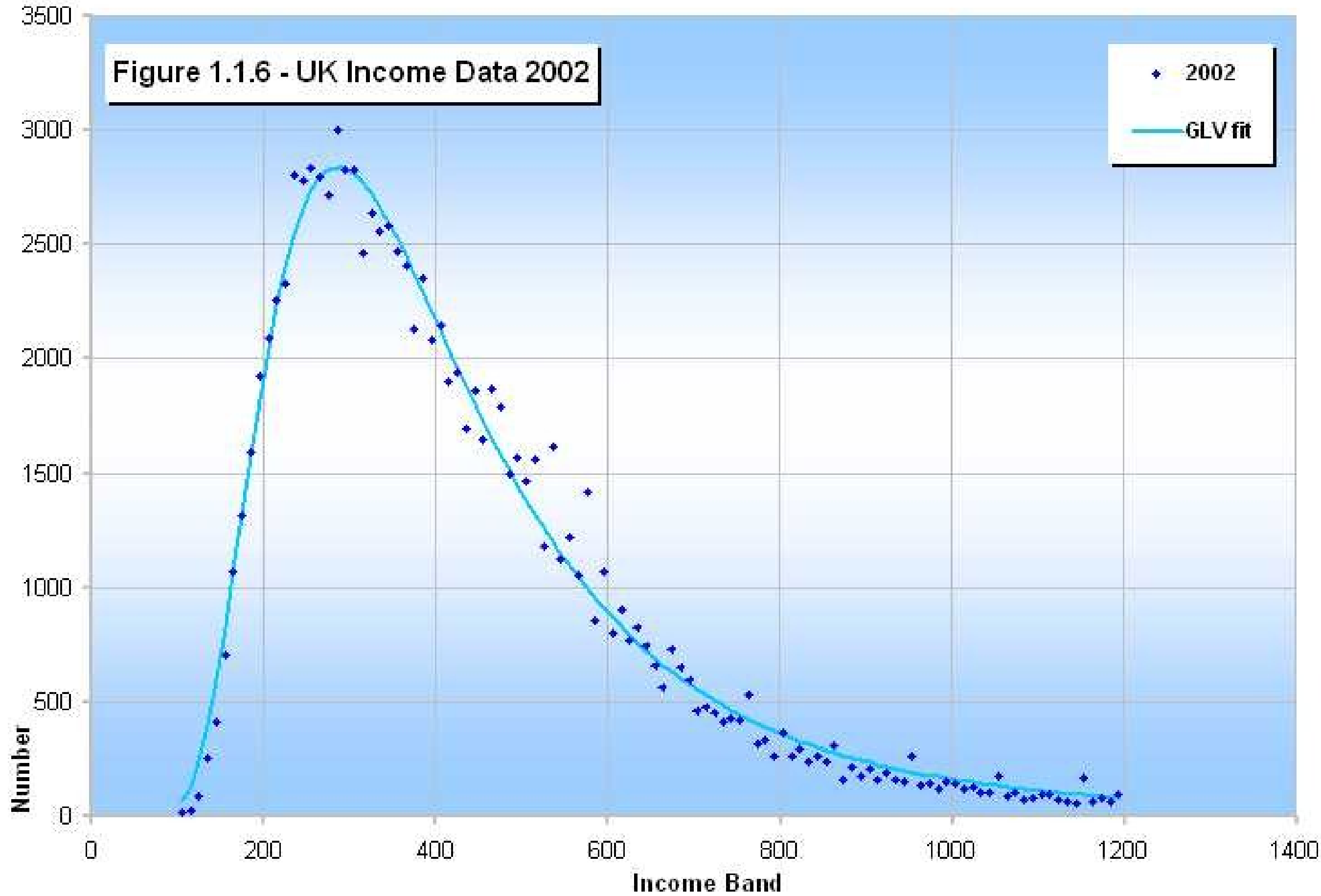


Figure 1.1.6 - UK Income Data 2002
2002
GLV fit
Number
Income Band
3500
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000
1200
1400

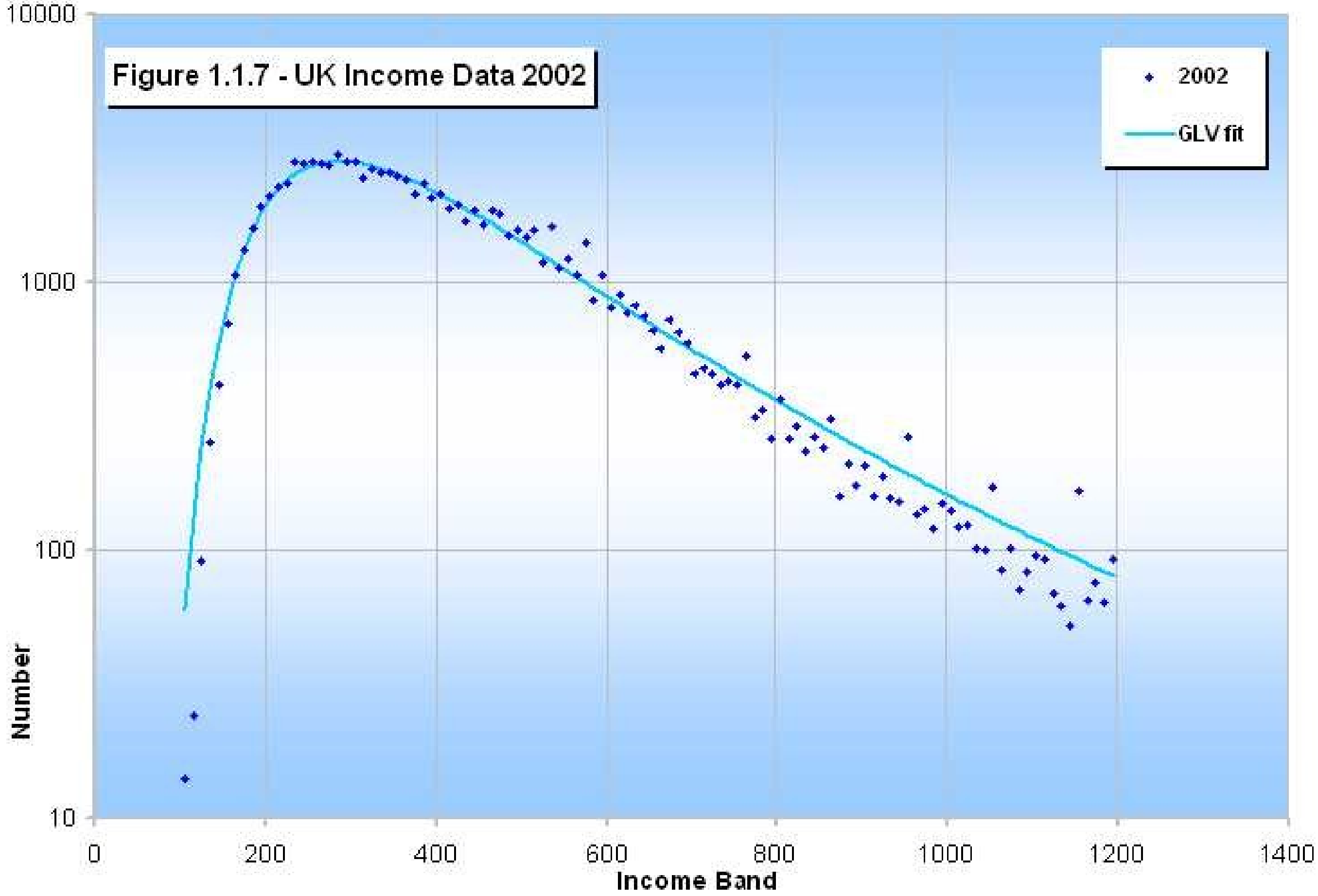
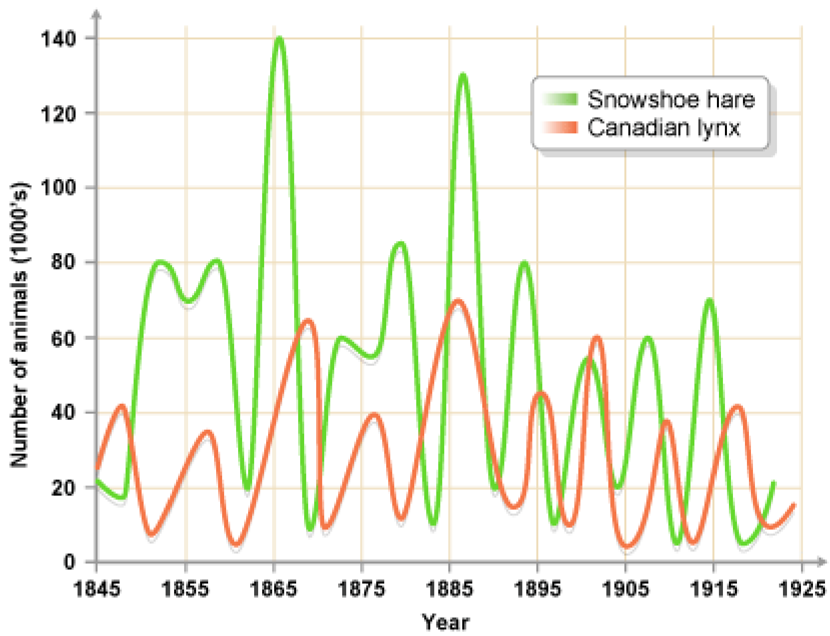


Figure 1.1.7 - UK Income Data 2002
2002
GLV fit
Number
Income Band
10000
1000
100
10
0
200
400
600
800
1000
1200
1400



Snowshoe hare
Canadian lynx
Number of animals (1000's)
Year
140
120
100
80
60
40
20
0
1845
1855
1865
1875
1885
1895
1905
1915
1925

Figure 1.2.1.1

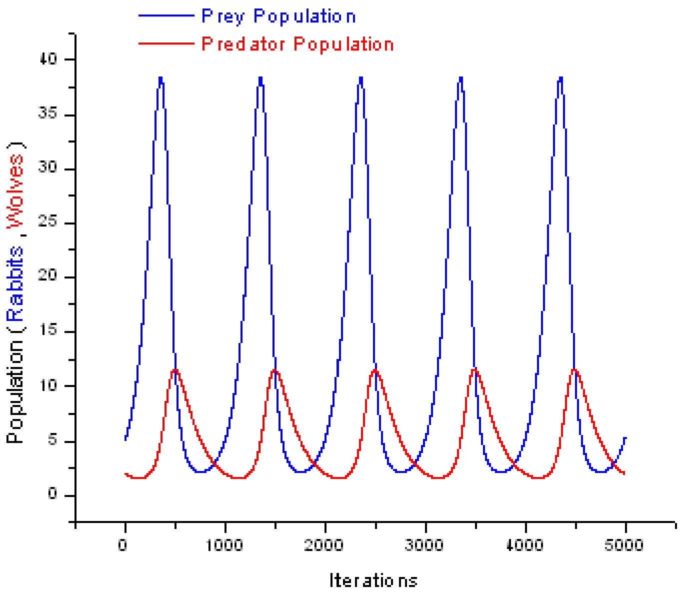


Figure 1.2.1.2

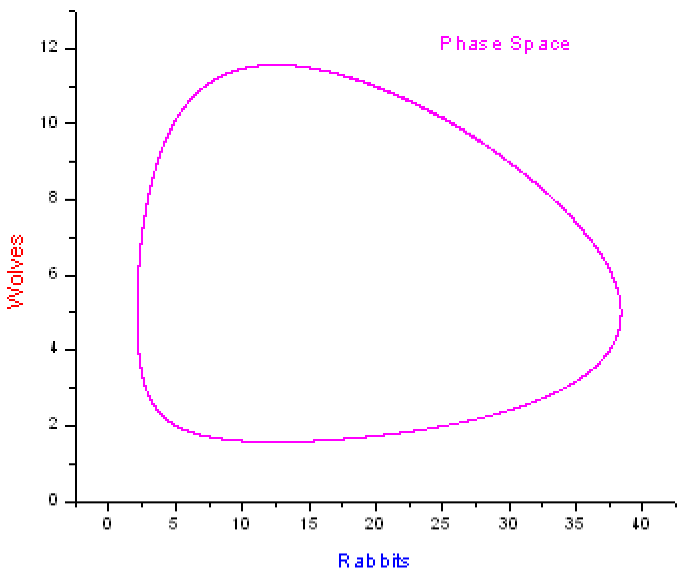


Figure 1.2.1.3

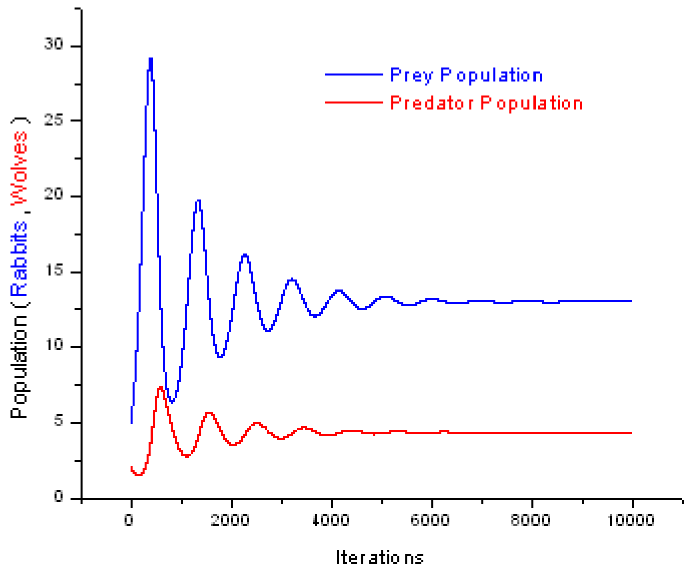


Figure 1.2.1.4

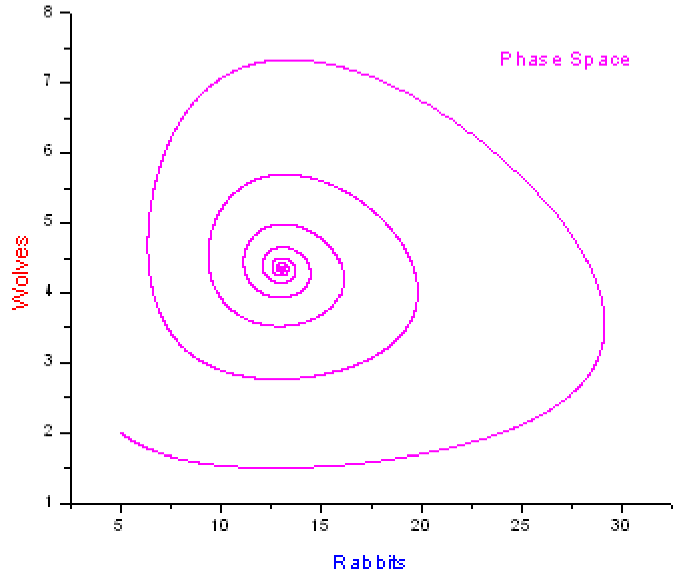


Figure 1.2.1.5

Figure 1.3.1

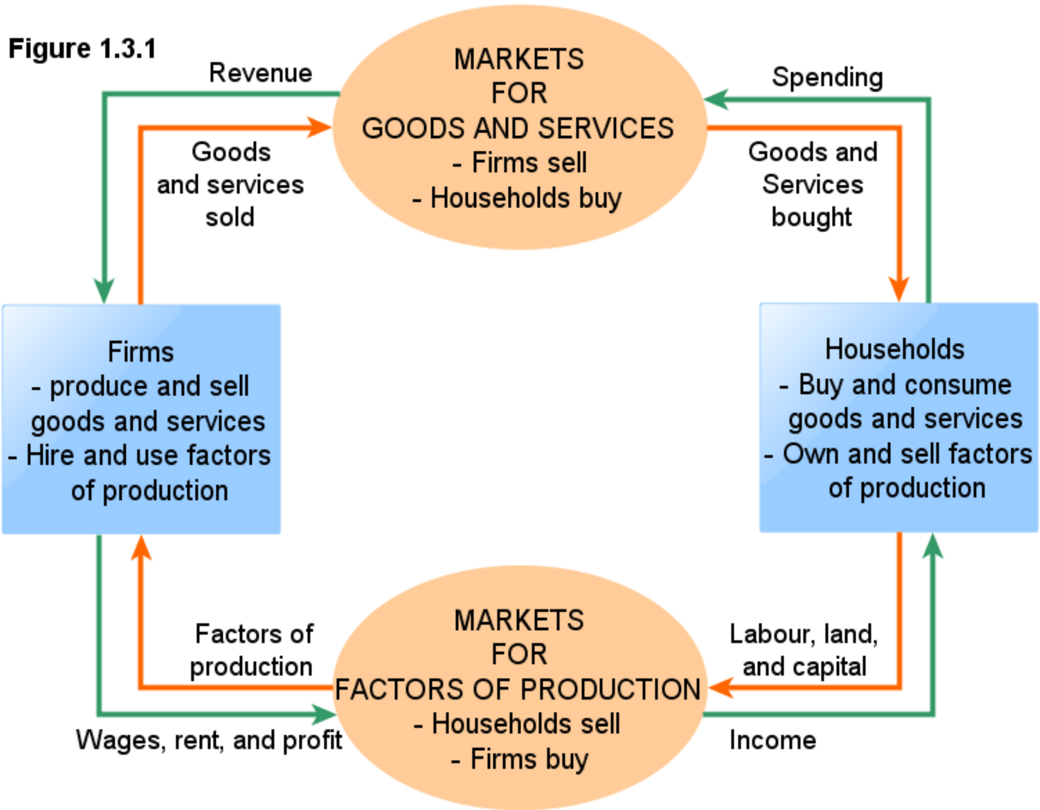


Figure 1.3.2

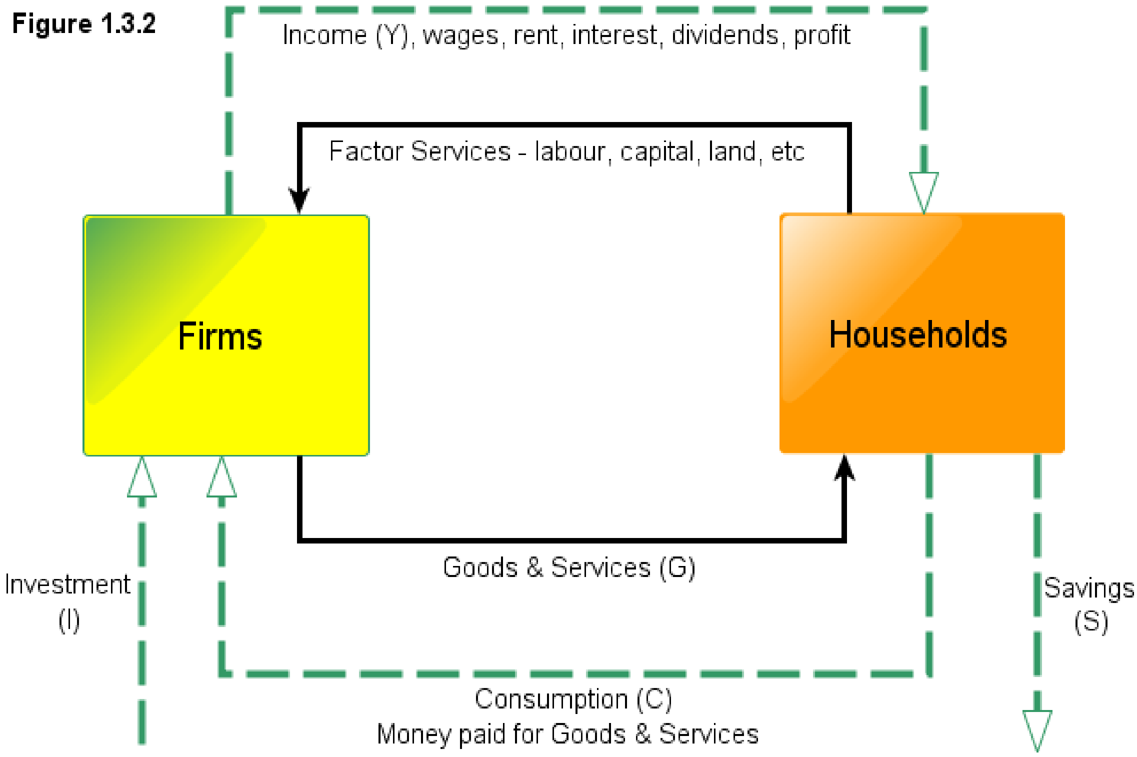


Figure 1.3.3

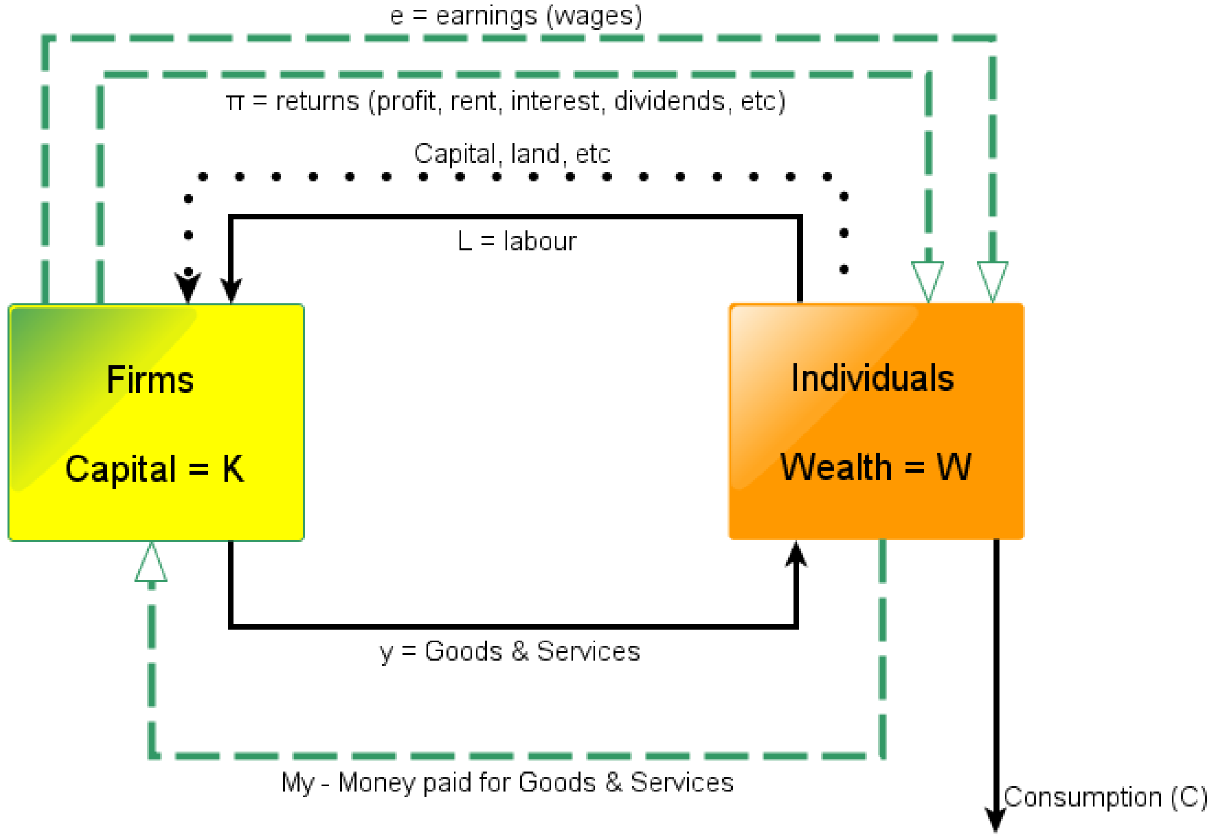


TABLE 14.1 The Financing of Investment: Flow-of-funds Estimated (%) (1970–1994)

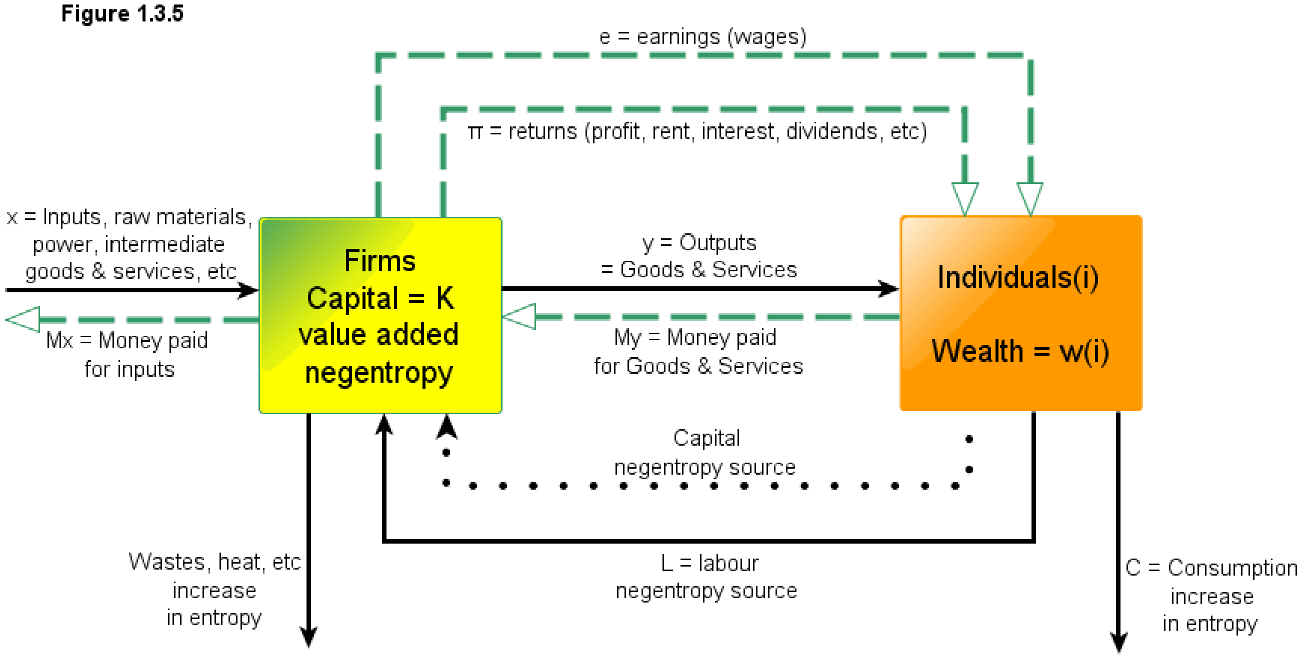
	Germany	Japan	UK	USA
Internal finance	78.4	69.9	95.6	94.0
Bank finance	12.0	30.1	15.0	12.8
Bond finance	−1.0	3.4	3.8	15.3
New equity	−0.02	3.4	−5.3	−6.1
Other	10.6	−6.8	−9.1	−16.0

Note: Internal finance comprises retained earnings and depreciation. The other category includes trade credit and capital transfers. The figures represent weighted averages where the weights for each country are the level of real fixed investment in each year in that country.

Source: Corbett and Jenkinson, "How Is Investment Financed?" *The Manchester School* (1996) vol. LXV, pp. 69–94.

