Issue no. 81
30 September 2017

support this journal and the WEA

visit the RWER Blog

Involuntary unemployment: a reminder

Roy H. Grieve [University of Strathclyde (retired)]

This note sets Keynes's theory of involuntary unemployment against neoclassical and New Classical approaches which fail to go beyond what Keynes described as voluntary unemployment.

Introduction

In an excellent paper Victor A. Beker ("Rethinking macroeconomics in the light of the US financial crisis," *Real-world Economics Review,* 2012) has drawn attention to the startlingly implausible explanation – offered by certain economic theorists – of why output and employment fell in the recent recession. He is referring to the ideas of the New Classical or Real Business Cycle school which holds that the economy, with rational optimizing agents operating in a context of price flexibility and constantly clearing markets, exhibits a constant state of full employment (albeit with the level of activity corresponding to "full employment" varying from time to time). Citing a recent analysis (Ohanian, 2010) of falling employment in the US, Beker comments:

> "Given the huge level of unemployment the crisis generated, it is not big news to know that the labour input sharply declined during that period. More surprising is the reason for that decline, according to Ohanian: the marginal rate of substitution between consumption and leisure was very low relative to the marginal product of labour. So, it seems that the crisis was caused by a sudden and mysterious increase in the preference for leisure. American workers suddenly decided to stay at home and watch TV instead of going to work. Of course, you are forced to reach that conclusion if you start assuming that the recession is an equilibrium outcome for agents who maximize their utilities. We are now again in a pre-Keynesian world where unemployment is always a voluntary decision by workers who have an increased preference for leisure compared with work" (Beker, 2012).

Beker's theme is that diagnoses of that of that degree of unrealism can only damage the reputation of macroeconomics and that it is high time therefore that this branch of economics was brought into closer contact with the real world. Accordingly he argues that there is a "need to rebuild macroeconomics" on a surer foundation than what would appear to be its present basis – the neoclassical faith that when rational agents operate under conditions of price flexibility the market mechanism can be relied upon to generate an optimal outcome. Beker's recommendation is that macroeconomics must again become "a discipline in which aggregate quantities play an essential role, while prices have only second order effects". Therefore, as he puts it in the concluding paragraph of his paper: "The first step in rethinking macroeconomics would be to rescue Keynes' original ideas".

The author of the present paper is wholly sympathetic to Beker's objective and to his suggested remedy: macroeconomics must be rescued from the trough into which it has fallen, and to achieve that, we need to return to Keynes's understanding of the working of the macrosystem. In fact the problem with contemporary macroeconomics is actually more widespread than might be inferred from Beker's focus on New Classical/RBC theory: mainstream macroeconomics – not only the more exotic New Classical variant – is itself in a poor state. While the New Classical set of ideas may be regarded by a majority of macroeconomists as rather extreme, the alternative "neoclassical" analysis more generally favoured and typically taught in the textbooks – an analysis very possibly described by its proponents as a form of "Keynesian macroeconomics" – is itself far from problem free. This mainstream or orthodox story, despite a superficial dressing in Keynesian language, can readily be shown to be, in essence, much more akin to pre-Keynesian theory than to the revolutionary thesis propounded in the *General Theory*. Essential elements of the Keynes theory disappear from view and in their place emerges a conception which recalls the conventional wisdom of times long past. It must be a potent source of misunderstanding that old "classical" ideas are surreptitiously rehabilitated and presented within a nominally "Keynesian" framework.

The purpose of the present note is to direct attention to Keynes's original analyses (1936; 1939) – nowadays all too frequently overlooked[1] – of the phenomenon of involuntary unemployment. At the present time it is very much worthwhile escaping from both the mainstream neoclassical thinking on the subject and from the more radical New Classical analyses. All such approaches completely ignore the essential – indeed revolutionary – point Keynes was making in the *General Theory* – that involuntary unemployment, as he defined it, resulted from a general deficiency of demand for the output that labour was potentially capable of producing. If workers were thus unemployed, it was not through their own doing: circumstances beyond their control were responsible.[2] It will be, we hope, instructive to notice how Keynes broke away from the conventional approach which focused attention on the conditions of labour supply, recognising instead that the state of demand in the labour market was the key determining factor, and that demand for labour was not simply an independent variable, but ultimately derived from demand for output in the commodities markets.

Involuntary unemployment

It was in the depths of the world-wide economic depression of the inter-war years that J. M. Keynes (in his *General Theory of Employment, Interest and Money*, published in 1936), identified the problem of abnormally high and persisting unemployment as one of *involuntary unemployment.*

He used the defining adjective "involuntary" to emphasise that, on his understanding, and contrary to the conventional view, the heavy unemployment of the time was not the fault of the unemployed themselves in demanding wages too high to permit their employment, but, on the

---

[1] On the drift away in recent years from Keynes's theory and the accompanying mainstream return to pre-Keynesian modes of thinking, see Grieve (2014).
[2] By "voluntary" unemployment Keynes meant unemployment which was attributable to the actions of the workforce itself. Thus (1936, p.6): "in addition to 'frictional' unemployment [I recognise the existence of] 'voluntary' unemployment due to the refusal or inability of a unit of labour, as a result of legislation or social practices or of combination for collective bargaining or of slow response to change or of mere human obstinacy, to accept a reward corresponding to the value of the product attributable to its marginal productivity.

contrary, of a lack of demand for the output their employment would have produced. If, as was case in the 1930s, employers feared that, for want of demand in the market, they would not be able to sell all the output they potentially could produce, they would not take on workers beyond the number whose output was expected to sell. In such circumstances, workers willing to work on terms compatible with their employment could, through no fault of their own, find themselves without a job. On Keynes's diagnosis, the root of the trouble lay *not* in the labour market, but in the markets for goods and services.

Thus Keynes explained the occurrence of the high level of contemporary unemployment as due to a falling-off in the desire to buy the products of industry, both producer and consumer goods. In particular he saw gloomy expectations of investment prospects and lack of confidence as responsible for a collapse in orders for capital goods, bringing unemployment; and, as employment and earnings in the capital goods sector fell, further contraction of demand and employment resulted throughout the economy. Individual national economies, of the UK, USA, Germany, and indeed the world economy as a whole, slid into a deep recession, reaching an equilibrium characterised by low output, underused productive capacity and high unemployment. The practical implication of the Keynesian analysis was that governments should direct their attention to stimulating demand for output, rather than attempting to boost employment via wage reductions. From the Keynes perspective cutting wages would, by further reducing incomes and spending, make the situation worse rather than better.

The "classical" theory of unemployment

In explaining general unemployment as being due to deficiency of demand Keynes was adopting a theoretical position which was, at that time, "revolutionary". In the dim and distant past, in the earlier days of industrialisation, observers of the economic scene such as the Rev Thomas Malthus, the Rev Thomas Chalmers, the Swiss thinker, J C L Sismondi, and indeed Karl Marx, had taken a similar view and had worried about the possibility of maintaining an overall balance between the ever-growing capacity of the economy *to produce* and the supporting demand of the community *to purchase* goods and services, raising the possibility of unemployment occurring if the desire to buy fell short of the system's ability to produce. But, long before Keynes's time, such fears had been dismissed by orthodox economic opinion as groundless; conventional wisdom came to believe that a general want of demand for output relative to productive capacity (except perhaps very temporarily on the occasion of a commercial crisis) was an impossibility. The observed unemployment of the inter-war years was interpreted as the consequence of excessive wages – of workers pricing themselves out of employment by demanding wages higher than could be afforded by employers. In fact, just prior to the publication of Keynes's *General Theory*, the eminent English economist Professor A. C. Pigou had expounded in his *Theory of Unemployment* (1933) the conventional (what Keynes called "classical") view that the problem was one of too high wages. He wrote:

> "Since the post-Armistice boom, however, the unemployment situation has been very different from what it was before the war. Instead of a percentage of unemployment amounting to an average over good and bad years, to around 41/2 per cent. post-war unemployment has moved around a mean from twice to three times as large as this. This circumstance suggests strongly that the goal of long-run tendencies in recent times has been a wage

substantially above that proper to nil unemployment, and that a substantial
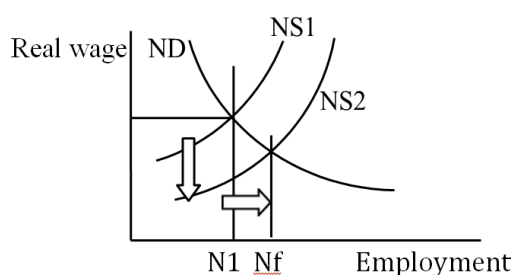part of post-war unemployment is attributable to that fact" (Pigou, 1933).

Pigou held that a reduction in money and (correspondingly, he believed) real wages would
give a powerful boost to employment. His conception of unemployment was as "voluntary" –
"voluntary" in the sense that the remedy for the situation lay in the hands of the workforce.
This thesis Keynes attacked and (at least for some decades) succeeded in thrusting from the
stage.

**The labour market (neoclassical account)**

The conventional (to Keynes, "classical" or as we would today call it "neoclassical") theory is
illustrated in Figure 1(a). Note that neoclassical theorists modelled the labour market exactly
in the same way as they would any market for a final consumption good: equilibrium is
established at the point of intersection of the demand and supply curves, demand and supply
each being a function of price (in this instance the real wage). ND is the neoclassical (so-
called) labour demand curve. NS is the usual, upward-sloping, labour supply curve. Pigou
himself [Fig. 1(b)], perhaps more realistically, assumed the NS curve to be of a reverse-L
shape, horizontal up to full employment.[3]) In the situation depicted in Figs.1(a) and 1(b)
unemployment exists, for the reason that the labour supply curve (NS1) happens to be set
"too high" to allow full employment. It is presumed that an appropriate change in the
conditions of labour supply (shifting from position NS1 to NS2) would cause employment to
increase to the full employment level (Nf).

Figure 1 Neoclassical representations of the labour market

(1a) The standard model   1(b) Pigou's model
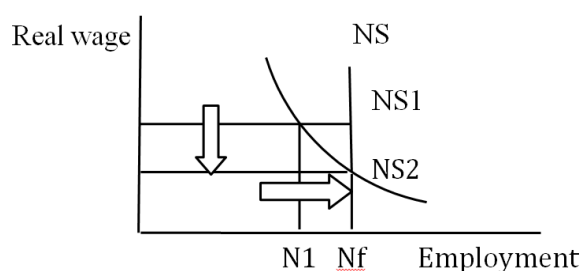
Consider this story more closely. It is supposed that when unemployment exists, if a lower
rate of wages were to be accepted, more labour would naturally be employed – on the tacit
assumption that the extra output thereby produced is guaranteed a market. That is to say,
underlying the conventional analysis of wages and employment is a Say's Law presumption
that intended aggregate demand for output can be expected to match the volume of output
offered for sale on the market; in other words that, as demand for output is "tame", there is no
need to worry about demand for the extra output produced.

---

[3] A.C.P. confirmed in a note to J.M.K. (May 1937) that he envisaged the labour supply curve as right-
angled with a reverse L-shape – the position of the vertical section indicating the number of "would-be
wage earners" and the height of the horizontal section indicating the wage for which labour "stipulates"
(Keynes, 1973, p.54).

It is important to recognise that there is something missing from this conventional representation of the labour market: the fact is that there is nothing in the standard diagram [Figure 1(a)] to show how demand for labour – as *derived* demand – may be affected by a change in demand for output. The so-called demand for labour curve – the downward-sloping (aggregate) marginal product of labour function – is in the short term fixed in position by the given conditions of technology: it cannot shift with changes in planned demand for output. That curve simply shows how the marginal productivity (by the neoclassical theory, the real wage) of labour varies with employment – whatever the level of employment may be. By contrast, in the short-term the labour supply curve alone can move (corresponding to changes in the terms on which labour is willing to work). Therefore, when utilising this diagram, the only feasible way of accounting for short term changes in the equilibrium level of employment is via shifts of the labour supply curve – *which is exactly how present day neoclassical theory explains changes in employment.*

All the familiar stories relying on misperceptions about the real value of money wages, on stickiness or on rigidity of money wages (as underlie the supply side of the commonly employed AD/AS analysis) explain changes in employment which follow from changes in expenditure as being due to shifts in the *real* terms on which labour is available for employment. (Figures 1(a) and 1(b)] That is to say, with respect to the neoclassical labour market diagram, the labour supply curve is understood to move relative to the demand for labour, thus altering the point of intersection of the two curves and implying a change in the level of employment. Unemployment emerging in such circumstances is what Keynes described as "voluntary" – and, unlike involuntary unemployment, is remediable by real wage adjustment.

What is wrong with the standard neoclassical labour market diagram is that the marginal product of labour schedule is incorrectly labelled as the "labour demand curve". That description ignores the fact that whatever the marginal product of labour may be, it cannot be profitable to employ and pay labour according to the notional value of its marginal product *if that product cannot actually find a market*. If we maintain, neoclassical fashion, the idea of diminishing marginal returns to labour in the short run, the reality (the satisfaction of this condition taken for granted by the conventional theory) is that if employment is to increase, it is necessary not only that the going wage falls as the marginal product of labour diminishes, but it is essential also that any increase in employment offered must be accompanied by a sufficient increase in demand for output to justify that extra employment.

If demand for output, and so for labour, increases or decreases, firms will move to the right or left along their short-run supply curves, with the marginal product of labour decreasing or increasing accordingly. In other words, with respect to the labour market diagram, the MPN curve should be read as a *real wage function*, showing how the marginal product of labour varies with employment – as employment (the dependent variable) varies according to changes in the demand for output. The proper designation of the MPN schedule in this context is therefore as a "real wage schedule showing real wages as a function of the level of employment not vice versa." But the MPN schedule itself cannot tell us – even if the conditions of labour supply are known - what the demand-determined volume of employment actually is.

The labour market: Keynes's vision (1936)

The essential novelty of Keynes's theory of employment is that he identified demand for output as the key factor determining employment. From the Keynes perspective, employment is, *ceteris paribus*, determined by conditions *outside* the labour market – as we have already mentioned - by the state of demand for output in the product markets. If, as Keynes believed, aggregate demand for output, in a world of uncertainty, depends on unstable expectations, hopes and fears, concerning an unknown future, the level of employment within the economy is liable to be affected by swings in expectations and business confidence. Persisting pessimism amongst businessmen implies persisting slump conditions: the labour market is then characterised by an excess supply of labour relative to demand – involuntary unemployment. How do we relate that understanding to the conventional model of the labour market?

As demand for labour depends crucially on conditions beyond the labour market, on the state of expected demand for output, we may, in order to depict the impact of demand for output on the labour market, impose on the conventional labour market diagram a vertical line indicating how much labour (corresponding to demand for output) is actually required in the labour market. This link between **the** markets we call the "derived demand for labour function" (DDN); see Figure 2. The DDN function shows how demand for labour in the labour market [see 2c)] derives initially from the demand for output in the products market, via the going conditions of production [see 2(a) and 2(b)].

Figure 2 The derived demand for labour function (DDN) and involuntary unemployment
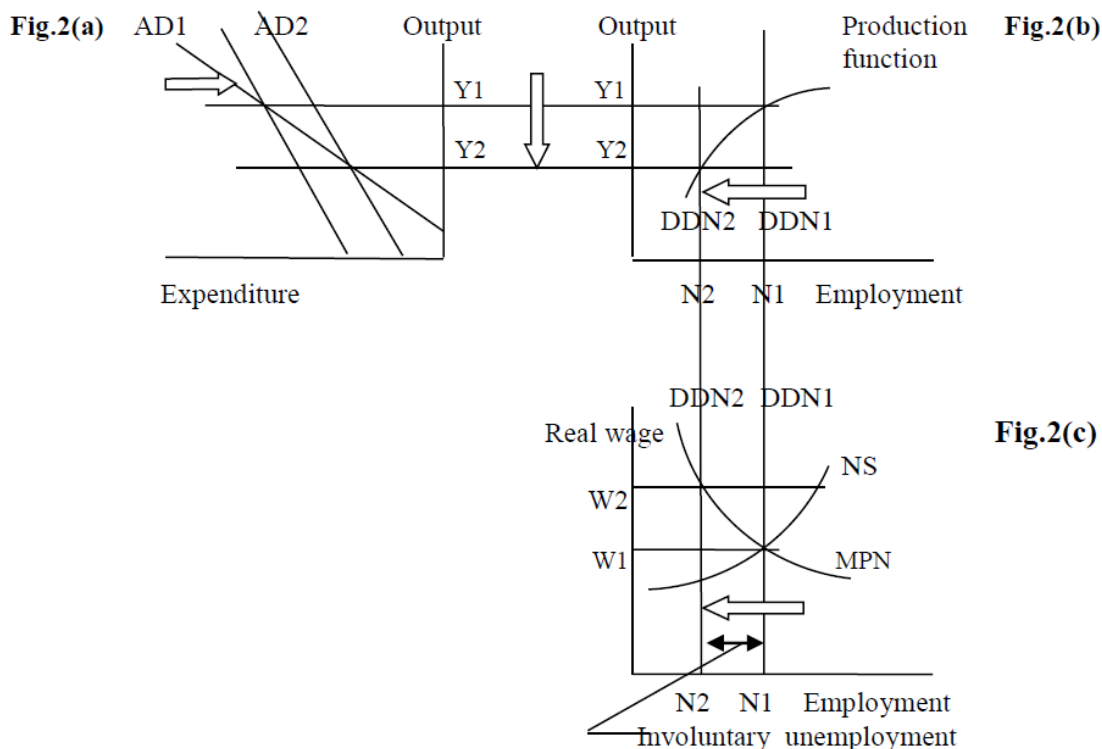
If we trace through the sequence of events depicted by this three-part diagram, the nature of Keynes's involuntary unemployment as demand-deficient employment should become clear. We consider first Keynes's 1936 account as presented in the *General Theory*. We then describe the simplified, but in basic principle *identical*, account advanced by Keynes in 1939. Consider the analysis of the *General Theory*. Start with the 45 degree line ("Keynesian cross") diagram in figure 2(a). We suppose that initially the aggregate demand function (AD1) lies in a position corresponding to full employment (income and output equal to Y1). The position of the "derived demand for labour" curve (DDN1) in diagrams 2(b) and 2(c) indicates, via the production function, that that level of output requires for its production the volume of employment N1 (corresponding to full employment) in the labour market.

Now suppose a fall occurs in aggregate demand, from AD1 to AD2. In response, with unsold inventories piling up, firms cut production and employment, so that output falls to Y2 and employment to N2. DDN shifts to position DDN2. In the labour market a gap emerges (N1 – N2) indicating the extent to which the contraction of final demand for output has brought about a decline in the demand for labour. (Note that equilibrium in the labour market is no longer established at the point of intersection of the MPN and labour supply curves.) This gap illustrates the presence of *demand-deficient* unemployment. It is also *involuntary* unemployment in that workers hitherto in employment have lost their jobs through no action or fault of their own. There has occurred no change either in the technological conditions of production and employment, nor in the terms on which labour is seeking employment: all that has happened is that conditions in the output markets have deteriorated, so that only a proportion of the previous volume of output can be sold.

But what about wages? If demand for labour has fallen and unemployment has emerged, the above model shows that real wages will have risen (W1 to W2). But Keynes in the *General Theory* makes the emphatic point that any such increase in real wages is the *consequence*, not the *cause*, of the rise in unemployment. What has happened is that, with demand for output falling firms have moved down their short run supply curves reducing employment (DDN moves to position DDN2) and implying, even with no alteration of money wages, a fall in commodity prices – and so some fall in the cost of living. In the face of falling demand, the workforce has certainly not pushed for an increase in *money* wages. The rise in real wages is simply an *incidental result* of the fall in demand for output, contraction of production and falling commodity prices; it is *not* the causative factor responsible for the fall inemployment.

Consider the opposite case – of a rise in aggregate demand - from AD2 to AD1. Output and employment increase, back, say, to Y1. DDN shifts from DDN2 to DDN1 and employment from N2 to N1. As output and employment rise, so does the cost of living, with prices rising (slightly) against unchanging money wages. Thus, real wages fall *permitting* movement down the MPN function, but that movement is initiated by, and occurs only with the rightward movement of the DDN curve. What do we make of this fact that real wages have fallen (W2 to W1)? Again, as in the equivalent case of a decrease in output and employment, the change in wages is the consequence, not the cause of the change in demand and employment. Such a reduction in real wages would not, *of itself*, have boosted demand to the full employment level. Keynes's reasoning here was that, when employment is increasing, even if the cost of living is rising a little, workers will not risk missing out on increased employment by insisting on higher money wages. In the *General Theory* Keynes put it thus:

> "[I]t is fortunate that the workers, though unconsciously, are instinctively more
> reasonable economists than the classical school, inasmuch as the resist

reductions of money-wages, which are seldom or never of an all-round character, even though the existing real equivalent of these wages exceeds the marginal disutility of the existing employment; whereas they do not resist reductions of real wages, which are associated with increases in aggregate employment and leave relative money-wages unchanged, unless the reduction proceeds so far as to threaten a reduction of the real wage below the marginal disutility of the existing volume of employment. Every trade union will put up some resistance to a cut in money-wages, however small. But since no trade union would dream of striking on every occasion of a rise in the cost of living, they do not raise the obstacle to any increase in aggregate employment which is attributed to them by the classical school" (pp.14-15).

In other words, it was Keynes's opinion in 1936 that if, with changes in aggregate demand, commodity prices alter relative to money wages, such increases or decreases in the price level, and so in real wages, while facilitating adjustment (expansion of employment) in the labour market, are not the operative cause of recovery. Note also that Keynes supposes that, even if *money* wages are sticky, *real* wages – over the relevant range - are not; they adjust as required allowing employers to respond to increases in demand for output by increasing the volume of employment offered. It was therefore reasonable, he believed, *to link changes in demand for output directly with changes in the demand for labour and with employment*, without any intervening complications on account of these real wage changes. (That of, course, is what is illustrated by the DDN, derived demand for labour curve, we have introduced.)

The message, therefore, of the *General Theory*, in a nutshell, is this: because demand for labour is derived demand depending on expected demand for output, fluctuations in effective demand for output give rise to corresponding fluctuations in the demand for labour. Demand for labour can (and does) vary relative to the given quantity of labour seeking employment. Unemployment so caused may be described as "demand-deficient" or "involuntary" unemployment. Under conditions of deficient demand for output, there is little possibility of the workforce being able to remedy the situation through their own efforts. Demand for output is what matters. If it were possible to reduce real wages (to a greater extent than required to accommodate a given increase of demand) that would actually tend to diminish effective demand, both directly via reduced consumption, and indirectly, thereby worsening rather than improving the employment situation. Likewise money wage reductions, leading to price reductions and general deflation, would also be more likely to damage rather than stimulate effective demand: negative wealth effects from the increasing real burden of debt, combined with destabilising expectations of continuing deflation might very well outweigh any notional positive real balance effect.[4]

At this stage in the discussion it is appropriate to note Keynes's own definition of involuntary unemployment:

---

[4] Note Patinkin (1959, pp.582-587) on the unreliability of deflation as a means of stimulating effective demand: "The economic adjustment process of the market is too unreliable to serve as the practical basis of a full-employment policy. In other words, though the real balance effect must be taken into account in our theoretical analysis, it is too weak – and, in some cases (due to adverse expectations) too perverse - to fulfil a significant role in our policy considerations."

"Men are involuntarily unemployed if, in the event of a small rise in the price of wage- goods relative to the money-wage, both the aggregate supply of labour willing to work for the current money-wage and the aggregate demand for it at that wage would be greater than the existing volume of employment" (Keynes, 1936, p.15).

On the basis of our discussion so far we can appreciate that Keynes is here proposing a thought experiment for the identification of involuntary unemployment. His point is that, if, starting with unemployment, in the event of increased spending and a small rise in the cost of living, more labour is willing to work, and becomes employed, than in the current situation, we observe that increased effective demand is bringing more people into employment. Alternatively, an increase in effective demand for output creates jobs for people hitherto without work but willing to work on terms consistent with their employment. In our Figure 2, a rightward shift of the DDN curve brings increased employment, with, at the same time, some rise in the cost of living (which, in the circumstances, is not resisted by the workforce).

Keynes's simplification (1939)

We have, however, not quite completed our exposition of Keynes's theory of effective demand and involuntary unemployment. What remains to be noticed is that, shortly after publication of the *General Theory* Keynes (1939) simplified the somewhat complicated story offered in the *General Theory* regarding wages and employment. On the basis of newly-gathered evidence (Dunlop, 1938 and Tarshis, 1939) he concluded that in reality (contrary to the standard neoclassical prediction) real wages did *not* fall as employment increased. In fact, on the evidence, no systematic short-term relationship appeared to exist between employment and real wages. Keynes decided therefore that it was unnecessary to offer the rationale presented in the *General Theory* as to why workers would not resist small reductions in real wages when employment was rising, meaning that, in the short run, output and employment could be understood to increase or decrease without any accompanying changes in real wages occurring. (See Figure 3).

Figure 3 A representation of Keynes's 1939 model of the labour market
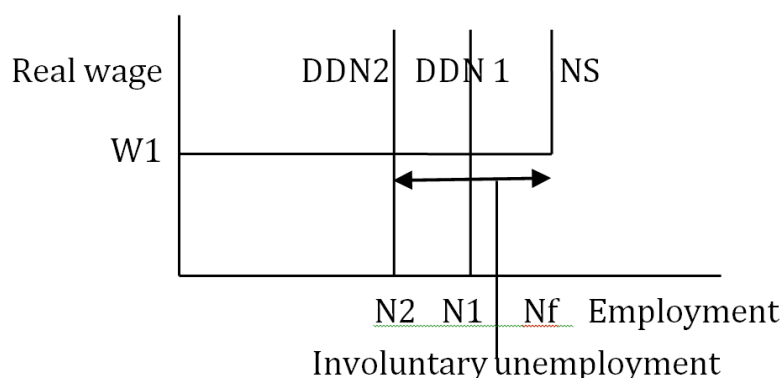
From this representation of the labour market, which no longer involves a downward-sloping MPN function, it is clearly seen that the going level of employment, and demand-led changes

in employment, are determined without reference to the current rate of real wages. Aggregate demand for output and employment vary together without any change in wages being implied. Involuntary unemployment is unambiguously attributable to want of demand for output.