Figure 1.3.4

Figure 1.3.5



e = earnings (wages)
π = returns (profit, rent, interest, dividends, etc)
x = Inputs, raw materials, power, intermediate goods & services, etc
Mx = Money paid for inputs
Firms
Capital = K
value added
negentropy
y = Outputs
= Goods & Services
My = Money paid
for Goods & Services
Individuals(i)
Wealth = w(i)
Capital
negentropy source
Wastes, heat, etc
increase
in entropy
L = labour
negentropy source
C = Consumption
increase
in entropy

Figure 1.3.6

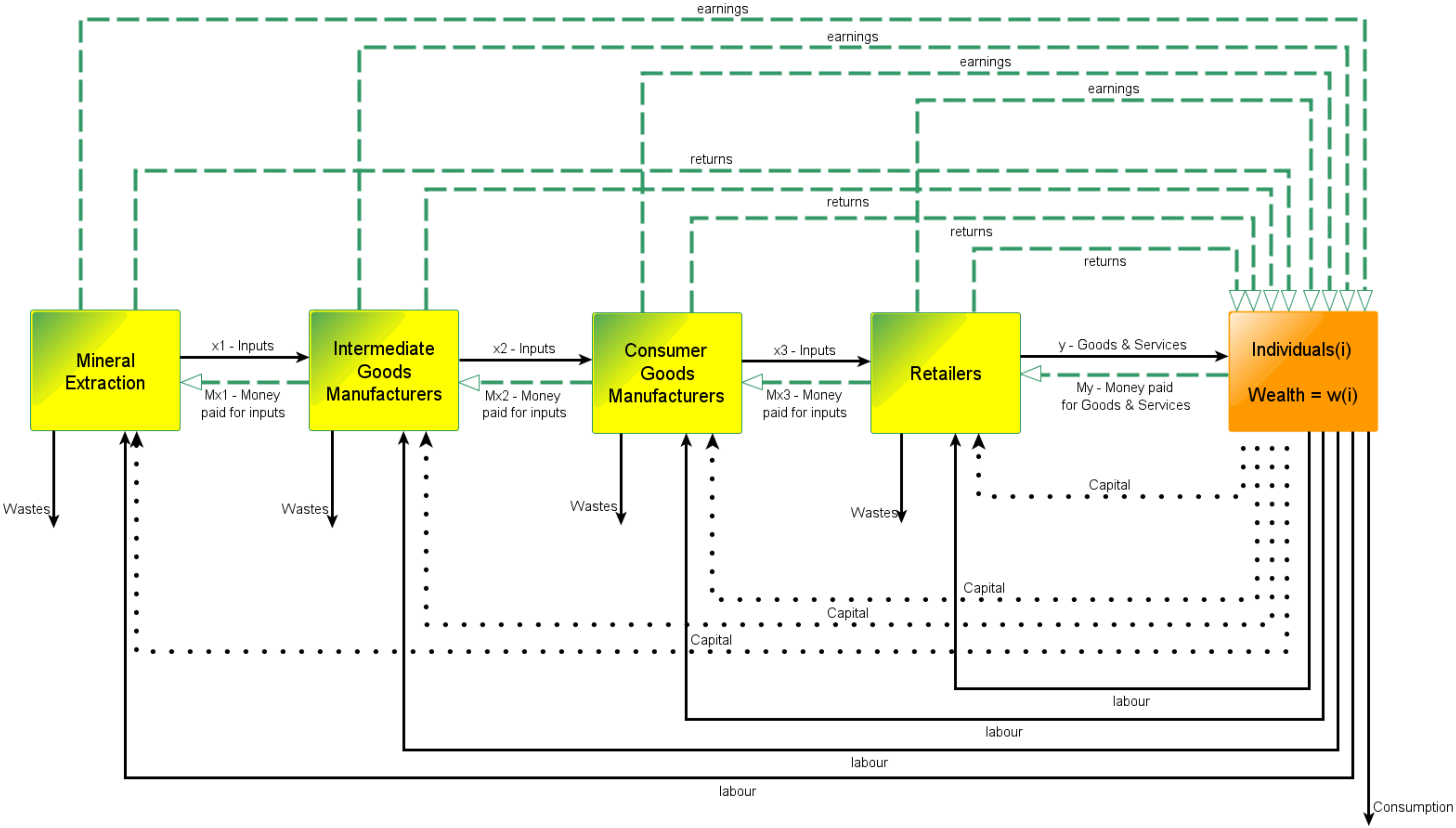
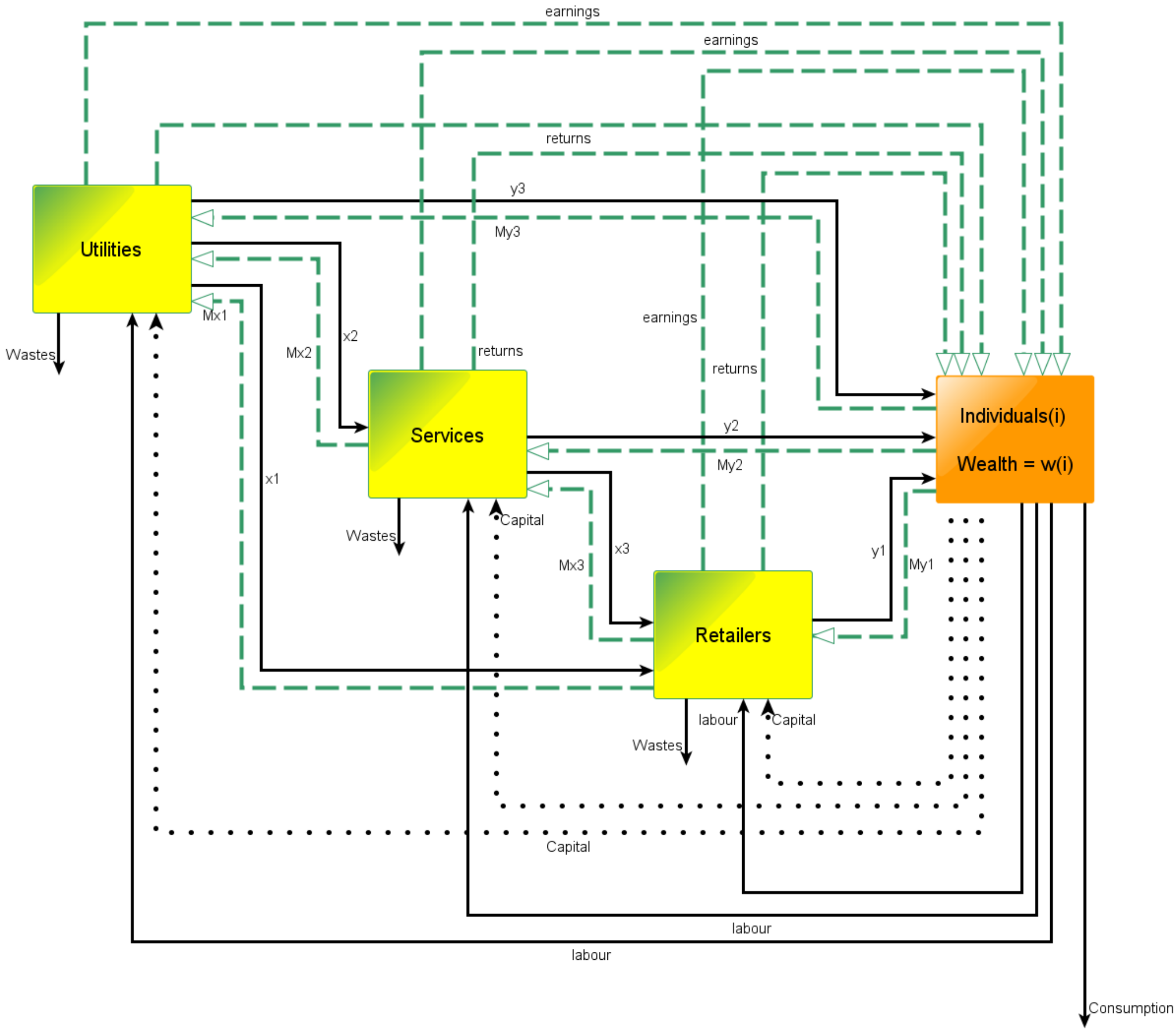

earnings
earnings
earnings
earnings
returns
returns
returns
returns
Mineral Extraction
Intermediate Goods Manufacturers
Consumer Goods Manufacturers
Retailers
Individuals(i)
Wealth = w(i)
x1 - Inputs
x2 - Inputs
x3 - Inputs
y - Goods & Services
Mx1 - Money paid for inputs
Mx2 - Money paid for inputs
Mx3 - Money paid for inputs
My - Money paid for Goods & Services
Wastes
Wastes
Wastes
Wastes
Capital
Capital
Capital
Capital
labour
labour
labour
labour
Consumption

Figure 1.3.7



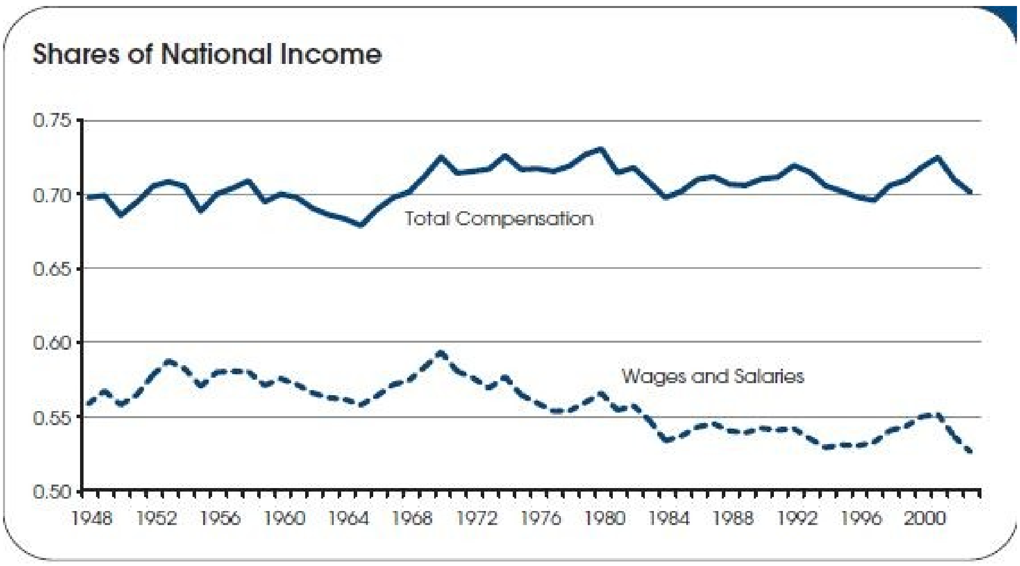


Figure 1.3.8

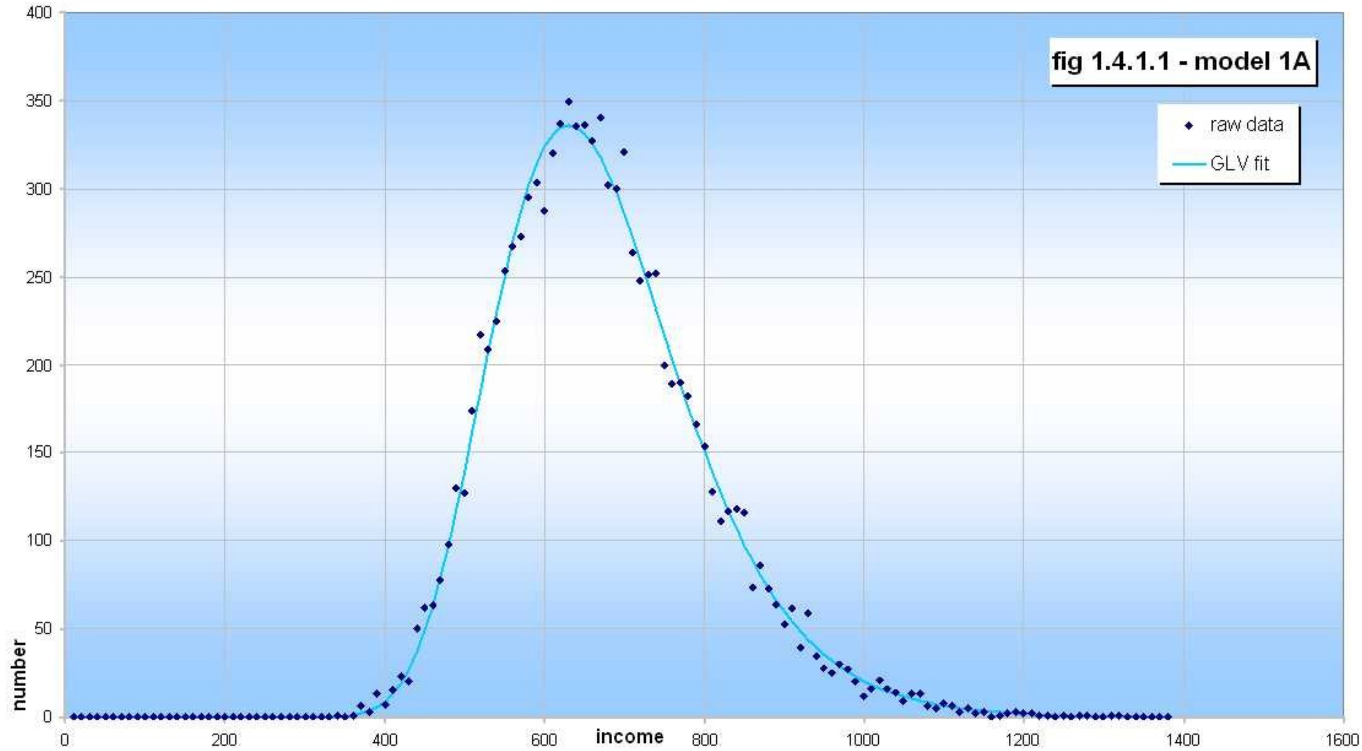


fig 1.4.1.1 - model 1A
raw data
GLV fit
number
income
0
50
100
150
200
250
300
350
400
0
200
400
600
800
1000
1200
1400
1600

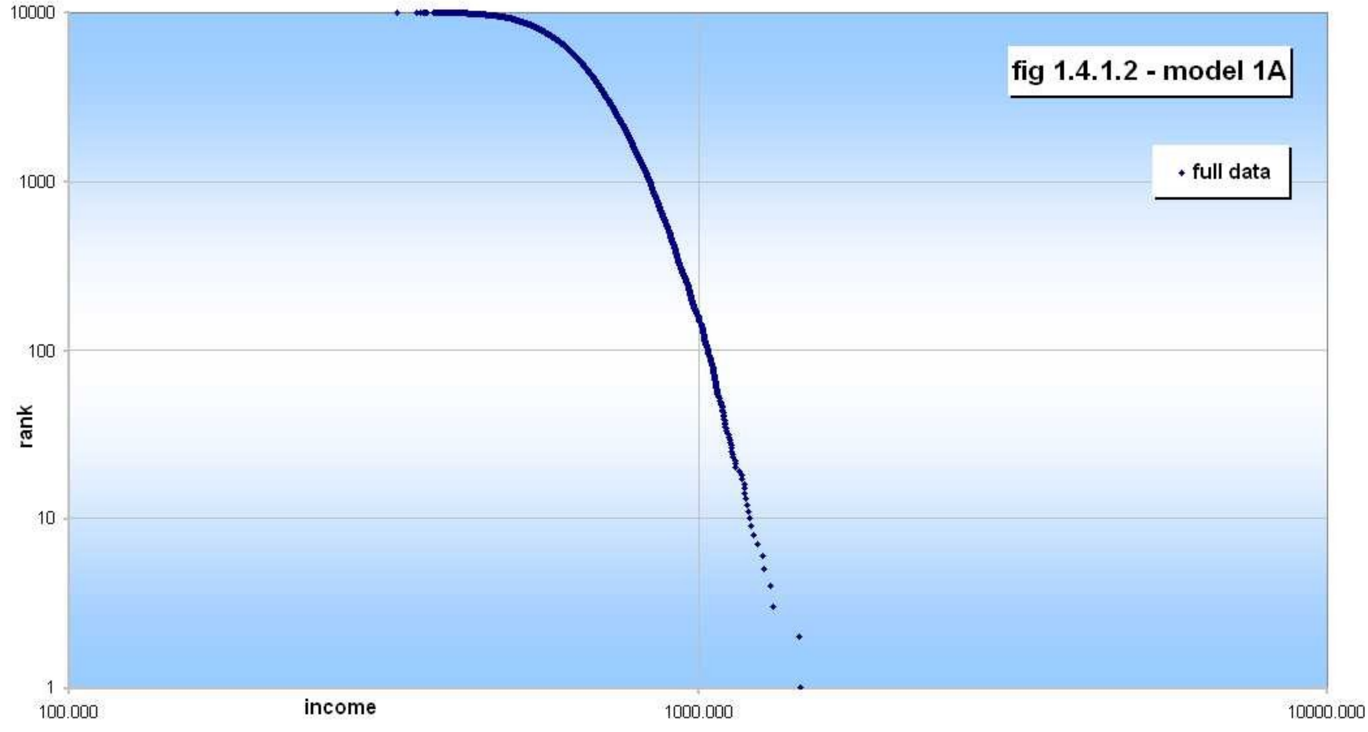


fig 1.4.1.2 - model 1A
full data
rank
income
1
10
100
1000
10000
100.000
1000.000
10000.000

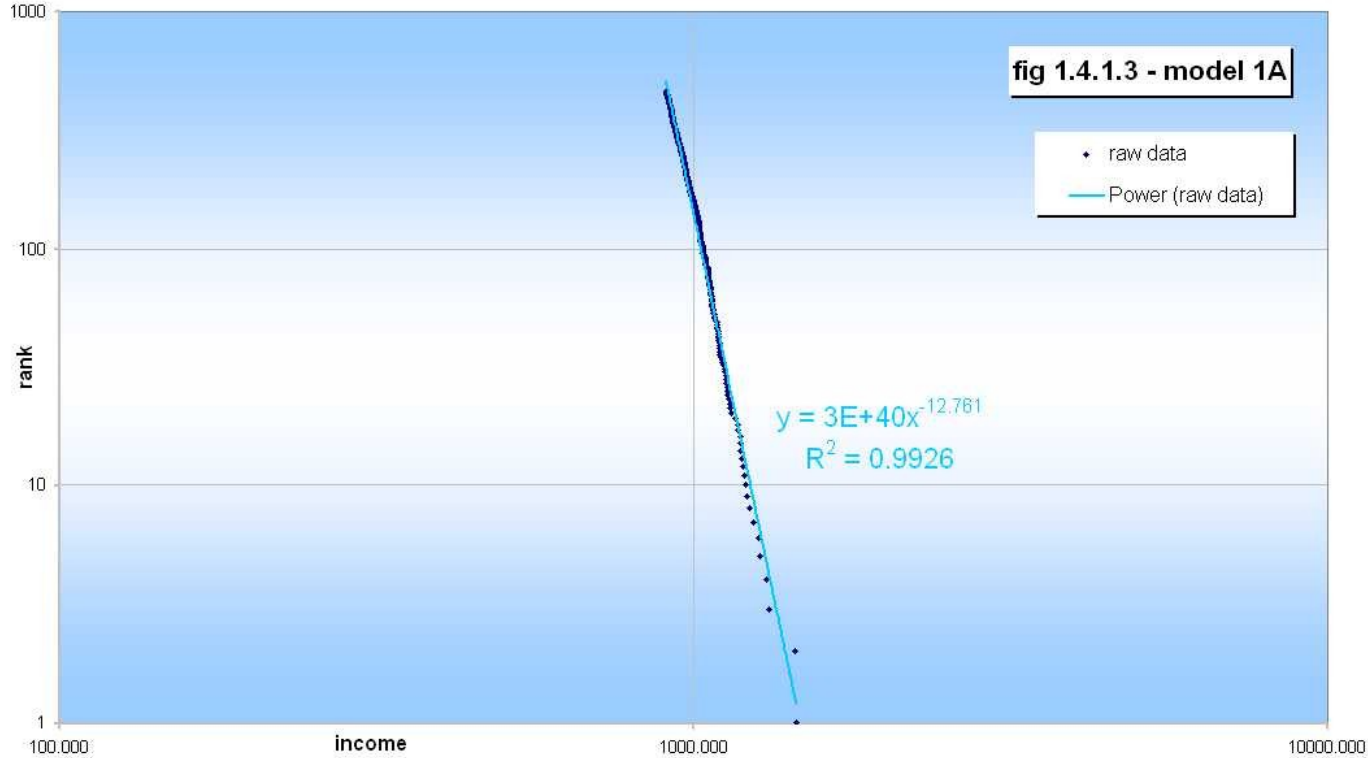


fig 1.4.1.3 - model 1A
raw data
Power (raw data)
y = 3E+40x^-12.761
R^2 = 0.9926
rank
income
1000
100
10
1
100.000
1000.000
10000.000

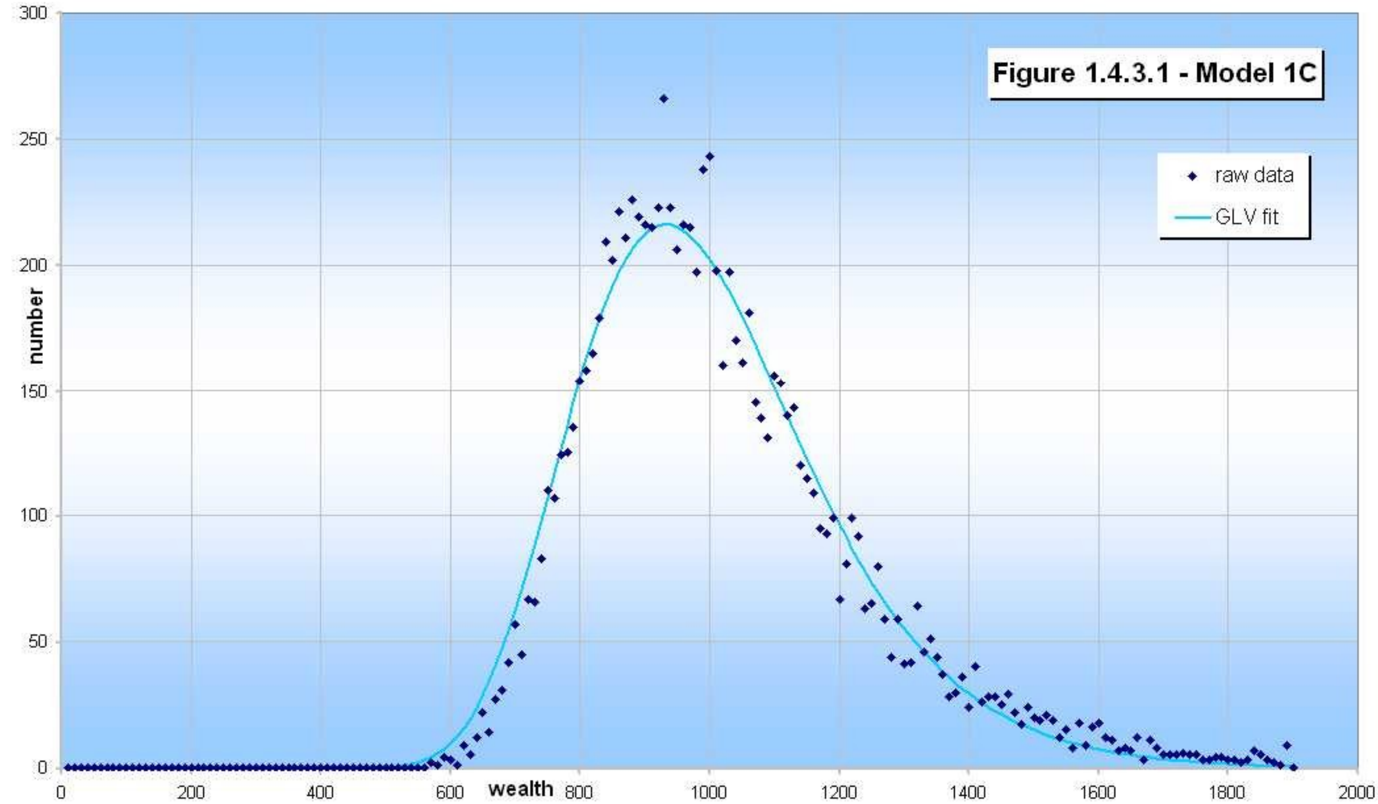


Figure 1.4.3.1 - Model 1C
raw data
GLV fit
number
wealth
300
250
200
150
100
50
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000

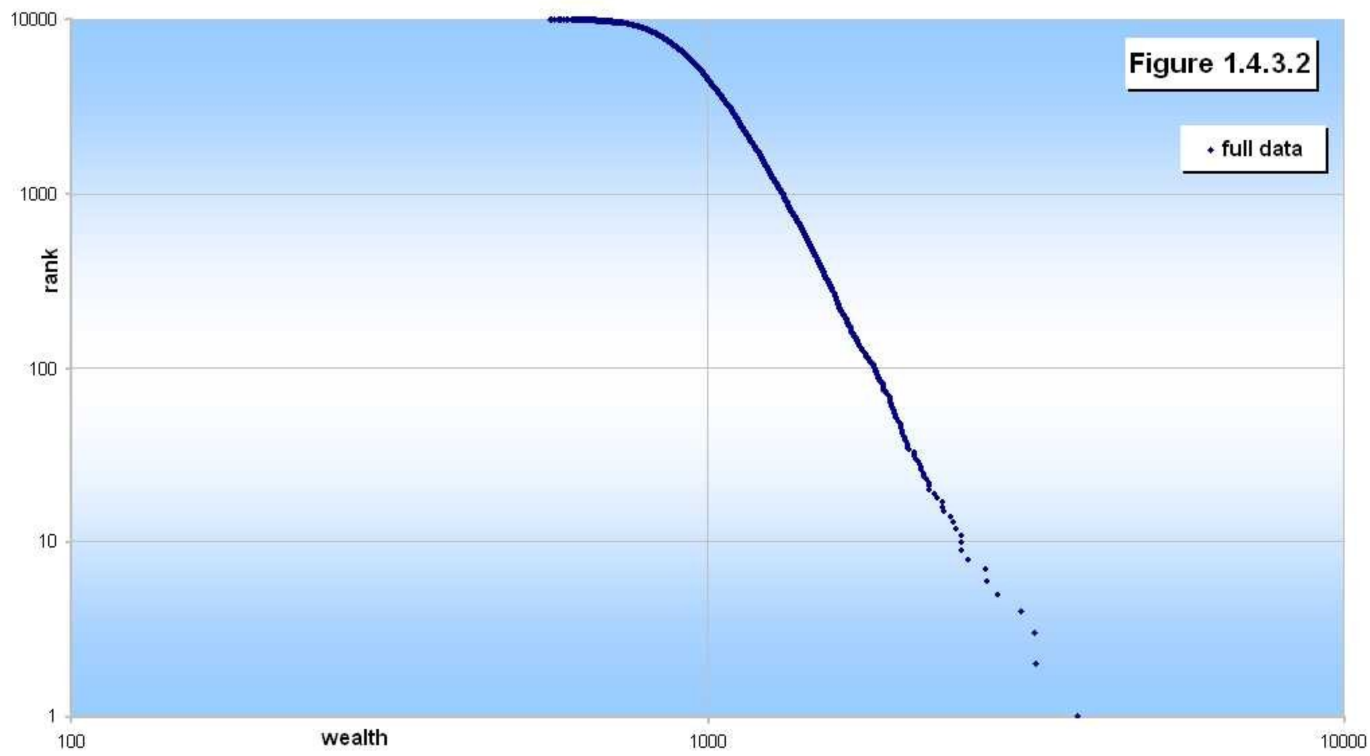


Figure 1.4.3.2
full data
rank
wealth
10000
1000
100
10
1
100
1000
10000

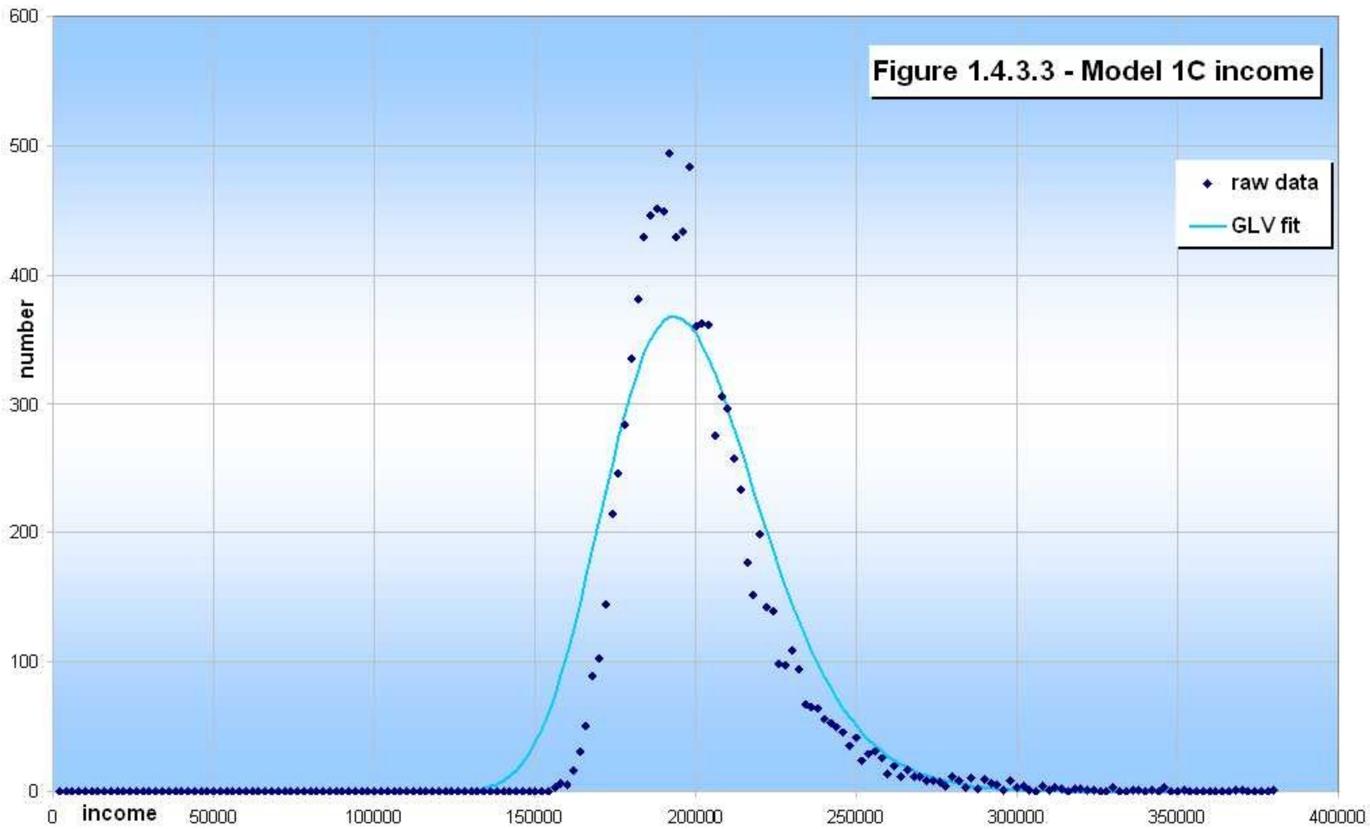


Figure 1.4.3.3 - Model 1C income
raw data
GLV fit
number
income
600
500
400
300
200
100
0
0
50000
100000
150000
200000
250000
300000
350000
400000