Figure 3 presents essentially the same understanding as does Figure 2 of how the situation in the labour market depends on the state of demand in the product markets. (Note we have borrowed the Pigouian reverse L-shaped labour supply curve.) The labour supply curve (NS) shows how much labour is available at the going (real) wage W1. It is understood that that wage corresponds to the (invariant) marginal value placed by employers on all labour available for employment (maximum labour supply, Nf). The vertical lines DDN1 and DDN2 show the derived demand for labour according to the levels of aggregate demand for output. The quantity of employment offered is determined at the point of intersection of the relevant DDN curve with the labour supply function. DDN moves rightwards or leftwards according to whether aggregate demand for output is rising or falling; the wage rate remains constant at W1. Any gap which exists between the equilibrium level of employment thus indicated and the level of full employment (Nf) represents demand-deficient unemployment, which, by its nature, is involuntary unemployment.

Other categories of unemployment

We should note here that a distinction may be drawn between involuntary and other – "frictional" and "structural" – types of unemployment. The former of these refers to the normal turnover of people who – for whatever reason – are temporarily between jobs; the latter denotes more difficult conditions due to a changing industrial structure, when workers redundant in a declining sector may not find it easy to secure employment in an expanding one. While not denying that these types of unemployment may well involve a painful involuntary element and do require remedial action, we distinguish these long-understood categories of unemployment from Keynes's demand deficient involuntary unemployment for the reason that these are problems best dealt with by policies other than those needed to cope with a general, economy-wide deficiency of demand.

Note also that it has been suggested that certain other theoretical models also describe situations of involuntary unemployment: "implicit contract theory" (Azariadis, 1975), "staggered wage setting" (Taylor, 1979) and "efficiency wages" (Shapiro and Stiglitz, 1984) have been mentioned in this context. But in so far as these involve price stickiness and artificially increased wages they do not seem to fall within the scope of Keynes's own definition of involuntary unemployment.

New Classical Macroeconomics: denial of involuntary unemployment

For 30 or so years after the publication of the *General Theory*, Keynes's understanding of the working of the macroeconomy essentially constituted the basis of a "Keynesian" orthodoxy, with involuntary unemployment understood as the consequence of deficient demand. But from the late 1960s or early 1970s, old ideas began to return to fashion, with a pre-Keynesian emphasis on conditions of labour supply as responsible for unemployment. What is known as the "New Classical Macroeconomics" has come to play a prominent role in current theoretical discussion. The theories advanced by this school of thought with respect to unemployment and the working of the labour market essentially represent a rehabilitation of the old classical

approach of Professor Pigou in that the focus is again on *conditions of labour supply* as critical to explaining what is happening to employment. From this perspective there is no such thing as involuntary unemployment, with labour's desire to work frustrated by a dearth of employment opportunities.

This New Classical analysis first appeared (Friedman, 1968; 1975; Phelps, 1975) in the context of attempts to prove that it is impossible, through macroeconomic management, to maintain over time a level of employment in excess of the so-called "natural rate of unemployment." The argument was that while, initially, as implied by the Phillips Curve, a higher level of demand could be bought at the price of a somewhat higher rate of inflation, in the longer term no such trade-off was possible as the "unnaturally" high level of activity could only be sustained by constantly increasing the rate of inflation. This scenario was derived via a return to the old classical model of the labour market, with employment determined at the intersection of the labour demand and supply curves, both labour demand and labour supply being taken as functions of the real wage. This analysis soon found its way into general use in macroeconomic theory. We can say that, with its introduction as a key element of the popular AD/AS macroeconomic model, the New Classical theory of the working of the labour market has become a well-established part of mainstream macroeconomics.

Consider how the system is supposed to operate. Starting from a situation of in the labour market of equilibrium at the natural rate (full employment), suppose a change in spending on goods and services – say, an increase in aggregate demand. According to the New Classical theory (which characteristically focuses on the efficacy of the price mechanism in a world without uncertainty) the sequence of events is as follows. An increase in spending causes commodity prices to rise: in the labour market the VMP (value marginal product) curve rises against the labour supply curve. Money wages increase and workers – not anticipating or perceiving a corresponding increase in commodity prices – misinterpret the increase in money wages for a sustained increase in real wages: the supply of labour therefore increases and employment rises. But once the workforce appreciates that goods prices have also risen, labour supply would return to its original level were it not that, with employment still in excess of the natural rate, money prices (and so money wages) continue to rise. This process of wage and price inflation, with wages playing catch-up on prices, continues until prices and wages have risen equi-proportionately; at which point the real wage has returned to its equilibrium value and employment is back at the natural rate. All the time, throughout this process of change, employment adjusts to accord with the wishes (even if ill-informed) of the workforce. An equivalent story could be told of the short-term and longer-term consequences of a decrease in aggregate spending on output: output and employment would fall below the natural rate as labour was withdrawn from employment on the "misperception" that real wages were falling, and employment would in time increase as, through the adjustment process, perceived real wages are restored to their "natural" value.

Note what this New Classical story implies. Output changes in response to changes in demand because, via the effects of these spending changes on commodity prices and thus on (perceived) real wages, employment is understood to rise or fall, thereby permitting output to change. From the point of view of labour, these changes in employment are voluntary: for instance, when employment is relatively low workers do not consider they have been forced out of work, rather their situation is that, in the circumstances, they prefer leisure to work. This is certainly not a situation of involuntary unemployment; labour is never "off its supply curve." We may add that, according to the New Classical theory, not only is this "unemployment" voluntary – it is also temporary and self-correcting: as, in time, with the confusion about the

real reward for working eliminated, employment returns to its normal (equilibrium) level. All in all, from this perspective, the effects of a downturn in aggregate demand for output do not seem to matter nearly as much as they do from the Keynesian viewpoint. As a recent critic asked, "do New Classical theorists really believe that in the years of the Great Depression, workers had simply chosen to enjoy a particularly long holiday?"

While the above "misperceptions" explanation of unemployment as resulting (temporarily) from changes in demand for output seems well embedded in present day mainstream macro theory, another strand of New Classical thought (but rather more on the fringe of the mainstream), the "Real Business Cycle Theory" (Stadler, 1994) likewise proposes that fluctuations in employment be read as corresponding to voluntary changes in the supply of labour offered for employment. This line of thought links changes in labour supply with changes in technology which are said to cause increases or decreases in the marginal product of labour, and so in the wages offered to workers. The theory is that workers will choose to work more (offer more labour), earn high wages and save in times of high productivity, in order to finance leisure from work at times of low productivity and low earnings. This appears to be a theory of the variations in the number of employed which, over time, constitutes "full employment." Perhaps, not surprisingly, the RBC theory remains a minority taste – one reason (not to mention doubts about the behaviour predicted) being that its proponents have found difficulty in identifying the technological changes said to engender the behaviour in question.

Conclusions

Within economics views differ on the subject of involuntary unemployment. The Keynesian side of the profession has no doubts that the concept is a realistic and relevant one, recognising involuntary unemployment as a damaging economic phenomenon which imposes high costs on society. Mancur Olsen (1982) condemned economists "who denied the concept of involuntary unemployment and put their theories ahead of 'common sense and the observations and experiences of literally hundreds of millions of people.'"[5] On the other hand some economists refuse to admit that such a phenomenon exists. Lucas (1978) claimed that "an unemployed worker at any time can always find a job at once."[6] The truth of that statement may be doubted, and even in the fortunate event of a marginally less unpalatable option – such as selling the *Big Issue* being available – that may properly be regarded not so much as a proper job, more of a desperate emergency measure.[7]

Although the idea of involuntary unemployment has largely disappeared from mainstream macroeconomics, that does not mean that the problem has disappeared from the real world; rather it points to the fact that much modern macrotheory, in assuming that agents possess virtually full knowledge of the future, and in losing sight of Keynes's penetrating understanding of the working of the macro economy, has altogether cut itself off from any hope of understanding fluctuations in economic activity and the causes of unemployment in the real world.

---

[5] Olsen (1982) observation quoted from Wikipedia entry on "Involuntary Unemployment".
[6] Likewise the Lucas (1978) quotation.
[7] Compare Joan Robinson (1936) on "disguised unemployment".

References

Azariadis, C. (1975) "Implicit contracts and underemployment equilibria," *Journal of Political Economy*, 83, pp.1183-202.

Beker, V. A. (2012) "Rethinking macroeconomics in the light of the US financial crisis," *real-world economics review*, 60, pp.120-138.

Dunlop, J. T. (1938) "The Movement of Real and Money Wages', Economic Journal, 48, pp.282-287.

Friedman, M. (1968) "The role of monetary policy," *American Economic Review*, 58, pp.1- 17.

Friedman, M. (1975) *Unemployment* versus *Inflation*? London: Institute of Economic Affairs.

Grieve, R.H. (2014) "Right back where we started from – from 'the classics' to Keynes and back again," *real-world economics review*, 68, pp.41-61.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan.

Keynes, J. M. (1939) "Relative Movements of Real Wages and Output," *Economic Journal*, 49, pp.34-51.

Keynes, J. M. (1973) *Collected Writings*, Vol XIV, London: Macmillan for the Royal Economic Society.

Lucas, R. E. (1978) "Unemployment Policy," *American Economic Review*, 68 (2), pp.353-7.

Olson, M. (1982) *The rise and decline of nations: economic growth, stagflation, and social rigidities*, New Haven: Yale University Press.

Ohanian, L. (2010) "The Economic Crisis from a Neoclassical Perspective," *Journal of Economic Perspectives*, 24 (4), pp.45-66.

Patinkin, D. (1959), "Keynes's Economics Rehabilitated: A Rejoinder to Professor Hicks," *Economic Journal*, 69 (2), pp.582-587.

Phelps, E. S. (1970), "Money-wage dynamics and labour market equilibrium" in E S Phelps *et al* (eds.), *Microeconomic Foundations of Employment and Inflation Theory*, New York: W. W. Norton.

Pigou, A. C. (1933), *The Theory of Unemployment*, London: Macmillan.

Robinson, J. (1936) "Disguised Unemployment," *Economic Journal*, 46, pp.225-37.

Shapiro, C. and Stiglitz, J. E. (1985) "Can Unemployment Be Involuntary? Reply," *American Economic Review*, 75 (5), pp.1215-17.

Stadler, G. W. (1994) "Real Business Cycles," *Journal of Economic Literature*, 32, pp.1750-83.

Tarshis, L. (1939) "Changes in Real and Money Wages," *Economic Journal*, 49, pp.150-154.

Taylor, J. B. (1979) "Staggered wage setting in a macro model," *American Economic Review, Papers and Proceedings*, 69 (2), pp.101-113.

Author contact: roygrieve@btinternet.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

Fixing the euro's original sins: the monetary – fiscal architecture and monetary policy conduct[1]

Thomas Palley   [Washington, DC]

Abstract

The euro zone (EZ) was created in January 1999. Its weak economic performance is significantly due to the euro's neoliberal monetary architecture and the design of monetary policy. Those features undermine national political sovereignty and consign the EZ to severe economic under-performance, which in turn fosters political demands for exit from the euro. Escaping this dynamic requires restoring fiscal space to EZ countries, and also changing the design of EZ monetary policy. The paper shows how this can be done. It decomposes the challenge of reform into generic problems related to the neoliberal construction of monetary policy, and specific problems concerning the euro as a currency union. The currency union problems are further decomposed into "money – fiscal policy" architecture problems and specific monetary policy conduct problems.

1. The euro's twin original sins

The euro is afflicted by twin original sins: rupture of the money – fiscal policy link and adoption of neoliberally designed monetary policies. Those twin sins have contributed to generating dismal economic outcomes, which have fostered ugly political conditions that echo the 1930s and risk causing the euro to disintegrate.

This paper shows the euro's twin original sins can be fixed in a politically viable manner. As regards economics, the euro is a monetary phenomenon, which means that getting the monetary architecture right is the *sine qua non* for success. Other economic policy adjustments can then further strengthen the euro zone's (EZ's) economic performance, but without the right monetary architecture economic success will inevitably prove elusive.

As regards politics, the fundamental problem is the EZ consists of national political sovereigns that have been required to surrender monetary sovereignty. However, those national political sovereigns need a degree of monetary sovereignty in order to defend their public finances and pursue expansionary fiscal policy in times of economic distress. The EZ's architecture makes little provision for this, because of a combination of fears of moral hazard from country bail-outs and intellectual blindness. Fixing the EZ's monetary architecture and restoring a degree of monetary sovereignty is essential for creating the policy space needed by national governments to make the euro politically viable.

2. Diagnosing the EZ's problems

The euro was introduced in January 1999. As shown in Table 1, its macroeconomic performance was barely satisfactory prior to the financial crisis of 2008, but it has been dismal since. Since peaking in the 1960s, EZ average GDP growth each decade fell steadily through

---

[1] An earlier version of this paper was published in H. Herr, J. Priewe, and A. Watt (eds.), *Saving the Euro: Redesigning Euro Area Economic Governance*, Social Europe Publishing, 2017.

the 1990s. The introduction of the euro saw a brief uptick, but growth has collapsed since the 2008 financial crisis. That story is mirrored in the unemployment rate which steadily increased through to the 1990s, then fell slightly with the euro's advent, but surged to sustained record highs after the financial crisis.

Table 1 EZ output growth and unemployment rate

	Region	1961-70	1971-80	1981-90	1991-98	1999-2008	2009-2015
Average GDP growth rate at 2010 market prices (%)	EZ-12	5.3	3.4	2.4	1.9	2.1	0.1
	EZ-19					2.1	0.1
Average unemployment rate (%)	EZ-12	2.3	4.0	8.5	10.0	8.5	10.8
	EZ-19					8.7	10.8

Source: Statistical Annex of the European Economy, Spring 2016 (Tables 3 and 10) plus author's calculations

Behind this data is a dismal economic policy history. That history begins with the adoption of tough anti-inflation policy in the late 1970s, which turned into neoliberalism in the early 1980s. Consequently, Europe never fully recovered from the dislocations of the 1970s. The neoliberal turn was further locked in place in the 1990s with the first steps to monetary union via the Maastricht Treaty and its imposition of strict euro zone economic convergence criteria, requiring a deflationary policy posture to meet them. Come the euro, there was a brief boom in the 2000s fueled by the intersection of low interest rates and speculation. However, when the bust arrived with the 2008 crisis, the design flaws in the euro's monetary architecture and policy conduct surfaced with a vengeance. Those flaws are systemic and remain largely unresolved. Consequently, they now pose an existential threat to the euro.

The weakness of the EZ's economic performance is significantly rooted in its monetary architecture and monetary policy conduct. As regards architecture, the design of the euro's monetary policy institutions has massively shrunk the space for national fiscal policy and also exposed government finances to market instability. Under the old system of national currencies, each country government had a central bank that acted as the "government banker". Thus, national central banks helped governments finance their budget deficit, and also defended government bonds against speculative attack. This government banker function was completely and mistakenly ignored by the euro's creators, thereby weakening governments' ability to finance fiscal policy and giving financial markets massive power over them (Palley, 2011a, 2011b).

Simultaneously, EZ monetary policy conduct has been sub-optimal. It was blind to asset price bubbles before the crisis; was slow to respond in the crisis; and the two percent inflation target risks being an unnecessary brake on performance if the EZ escapes the current stagnation.

Figure 1 outlines the nature of the problem. It decomposes the challenge of EZ monetary reform into generic problems related to the neoliberal construction of monetary policy, and specific problems concerning the euro as a currency union. The currency union problems are then further decomposed into architecture problems and conduct of policy problems.

Figure 1 A diagnosis of the failings afflicting the EZ's monetary architecture and policy
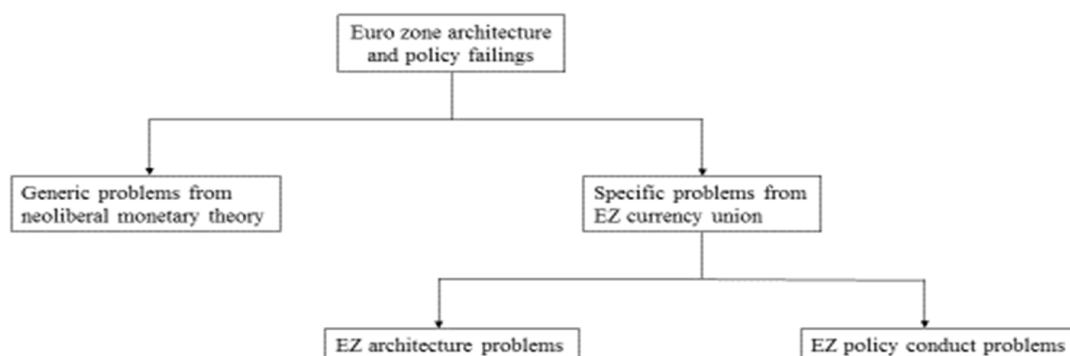
3. New Classical economics and the origins of the euro zone's monetary architecture and policy failings.

To understand the EZ's failings and the case for reform, it is necessary to begin with new classical economics which inspired and underlies the EZ's architecture and policy conduct. New classical macroeconomics (i.e. Chicago School macroeconomics) has under-pinned neoliberal economic policy, and it asserts:[2]

A) Money and inflation are neutral and have no effect on the real economy;

B) Inflation is caused exclusively by money supply growth;

C) The real economy automatically and quickly returns to full employment in response to negative shocks via price and nominal wage adjustment;

D) Financial markets are efficient and stable and determine a natural interest rate that delivers full employment;

E) Fiscal policy is ineffective.

Given the above theoretical framework, optimal policy involves having an independent central bank implement a credible transparent interest rate rule aimed at targeting stable low inflation. According to the policy rule, the equilibrium short-term interest rate should equal the inflation target plus the estimated natural real rate of interest. Furthermore, inflation targeting, implemented via the interest rate rule, is all that is needed to secure full employment because the economy goes there automatically and quickly.

This view of economic theory and optimal policy was hegemonic in the 1990s when the euro was designed and implemented, and it remains hegemonic today – albeit with less self-confidence. Its hegemonic standing meant that Social Democrats (like Jacques Delors and Wim Duisenberg) also accepted it. Consequently, it provided the theoretical template for designing the euro zone's architecture and policy conduct.

[2] Though somewhat more caveated today, new classical macroeconomics remains mainstream economists' dominant theoretical frame, which explains their incapacity to understand the problems of the EZ and resistance to reform. New classical macroeconomics' standing in relation to mainstream macroeconomics parallels the standing of neoclassical competitive general equilibrium theory to mainstream microeconomics.

3.a) Generic problems of new classical monetary policy

The new classical construction of monetary policy and central banking produces three grave generic problems that have afflicted monetary policy in both Europe and elsewhere. The first problem concerns mistakenly low inflation targeting. The problem stems from Milton Friedman's (1968) natural rate of unemployment hypothesis which claims money and inflation have no permanent real effects. Consequently, there is no trade-off between inflation and unemployment so that the long-run Phillips curve is vertical. This contrasts with the Keynesian view that a trade-off exists and the Phillips curve is negatively sloped because modest inflation helps grease the wheels of labor market adjustment (Tobin, 1972; Palley, 1994, 2012).

Figure 2 shows the new classical and Keynesian Phillips curves. Neoliberal macroeconomics recommends an ultra-low inflation target (π^*). The argument is that inflation is undesirable and confers no unemployment gain because the economy always gravitates to the natural rate of unemployment (u^*). From a Keynesian perspective, that will cause significant unnecessary unemployment as inflation of π^* implies a higher unemployment rate ($u^1 > u^*$) according to the Keynesian Phillips curve.

Figure 2 New classical (neoliberal) vs. Keynesian Phillips curves
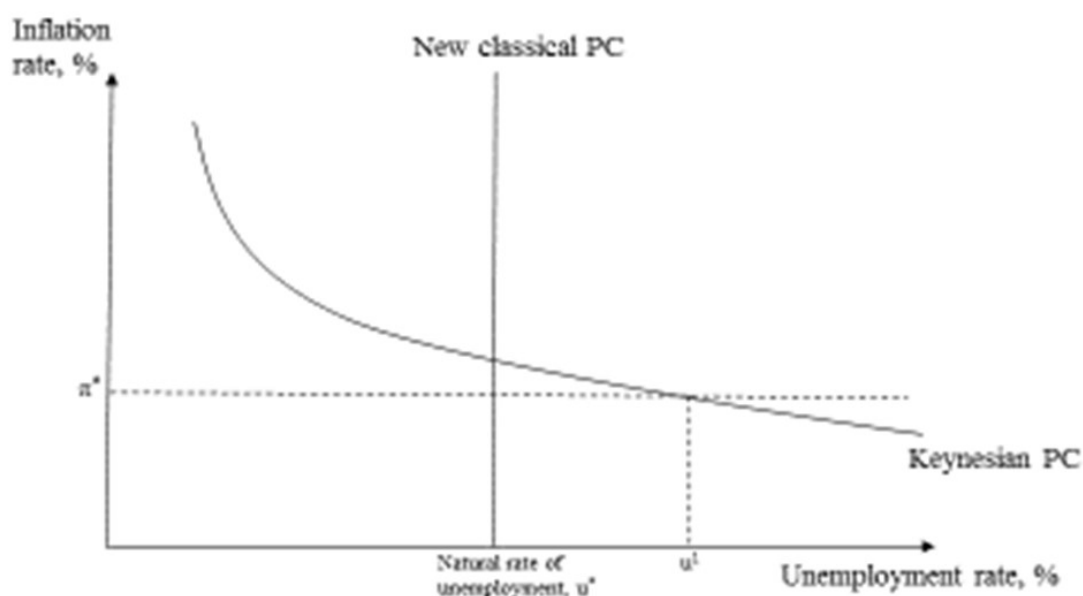
A second generic problem concerns central bank support for the so-called "labor market flexibility agenda" which aims to diminish workers' rights, protections and bargaining power. Natural rate theory argues the natural rate of unemployment is determined by frictions and rigidities within the labor market. Those frictions and rigidities are argued to include trade unions, minimum wages, unemployment insurance, and worker rights and protections. Since central banks believe in natural rate theory, that explains why they have persistently and vigorously lined up in support of the "labor market flexibility agenda" which has contributed to wage stagnation and increased income inequality.

The third generic problem of neoliberal economics is its belief that "flexible" labor markets and interest rate policy, targeted on low stable inflation, are all that is needed to secure full

employment. This belief stems from the assumptions of new classical economics about the economy's adjustment capacities and the character of financial markets. The important implication is it predisposes central banks against the need for financial market regulation or the need to intervene in asset markets to address asset price bubbles (Palley, 2003, 2006a). It also explains the retreat from and resistance to quantitative monetary policy (e.g. regulation of the asset side of banks' balance sheets), which was an important component of policy in the "golden age" three decades after World War II.

3.b) EZ monetary architecture problems: the rupture of the money – fiscal policy link

The major monetary architectural problem of the EZ concerns its divorce of the monetary authority from national fiscal authorities (Goodhart, 1998). From a new classical perspective, this divorce is inconsequential because fiscal policy is ineffective and increases in the money supply only cause inflation. Consequently, there is no need for money-financed fiscal policy and a hard divorce of the monetary and fiscal authorities is desirable.

According to new classical economics, if governments want to run budget deficits they should compete for finance with the private sector in financial markets. That is the efficient way to allocate capital. Additionally, in the context of a currency union, divorce of the monetary and fiscal authority is needed to prevent fiscal moral hazard. If member countries know the central bank will step in and finance their deficits, that would provide an incentive for countries to run larger and larger deficits.

The divorce of the monetary authority (i.e. the central bank) from the fiscal authority (i.e. the national state) is predicated on the assumptions that fiscal policy is ineffective, money financed deficits only cause inflation, and financial markets are stable and efficient. Once those assumptions are rejected, the new classical monetary architecture becomes dangerously dysfunctional.

The loss of national central banks and the divorce between monetary policy and fiscal policy leave national governments dependent on financial markets for their budget deficit financing needs. Consequently, governments may be unable to finance needed expansionary fiscal policy (Goodhart, 1998). Additionally, financial markets will have the power to veto fiscal policy via bond market sell-offs, and governments will also lack the means (i.e. a central bank under their control) to intervene and stabilize national financial markets in the event of financial panic (Palley, 1997). That is exactly what has happened in the EZ after the financial crisis of 2008.

3.c) EZ monetary policy conduct problems: too low an inflation target

As regards the conduct of EZ monetary policy, the generic policy problem of excessively low inflation targeting is amplified in a currency union (Palley, 1997, 2006b). This is illustrated in Figure 3. For new classical economists, a non-optimal currency union may increase the natural rate of unemployment for the currency union as a whole ($u^*_{PRE} < u^*_{POST}$). However, from their perspective, there is no cost in sticking with the pre-existing inflation target since monetary policy cannot affect the new natural rate of unemployment. In sharp contrast, a Keynesian perspective counsels differently. The Phillips curve shifts right from KPC_{PRE} to KPC_{POST}, so that preventing further increased unemployment requires the currency union to adopt a higher inflation target. If the target is unchanged and held at π^* after monetary union, the unemployment rate will rise to $u^2 > u^1$.

Figure 3 The effect of currency union on the Phillips curve
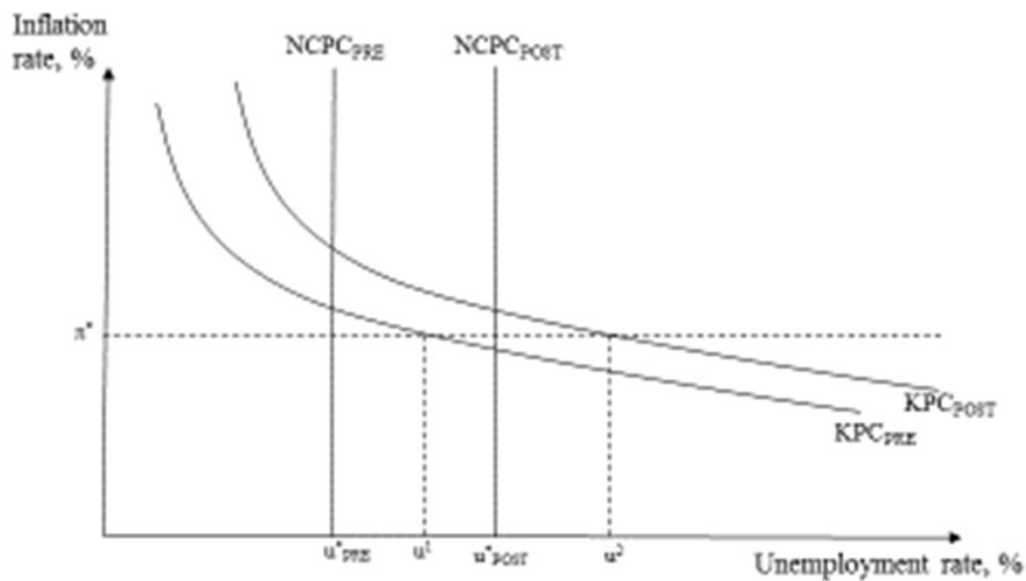
Additionally, the higher unemployment caused by the EZ's low inflation targeting problem has been further compounded by the fact that Germany's Bundesbank monetary policy was adopted as the template for the euro. The Bundesbank has long been dominated by monetarist thinking that is staunchly opposed to inflation. Its monetarist approach was imported into the ECB in the form of an inflation target mandating less than 2 percent inflation. In effect, the creation of the euro was used to lower the EZ's overall inflation target ($\pi_{POST} < \pi_{PRE}$) as shown in Figure 4. That caused an even larger increase in EZ unemployment to $u^3 > u^2$.