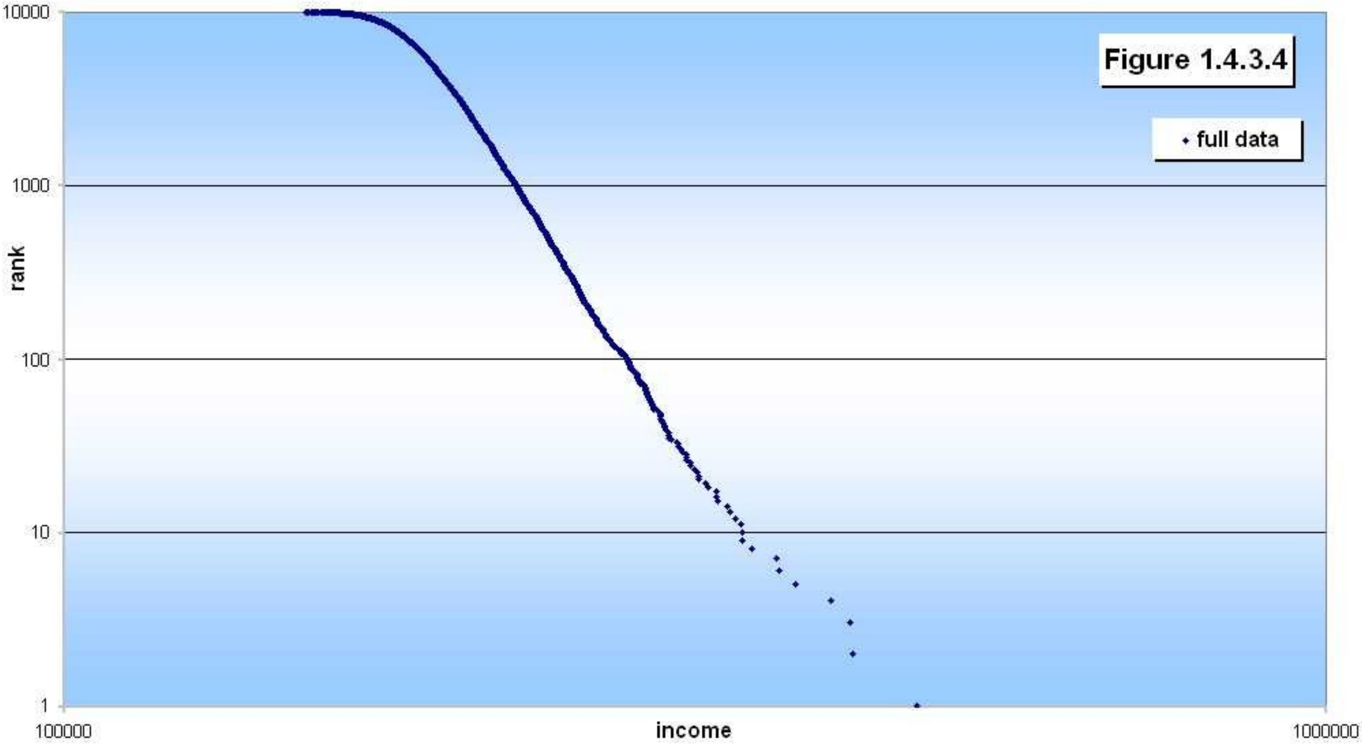


Figure 1.4.3.4
full data
rank
income
10000
1000
100
10
1
100000
1000000

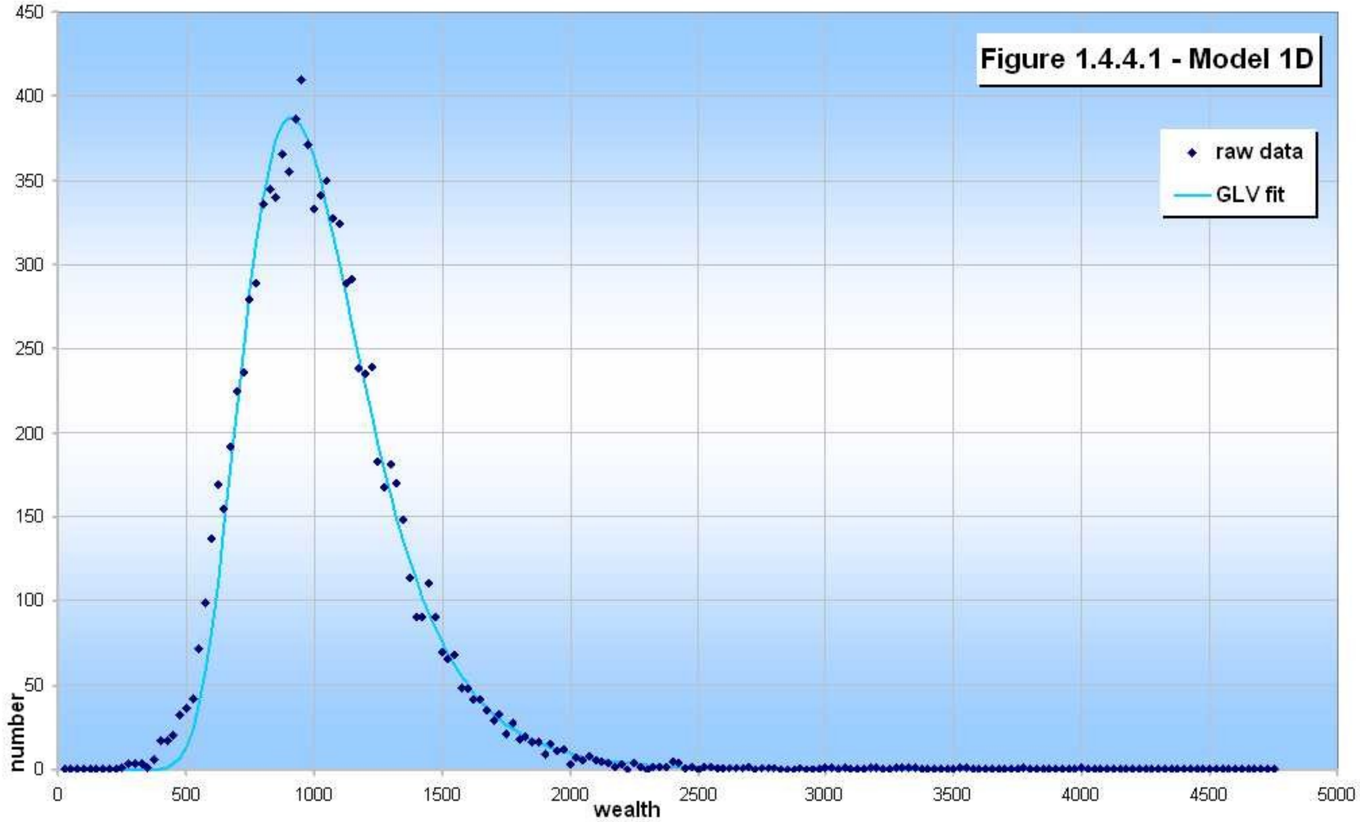


Figure 1.4.4.1 - Model 1D
raw data
GLV fit
number
wealth
450
400
350
300
250
200
150
100
50
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

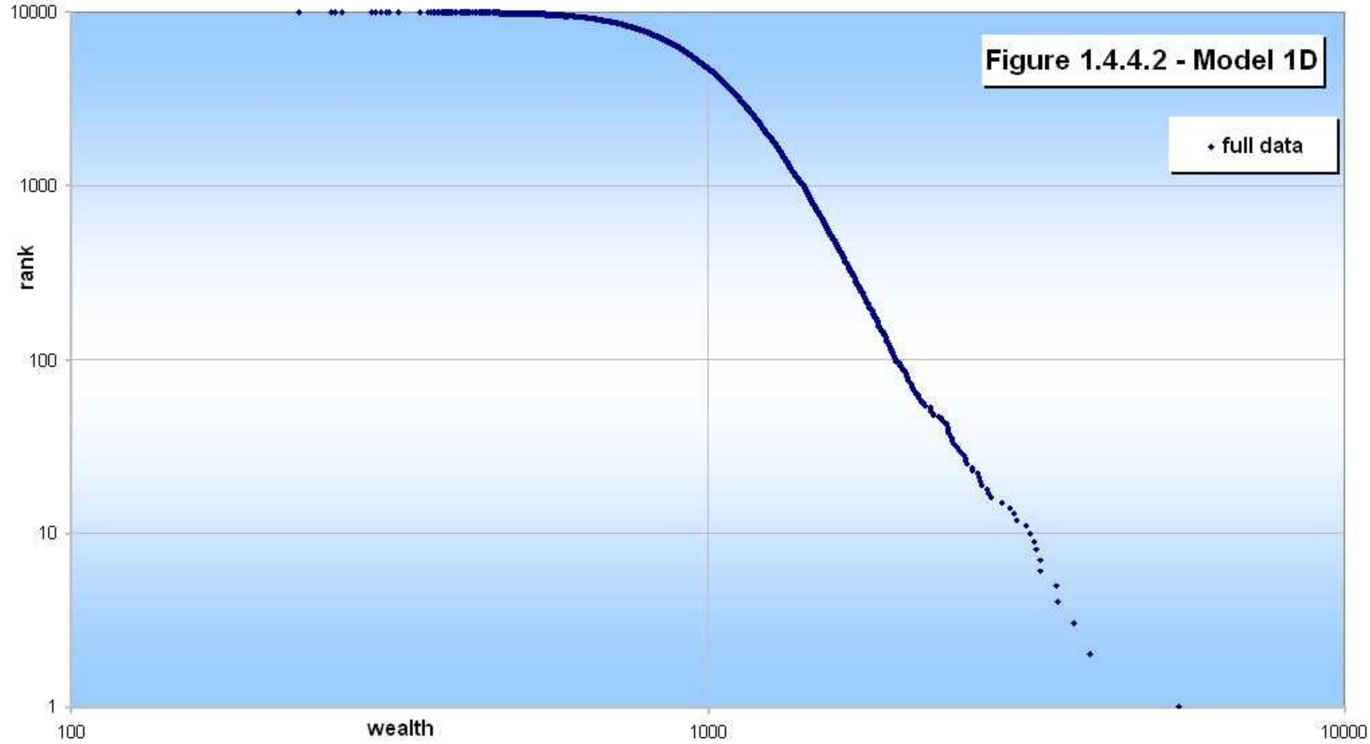


Figure 1.4.4.2 - Model 1D
full data
rank
wealth
10000
1000
100
10
1
100
1000
10000

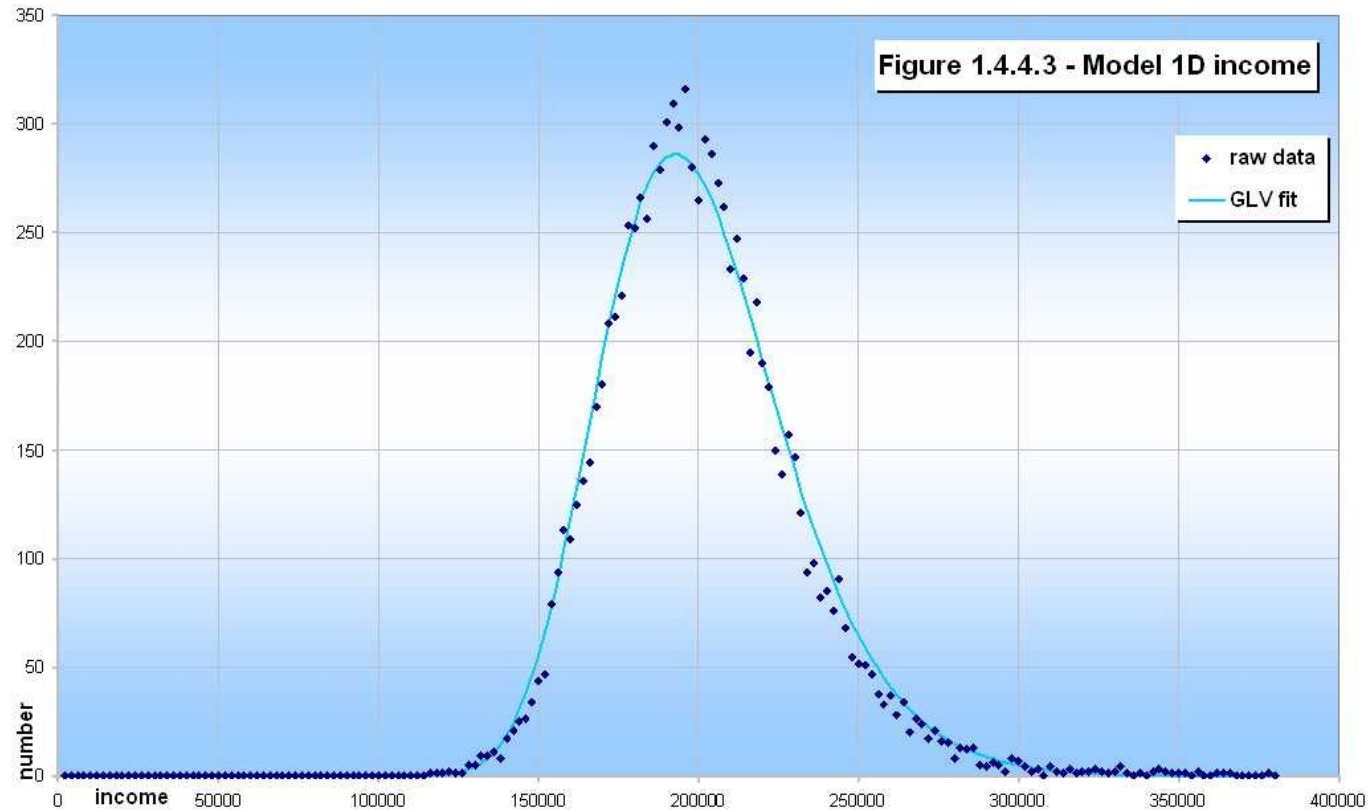


Figure 1.4.4.3 - Model 1D income
raw data
GLV fit
number
income
350
300
250
200
150
100
50
0
0
50000
100000
150000
200000
250000
300000
350000
400000

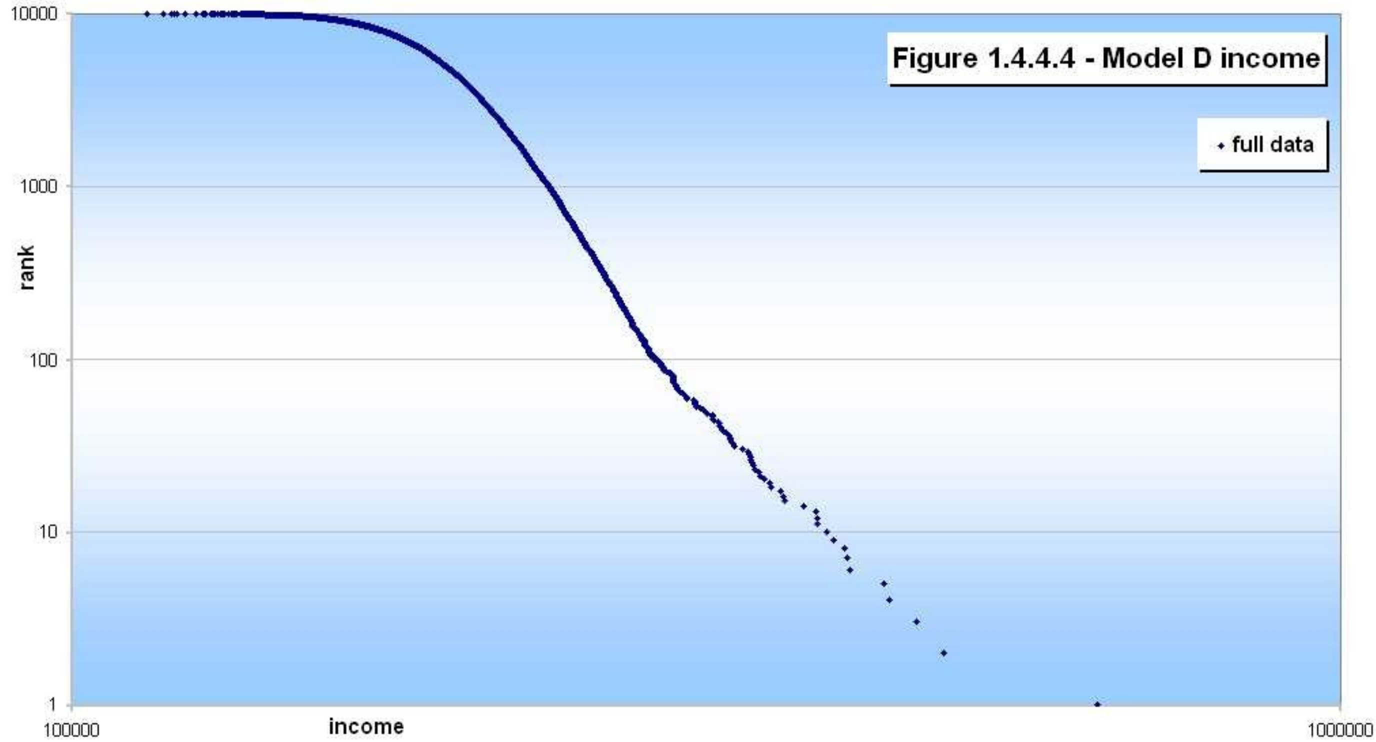


Figure 1.4.4.4 - Model D income
full data
rank
income
10000
1000
100
10
1
100000
1000000

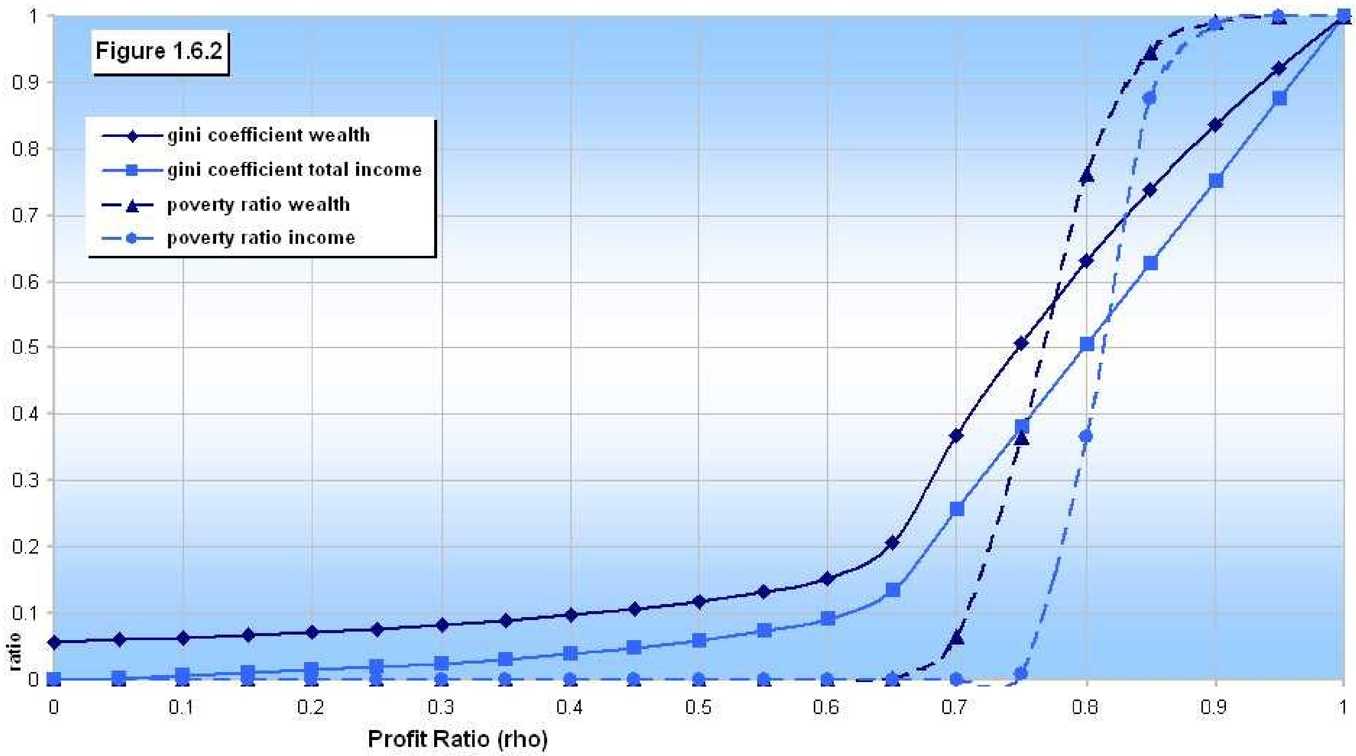


Figure 1.6.2
gini coefficient wealth
gini coefficient total income
poverty ratio wealth
poverty ratio income
ratio
Profit Ratio (rho)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

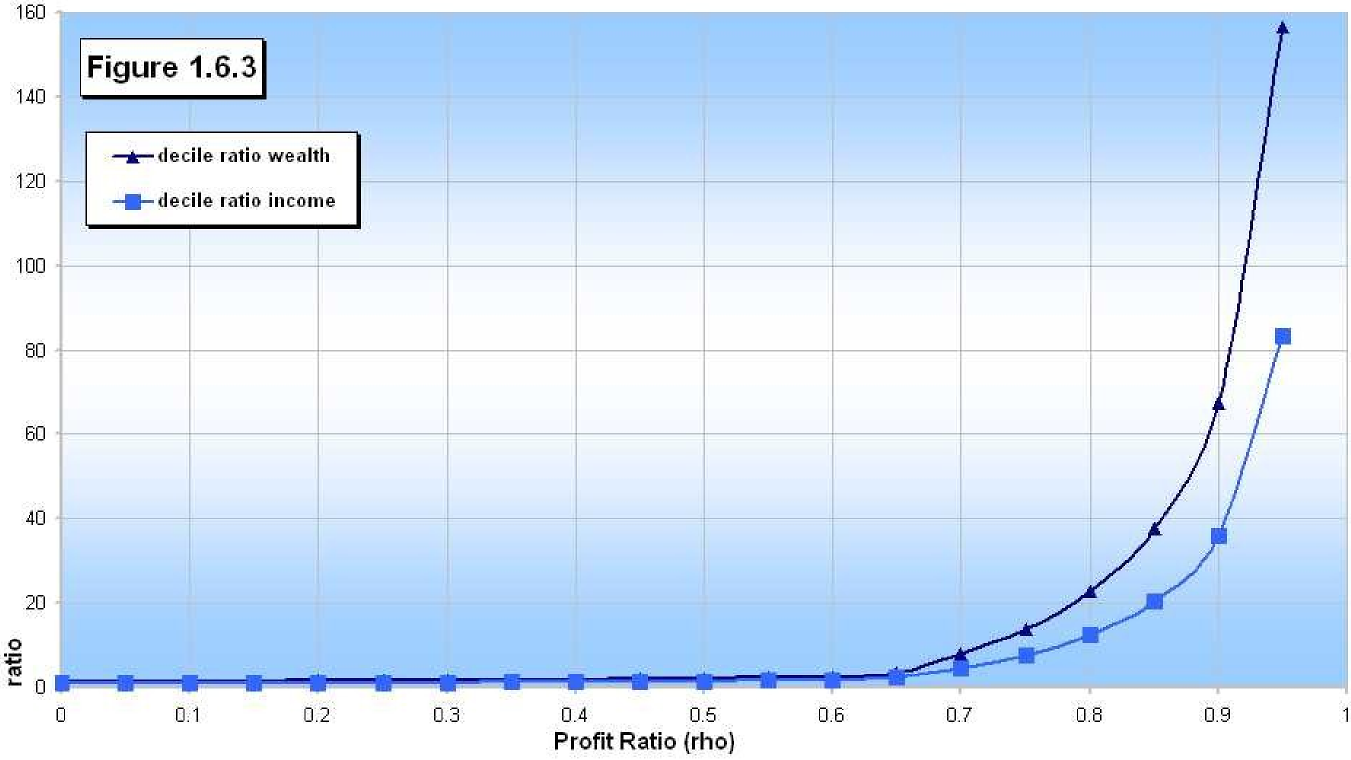


Figure 1.6.3
decile ratio wealth
decile ratio income
ratio
Profit Ratio (rho)
160
140
120
100
80
60
40
20
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

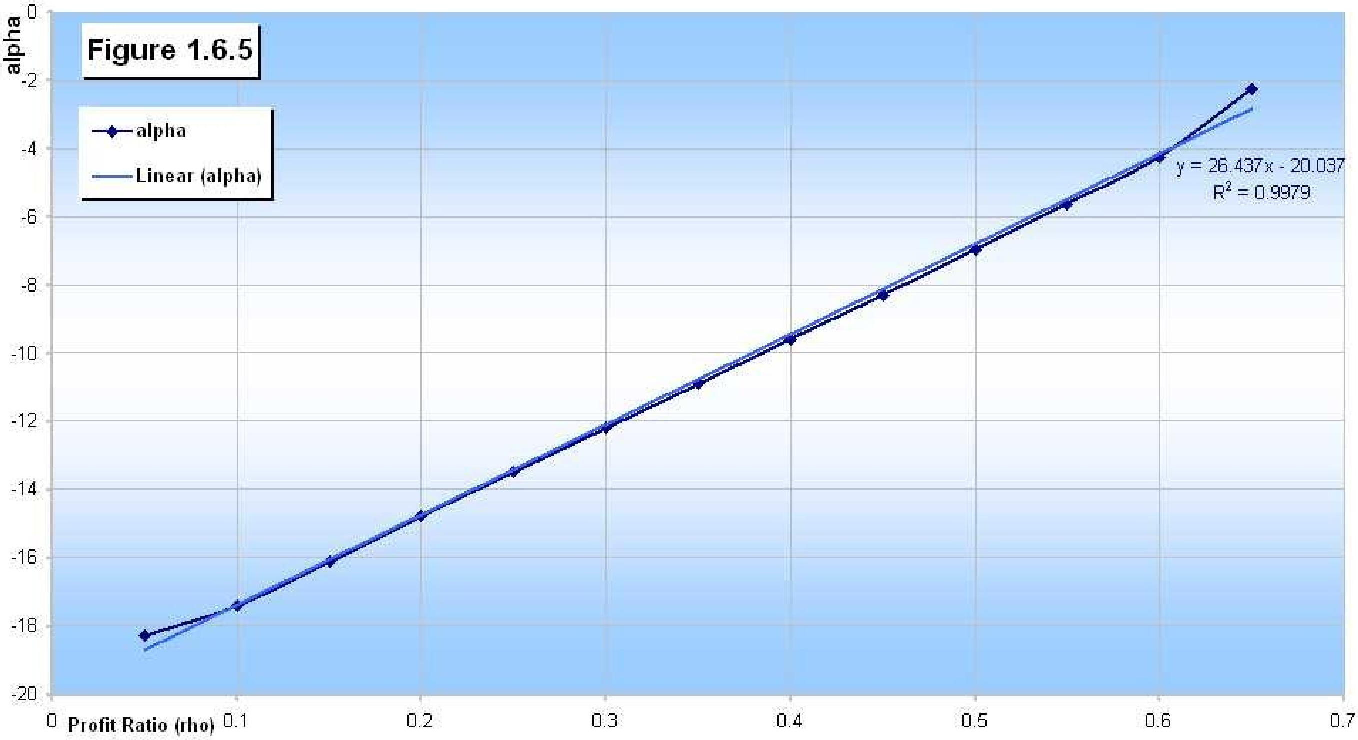


Figure 1.6.5
alpha
Linear (alpha)
y = 26.437x - 20.037
$R^2 = 0.9979$
alpha
Profit Ratio (rho)
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

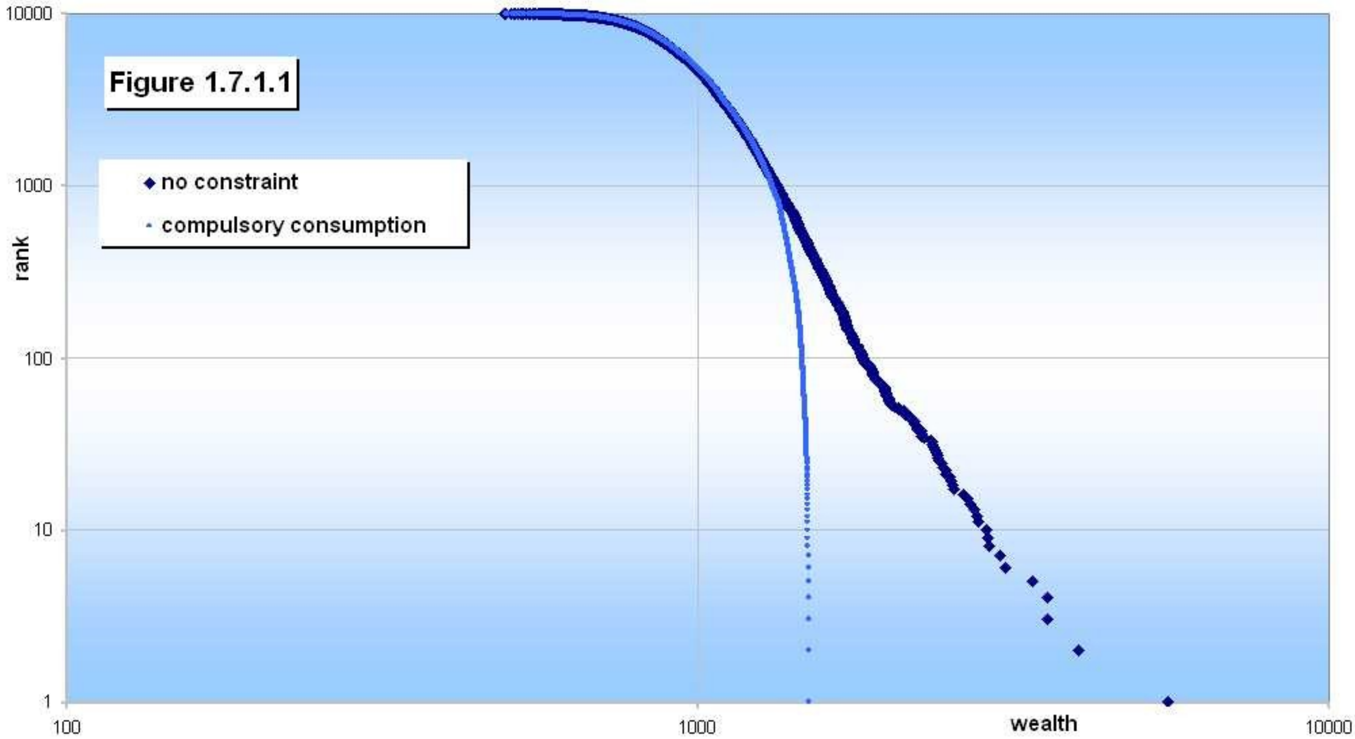


Figure 1.7.1.1
no constraint
compulsory consumption
rank
wealth
10000
1000
100
10
1
100
1000
10000