Figure 4 The effect of adopting the Bundesbank's ultra-low inflation target
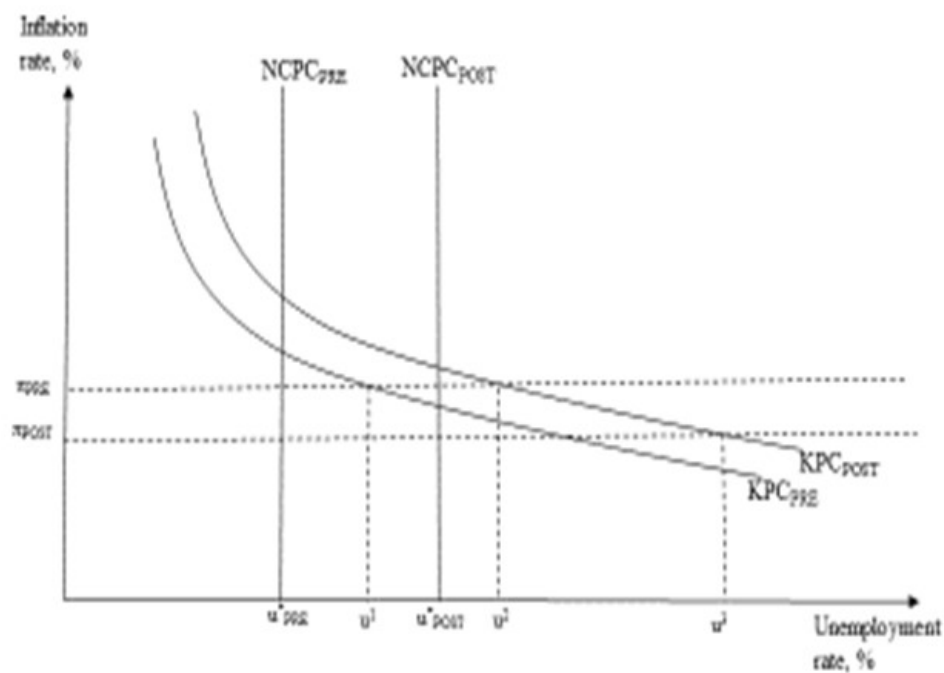
In sum, from a Keynesian perspective, not only did monetary policy fail to raise the inflation target to combat higher unemployment caused by monetary union creating a more diverse economy with more dispersed economic outcomes, it lowered the inflation target for many member countries which had higher targets prior to the euro. That made for a double failure in the conduct of monetary policy.

4. The crisis and the failure of neoliberal economics

The financial crisis of 2008, the Great Recession, and the ensuing stagnation should have entirely discredited neoliberal economics. These events have shown financial markets can be unstable and can greatly misprice assets; economies do not automatically and quickly rebound to full employment; fiscal policy can be highly effective; and inflation is not exclusively and automatically generated by money supply growth. That speaks to remaking the EZ's monetary architecture and redesigning the conduct of monetary policy as events have shown the current architecture and policy design are founded on flawed theory.

5. Remedying the EZ's monetary architecture and monetary policy

5.a) Repairing the money – fiscal policy link via a financing union

The euro's divorce of the monetary and fiscal authorities has created grave problems for governments' ability to finance fiscal policy and defend against financial market speculators. The conventional wisdom is the EZ needs "fiscal union" to overcome these architectural failings, but EZ countries do not politically want that. Instead, I (Palley, 2011a; 2011b; 2016) have argued for a "financing union" that involves collective issuance of debt, the proceeds of which are distributed among members on a per capita basis.

A financing union would require establishment of a European Finance Authority (EFA) governed by the finance ministers of euro zone countries. The Finance Authority would issue bonds jointly and severally backed by all member countries, which the ECB could buy.

The Authority would engage in no spending, and would simply pay issue proceeds to member countries on a per capita basis, with countries liable for debt service on the same per capita basis. Each year the EFA would determine the appropriate budget deficit for the euro zone, issue bonds, and distribute the proceeds to member countries to use as they deemed fit.

Those countries wanting fiscal stimulus could spend the proceeds: others could use them to buy EFA bonds, thereby covering their obligation and leaving their net debt position unchanged.

Countries could also issue their own national bonds to finance additional stimulus over and above that financed by EFA, and these national bonds would constitute a form of junior national debt.

Lastly, an accompanying bankruptcy mechanism would be established. Country national debt would be subject to a junior bankruptcy mechanism similar to the Chapter 9 provision in US law for states and municipalities. EFA debt would be subject to a senior sovereign bail-out mechanism that could permit conditionality arrangements.

The financing union proposal has many significant advantages, but three stand out. First, it permanently remedies the euro's original sin, creating both a permanent policy mechanism for deficit financing and a bond that can be bought without qualification by the ECB. Second, it avoids the great political pitfall of fiscal unions regarding usurping control of the purse from the state or imposing transfers between countries. Countries choose how they spend EFA proceeds. Third, it reconnects money and the state without creating fiscal moral hazard as countries are not bailed out by the EFA or ECB.

5.b) Conduct of monetary policy: a higher inflation target

With regard to conduct of monetary policy, the first change should be a higher inflation target in the region of 3 – 5 percent. Some mainstream economists (Blanchard et al., 2010) are also moving in this direction. Their argument is that a higher equilibrium inflation rate is needed to raise nominal interest rates, thereby providing space for the central bank to lower interest rates if the economy gets in trouble.

Such support is welcome, even if the reasoning is stuck in failed monetary theory. However, it would be far better if the Keynesian Phillips curve rationale were adopted as that would also bury the natural rate of unemployment hypothesis. As long as central banks hold to that hypothesis, there will be a perennial risk that central banks are drawn back into actively supporting the mistaken and damaging labor market "flexibility" agenda.

5.c) Conduct of monetary policy: target the bond rate on newly issued EFM bonds.

A financing union would create a steady growing supply of EFA bonds, and the ECB could then target the long bond rate as well as set the short-term interest rate. Neoliberal monetary theory recommends targeting just the short-term interest rate. The assumption is the combination of efficient financial markets plus a credible transparent interest rate rule ensures long term interest rates reflect expectations of future short-term interest rates markets. Consequently, there is no need to target the long rate.

Such indirect management is unreliable and imprecise as it rests on markets having correct expectations and understandings of future policy. The behavior of financial markets should have punctured that belief long ago. In future, rather than relying on market expectations to determine long rates, the ECB should directly target long rates using EFA bonds as the benchmark (Palley, 2013).

5.d) Asset based reserve requirements (ABRR)

Interest rate targeting should be supplemented by a system of ABRR which would extend margin requirements to a wide array of assets held by financial institutions (Palley, 2000; 2003; 2004; 2006b; 2010). ABRR require financial firms to hold reserves against different classes of assets, with the regulatory authority setting adjustable reserve requirements on the basis of its concerns with each asset class. One concern may be that an asset class is too risky; another may be that an asset class is expanding too fast and producing inflated asset prices.

A system of ABRR that covers all financial firms has multiple policy benefits. Most importantly, it enables central banks to target sector imbalances without recourse to the blunderbuss of interest rate increases. For example, if a monetary authority was concerned about a house

price bubble generating excessive risk exposure, it could impose reserve requirements on new mortgages. This would force mortgage lenders to hold some cash to support their new loans, raising the cost of such loans and cooling the market.

For the EZ, ABRR are additionally attractive because they can help address the policy instrument gap at the national level created by the euro's introduction (Palley, 2006b). That can be done by implementing ABRR on a geographic basis. For instance, requirements on new mortgage loans can vary by country, or even by region within countries.

5.e) Banking union

Just as the design of the EZ neglected fiscal policy and the need for a government banker, so too it neglected the problem of cross-country bank runs (as has happened with money fleeing from the EZ periphery crisis countries to Germany).

The ECB's TARGET2 balance system has plugged the hole by making liquidity available to banks losing deposits. However, it is an inefficient system that recycles liquidity *ex-post* rather than preventing its flight *ex-ante*. It also creates banking regulatory moral hazard across countries, since countries know their banks have access to emergency liquidity from the ECB. That speaks to the need for full banking union with deposit insurance and common regulatory standards and capital requirements for bank asset and liability structures.

6. Radical reform of central bank thinking: bring back pluralism and Keynesianism.

Lastly, there is need for profound radical reform of ECB thinking and practice. Over the last three decades, central banks have been arrogant and closed minded, ignoring all economists outside central banks' narrow sociological circle, and dismissing all who disagreed with their belief that low inflation targeting was sufficient. Events have proved central bank economists wrong and shown the assumptions of neoliberal monetary theory to be disastrously flawed.

At the euro's outset, the focus of mainstream economics was the EZ's properties as an optimal currency area (OCA), and mainstream discussion was conducted exclusively through that lens. The principal concern was the euro was not an OCA (see for instance Bayoumi and Eichengreen, 1992; 1994) The fear was individual countries within the EZ would suffer macroeconomic losses from giving up their own currency and surrendering the exchange rate and interest rate as tools of country economic policy. Those losses from not having one's own currency would outweigh trade and capital flow gains. Feldstein (1997) argued those costs of not being an OCA would cause the euro to ultimately fail in a few decades, possibly even generating military conflict within the EZ.

These mainstream concerns were generic and not policy helpful. They contrast with the concerns of Keynesians (Godley, 1992; Palley, 1997; Goodhart, 1998) who, not only identified the OCA aspects, but also correctly identified and emphasized specific flaws in the euro's neoliberal monetary architecture and monetary policy design.

Godley (1992) argued the euro had a blind spot regarding need for a European federal institution to undertake counter-cyclical fiscal policy:

> "The incredible lacuna in the Maastricht programme is that, while it contains the blueprint for the establishment and modus operandi of an independent central bank, there is no blueprint whatever of the analogue, in community terms, of a central government" (Godley, 1992, p.3).

Goodhart (1998) emphasized the importance of the link between the fiscal authority (i.e. the state) and the monetary authority (i.e. the central bank), and identified the dangers for financing fiscal policy of divorcing the monetary and fiscal authorities:[3]

> "In particular, the participating nation states will continue to have the main fiscal responsibilities; but in the monetary field, their status will have changed to a *subsidiary* level, in the sense that they can no longer, at a pinch, call upon the monetary authority to create money to finance their domestic national debt. There is to be an unprecedented divorce between the main monetary and fiscal authorities" (Goodhart, 1998, p.410).

Palley (1997, 2006a) identified the importance of the central bank's policy preferences and the interaction of those preferences with economic understandings of the Phillips curve. Making the euro successful required a higher inflation target. It also required introduction of quantitative monetary policy and ABRR to supplement interest rate inflation targeting policy, thereby giving member countries additional policy instruments to replace those lost owing to currency union.

Additionally, Palley (1997) argued the divorce of the monetary and fiscal authorities would give bond markets the power to discipline governments who pursue economic policies that financial markets dislike. That is because governments would no longer have a central bank to buy their bonds and protect against capital flight:

> "Thus, if financial capital dislikes the stance of national fiscal policy, there could be a sell-off of government bonds and a shift into bonds of other countries. This would drive up the cost of government borrowing, putting a break on fiscal policy" (Palley, 1997, p.156).

This feature is cruelly ironic as part of the intention of the European monetary union was to protect against capital market flight, such as had undermined the policies of France's President Mitterrand in the early 1980s.

In sum, the record clearly shows Keynesians had a far superior understanding of the monetary macroeconomics of currency unions and anticipated many of the operational problems of the euro. That suggests it is time to heed the Keynesians by reforming the EZ along the lines they have advocated.

More broadly, the superior analytical insights of Keynesians regarding monetary union provides another case study of mainstream economics failure, and adds to the record of failure which has been accumulating for three decades. Given that record, it is also time to

---

[3] Goodhart is perhaps the only establishment economist to have anticipated specific structural problems of the euro, as against generic concerns regarding the euro being a non-optimal currency area. That said Goodhart is a distinguished "grey beard" who was admitted to the circle of central bankers before the ideological boom came down in the 1980s and put an end to pluralism in economic thought.

break the new classical monopoly on monetary theory and monetary policy, and to open central banking to a wider range of perspectives.

References

Bayoumi, T. and Eichengreen, B. (1992), "Shocking Aspects of Monetary Union," Working Paper 3949, NBER, Cambridge: MA, January.

Bayoumi, T. and Eichengreen, B. (1994) "One Money or Many? Analyzing the Prospects for Monetary Unification in Various Parts of the World," *Princeton Studies in International Finance*, no. 76, September.

Blanchard, O., Dell'Arica, G., and Mauro, P. (2010) Rethinking Macroeconomic Policy, in Research Department, IMF, Washington DC, February, available at http://www.imf.org/external/pubs/ft/spn/2010/spn1003.pdf.

Godley, W. (1992), "Maastricht and All That," *London Review of Books*, Vol. 14 No. 19, 8 October.

Goodhart, C.A.E. (1998) "The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas," *European Journal of Political Economy*, 14 (3), 407-432.

Feldstein, M. (1997) "EMU and International Conflict," *Foreign Affairs*, Vol. 76, no. 6, November / December .

Friedman, M. (1968) "The Role of Monetary Policy", *American Economic Review*, 58 (May), 1-17.

Palley, T. I. (1994) "Escalators and Elevators: A Phillips Curve for Keynesians", *Scandinavian Journal of Economics*, 96 (1), 117-123.

Palley, T. I. (1997) "European Monetary Union: An Old Keynesian Guide to the Issues", *Banco Nazionale del Lavoro Quarterly Review*, vol. L, no. 201 (June), 147 164.

Palley, T. I. (2000) "Stabilizing Finance: The Case for Asset Based Reserve Requirements", *Financial Markets and Society*, The Financial Markets Center, Philomont, VA, August. http://www.thomaspalley.com/docs/articles/macro_policy/stabilizing_finance.pdf

Palley, T. I. (2003) "Asset Price Bubbles and the Case for Asset Based Reserve Requirements," *Challenge*, 46 (May–June), 53-72.

Palley, T. I. (2004), "Asset Based Reserve Requirements: Reasserting Domestic Monetary Control in an Era of Financial Innovation and Instability," *Review of Political Economy*, 16 (January), 43-58.

Palley, T. I. (2006a) "Currency Unions, the Phillips Curve, and Stabilization Policy: Some Suggestions for Europe," *Intervention: European Journal of Economics and Economic Policies*, 3(2), 351-69.

Palley, T. I. (2006b) "Monetary Policy in an Endogenous Money Economy", in Arestis and Sawyer (eds.), *A Handbook of Alternative Monetary Economics*, Edward Elgar: Cheltenham, U.K., p.242-257.

Palley, T. I. (2010) "Asset Price Bubbles and Monetary Policy: Why Central Banks Have Been Wrong and What Should Be Done", *Intervention: European Journal of Economics and Economic Policies*, 7(1), 91-107.

Palley, T. I. (2011a) "Monetary Union Stability: The Need for a Government Banker and the Case for a European Public Finance Authority", in IMK Working Paper 2, Dusseldorf, Germany, and published as "The European Union Needs a Government Banker", *Challenge*, 54 (July-August 2011), 5-21.

Palley, T. I. (2011b) „Stabiltät in der Währungsunion: Plädoyer für eine Staatsbank und eine Europäische Finanzagentur", *WSI Mitteiliungen*, 6 (June).

Palley, T. I. (2012) "The Economics of the Phillips Curve: Formation of Inflation Expectations versus Incorporation of Inflation Expectations", *Structural Change and Economic Dynamics*, 23(3) (September), 221-230.

Palley, T. I. (2013) "Monetary Policy and Central Banking after the Crisis: The Implications of Rethinking Macroeconomic Theory", in *The Handbook on Political Economy of Financial Crises*, Editors: Gerald Epstein and Martin Wolfson, Oxford: Oxford University Press, Chapter 31, p.624-643.

Palley, T. I. (2016) "Financing Unions vs. Spending Unions: How to Remedy the Euro Zone's Original Sin", *Social Europe*, 14 July, https://www.socialeurope.eu/2016/07/financing-vs-spending-unions-remedy-euro-zones-original-sin/

Tobin, J. (1972) "Inflation and Unemployment", *American Economic Review*, 62 (1972), 1-26, reprinted in *Essays in Economics: Volume 2*, New York: North-Holland, 1975, p. 33-60.

Author contact: mail@thomaspalley.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

Missing from the mainstream: the biophysical basis of production and the public economy

June Sekera [Tufts University, Medford. MA, USA]

Abstract
Just as mainstream economics neglects the biophysical basis of production and disregards energy as the most fundamental input, it likewise ignores the existence of the public economy. Both types of denialism threaten the ability of societies to develop energy solutions that can meet the needs of the polity. This article calls for a new theory of the public economy and it outlines elements of such a theory. Both a biophysical economics and a new public economics are needed to address the energy challenges confronting modern societies.

One of the most important contributions of biophysical economics is its critique that mainstream economics disregards the biophysical basis of production (Hall et. al., 2001), and energy in particular (Hall and Klitgaard, 2012).

Likewise, mainstream economics ignores the existence of the public economy. The public economy is a vital system of production and delivery that produces scores of products: goods, services, benefits and innovations. Yet, standard economics lacks a theory of this non-market system.

To miss or minimize basic empirical verities – both the biophysical basis of production and the existence of the public non-market production economy – is not only astonishing denialism.[1] Such obstinate myopia within economics may foreclose the development of solutions, such as alternative sources of high-EROI[2] energy reliably produced and affordably supplied on a planetary scale.

1.   Denial of the public non-market system, and the consequences

Public non-market production makes up a quarter to a half or more of all economic activity among advanced democratic nation-states. Yet the public economy's ability to function on behalf of the populace as a whole is seriously imperiled in many western democracies, and particularly jeopardized in the United States. The surging influence of mainstream economics has been a prime factor in the degradation of the public domain over the last several decades – a phenomenon that James Galbraith (2008) has called "the collapse of the public governing capacity." Market advocates, exploiting neoclassical economic theory, have foisted market axioms and precepts onto government, intent on transforming public goods production in imitation of an idealized and idolized market model. The ravaging of government in the interests of ideology and private profit has proceeded largely unhampered because we have no adequate theory to explain the nature and dynamics of the non-market public economy, no intellectual infrastructure to explain how its purposes and processes differ crucially from those of the market, and no effective explanatory model that shows why such differences matter substantially for democratic governance and the well-being of the populace.

---

[1] "Denialism: refusing to accept an empirically verifiable reality"
[2] "Energy Return On Investment"

Government produces its outputs in a non-market environment. Its resource inputs are supplied collectively: from the authority of the people (their votes for elected representatives) and from their aggregate financing (taxes). The mission and the result of government's distributed decision-making, collective-financing system of production is that goods, services, benefits, and protection are supplied for the wellbeing of the society as a whole, and can be accessed regardless of personal wealth because they are provided free or below cost at the point of usage. Economic theory today lacks any cogent theory of this non-market system.

Public choice theory, to which many contemporary economists default for a "public economics," draws its lifeblood from market-centric ideology. The public choice school holds that the axioms and assertions of market-model economics apply to the public economy. Simply put, there are two fundamental problems with this school: 1) it fails to recognize that the public economy is *non-market*; and 2) many of the basic assumptions and assertions of market economics have been challenged and disproven by pluralist economists regarding their applicability to the *market* (e.g., see Fullbrook, 2007), nevermind the non-market.

A myopic market-centric view of the public economy prevails in textbooks, in university classrooms, in the documents and debates shaping public policy and in the current practice of public administration. As it stands now, students in university economics courses learn about the superiority of markets from a professoriate that transmits the reigning market-centric economics, that speaks regularly of government as little more than an impediment to "efficient markets," and that understands public goods as a problem of "market failure." In the United States, about 40% of college students take at least one economics course (Goodwin, 2014); after graduation, more than half of economics majors go to work in government (Kalambokidis, 2014).

My argument is that mainstream, market-centric economics has been broadly and dangerously transformative within government and public institutions. Market-centric economics is the smog that pervades the atmosphere of public policy and public administration, a smog that has at once caused and obscured many of the failures of what some say is a "broken government" (Schuck, 2014; Howard, 2014; T. Smith, 2014; Fahrenthold, 2014; Luntz, 2014). "Economic abstraction has been coupled with power to impose that abstraction throughout [the nation]. The result has been a political economy that generates the conditions for its own failure..."[3]

The consequences of the contrived and contorted imposition of market-model economics on the public domain range from the unfortunate to the disastrous. Agencies originally created to meet a public need are being warped into entities whose purpose is to generate revenue and deliver private profits at public expense. National parks are selling naming rights to corporations who will rebrand Yellowstone and Yosemite in their corporate images (Rein, 2016; Olorunnipa, 2016). The "policing-for-profit" model in criminal justice results in officers stopping motorists for minor infractions in order to make fee-and-fine quotas (U. S. Dept. of Justice, 2015; anon, Harvard Law Review 2015; Zapotosky, 2016). Public education – today being relabeled "government education" by those on the right – is being taken over by Wall Street, which has targeted "the education industry" as a new profit center through the spread of private "charter schools" funded by taxpayers, but shown in multiple studies to arrive at widely inferior results (Persson, 2015; Losen et al., 2016). Through "public-private-

---

[3] Bowman et.al., 2014. The authors write principally about the UK, but their argument brilliantly captures the American reality too.

partnerships," multinational corporations build toll roads that go bankrupt, leaving taxpayers holding the bag. Private collection companies, contracted by government agencies, are being granted the sovereign power of the state to garnish wages of students, the poor, and other citizens in order to collect overdue debt AND fees and fines imposed by the companies themselves (Choudhury, 2014; Edsall, 2014; Shapiro, 2014; Stillman, 2014). Privatizers are very close to turning the venerable Veterans Health Administration into an ATM for the private healthcare industry, despite studies that have consistently shown that the VHA provides health care superior to private care systems (Farmer et al., 2016; MITRE, 2015; Gordon, 2015; 2016; Mundy, 2016; Kime, 2016). The result is a subversion and erosion of the capabilities of the public system of production, such that it can no longer deliver its intended results. A mission-model economic system, in which meeting public needs was the guiding purpose, is being distorted into a faux market-model system, in which revenue-raising becomes the goal.

While this transformation debilitates government overall, there is specific relevance to the growing energy challenges related to the biophysical constraints on economic activity and production. The historic role of government in leading and supporting basic scientific innovation is being hobbled. Regarding solutions to energy challenges in particular, we are confronted with what seem to be hopelessly complex problems that require: a long-term view; basic research financed by investments not tied to quarterly profits; breakthrough innovation; and development of society-wide solutions. These are the attributes not of the market, but rather, of the public non-market.

Indeed, the public non-market is the unrecognized innovator in our nation. Government has been the source – through its investments and leadership – of scores of breakthroughs that people often assume came from the private sector. Government's role in innovation has been documented by Mariana Mazzucato and Fred Block, among others, who have exploded the myth that all innovation is market-driven. A sampling includes:

Debunking the Narrative of Silicon Valley's Innovation Myth
Forbes | Bruce Upbin

"The real innovation engine in the global economy is not the entrepreneurial class blazing capitalist trails through the thicket of government red tape and taxation. No. The real engine of innovation is government." Economist Mariana Mazzucato's "case study for myth-debunking is the iPhone, that icon of American corporate innovation. Each of its core technologies–capacitive sensors, solid-state memory, the click wheel, GPS, internet, cellular communications, Siri, microchips, touchscreen—came from research efforts and funding support of the U.S. government and military. Did the public see an iPhone dividend? Not really."

The High Return on Investment for Publicly Funded Research
Center for American Progress | Sean Pool and Jennifer Erickson

In order for the U.S. to maintain its role as an innovation-driven economy, "government must provide three key public-good inputs that allow innovation to blossom: investments in human capital, infrastructure, and research."

The authors cite and summarize the contributions of influential research funded by the U.S. Government through the Dept. of Energy Labs, The National Science Foundation, The Human Genome Project, The Defense Advanced Research Projects Agency and the Apollo Space Program.

Markets, States, and the Green Transition
The American Prospect | Fred Block

"... [U]nder-appreciated state involvement is true of many new technologies and sectors, but it emphatically describes the necessary transition to renewable energy. Private entrepreneurs contemplating investment in green energy face a chicken-and-egg problem. Technologies either do not yet exist, or they do not exist at a competitive price ... Unless government intervenes on the supply side—to promote the innovation that is too risky for private entrepreneurs—and on the demand side—to accelerate creation of mass markets for green sources of energy—private industry cannot get the job done."

Innovation: let the good risk-takers get their reward
The Guardian | Mariana Mazzucato and William Lazonick

Mazzucato and Lazonick write that, "the advanced economies of the west are in deep trouble. Growth is slow or non-existent, income distribution is highly unequal …[and] the crucial question is how to reform policy so that the relationship between risk and reward is one that supports long-run growth rather than undermining it."

They point out that taxpayers are the real venture capitalists; taxpayers fund the riskiest investments in the "knowledge economy," but it is shareholders who receive recognition and profit for reputedly bearing the risk.

The Seeds That Federal Money Can Plant
The New York Times | Steve Lohr

"Government support plays a vital role in incubating new ideas that are harvested by the private sector, sometimes many years later, creating companies and jobs."

The author cites a report from the National Research Council that finds nearly $500 billion a year of revenue at "30 well-known corporations ... [can] be traced back to the seed research backed by government agencies."