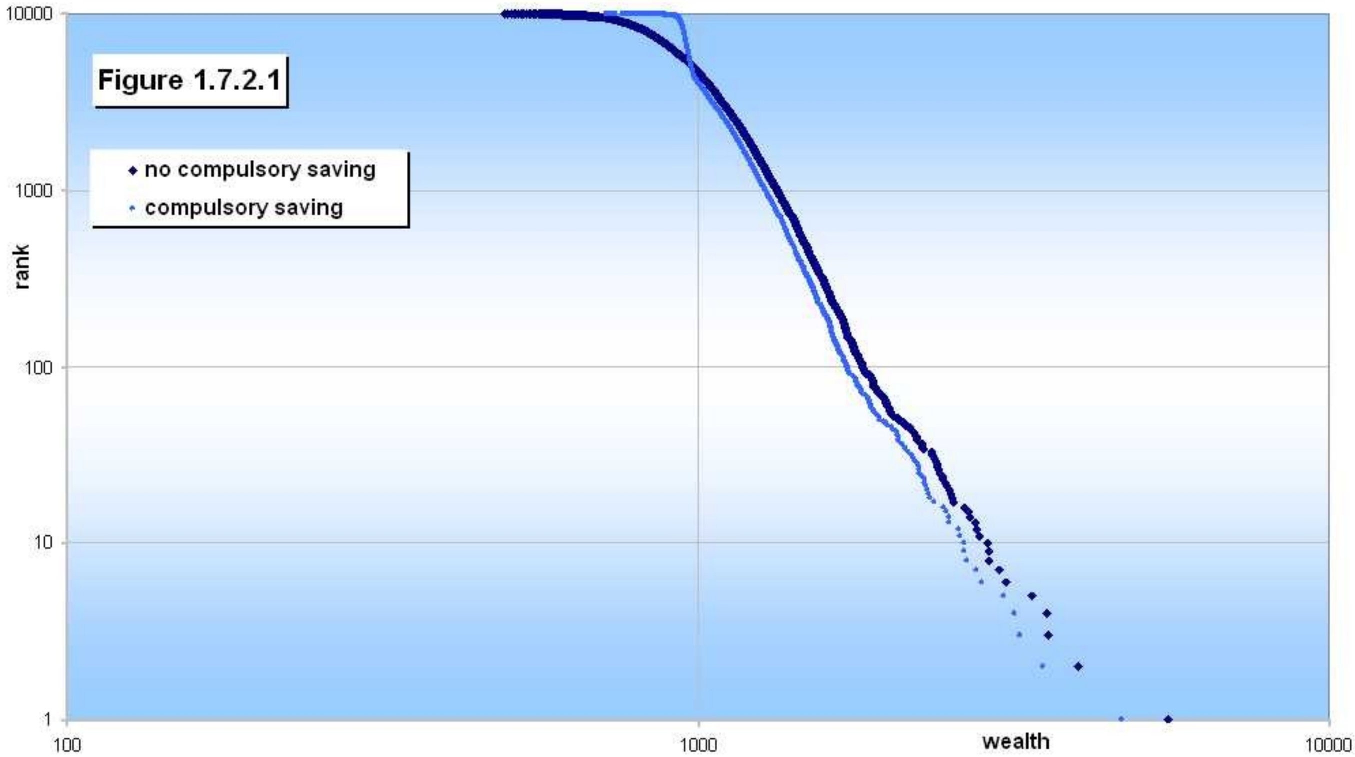
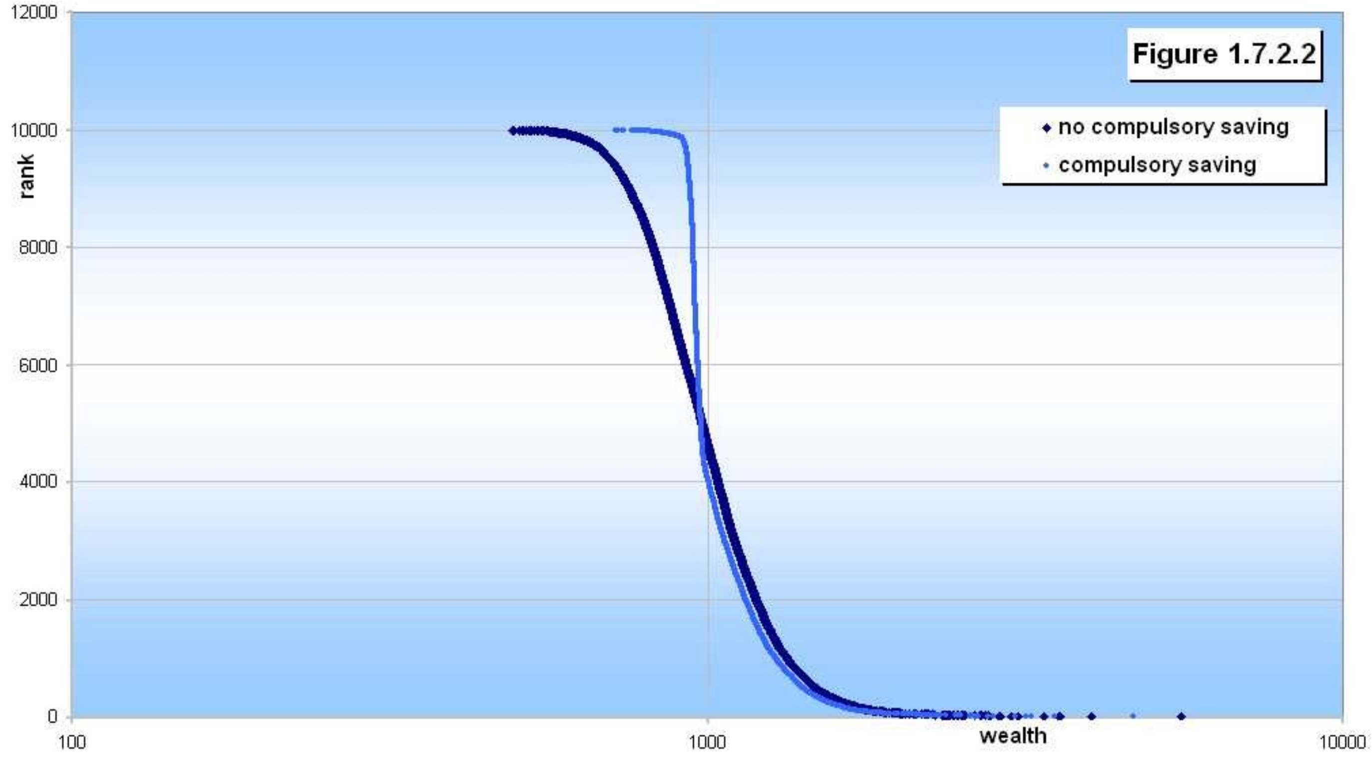


Figure 1.7.2.1
no compulsory saving
compulsory saving
rank
wealth
10000
1000
100
10
1
100
1000
10000



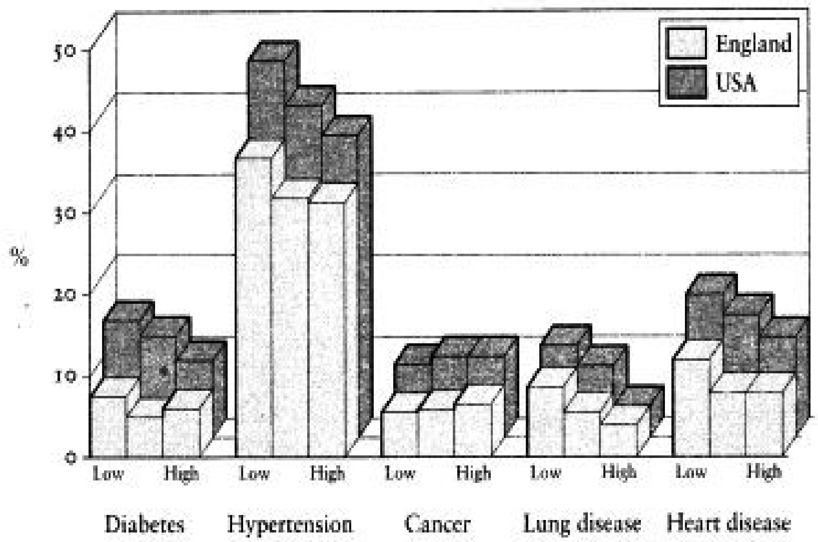


Figure 13.2 *Rates of illness are lower at both low and high educational levels in England compared to the USA.*[315]

Figure 1.8.1

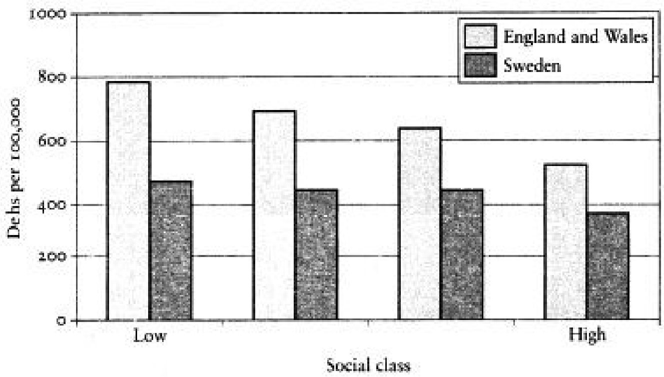


Figure 13.3 *Death rates among working-age men are lower in all occupational classes in Sweden compared to England and Wales.*[317]

Figure 1.8.2

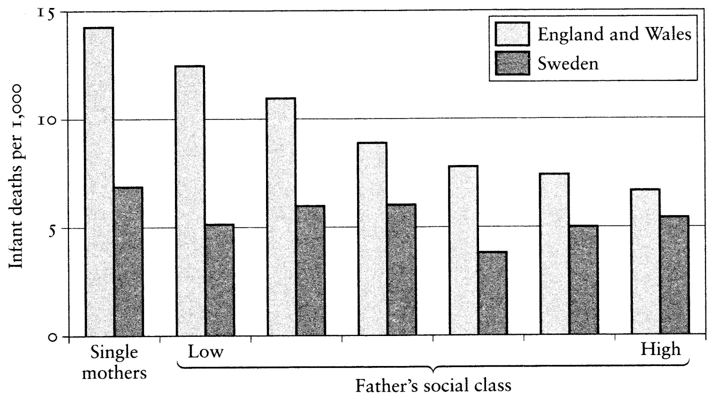
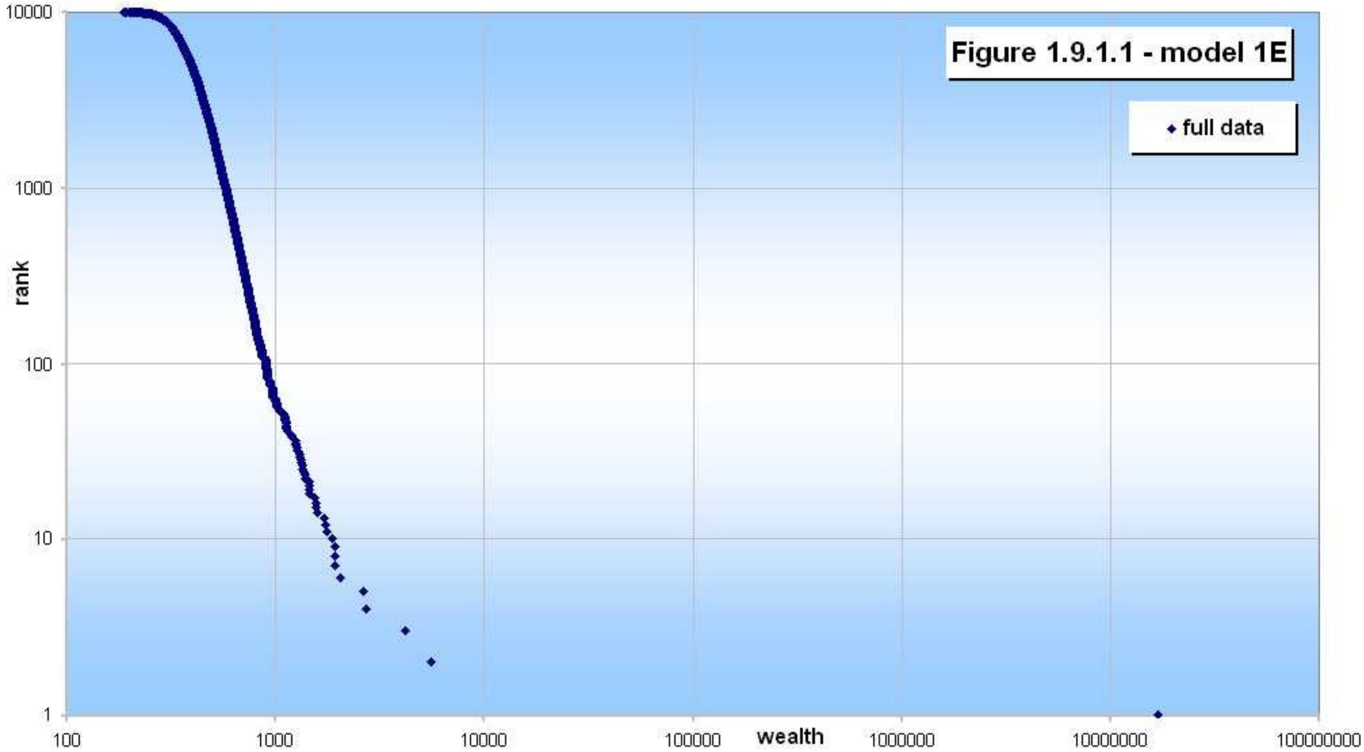
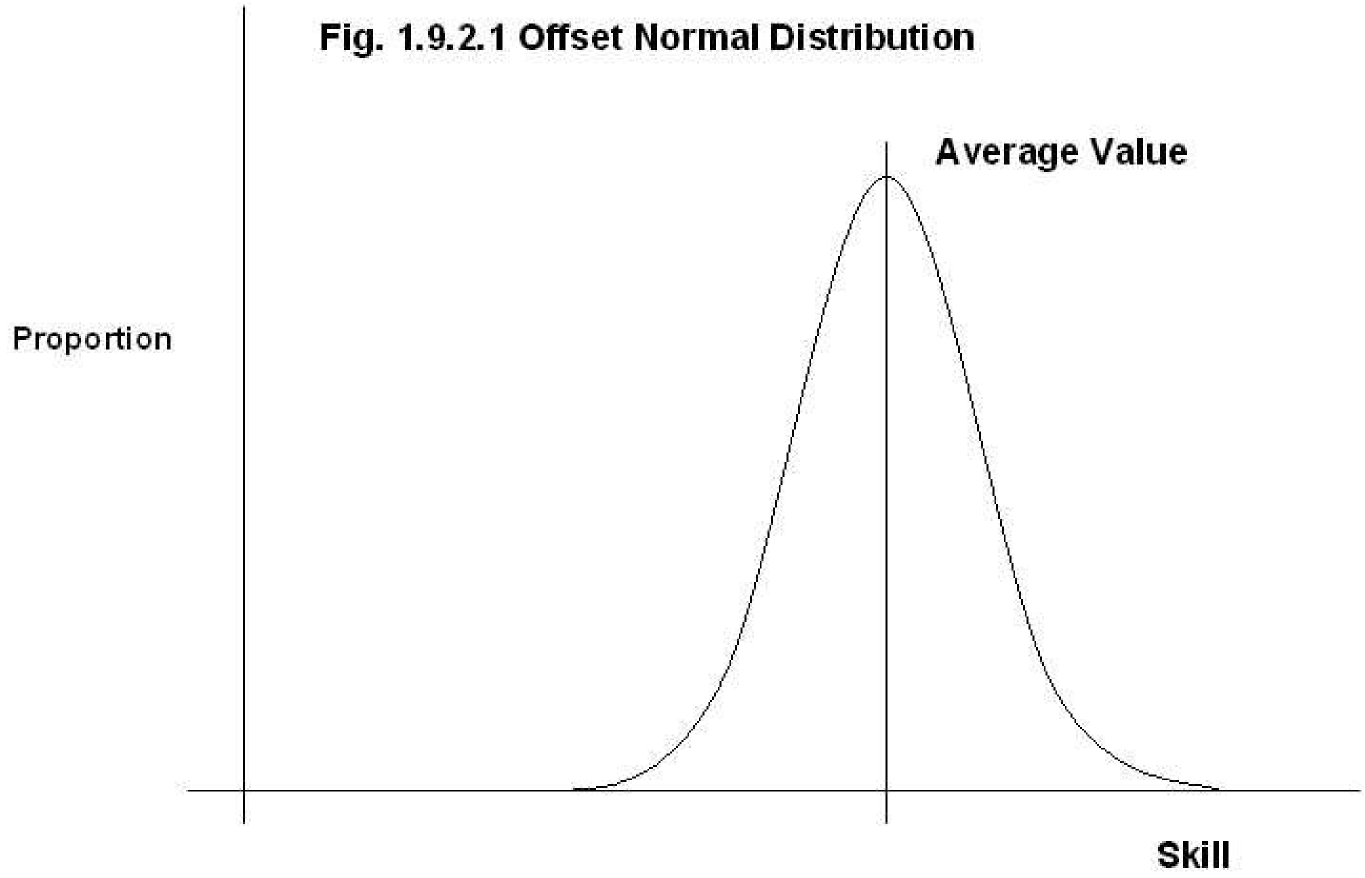


Figure 13.4 *Infant mortality rates are lower in all occupational classes in Sweden than in England and Wales.*[318]

Figure 1.8.3





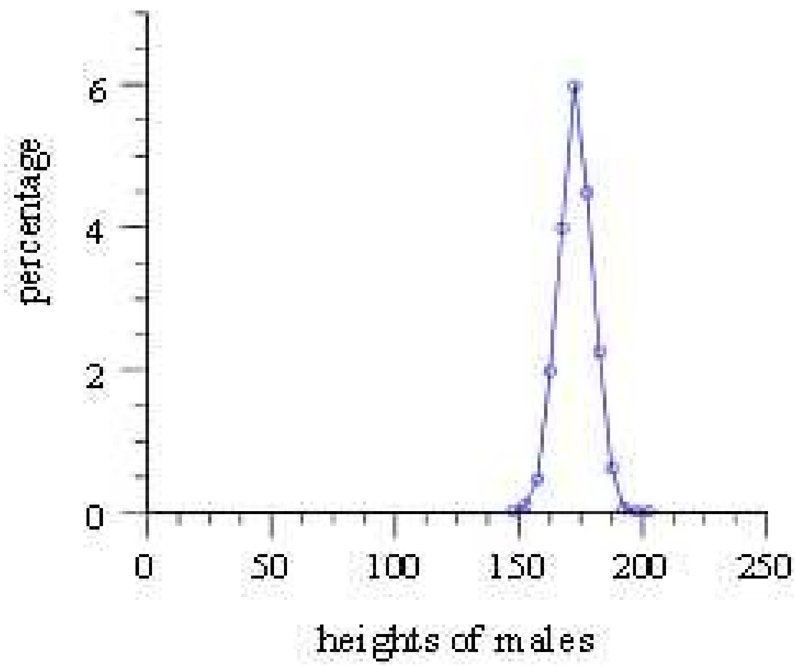
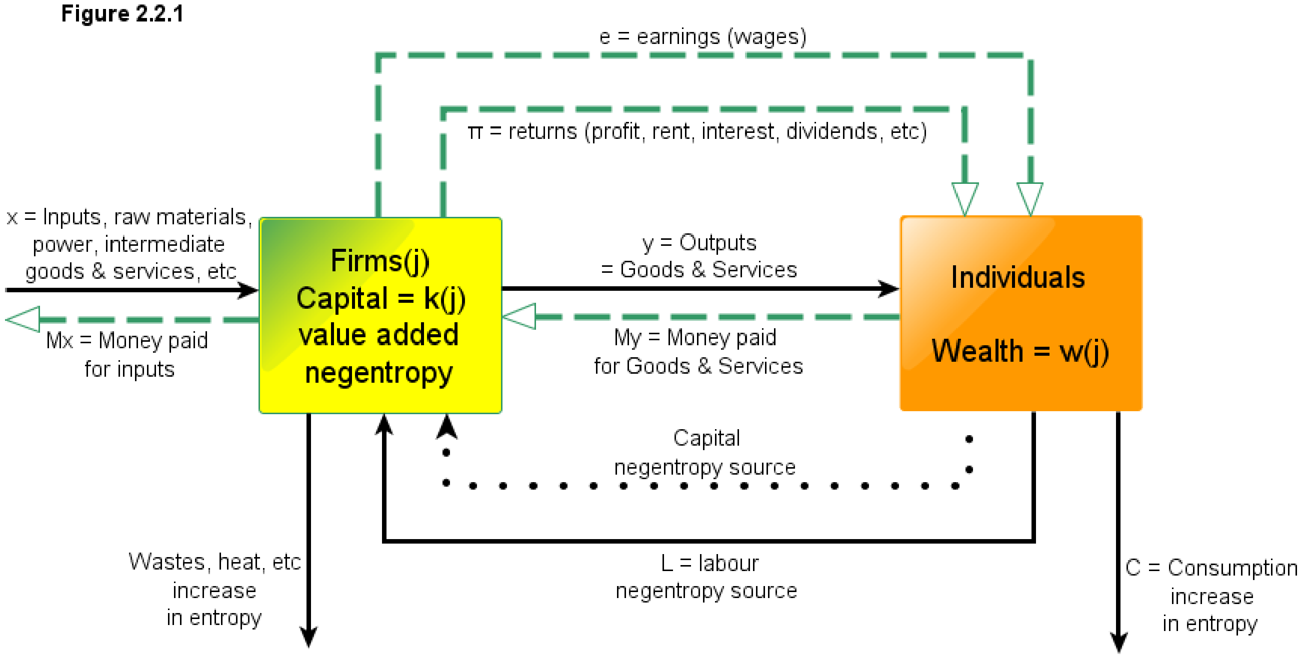


Figure 1.9.2.2 [Newman 2005]

Figure 2.2.1



e = earnings (wages)
π = returns (profit, rent, interest, dividends, etc)
x = Inputs, raw materials, power, intermediate goods & services, etc
Mx = Money paid for inputs
Firms(j) Capital = k(j) value added negentropy
y = Outputs = Goods & Services
My = Money paid for Goods & Services
Individuals
Wealth = w(j)
Capital negentropy source
Wastes, heat, etc increase in entropy
L = labour negentropy source
C = Consumption increase in entropy

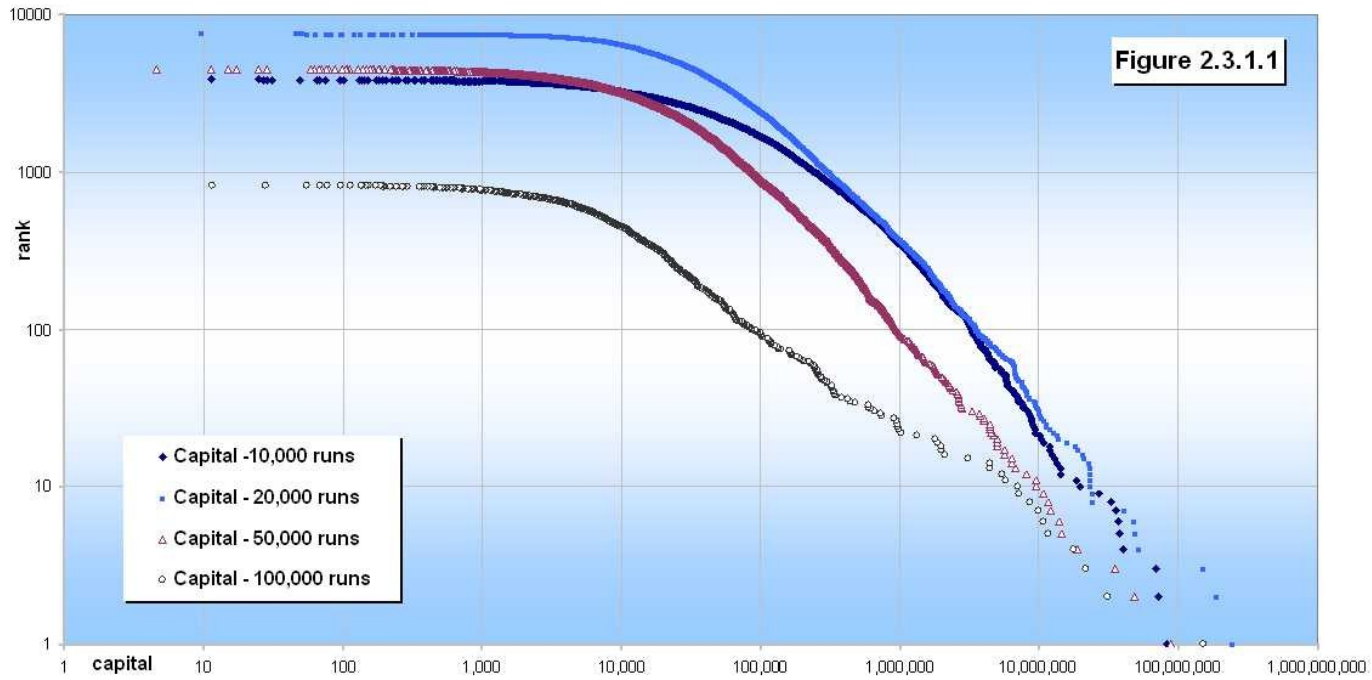


Figure 2.3.1.1
10000
1000
100
10
1
rank
1
capital
10
100
1,000
10,000
100,000
1,000,000
10,000,000
100,000,000
1,000,000,000
Capital -10,000 runs
Capital - 20,000 runs
Capital - 50,000 runs
Capital - 100,000 runs

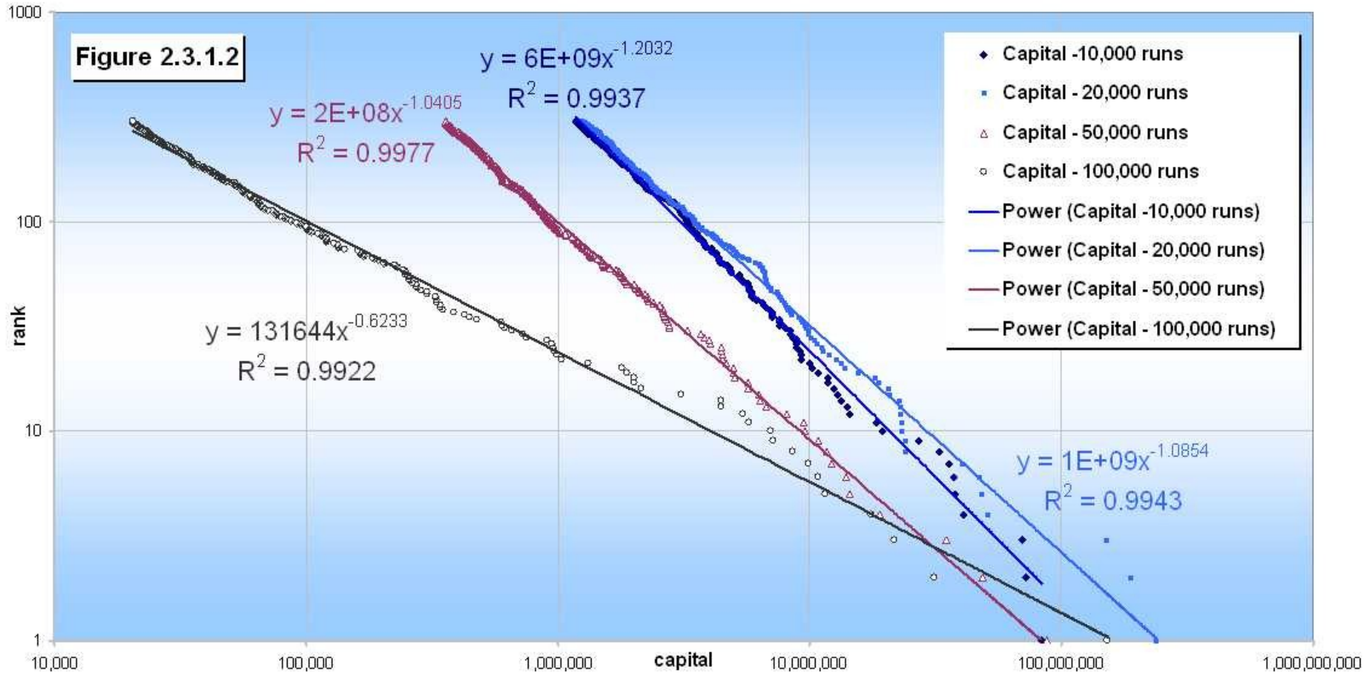


Figure 2.3.1.2
y = 6E+09x-1.2032
R2 = 0.9937
y = 2E+08x-1.0405
R2 = 0.9977
y = 131644x-0.6233
R2 = 0.9922
y = 1E+09x-1.0854
R2 = 0.9943
Capital -10,000 runs
Capital - 20,000 runs
Capital - 50,000 runs
Capital - 100,000 runs
Power (Capital -10,000 runs)
Power (Capital - 20,000 runs)
Power (Capital - 50,000 runs)
Power (Capital - 100,000 runs)
rank
capital
1000
100
10
1
10,000
100,000
1,000,000
10,000,000
100,000,000
1,000,000,000