Imagine spending a day without the Internet and GPS
Continuing Innovation in Information Technology | National Research Council

The internet and GPS (a U.S.-owned utility) are among many innovations that have been funded by the U.S. Government. The authors of *Continuing Innovation in Information Technology* write, "fundamental research in IT,

conducted in industry and universities, has led to the introduction of entirely new producer categories that ultimately became billion-dollar industries."

Underscoring the impact of government's outsized role in creating the dominant technologies of the 21st century, the authors of this report ask readers to imagine a day without information technology. "This would be a day without the Internet and all that it enables ... A day without digital media ... A day during which aircraft could not fly, travelers had to navigate without benefit of the Global Positioning System (GPS), weather forecasters had no models, [and] banks and merchants could not transfer funds electronically..."

The economic system that produced these innovations – the public non-market economy – of which government is the agent, remains unrecognized in contemporary mainstream economics, which is blind to it as a valid, viable, essential production system.
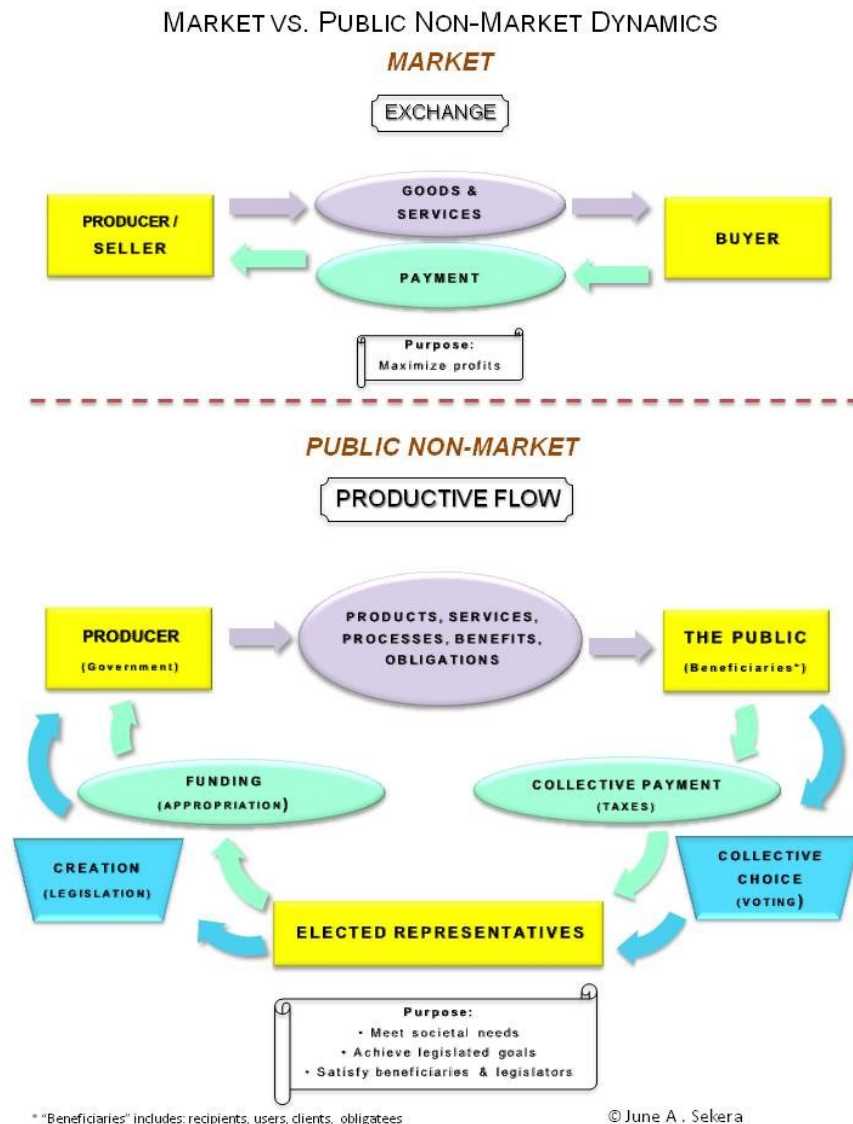

2. Elements of the public nonmarket

In the public non-market, the most basic constructs of mainstream economics do not apply. There is no "exchange" between "buyers," and "sellers." There is no market-model competition, only "pseudo-privatization" (Siltala, 2013). The driver is not demand but identified societal need. Satisfying "customers" does not produce revenue. The monopsonist is often rendered powerless to set prices. Government expenditure actually results in "crowding-in," boosting rather than curtailing growth. In a non-market, outcome goals are devilishly difficult to define—unlike the simple market goal of maximizing profit. Results are often obscured because of factors unique to non-markets, where *invisibility* of outputs and *absence* of harmful conditions are hallmarks of success.

The public non-market is the economy in which the production of goods, services and other products is capitalized collectively (through taxes), and is empowered through collective choice (voting), and in which products are provided free or below cost at the point of receipt or usage. In *The Public Economy in Crisis: A Call for a New Public Economics*, (Sekera 2016) I outline the elements of a new theory of the public non-market economy. In summary, these include:

- The systemic purpose is meeting unmet societal needs; not maximizing profits.

- The public nonmarket is need-driven, not demand-driven. Collective choice replaces demand.

- The two fundamental systemic drivers are collective choice and collective payment.

- Flow relationships and dynamics are intrinsically different in the two economic systems. The market is an exchange; the public nonmarket is a three-node flow. See Figure 1.

Figure 1

MARKET VS. PUBLIC NON-MARKET DYNAMICS

MARKET

EXCHANGE

PRODUCER / SELLER → GOODS & SERVICES → BUYER
PAYMENT

Purpose: Maximize profits

PUBLIC NON-MARKET

PRODUCTIVE FLOW

PRODUCER (Government) → PRODUCTS, SERVICES, PROCESSES, BENEFITS, OBLIGATIONS → THE PUBLIC (Beneficiaries*)

FUNDING (APPROPRIATION) COLLECTIVE PAYMENT (TAXES)

CREATION (LEGISLATION)   COLLECTIVE CHOICE (VOTING)

ELECTED REPRESENTATIVES

Purpose:
• Meet societal needs
• Achieve legislated goals
• Satisfy beneficiaries & legislators

* "Beneficiaries" includes: recipients, users, clients, obligatees   © June A. Sekera

And there are other intrinsic differences:

- *Products.* The market produces goods and services; the public non-market produces goods, services, benefits and obligations. "Obligations," such as the obligation of drivers to obey speed limits and of factories to obey pollution regulations, are a unique product of the public nonmarket (Moore 2014). In a democratic nation-state, such power is conferred by the polity through voting.

- *Invisibility*. In the public non-market invisibility is a hallmark of effectiveness: needs met; problems solved; harms that do not happen because of effective protection.

- *Results measurement*. Businesses' success is measured by profitability (since profits are required for survival), which is quite simple in comparison to the non-market. In the public non-market defining measurable outcomes in a way that obtains valid, useful

measurements, and that avoids unintended consequences is extraordinarily complex and confounding difficult. Exhibit A is public education: measuring the effectiveness of education through student testing alone. There are dozens more examples of the dismaying, and often disturbing, techniques that have been applied all across government and backfired, underscoring the difficulty of measuring nonmarket outcomes in a meaningful, useful and valid way.

Each of the characteristics listed above is discussed in detail in my book, *The Public Economy in Crisis: A Call for a New Public Economics*.

3. What happened? A transformation within economics and an absence of theory today

More than a century ago, the effective operation of the public economy was a significant, active concern of economists. With the insurgence of market-centrism and rational choice economics, however, government was devalued, its role circumscribed and seen from a perspective of "market failure." As Backhouse (2005) has shown, the transformation in economic thinking in the latter half of the 20th century led to a "radical shift" in worldview regarding the role of the state. The very idea of a valid, valuable public non-market has almost disappeared from sight.

In 18th and 19th century Germany, Kameralwissenschaft ("Cameralism") represented a form of public economics. Backhouse (2002, p. 166), describes this school as the era's "science of economic administration," which had three components: public finance, economics, and public policy. The "Historical School" of economics emerged in later 19th century Germany and viewed government positively as a system for promoting social well-being (Bogart, 1939; Shionoya, 2005). It stopped short, however, of explaining the operational or production aspects of the system. During the late 19th and early 20th centuries, economists wrestled with the question of how the "public economy" operates. A "voluntary exchange" theory of the public economy was advanced by Emil Sax, DeViti De Marco, Knut Wicksell and Erik Lindahl (Sekera, 2016). During the 1940s–50s, Richard Musgrave argued against the voluntary exchange concept and pursued a line of thinking that led to the construction of a concept of "public goods" that was eventually adopted, mathematicized and popularized by Samuelson (Desmarais-Tremblay, 2013). Samuelson's widely-disseminated 1950s formulation of public goods as stemming from market failure (following Musgrave) soon led to their devaluation, and a wholesale devaluation of government, by market centrists and libertarians, eventually by all tributaries of mainstream economics. What had begun as a serious effort to understand the important role of public sector production ended in its willful neglect.

In an important paper, Roger Backhouse (2005) describes the "profound changes in economic theory" that took place between 1970 and 2000. With the triumph of rational-choice economics came "a radical shift of worldview" and a "remarkable and dramatic change in attitudes toward the role of the state in economic activity." The rise of "free market" economics and the "ideology of rational choice" created a "climate of opinion" that seriously biased economics against government and led to a view of the state as an agent whose actions lead to perverse outcomes. As Backhouse shows, however, "the shift toward market solutions did not occur spontaneously: it was actively promoted by groups of economists committed to opposing socialism [and] making the case for free enterprise."

In his landmark book, *A Perilous Progress: Economists and Public Purpose in Twentieth-Century America* (2001), Michael Bernstein explores the evolution of economics from an academic field marginal to public policy into a powerhouse that influenced and oriented government decision-making. Economists in the late 19th and early 20th centuries ardently sought to cultivate influence with elected and appointed officials to shape public policy and contribute to "purposeful management" and "statecraft." These were among the driving ambitions of the economists who led the American Economics Association after its founding in 1885. Seeking respect for economics as a new "scientific" field (no longer framed philosophically as "political economy"), "scholars sought a privileged and powerful access to public policy debate, formulation and implementation." Once the influential Cambridge University economist Arthur C. Pigou asserted in 1922 that it was not the business of economists to tell businessmen how to run their companies, it became all the more critical that economists claim for their discipline a legitimate role in statecraft. And they got their big chance in war. Tracing the many roads by which economists entered the public arena, Bernstein finds that the profession came fully into its own through its impact on national decision-making during World War II. Ironically, "Not individualism but rather statism provided the special circumstances" for American economists to obtain prestige and power (p. 89). "In point of fact, it was statism and centralized economic policy practice that had brought economists and their discipline to the prominence and influence they [came to] enjoy (p. 194)."

Yet even when applying their theories and practices to the non-market environment of government, mainstream economists have relied insistently on the market model. Because mainstream economists in the U.S. and elsewhere have been so market-focused for so long, production outside the market has been erased from the equations of economics. So now, government action is regarded as an "intervention" that "distorts" smooth operation of an otherwise beneficent market. Government is considered to have an economic role only (or primarily) in cases of so called "market failure." Consequently, there is no viable and explanatory concept of an actual, let alone a legitimate, public non-market economy. So pervasive is the creed that government only "intervenes" in what is thought to be the valid, market economy that even literature from the Congressional Research Service (Labonte, 2010) relegates government to an outsider role.

The term "non-market" and its meaning remain elusive. For example, Karl Polanyi wrote extensively about the differences between markets and non-markets but did not deal with the dynamics and forces of production in the non-market public economy (Krippner, 2001; Mayhew, 2016; Zaman, 2016). Polanyi argued that the market was embedded within, and enabled by, the public sector, but did not concern himself with the operations – forces, dynamics, drivers – of the public non-market system itself. Neither do such widely-cited economists of the public sector as Robert Dahl and Charles Lindblom, Charles Wolf or Kenneth Arrow (Sekera, 2016). Joseph Stiglitz produced an entire textbook on "the economics of the public sector" (the latest edition in 2000) without recognizing the distinctive characteristics of a public non-market.

As I noted earlier, the "public choice" school has become the framework to which economists default for an explanation of the public economy. Backhouse (2005) outlines the development of the public choice school, which stems from a cluster of works published in the 1950s and 1960s by James Buchanan, Gordon Tullock, Mancur Olson, and Anthony Downs. It became a school, and a movement, when James Buchanan and Warren Nutter found a home for their efforts at George Mason University in Virginia. In the mid-1980s George Mason opened the

Center for the Study of Market Processes, with its largest supporter being the Koch Family Foundations. Stretton and Orchard (1994) have demonstrated the anti-government, anti-democratic stance of public choice theorists in their extensive treatment of the school in *Public Goods, Public Enterprise, Public Choice; Theoretical Foundations of the Contemporary Attack on Government*. After critiquing the theory in economics terms, they suggest that public choice "reasoning seems to arise from the theorists' reluctance to 'come out' and identify themselves as open enemies of democracy or at least of universal suffrage…Governments are viewed as exploiters of the citizenry, rather than the means through which the citizenry secures for itself goods and services that can best be provided jointly or collectively."