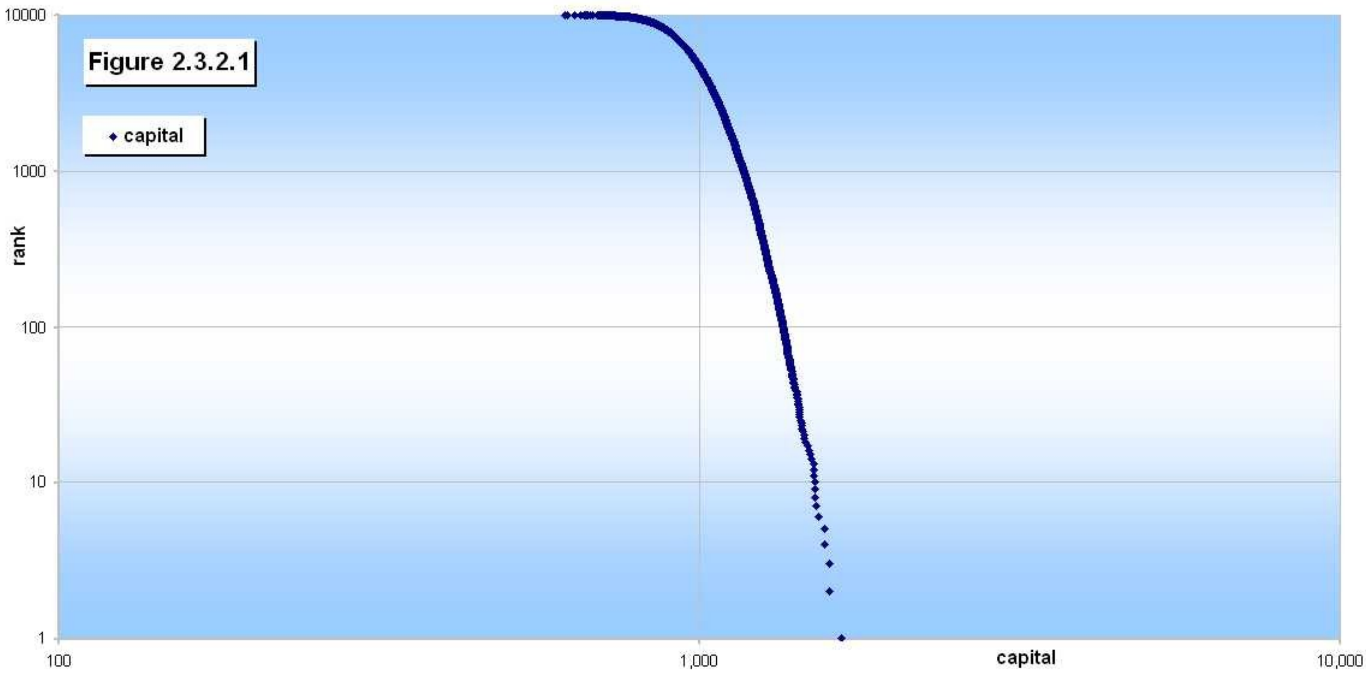


Figure 2.3.2.1
capital
rank
capital
10000
1000
100
10
1
100
1,000
10,000

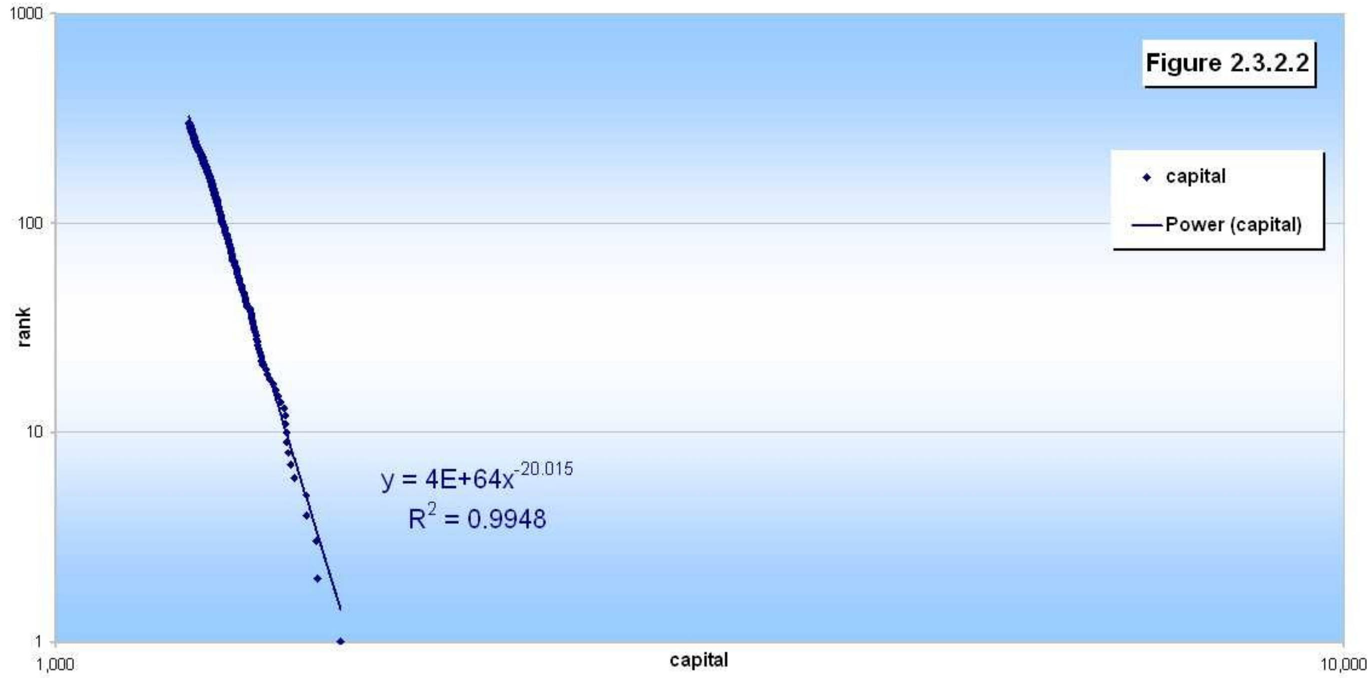


Figure 2.3.2.2
capital
Power (capital)
y = 4E+64x-20.015
R2 = 0.9948
rank
capital
1000
100
10
1
1,000
10,000

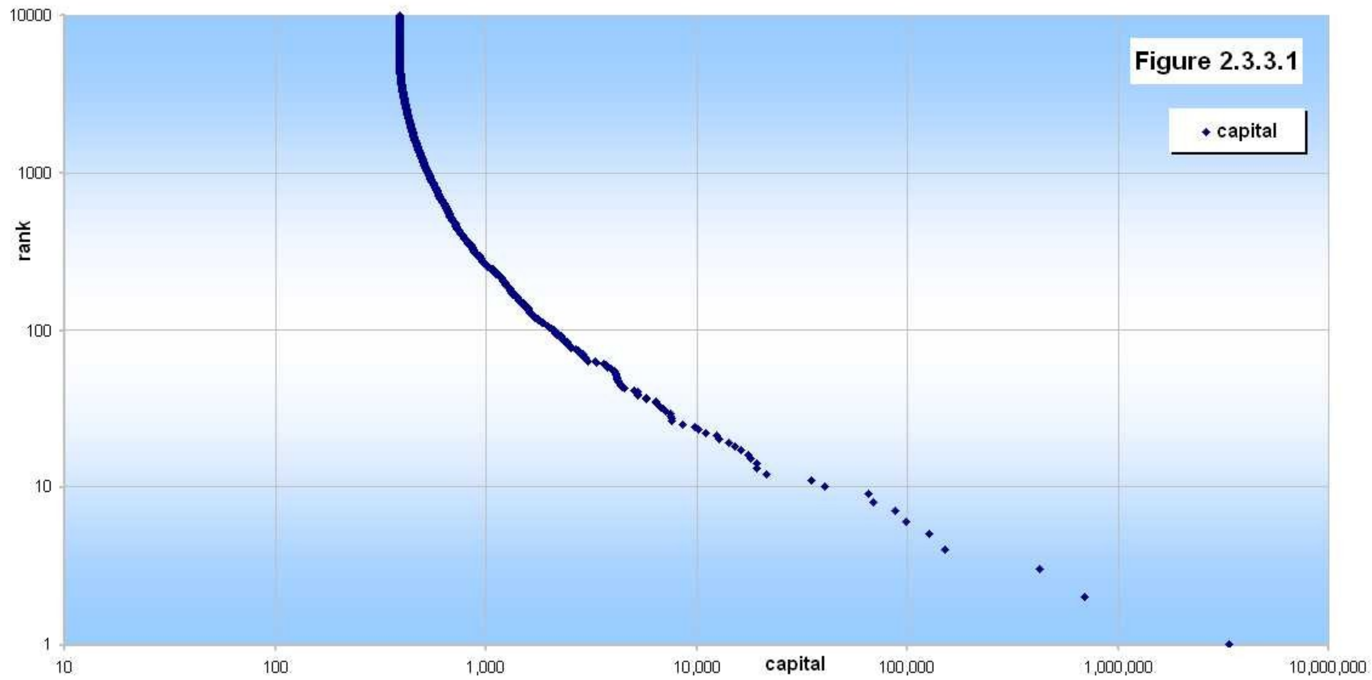


Figure 2.3.3.1
capital
rank
capital
10000
1000
100
10
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000

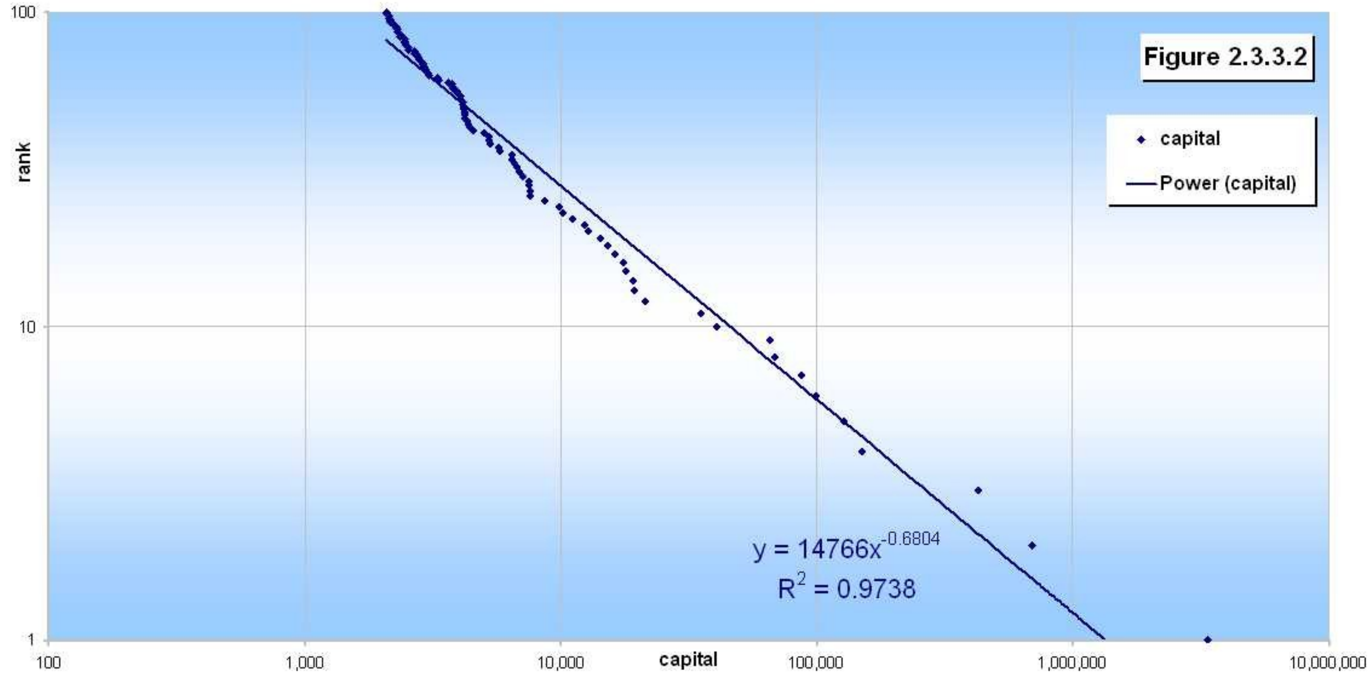


Figure 2.3.3.2
capital
Power (capital)
y = 14766x^-0.6804
R^2 = 0.9738
rank
100
10
1
capital
100
1,000
10,000
100,000
1,000,000
10,000,000


Figure 3.1.1 Copper Price (1966 $)
price
price
100
90
80
70
60
50
40
30
20
10
0
year
65
70
75
80
85
90
95
00
05
10

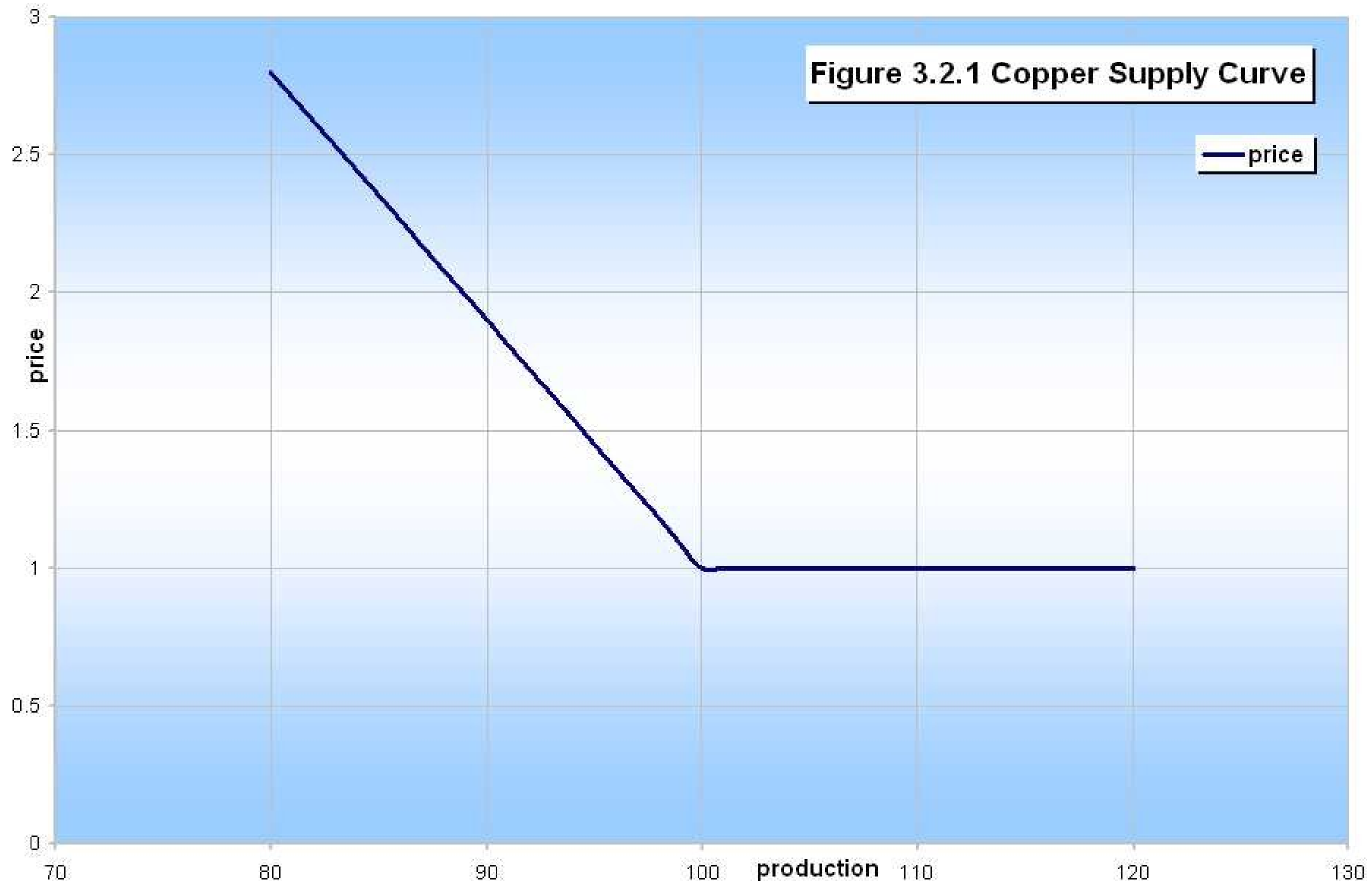


Figure 3.2.1 Copper Supply Curve
price
price
production
3
2.5
2
1.5
1
0.5
0
70
80
90
100
110
120
130

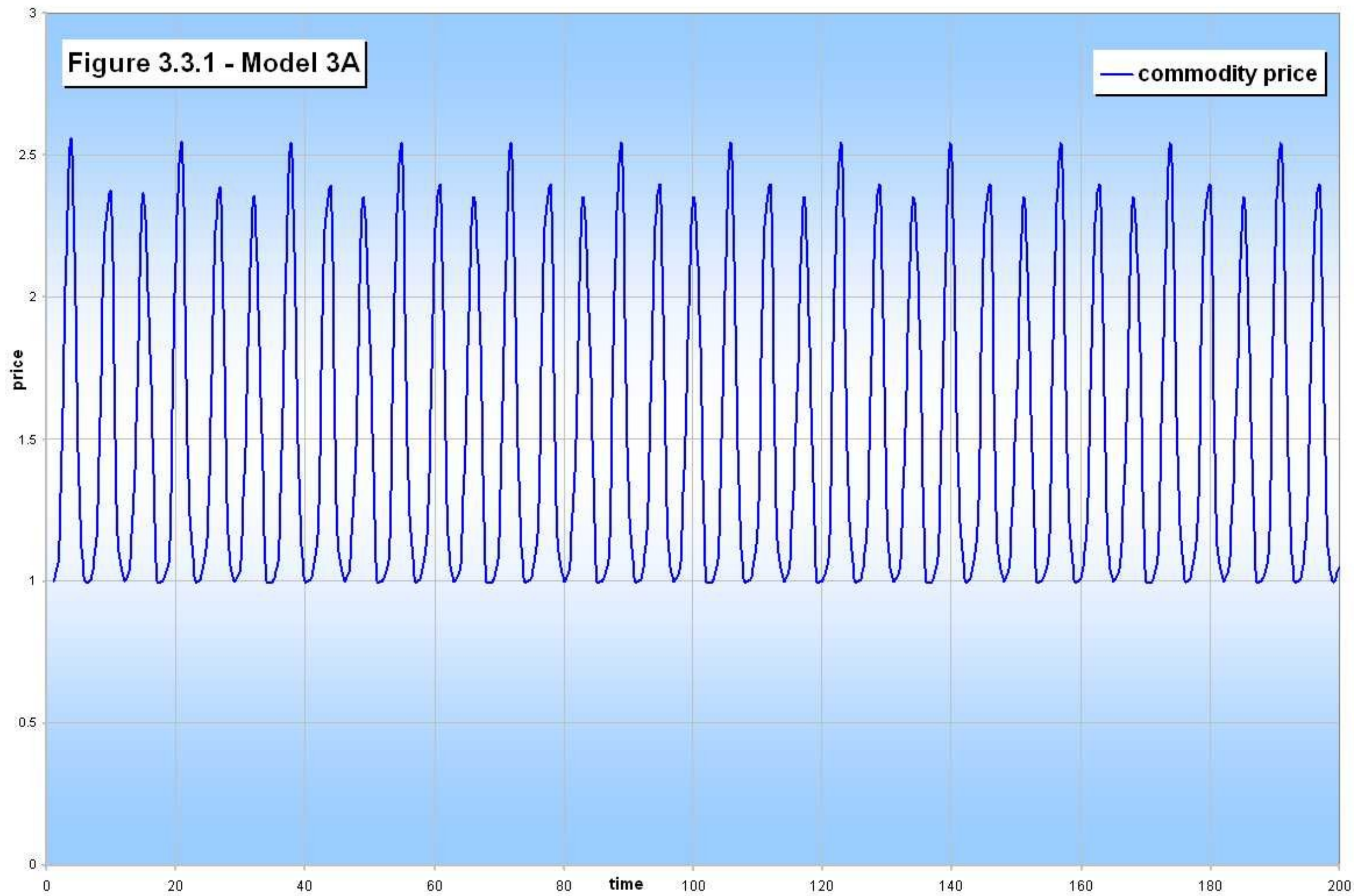


Figure 3.3.1 - Model 3A
commodity price
price
time
3
2.5
2
1.5
1
0.5
0
0
20
40
60
80
100
120
140
160
180
200

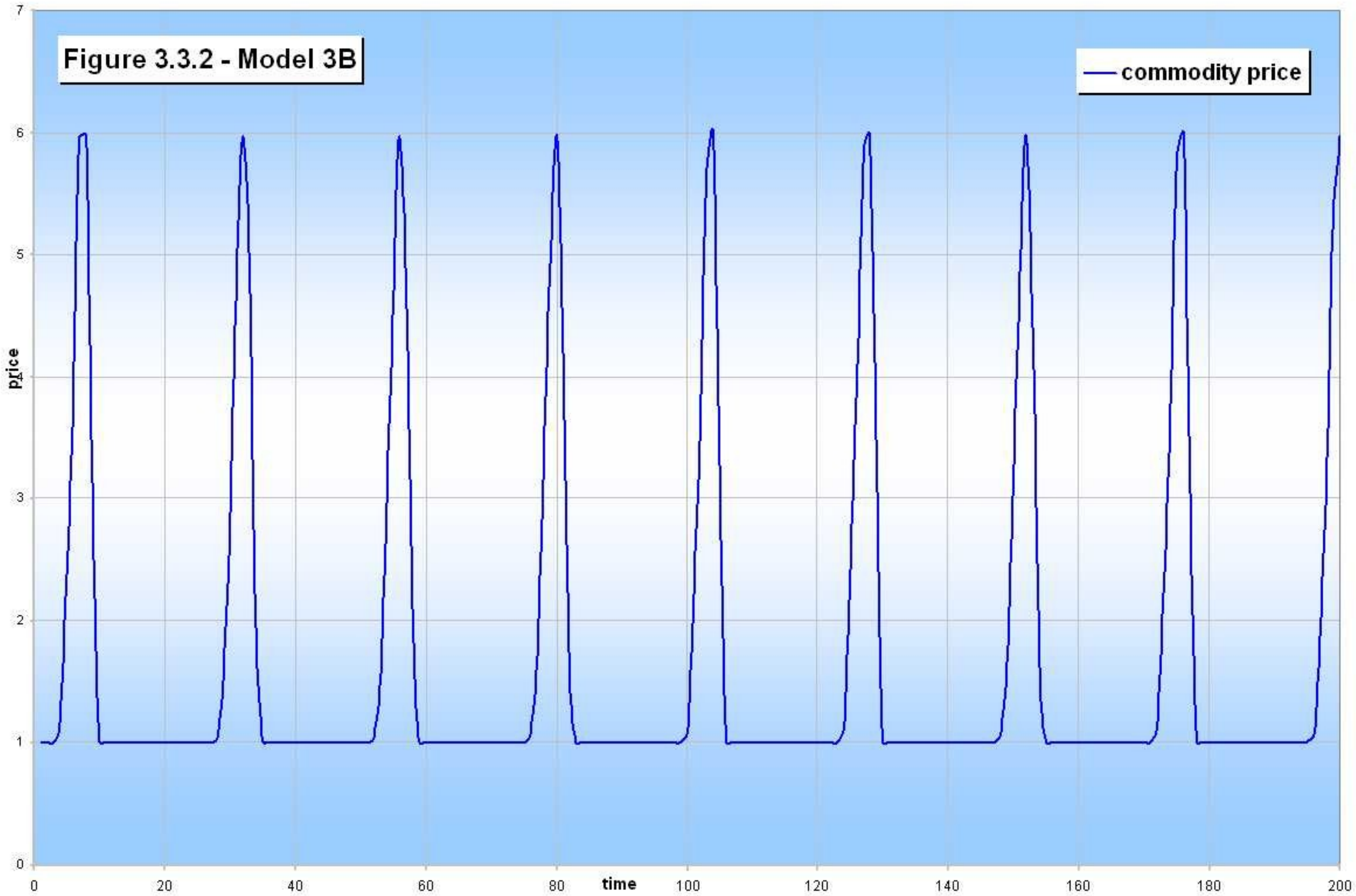


Figure 3.3.2 - Model 3B
commodity price
price
time
0
1
2
3
4
5
6
7
0
20
40
60
80
100
120
140
160
180
200

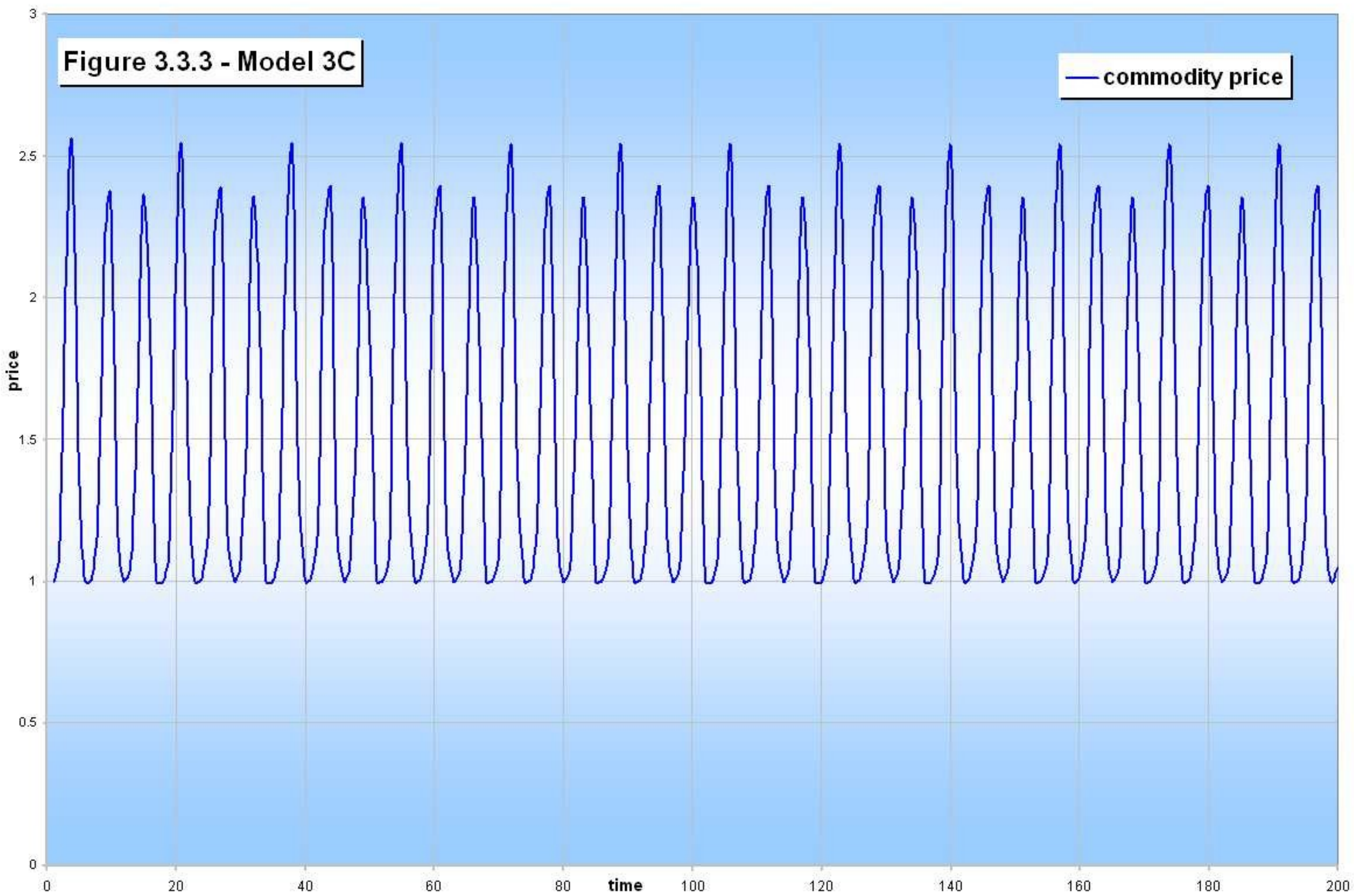


Figure 3.3.3 - Model 3C
commodity price
price
time
0
0.5
1
1.5
2
2.5
3
0
20
40
60
80
100
120
140
160
180
200

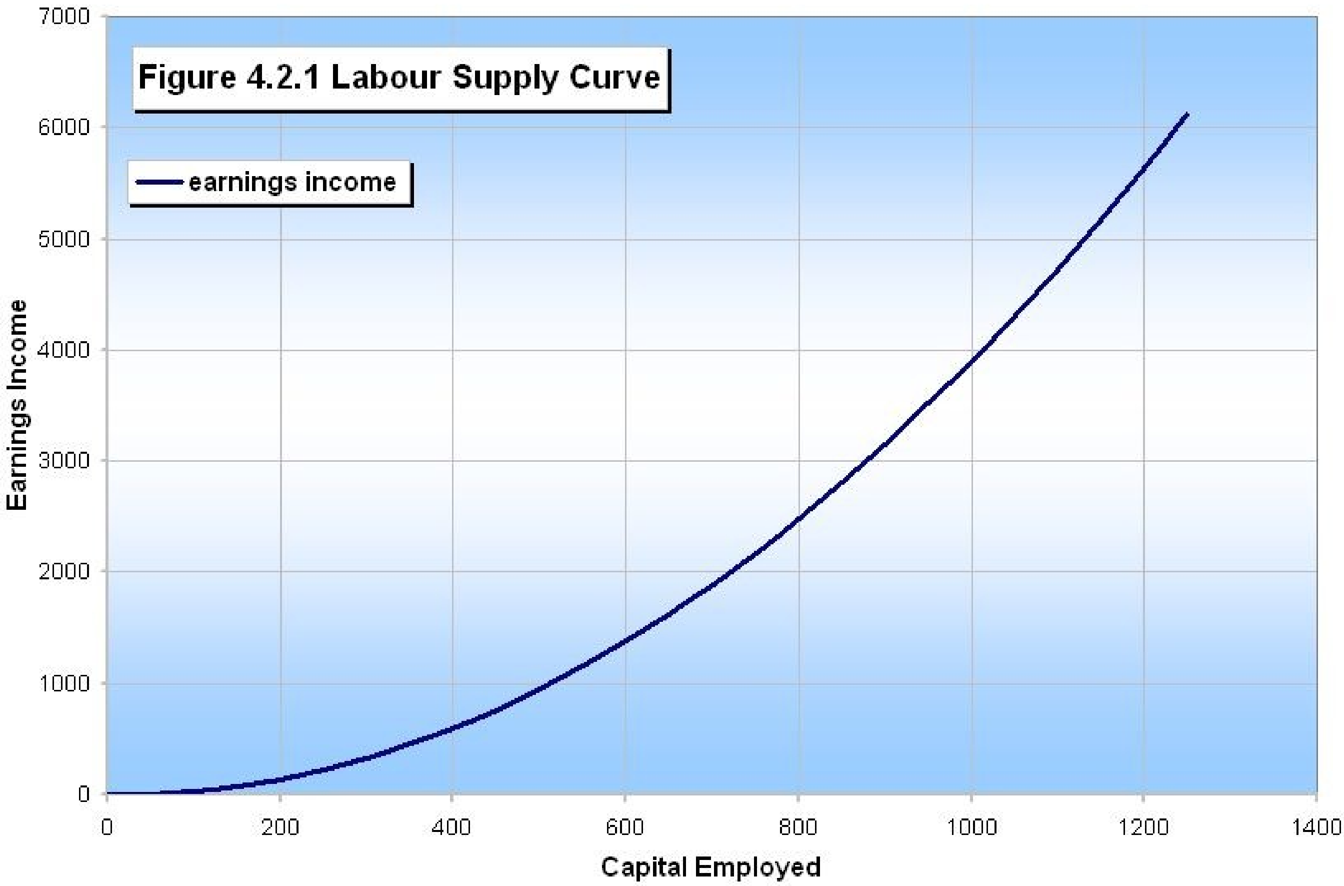


Figure 4.2.1 Labour Supply Curve
earnings income
Earnings Income
Capital Employed
7000
6000
5000
4000
3000
2000
1000
0
0
200
400
600
800
1000
1200
1400

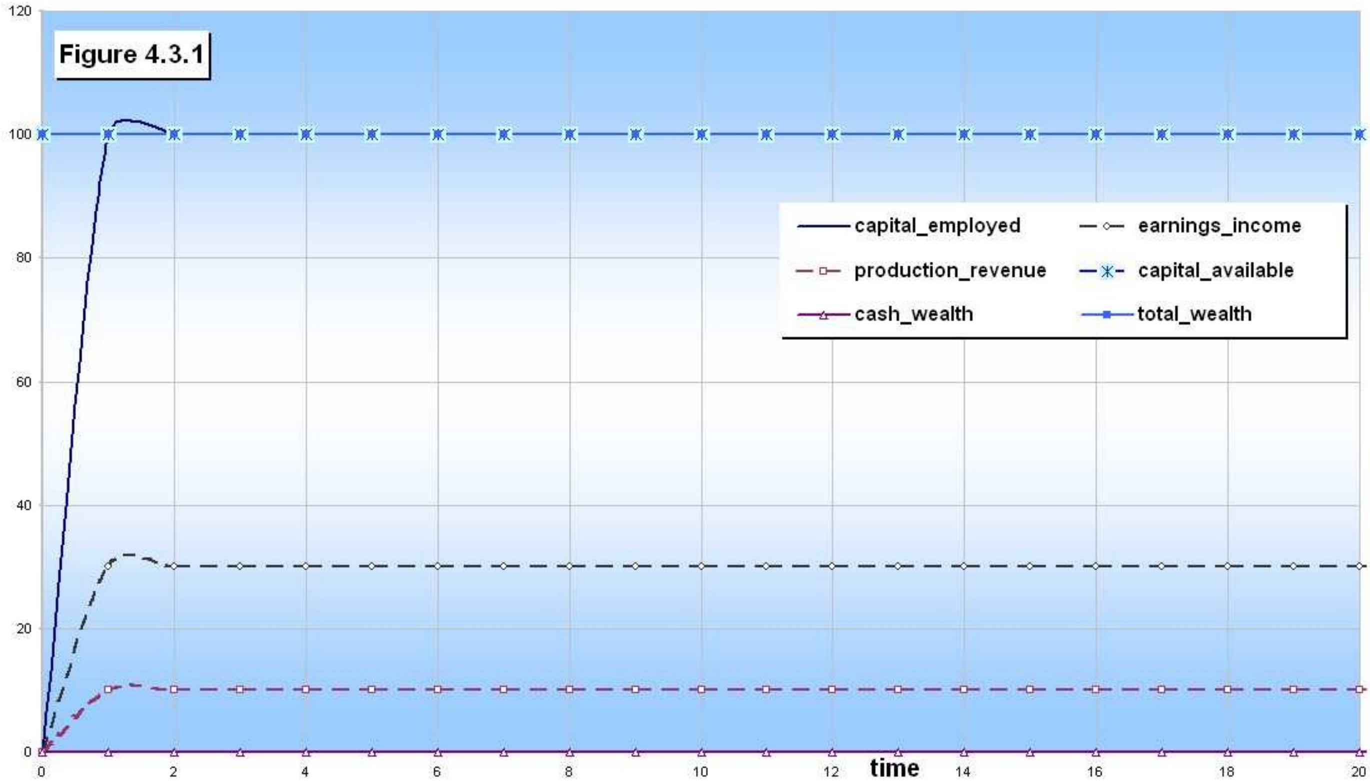


Figure 4.3.1
capital_employed
earnings_income
production_revenue
capital_available
cash_wealth
total_wealth
time
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
18
20

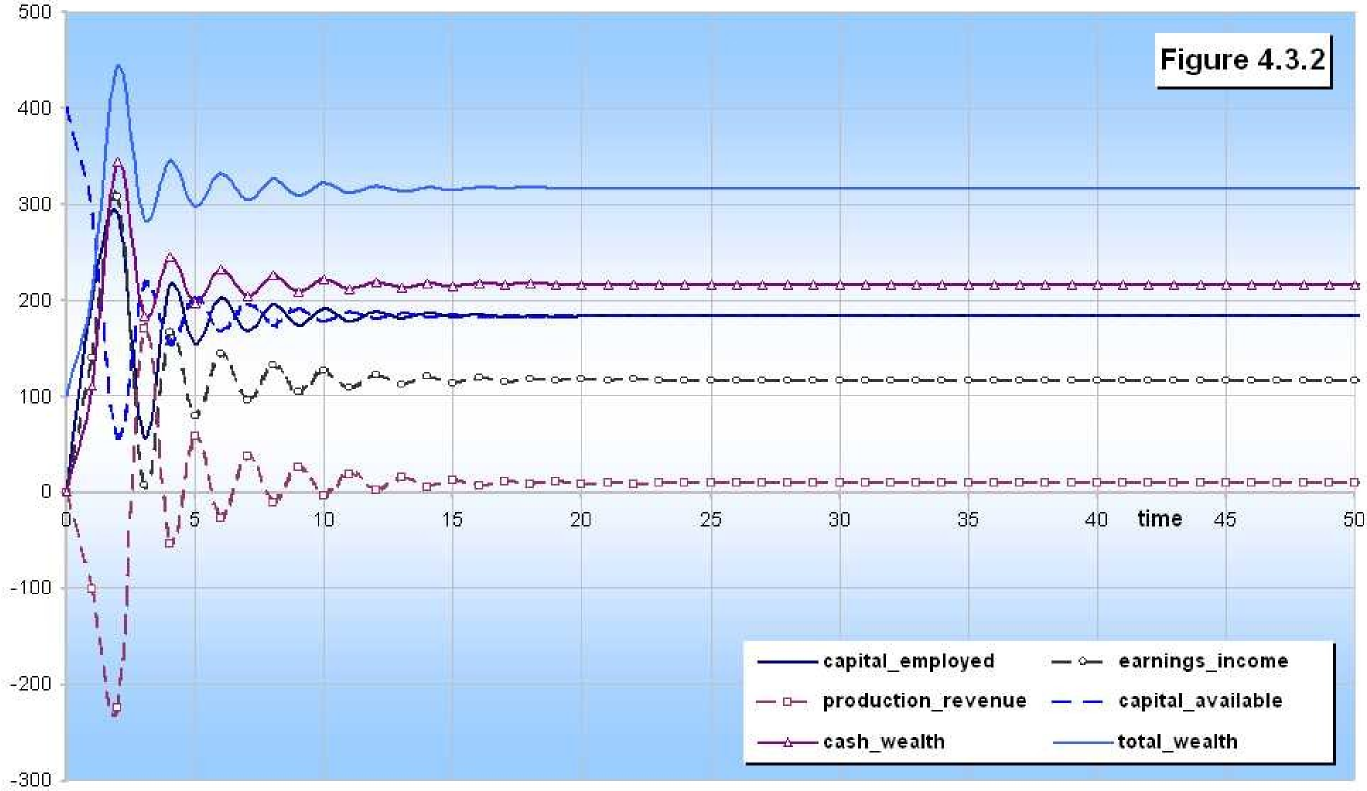


Figure 4.3.2
500
400
300
200
100
0
-100
-200
-300
0
5
10
15
20
25
30
35
40
time
45
50
capital_employed
earnings_income
production_revenue
capital_available
cash_wealth
total_wealth

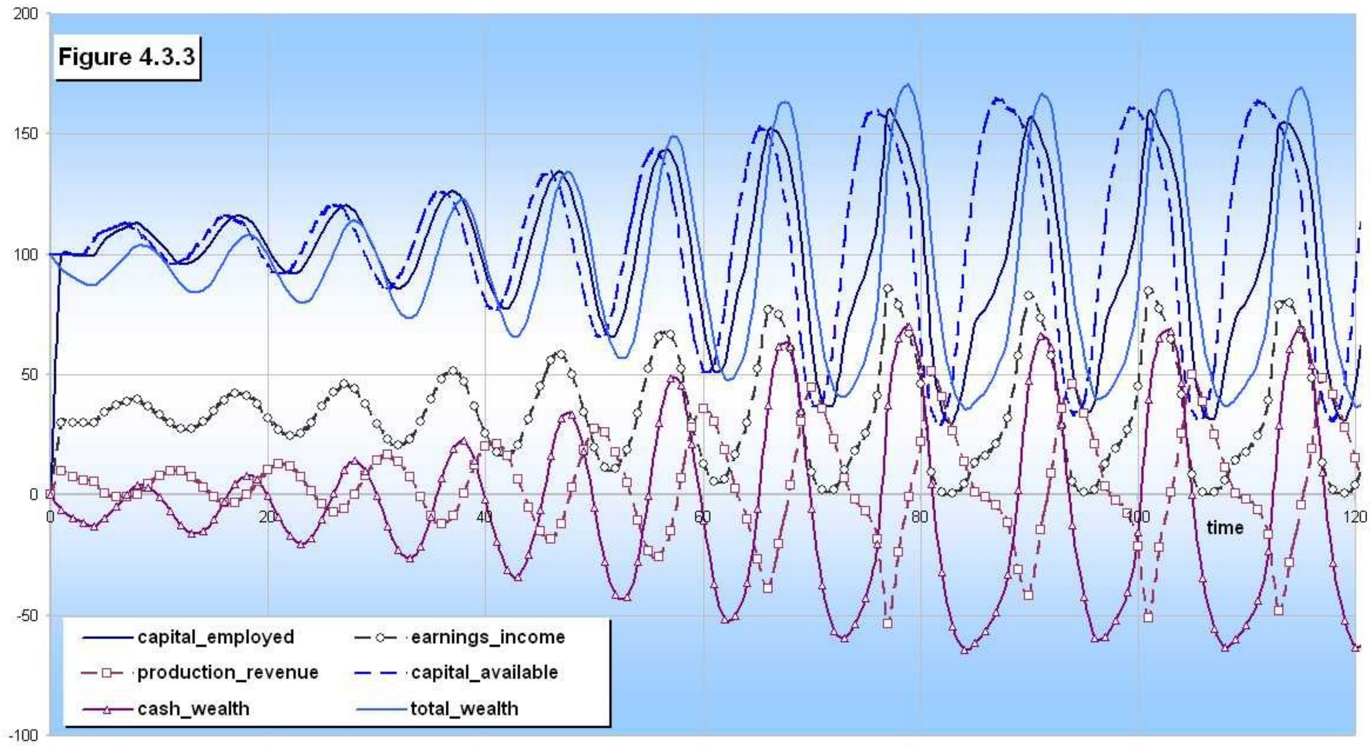


Figure 4.3.3
200
150
100
50
0
-50
-100
0
20
40
60
80
100
time
120
capital_employed
earnings_income
production_revenue
capital_available
cash_wealth
total_wealth

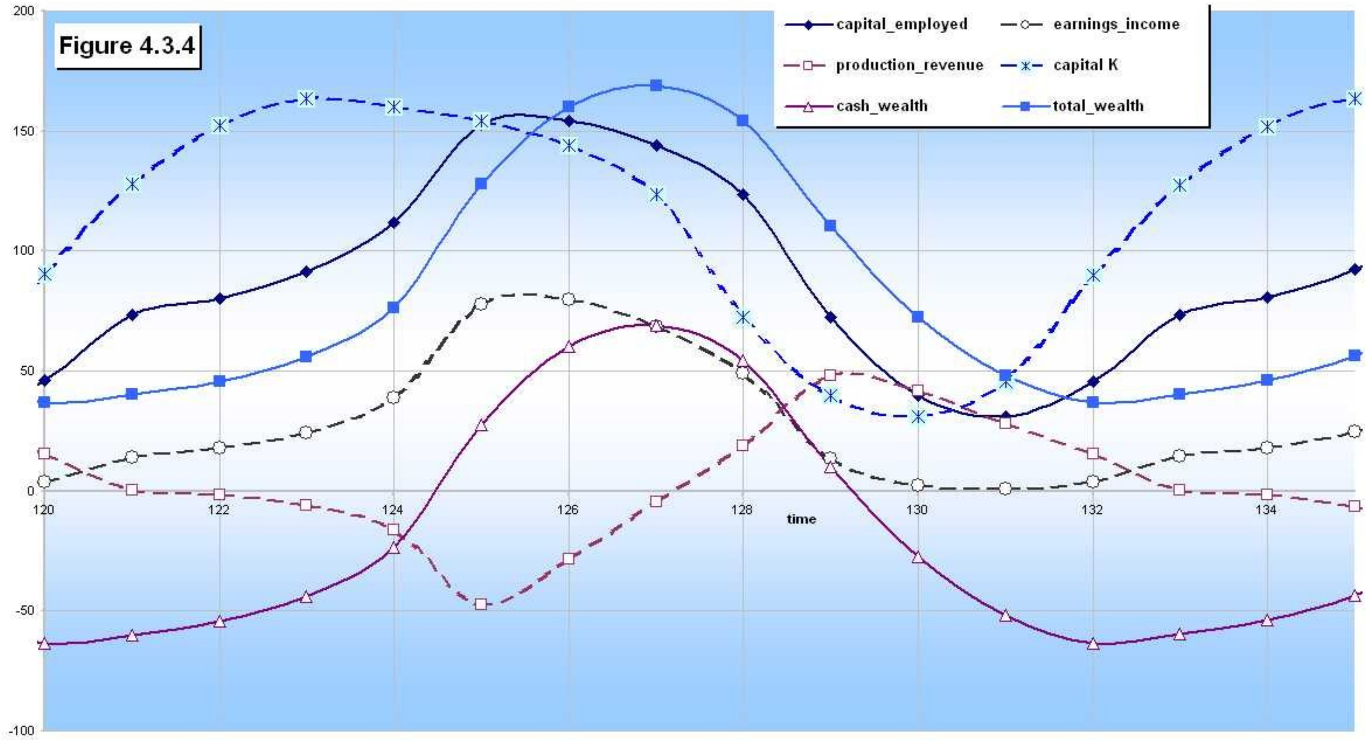


Figure 4.3.4
capital_employed
earnings_income
production_revenue
capital K
cash_wealth
total_wealth
time

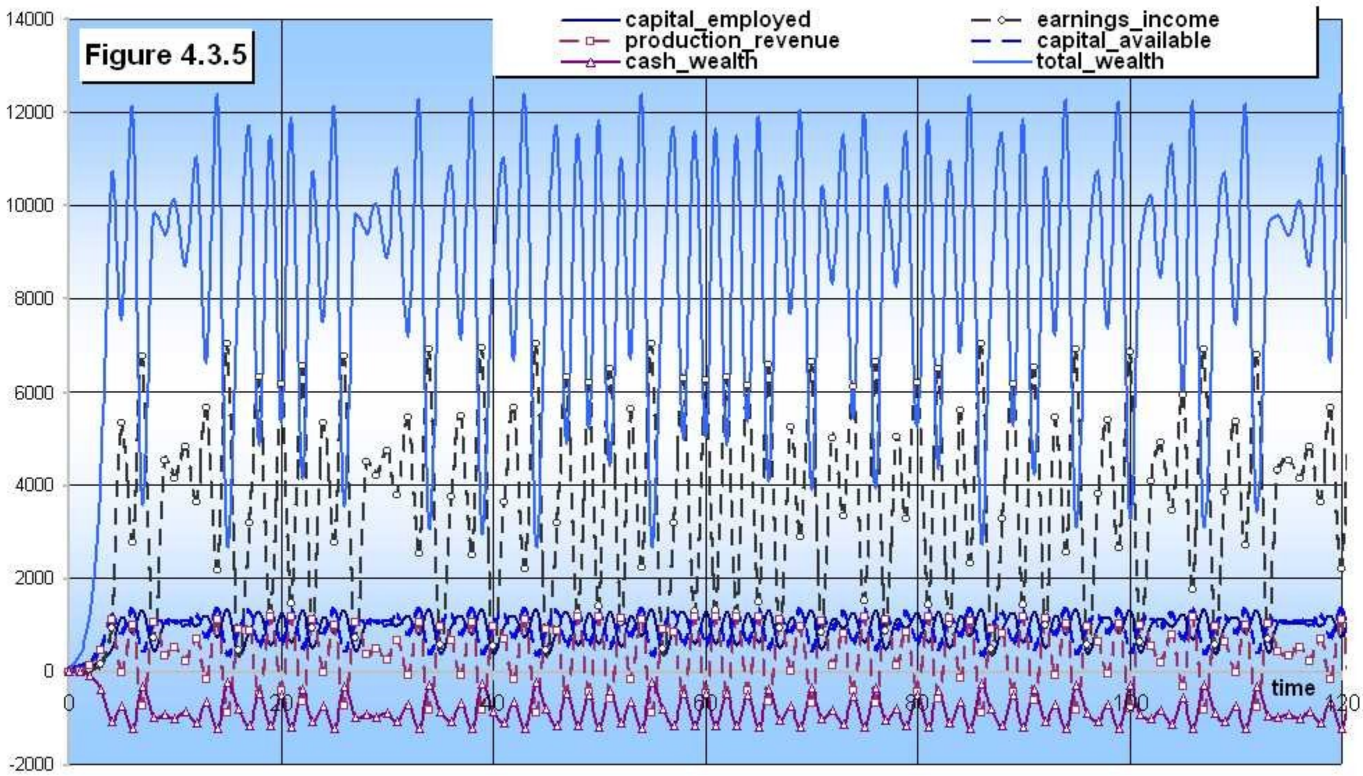


Figure 4.3.5
capital_employed
earnings_income
production_revenue
capital_available
cash_wealth
total_wealth
time

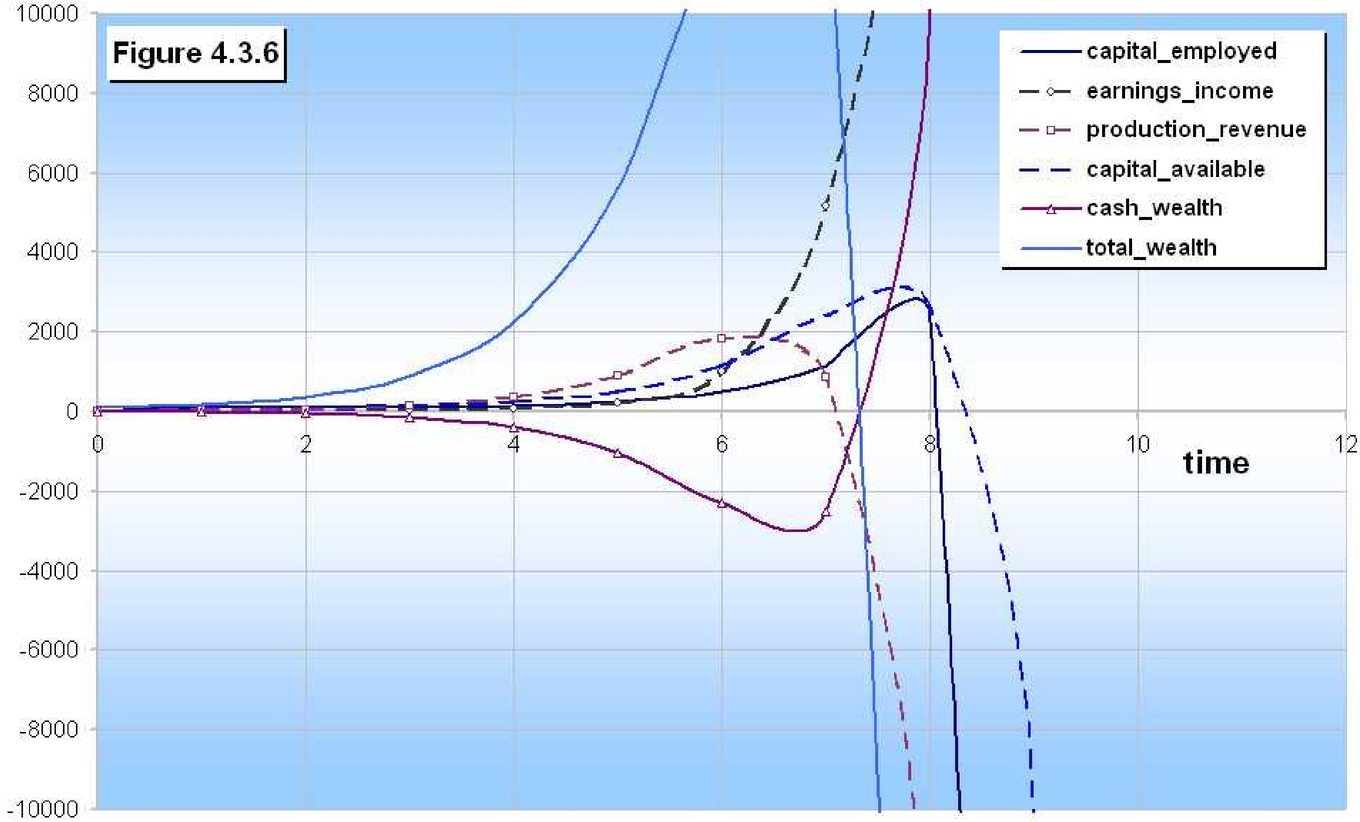


Figure 4.3.6
capital_employed
earnings_income
production_revenue
capital_available
cash_wealth
total_wealth
time
10000
8000
6000
4000
2000
0
-2000
-4000
-6000
-8000
-10000
0
2
4
6
8
10
12

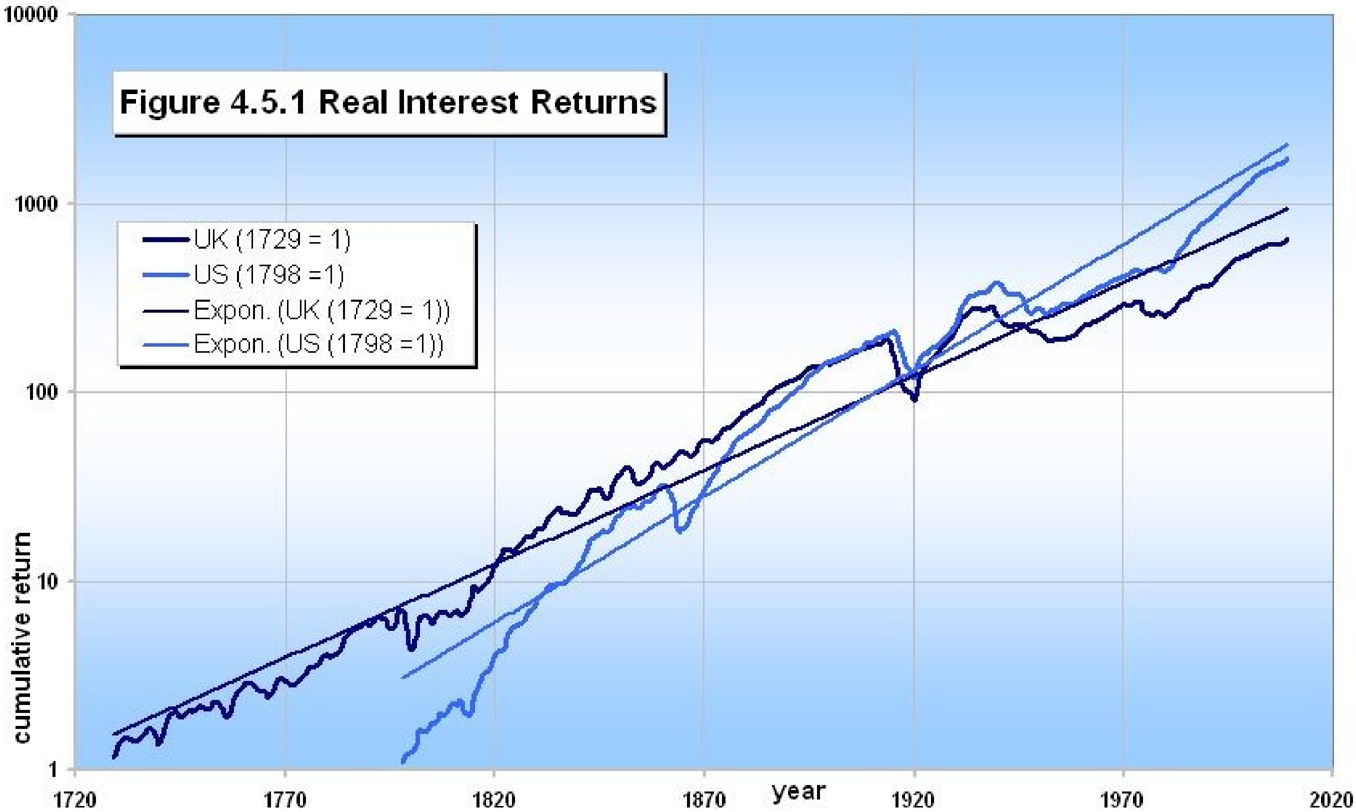


Figure 4.5.1 Real Interest Returns
UK (1729 = 1)
US (1798 =1)
Expon. (UK (1729 = 1))
Expon. (US (1798 =1))
cumulative return
year
10000
1000
100
10
1
1720
1770
1820
1870
1920
1970
2020

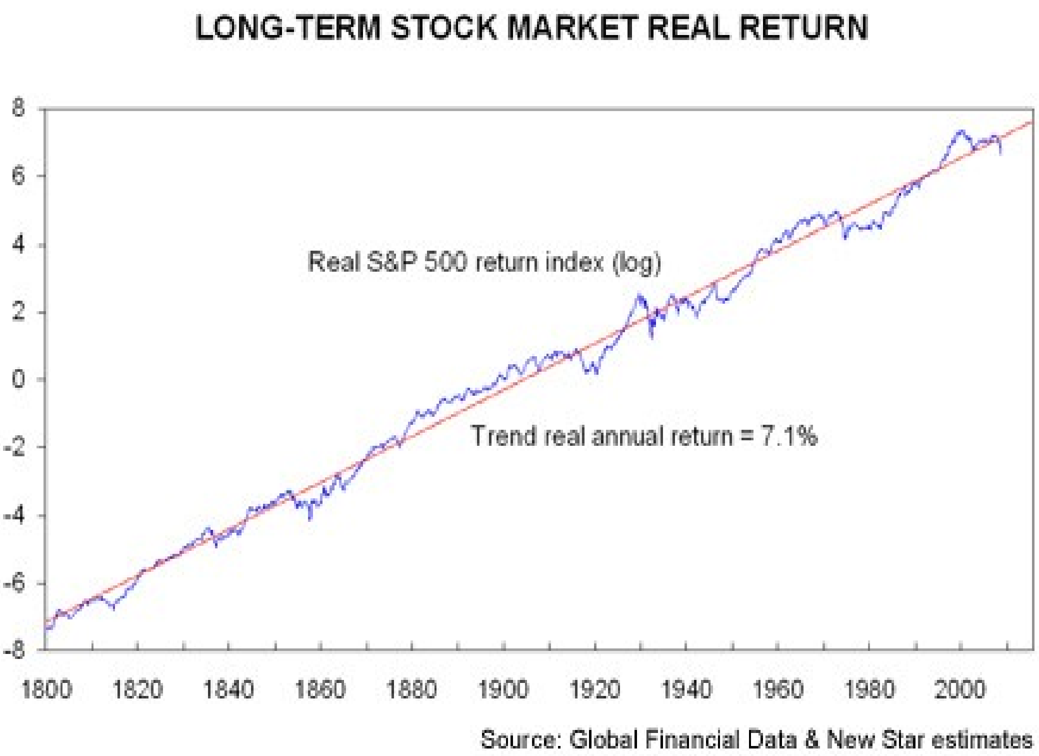

LONG-TERM STOCK MARKET REAL RETURN
8
6
4
2
0
-2
-4
-6
-8
Real S&P 500 return index (log)
Trend real annual return = 7.1%
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Source: Global Financial Data & New Star estimates

Figure 4.5.2

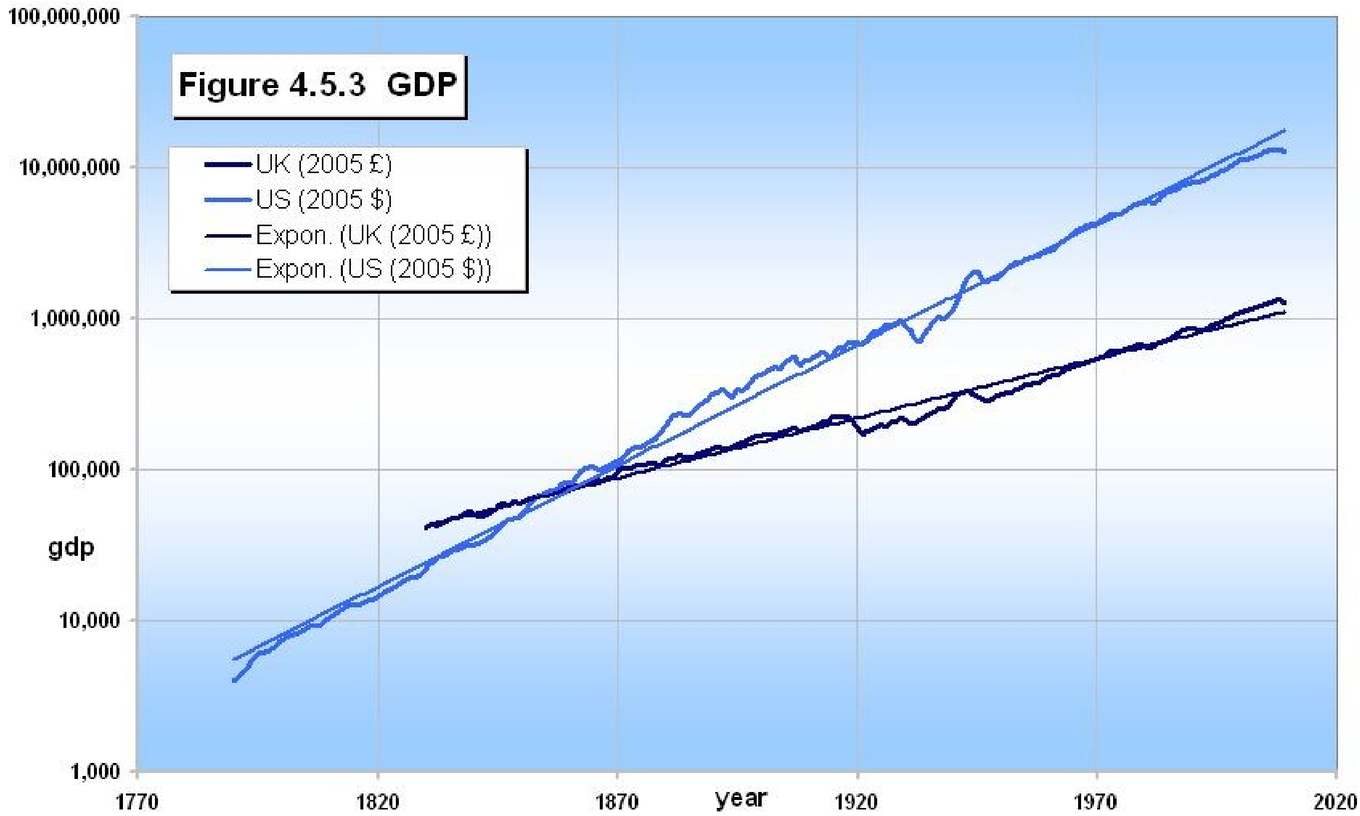
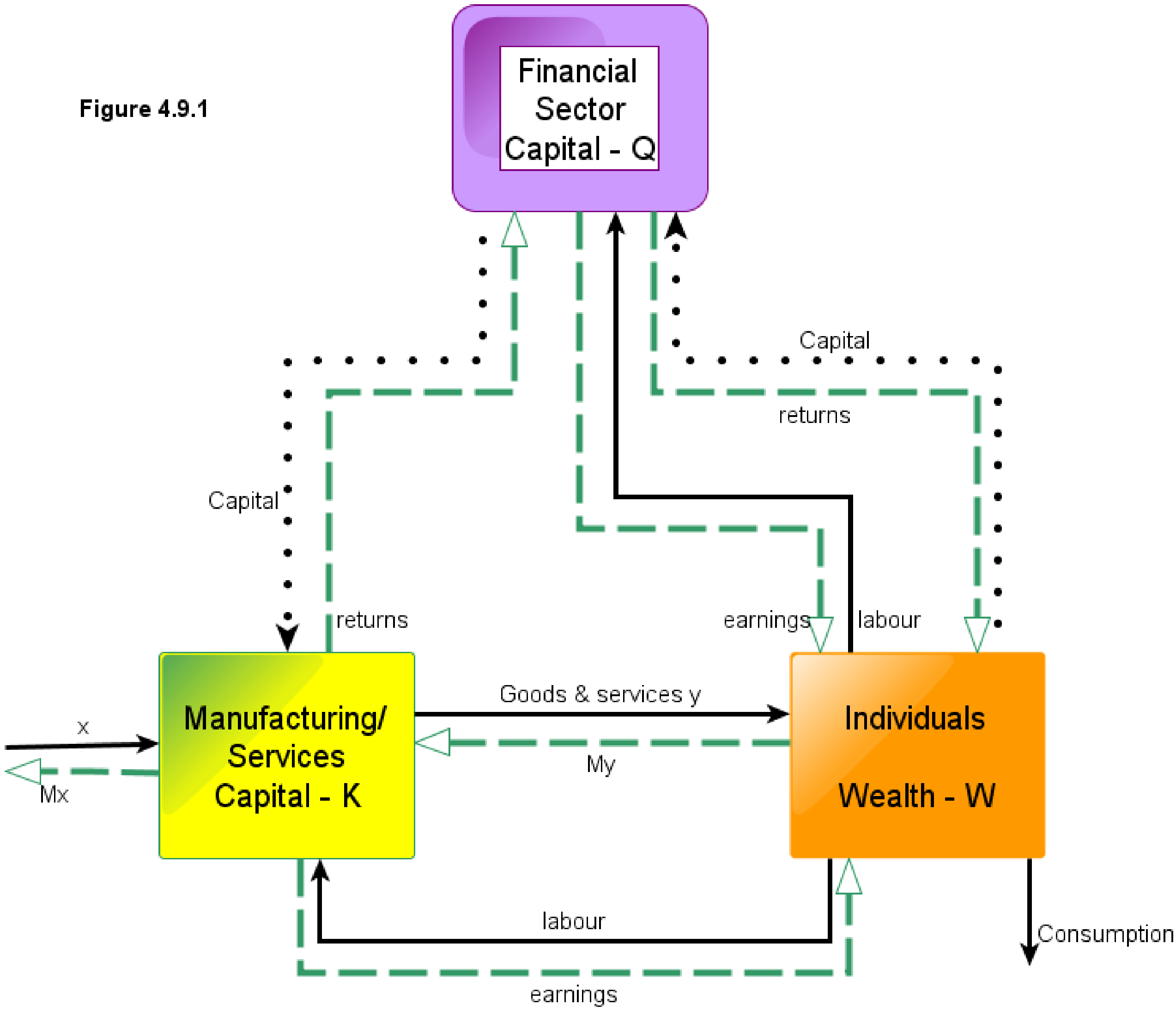


Figure 4.5.3 GDP
UK (2005 £)
US (2005 $)
Expon. (UK (2005 £))
Expon. (US (2005 $))
100,000,000
10,000,000
1,000,000
100,000
gdp
10,000
1,000
1770
1820
1870
year
1920
1970
2020

Figure 4.9.1



Financial Sector Capital - Q
Capital
returns
Capital
returns
returns
earnings
labour
Manufacturing/ Services Capital - K
Goods & services y
My
Individuals
Wealth - W
x
Mx
labour
earnings
Consumption

Figure 4.10.1

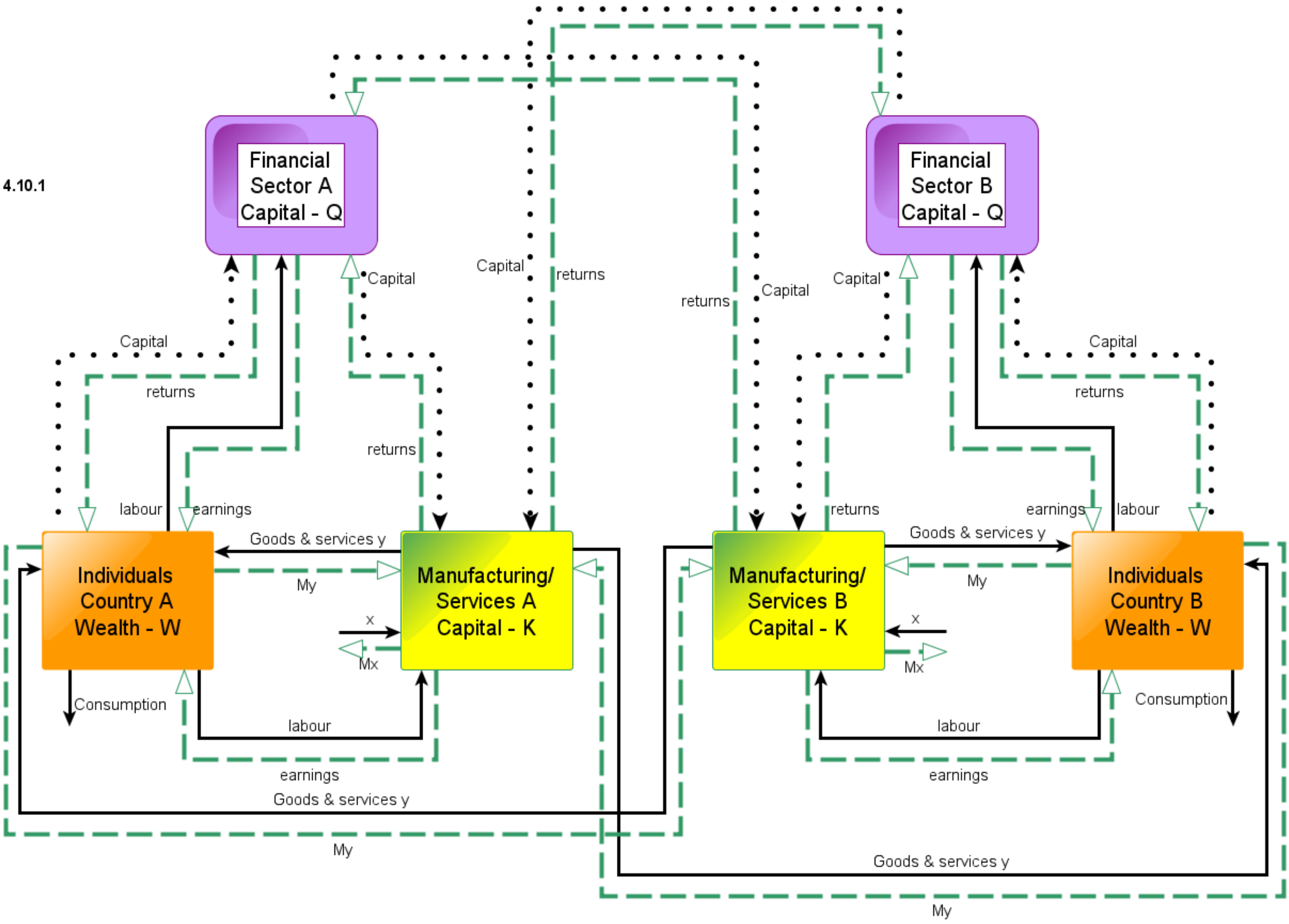


Figure 4.11.1

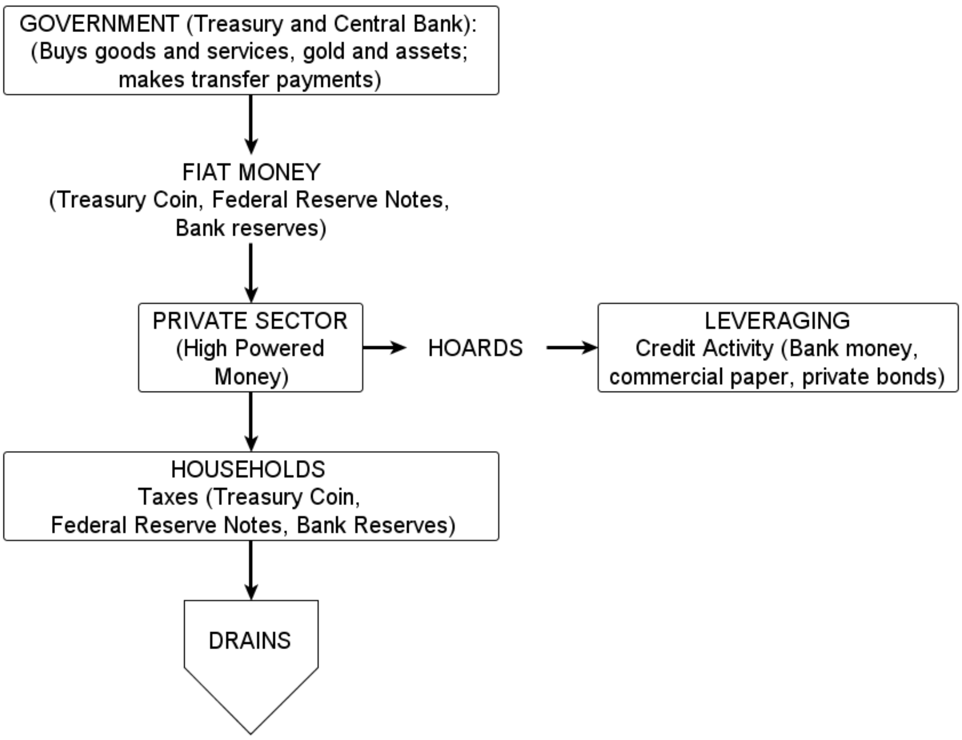
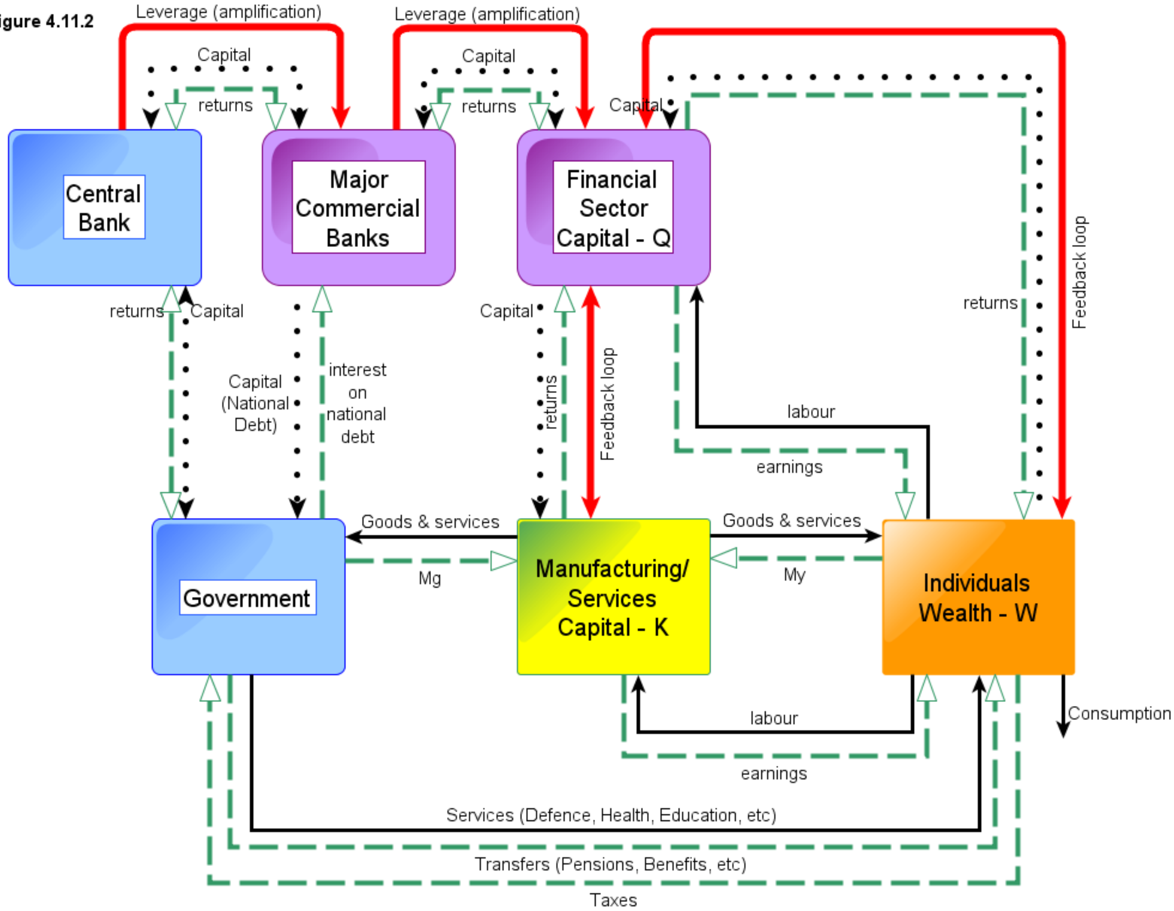


Figure 4.11.2



Leverage (amplification)
Leverage (amplification)
Capital
returns
Capital
returns
Capital
Central Bank
Major Commercial Banks
Financial Sector Capital - Q
returns
Capital
Capital (National Debt)
interest on national debt
Capital
returns
Feedback loop
labour
earnings
returns
Feedback loop
Government
Goods & services
Mg
Manufacturing/ Services Capital - K
Goods & services
My
Individuals Wealth - W
labour
earnings
Consumption
Services (Defence, Health, Education, etc)
Transfers (Pensions, Benefits, etc)
Taxes

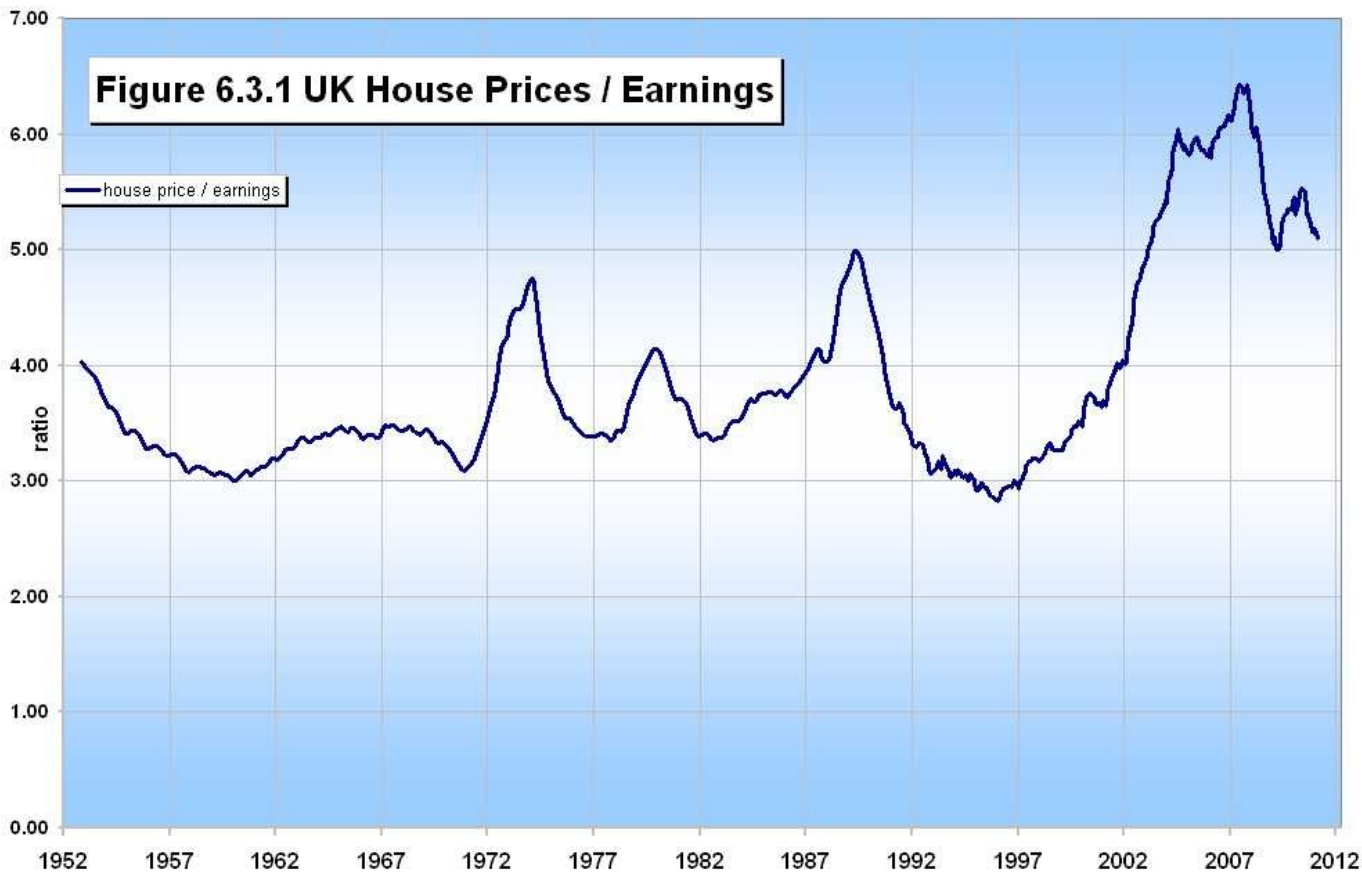


Figure 6.3.1 UK House Prices / Earnings
house price / earnings
ratio
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
1952
1957
1962
1967
1972
1977
1982
1987
1992
1997
2002
2007
2012

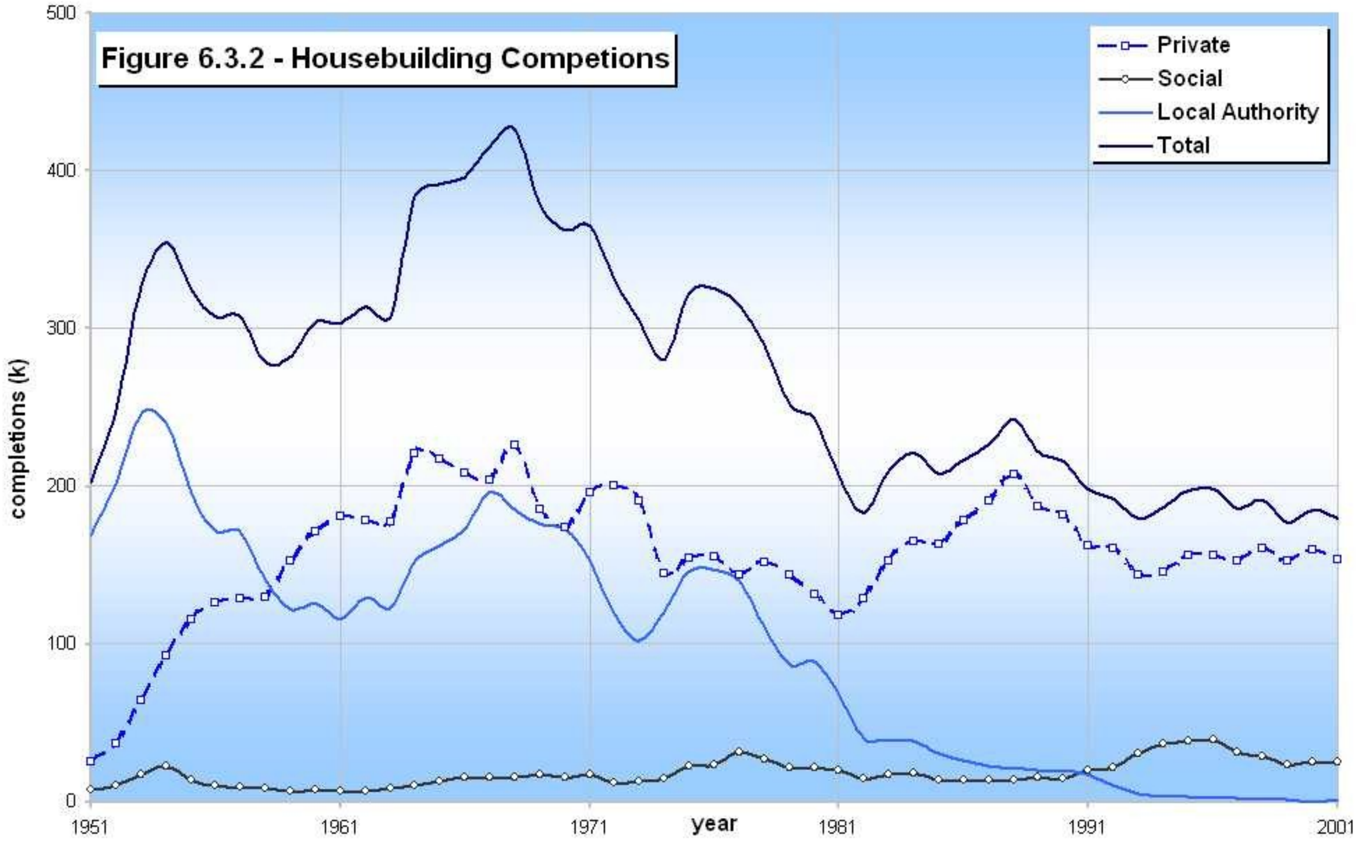


Figure 6.3.2 - Housebuilding Competions
Private
Social
Local Authority
Total
completions (k)
500
400
300
200
100
0
1951
1961
1971
year
1981
1991
2001


Figure 6.3.3 - US House Prices
Home Price Index
Index
250
200
150
100
50
0
1885
1905
1925
1945
Year
1965
1985
2005

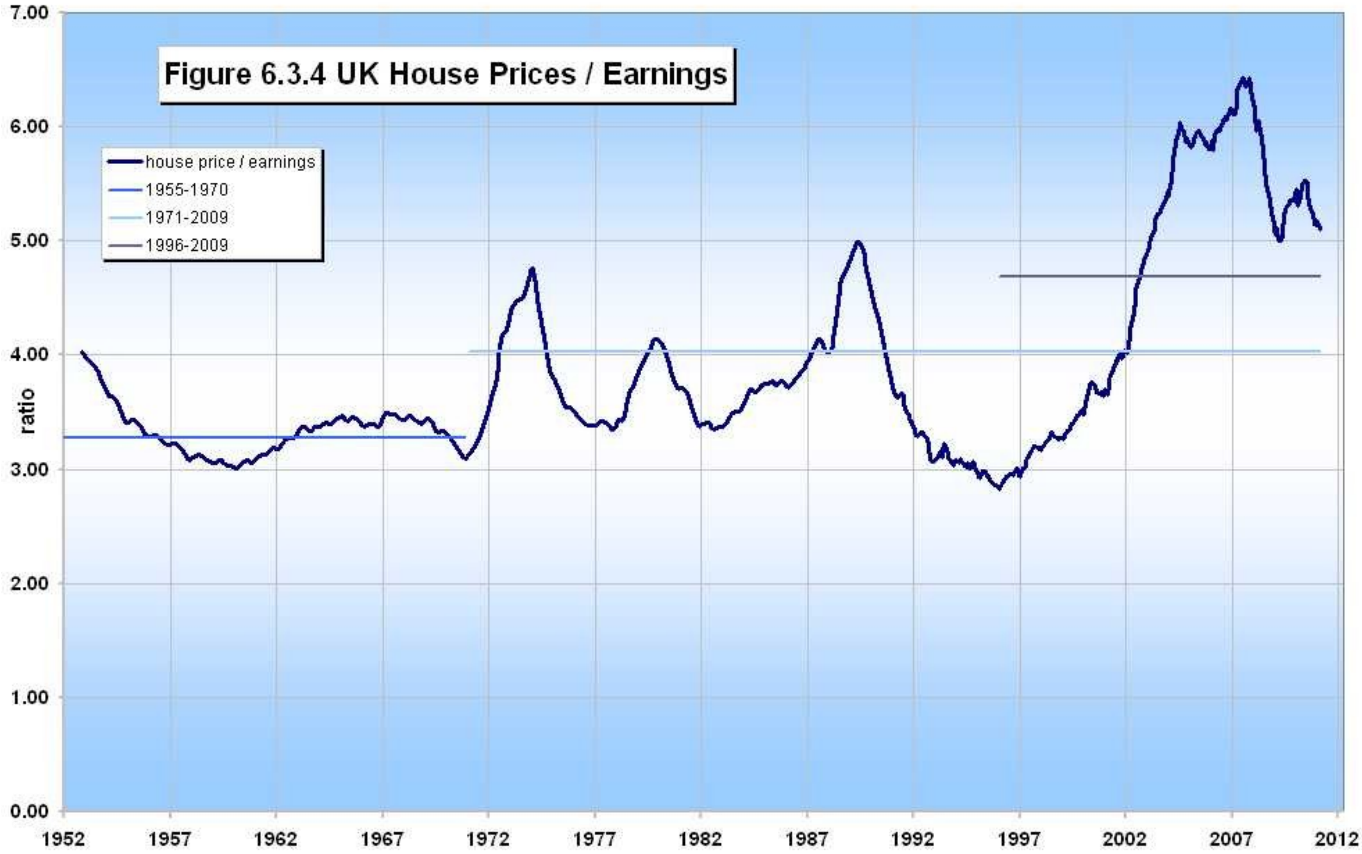


Figure 6.3.4 UK House Prices / Earnings
house price / earnings
1955-1970
1971-2009
1996-2009
ratio
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
1952
1957
1962
1967
1972
1977
1982
1987
1992
1997
2002
2007
2012

Fixed-rate vs. floating rate systems	
Land	**Rate adjustment (percentage of new business)** *
Belgium	F (75%), M (19%), V (6%)
Denmark	F (75%), M (10%), V (15%)
Germany	Mainly F and M
Greece	F (5%), M (15%), V (80%)
Spain	V (more than 75%)
France	F/M/O (86%), V (14%)
Ireland	V (70%), otherwise mainly M
Italy	F (28%)
Luxembourg	V (90%)
Netherlands	F (74%), M (19%), V (7%)
Austria	F (75%), V (25%)
Portugal	Mainly V
Finland	F (2%), V 97%), O (1%)
United Kingdom	V (72%), M (28%)

* Fixed (F): interest rate fixed for more than five years or until final maturity;
Mixed (M): interest rate fixed for one to five years;
Variable (V): interest rate renegotiable after one year or tied to market rates or adjustment at the lender's discretion
Other (O)

Source: ECB (2003)

Figure 6.3.5 [Hess & Holzhausen]

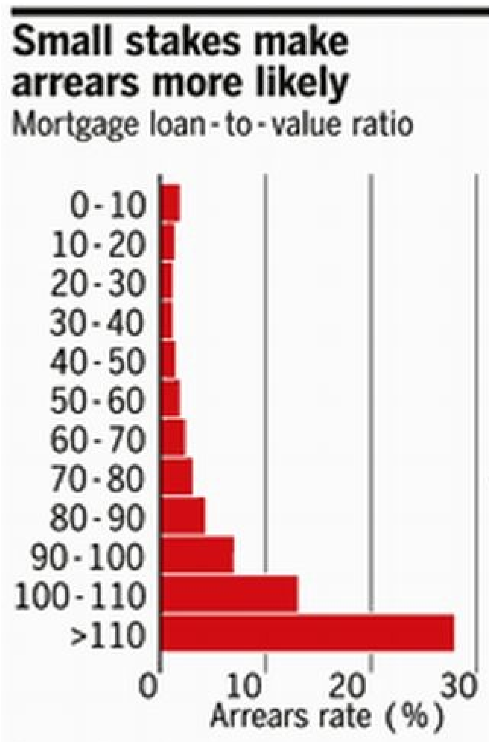
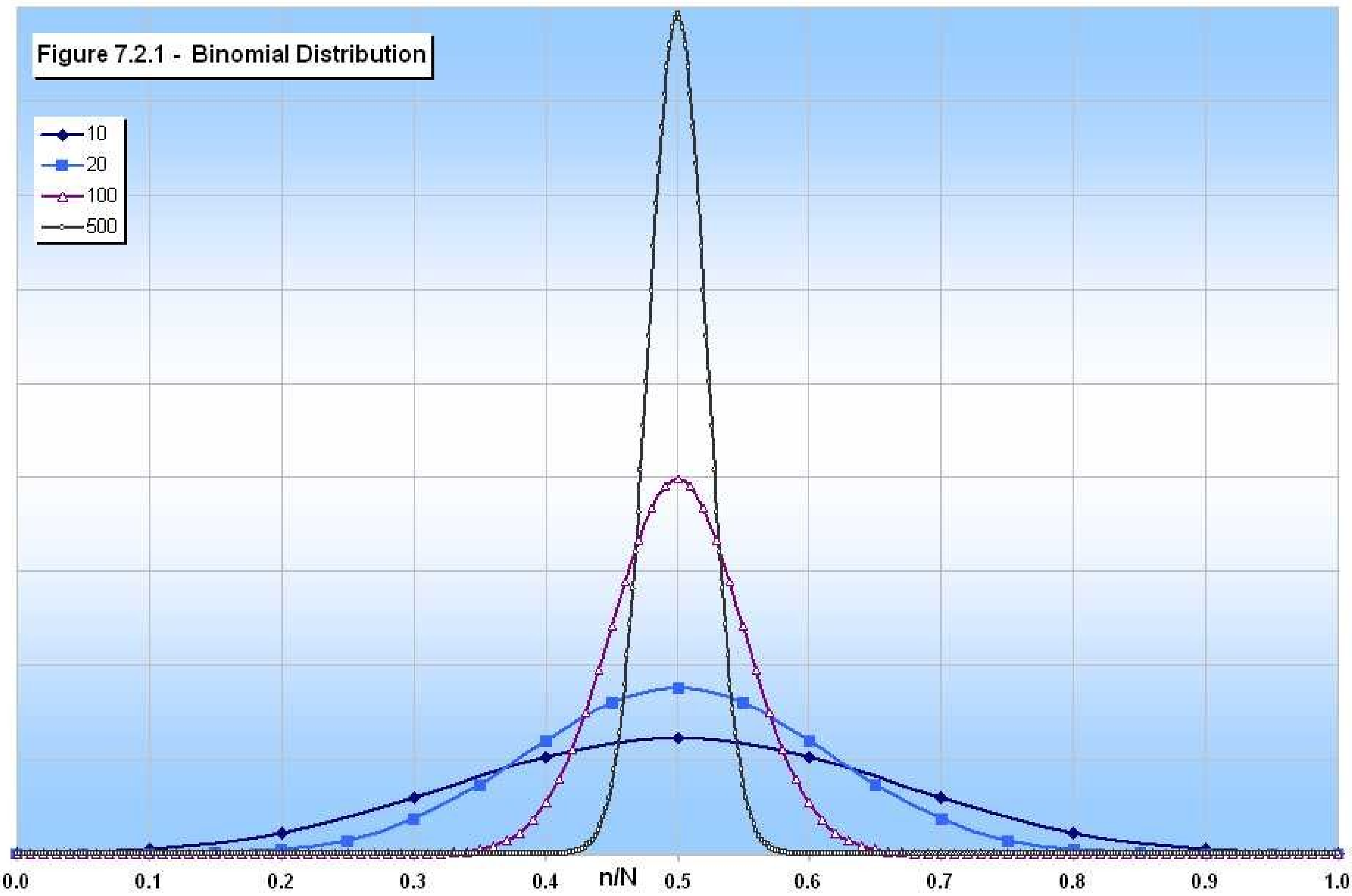


Figure 6.3.6



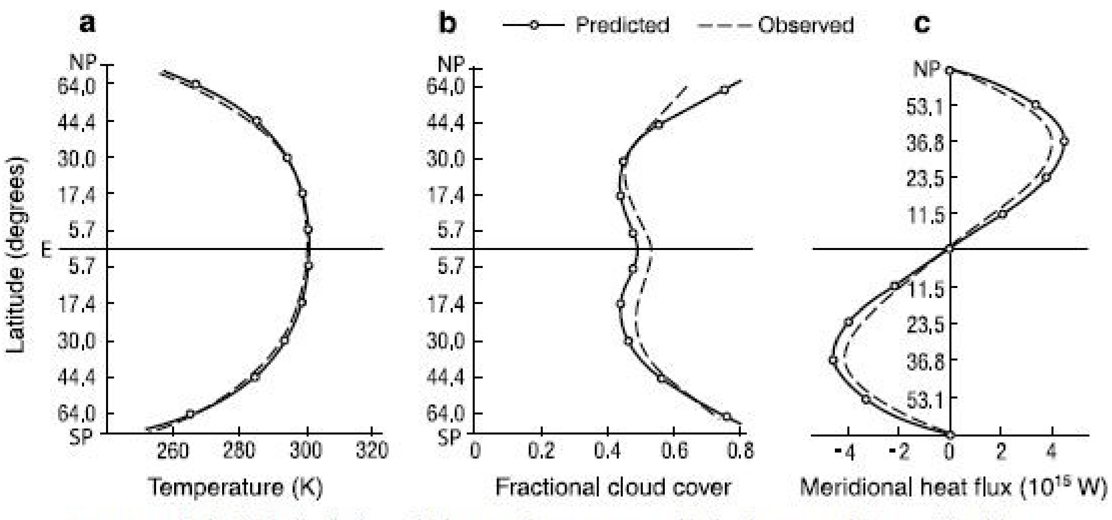


Figure 2. Latitudinal distributions of (a) mean air temperature, (b) cloud cover, and (c) meridional heat transport in the Earth. Solid line curves indicate those predicted with the constraint of maximum entropy production (equation (9)), and dashed lines indicate those observed. Reprinted from *Paltridge* [1975] with permission from the Royal Meteorological Society.

Figure 7.3.1

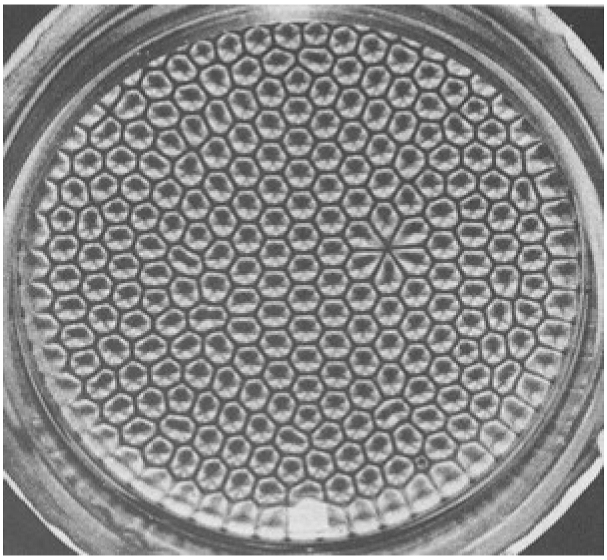


Figure 7.3.2

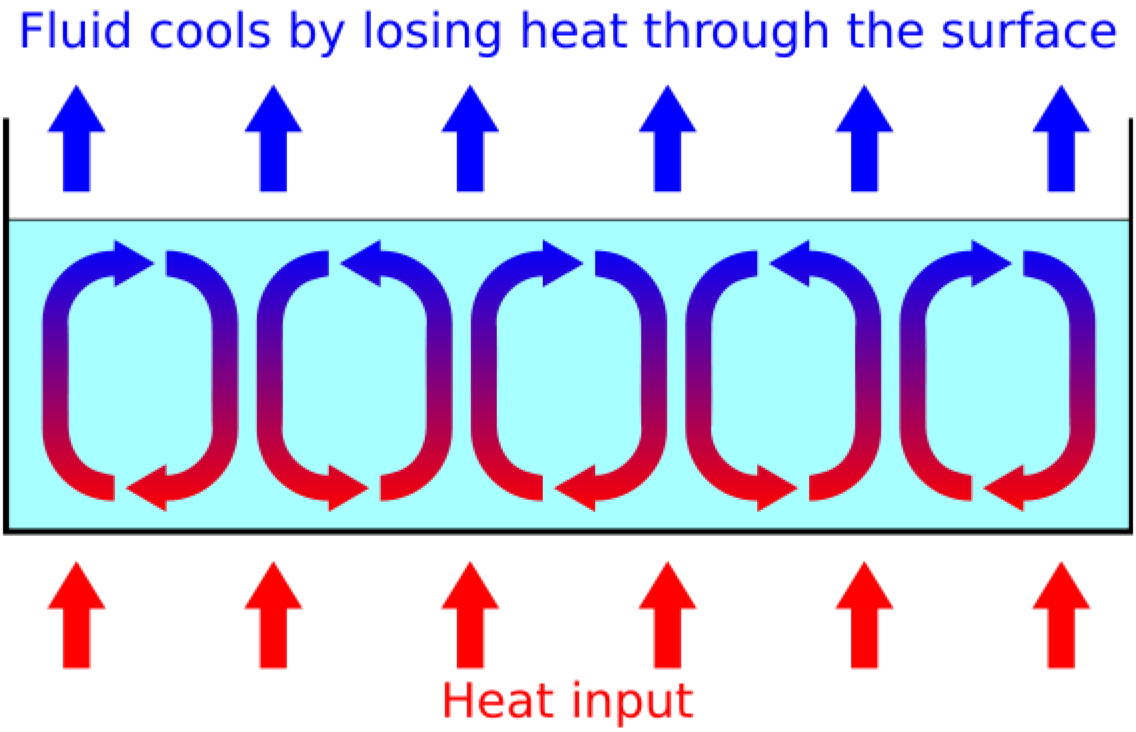


Figure 7.3.3

Figure 9.2.1

Figure 9.2.2

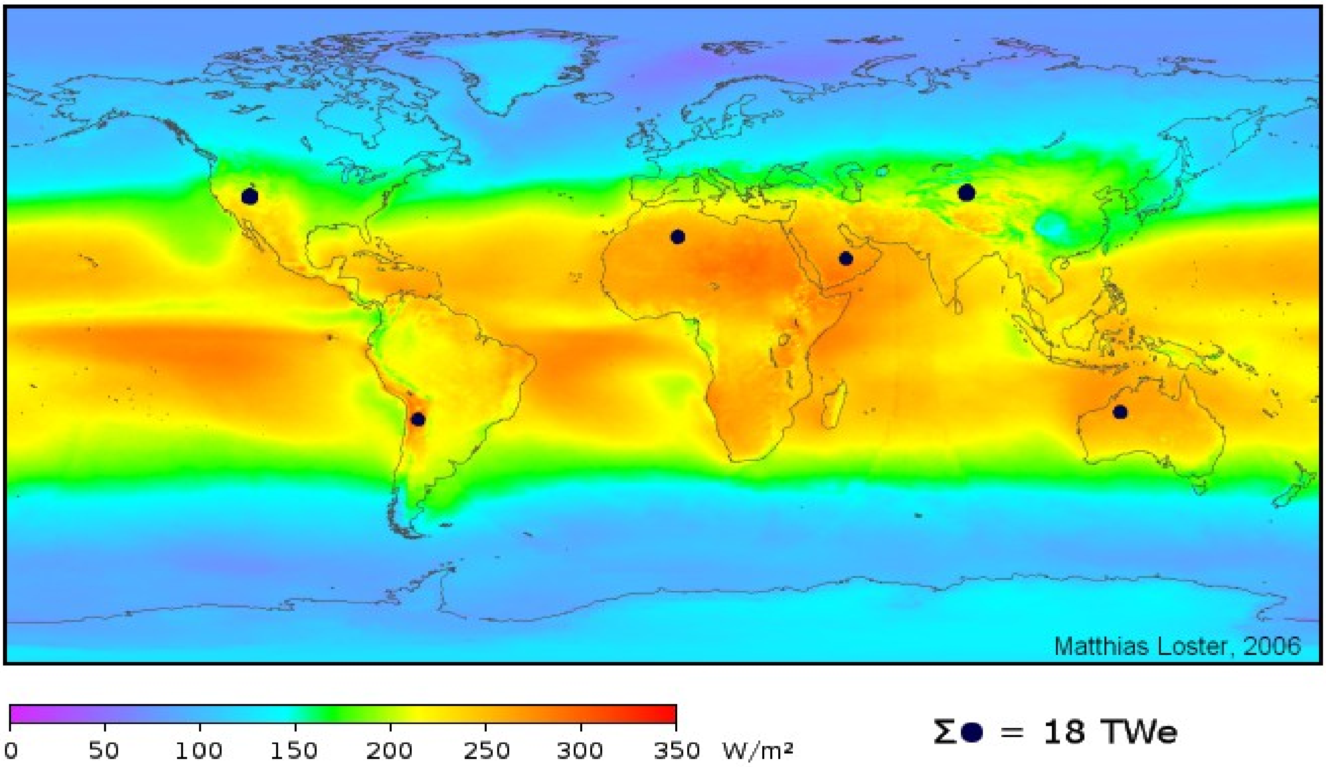


Figure 9.3.1

Figure 9.3.2

cleared of parasites.

With the increasing number of people travelling it might be a good idea if pharmacists made a habit of dispensing warning leaflets along with anti-malarials, describing the symptoms of the disease and urging sufferers to get themselves to a doctor.

And roll on the day someone develops an effective vaccine against malaria.

Sue Birchmore
Sparkhill, Birmingham

Crystal clear

In "Not liquid gold" (Technology, 13 February) a basic error was made in understanding the physics of ferroelectric liquid crystals (FLCs). The article attributed their behaviour to them "generating a magnetic field". In fact, magnetic fields have almost no part to play in detailing the behaviour of such systems.

In an FLC the molecules self-organise in such a fashion that individual molecular dipoles (electrical not magnetic) reinforce to produce a macroscopic electric polarisation. This is then switched in an external electric field. In a flat-panel display surface alignment forces are used to store this induced ordering in the absence of an applied electric field; hence the so called bistability or ferroelectric behaviour.

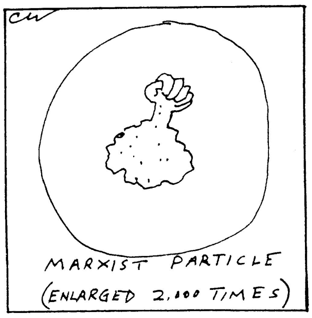
Harry Walton
Schuster Laboratory
University of Manchester

People and particles

If Professor Morrison is interested in applying ideas from modern physics to economics ("Complexity: Beyond chaos", special supplement, 6 February), he may be interested in looking at the clear parallels between thermodynamic systems and economic systems.

For example, the distribution of wealth between the freely interacting people of a state—typically, a lot of people without very much and fewer people with quite a lot—closely resembles the distribution of energy between freely interacting particles in an open thermodynamic system.

This may point to solutions of the eternal problem of capitalism, that although it is



good at creating wealth, it seems very poor at distributing it. For example, a mathematically trivial (but politically difficult) solution for the elimination of poverty would be the creation of a legally enforceable maximum wealth. If this was set low enough—at double the average wealth—this would produce a statistical distribution close to the distribution of abilities found in human societies.

Clearly such a solution is impractical as Britain is not a closed system and high energy particles would migrate to the Bahamas, Channel Islands, etc. However, it does point the way to more subtle ways of using statistical theory to create closed systems at low wealth levels which could be "pressurised" by governments, rather than using present systems of taxation and welfare which merely fight the existing statistical distribution head on and are therefore doomed to failure.

Mike Willis
Kirkby Stephen, Cumbria

Giants and dwarfs

John Gribbin's article, "On the shoulders of giants" (Forum, 13 February) implies that Newton was the originator of this famous remark. In their informative book, "Blueprints: Solving the mystery of evolution", Maitland A. Edey and Donald C. Johanson give the source as Bernard of Chartres (circa 1100): "We, like dwarfs on the shoulders of giants, can see more and farther, not because we are keener and taller, but because of the greatness by which we are carried and exalted."

No mention of dwarfs in Newton's remark but then he wouldn't want to give succour to the vertically-challenged Hooke, would he?

John Hunter
Fauldhouse, West Lothian

Totally phased

Tony Lang's article "Through a glass strangely" (Forum, 23 January) on prisms that refract the "wrong way" was no doubt intended to be humorous, but I found it quite disappointing and indicative of a lack of understanding of basic principles in optics.

Lang's entire thesis is based on the erroneous supposition that the index of refraction is the ratio of the speed of light in a vacuum to the optical group velocity. From this he concludes that materials making up the wrong-way prisms have index less than one and hence one could use such materials for faster-than-light communication and so forth. As every student knows, the index is the ratio of the vacuum velocity to the phase velocity. Lang's conclusion holds only for materials in which the phase and group velocity are the same, materials in which the frequency is proportional to the wave number. This condition in fact holds for no material, but can model dielectrics over a limited frequency range.

There are many materials where this is not so and that have index less than one. I have, sitting on my desk, a chunk of photonic band gap material for which the index of refraction is zero at a certain frequency. In this case the phase velocity is infinite! However, no instantaneous communication is possible because—as every student knows—it is the group velocity and not the phase velocity that determines the maximum speed of signal transmission. In fact the group velocity is zero at infinite phase velocity in my chunk of stuff.

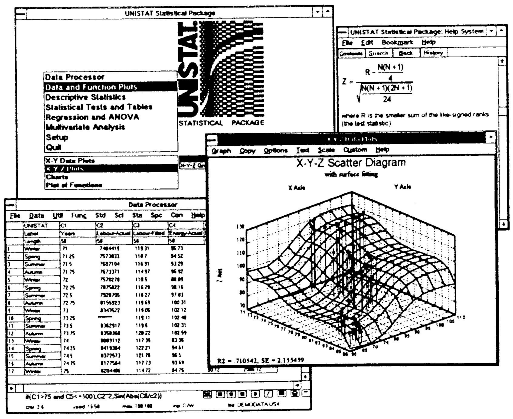
If I made a prism out of this material, it would refract electromagnetic waves "the wrong way" just as on the Pink Floyd album—one of my favourite records, by the way.

Jonathan Dowling
US Army Missile Command
Redstone Arsenal, Alabama

New names

Work on a *New Dictionary of National Biography* has now begun, funded by a grant from the British Academy. The *New DNB*, which will be published

UNISTAT for Windows



*** NEW * Version 1.2 ***

UNISTAT for Windows is a complete data handling, analysis and presentation package bringing together a powerful spreadsheet, a wide range of statistics and presentation quality 2D and 3D graphics.

UNISTAT features a true spreadsheet for scientific data handling. Export/import Lotus 1-2-3, Excel, dBase 2-3-4, .DIF, Sylk, fixed and free ascii, DDE with Excel. 2D and 3D colour plots, pie, bar, line, ribbon charts, error bars, confidence intervals, two y-axes, log options, distribution, curve and surface fitting, box and whisker plots, 2D and 3D histograms, normal plots, 2D-3D function plots with full resolution printer support. Descriptive statistics, histograms, 19 distribution functions, more than 40 parametric and non parametric tests, goodness of fit tests, correlation coefficients, contingency tables, cross-tabulation and break-down, regression, ANOVA, ANCOVA, cluster, discriminant, principal components, factor analyses, multi-dimensional scaling, canonical correlation.

UNISTAT Ltd, Unistat House, 4 Shirland Mews, Maida Vale, London W9 3DY
Tel: 081 964 1130, Fax 081 964 0531. Price: £495 + P&P + VAT.

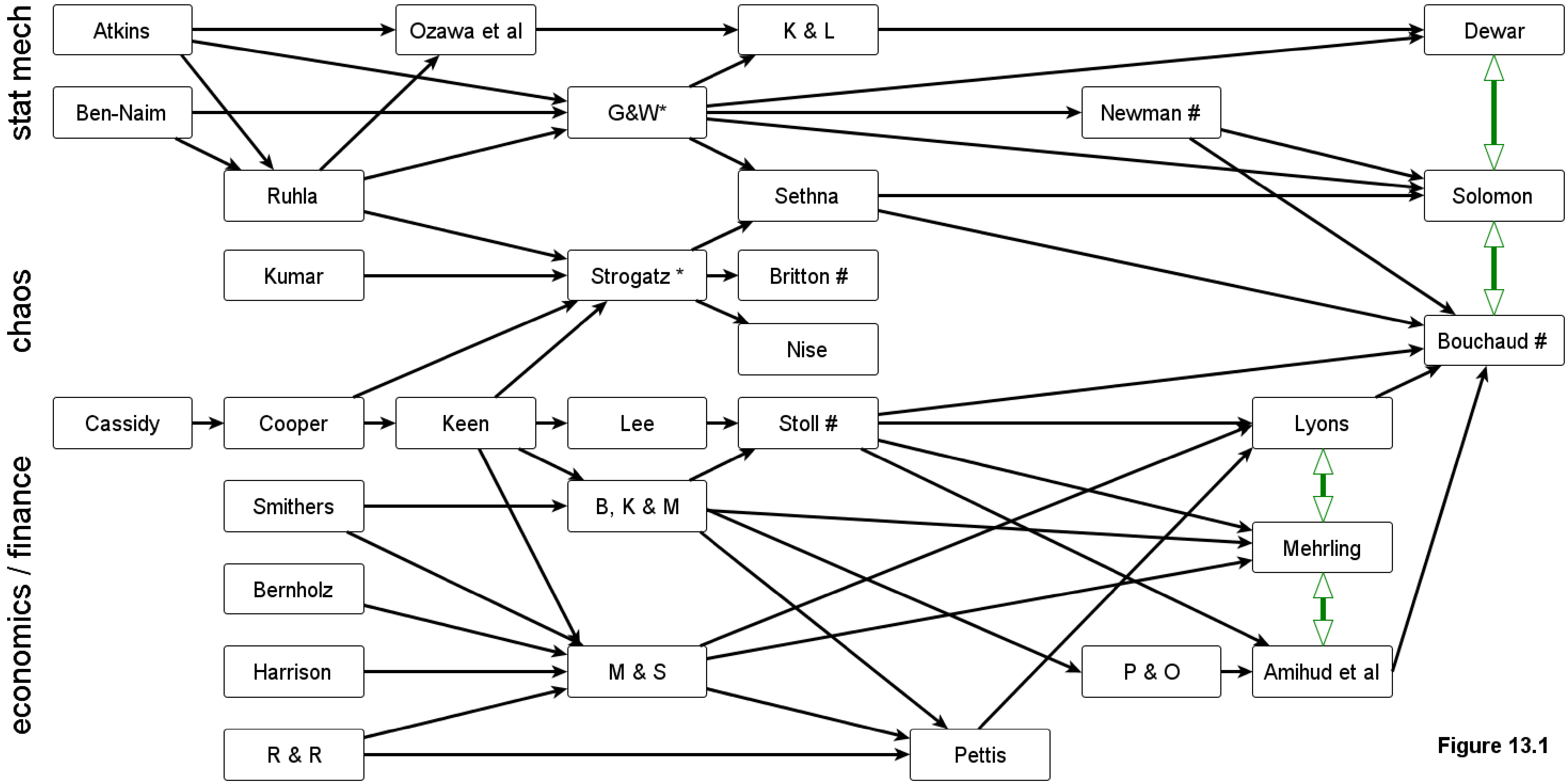

stat mech
chaos
economics / finance
Atkins
Ben-Naim
Ruhla
Kumar
Cassidy
Cooper
Smithers
Bernholz
Harrison
R & R
Ozawa et al
G&W*
Strogatz *
Keen
Lee
B, K & M
M & S
K & L
Sethna
Britton #
Nise
Stoll #
Pettis
Newman #
P & O
Lyons
Mehrling
Amihud et al
Dewar
Solomon
Bouchaud #

Figure 13.1