A theory of the public nonmarket remains woefully lacking. The absence is not just an academic gap; it leaves a vacuum that undermines the public provisioning required to meet societal needs and to develop solutions to pressing common problems, including the depletion of high-EROI energy sources.

4. Dealing with energy challenges: the connection between public economics and biophysical economics

Orthodox economics "posits that marketplace dynamics will determine the energy transition from fossil fuels to something else through the price mechanism" and "assumes that innovation will appear as needed" (Cobb, 2010). There is little evidence to support either the postulate or the assumption.

A new economics is called for. In fact, two: biophysical economics and a new public economics. Readers familiar with environmental economic dynamics no doubt understand the need for a biophysical economics that recognizes and takes account of the inherent limits of the biophysical world. Such an economics, while essential, is not sufficient. Solutions to the energy challenges we face will necessitate both a new, biophysical understanding of production and a new understanding and conceptual model of the public economic system. Neither is entirely sufficient without the other.

The energy transition will require breakthrough innovation. Both theory and the history of recent decades demonstrate that solutions to technologically complex, common-need problems require scientific breakthroughs that come through distributed decision-making and collective action[4] and that will not come from market forces alone, if at all. Such solutions require long time-horizon investment: investments with no immediate payoff in terms of saleable products, no visible ROI (return on investment), no profit-making in the near-term. Such investment can be generated only in a non-market environment, in which payment is collective and financial profit is not the point.

Moreover, in the modern market, businesses, backed by profit-driven investors, intentionally produce products with a surfeit of waste baked in, a characteristic inimical to an energy transition that seeks to minimize energy waste. The production of extraordinary waste is inherent to the modern, market business model (MacKinnon, 2016; Arieff, 2016). Today's

---

[4] Collective action in a democratic society is exercised through distributed decision-making (voting) and is financed through collective payment (taxes). Note that collective action and collective payment are not synonymous with state-ownership of enterprises. The matter of state-owned enterprises that respond to market forces is not material to the issues I am describing here.

business model, particularly that of technology corporations, has created a "throw-away" culture with a value system that fosters the discard of millions of electronic devices or components (Urry, 2016) and adheres to a design formula that intentionally makes repair impossible or difficult (Matchar, 2016). While laying claim to "efficiency" as an alleged attribute (often achieved merely through labor cost reductions), the American business model is geared to "repetitive consumption" and "planned obsolescence" (MacKinnon, 2016). The second law of thermodynamics tells us that waste is an intrinsic feature of the use of energy in production. But there is a qualitative, and controllable, difference between the level of necessary "waste" generated by energy conversion and that gratuitous waste inherent in the modern American business model, a wastefulness that goes unaddressed by market economics. Ecological economist Herman Daly (1998a; 1998b; 2003), having shown how "The concepts of throughput, of entropy…are foreign" to "mainstream neoclassical economists," argues for an economic policy of "frugality first." He defines frugality as "non-wasteful sufficiency". Such frugality is not a characteristic of modern market production.

The market is not constituted to produce solutions to extraordinarily complex, technological common-need problems. This is so even if market actors start to perceive the biophysical basis for production. The inherent, driving forces and dynamics of the modern market – short time horizons, growth as a requisite, gratuitous waste baked-in, profits as life-blood – render it incapable of producing solutions that demand long-view investment without profits. The challenges we face may be unprecedented. In a paper on "EROI of Different Fuels and the Implications for Society," Hall, Lambert and Balogh (2014) conclude:

> "The decline in EROI among major fossil fuels suggests that in the race between technological advances and depletion, depletion is winning. …. Thus society seems to be caught in a dilemma unlike anything experienced in the last few centuries. During that time most problems (such as needs for more agricultural output, worker pay, transport, pensions, schools and social services) were solved by throwing more technology investments and energy at the problem... We believe that the future is likely to be very different, for while there remains considerable energy in the ground it is unlikely to be exploitable cheaply, or eventually at all, because of its decreasing EROI."

Many will advocate "market-based solutions" or "public-private-partnerships" as the route to take, based on a misplaced faith or ideological belief in the market. Kate Aronoff, (2015) a former organizer with the fossil fuel divestment movement, sees evidence that the positioning of corporate leaders will enable them to make the case that "the free market is better suited than the state to take on the climate crisis." Indeed, investors are hovering, anxious to profit from EROI decline. Wall Street and private equity investors see new opportunities for profit, often at the public's expense. Libertarian venture capitalist Peter Theil (2015) (PayPal cofounder) wrote about the opportunity he sees as an investor: "We already know that today's energy sources cannot sustain a future we want to live in…The need for energy alternatives was already clear to investors a decade ago…" Henry Paulson (Paulson, 2016), former Chairman and Chief Executive Officer of Goldman Sachs, former Secretary of the Treasury and now head of the Paulson Institute, says "Saving our planet from the worst effects of climate change won't be cheap…governments must create conditions that encourage private investment in clean technologies and sustainable development… **incentives** and **subsidies** for clean energy investments" are needed. [Emphasis added]. For Theil, atomic energy is the solution. Americans need not suffer a decline in living standards or businesses sacrifice growth, he implies. But government regulation and "liberals" with a "fear of technology" stand

in the way of "venture capitalists like me ready to put money behind nuclear power." However, muting regulators or mooting regulations so that investor money can flow may not be the path to the optimal energy solution for the polity, or perhaps for the planet. Nevertheless, if the history of recent decades is a guide, corporations will seek, and obtain, public subsidies to underwrite their technology development process.

One way or another there *will* be government investment. The only question is – who will control the use of those investments? Market actors whose purpose is private profit-making? Or the collective choice of the polity, who, long term, will be the beneficiary or the victim of those investments? Is government up to the work that needs to be done? Given the scant attention paid to biophysical constraints among economists who advise government and even much of the scientific community, given the proclivity on all sides to look to market-like solutions, and given a hollowed-out, outsourced and degraded government, that question is scarcely rhetorical.

A paper by Day et al. (forthcoming) on "The Energy Pillars of Society" observes that new "policy" is needed, and expresses concern that the policy makers won't get it right.

> "[The] issue is that societal net energy yield is falling. Adopting growth-based economic policy without consideration of net energy yield is likely to leave society vulnerable to a future without sufficient energy to provide for basic needs. More careful analysis of resources and economy, incorporating net energy, is necessary to inform policy and management during the coming energy transition.
>
> Many crucial aspects of the proposed transition, net energy and resource constraints in particular, have been largely overlooked by policymakers and much of the scientific community. Our central thesis here is that proposed climate-related energy policies will be impacted and often restricted by biophysical constraints, especially net energy and total production. Serious economic and societal displacements will occur if the existing energy system is disrupted.
>
> Governments and major policy agencies must recognize how biophysical constraints will impact plans for the future and develop research programs that are aimed at investigating the tradeoffs of society's energy investments within the context of net energy and resource constraints. Obviously further energy research in all sectors is warranted. Policy must also prioritize fossil fuel conservation and system wide efficiency. **We cannot stress enough the importance that all energy research and policy is guided by a systems based understanding of the biophysical constraints (especially, *net energy*) that govern the natural world"** [Emphasis in original].

Yes, policy makers need to understand the biophysical imperative: that societal net energy yield is falling. Hence the need for a biophysical economics, and for policymakers to comprehend its central messages. But the other major issue is that policy-makers – both the leaders and the public servants who write policy "options papers" for them -- have been taught to embrace "market solutions" for every sort of societal need, from education, to infrastructure, to water supply and national security. "Market solutions" is the tide that has swept in across the public sector, "public-private-partnerships" the wave that has been

flooding all parts of government for more than 30 years, with ever-more destructive force. We need a new public economics that comprehends and embraces the public purposes of the public domain, and which supports long-view policies that both solve the problem and serve the public

Only an economic system that can spawn breakthrough innovations with no profit in sight, and only one in which gratuitous waste is not intentionally baked-in (for the sake of future profits) can come up with the solutions. In current nation-states, that system is the public nonmarket economy, of which government is the agent. But the public nonmarket is being dismantled, hollowed-out, outsourced, privatized. If this degradation continues, if the public economy continues to be contorted into a faux-market system, if collective payment as the financing method continues to be choked off or diverted to private profits, if government is increasingly forced to make revenue-generation a goal, and if large swaths of the public nonmarket continue to be captured by profit-maximizing corporations, then the public nonmarket will have ceased to become the source of the solution. Indeed, it will have ceased to exist.

References

Anon. (2015) "Policing and Profit" *Harvard Law Review*, Vol 128, No. 6, April 10, 2015. http://harvardlawreview.org/2015/04/policing-and-profit/

Arieff, Allison (2016) "Solving All the Wrong Problems" *New York Times,* July 10, 2016.

Aronoff, Kate (2015) "The Death of Climate Denialism", *In These Times,* June 15, 2015 http://inthesetimes.com/article/18040/the-death-of-climate-denialism

Backhouse, Roger E (2002) *The ordinary business of life: a history of economics from the ancient world to the twenty-first century*. Princeton University, Princeton and Oxford.

Backhouse, Roger E (2005) "The rise of free market economics: economists and the role of the State since 1970." *Hist. Polit. Econ.* 37(Suppl 1): 355-392. doi:10.1215/00182702-37-Suppl_1-355.

Bernstein, Michael A (2001) *A perilous progress: economists and public purpose in twentieth-century America*. Princeton University, Princeton.

Bogart, Ernst L (1939) "Government expansion in the economic sphere." *Annals Amer. Acad. Polit. Soc. Sci.* 206: 1-5.

Bowman, Andrew, Julie Froud, Sukhdev Johal, John Law, Adam Leaver, Mick Moran, and Karel Williams (2014) *The end of the experiment? From competition to the foundational economy.* Manchester University, Manchester; Palgrave Macmillan, New York.

Choudhury, Nusrat (2014, February 6) "Court-sanctioned extortion by private probation companies: modern debtors' prisons." ACLU blog, concerning the Human Rights Watch report, "Profiting from probation." https://www.aclu.org/blog/court-sanctioned-extortion-private-probation-companies-modern-debtors-prisons

Cobb, Kurt (2010) "Biophysical economics: Putting energy at the center"; *Scitizen*, Jan. 21, 2010.

Daly, Herman E. (1998a) "Consumption: Value added, physical transformation, and welfare." In David A. Crocker and Toby Linden, eds., *Ethics of Consumption: The Good Life, Justice, and Global Stewardship*. Lanham: Rowman and Littlefield, pages 19-27.

Daly, Herman E. (1998b) "Reply to Mark Sagoff's 'Carrying Capacity and Ecological Economics." In David A. Crocker and Toby Linden, eds., *Ethics of Consumption: The Good Life, Justice, and Global Stewardship*. Lanham: Rowman and Littlefield, pages 53-62.

Daly, Herman E. (2003) "The Illth of Nations and the Fecklessness of Policy: An Ecological Economist's Perspective" *Post-Autistic Economics Review*, No. 22, Nov. 24, 2003, article 1.

Day, John W., Christopher F D'Elia, Charles Hall, Adrian R.H. Wiegman, Jeffrey S Rutherford, Robert L Lane, David E Dismukes, "The Energy Pillars of Society and the Transition to Renewables," forthcoming.

Desmarais-Tremblay, Maxime (2013, May) "On the definition of public goods: assessing Richard A. Musgrave's contribution." Paper presented at 17th Annual Conf Europ Soc Hist Econ Thought, Kingston University, London. Also available at Documents de travail du Centre d'Economie de la Sorbonne. https://halshs.archives-ouvertes.fr/halshs-00951577/document

Edsall, Thomas B (2014, August 26) *The expanding world of poverty capitalism.* New York Times, Opinion Pages.

Fahrenthold, David A (2014, May 30) "Breaking points: where government falls apart; how the VA developed Its culture of coverups." *Washington Post*.

Farmer, Carrie M., Susan D. Hosek and David M. Adamson. Balancing Demand and Supply for Veterans' Health Care: A Summary of Three RAND Assessments Conducted Under the Veterans Choice Act. Santa Monica, CA: RAND Corporation (2016). http://www.rand.org/pubs/research_reports/RR1165z4.html

Fullbrook, Edward (ed.) (2007) *Real world economics; a post-autistic economics reader*. Anthem Press, London

Galbraith, James K (2008) *The predator state: how conservatives have abandoned the free market and why liberals should too*. Free Press, New York

Goodwin, Neva R (2014) "The human element in the new economics: a 60-year refresh for economic thinking and teaching." *Real-World Economics Review,* Issue 68: 98-118. http://www.paecon.net/PAEReview/

Gordon, Suzanne (2015) "Report: VA Outperforms Private Sector on Key Measures" *The American Prospect,* Nov. 20, 2015 http://prospect.org/article/report-va-outperforms-private-sector-key-measures

Gordon, Suzanne (2016) "Congressional Commission Moves Toward Privatization of VA," *American Prospect*, Feb. 9, 2016. http://prospect.org/blog/tapped/congressional-commission-moves-toward-privatization-va

Gowdy, J., Hall, C., Klitgaard, K., and L. Krall. 2010. "The End of Faith Based Economics." *The Corporate Examiner.* 37: No. 4-5: 5-11.

Hall, C.A.S., and K. Klitgaard. 2012. *Energy and the Wealth of Nations: Understanding the Biophysical Economy*. Springer, NY.

Hall, Charles, D. Lindenberger, Reiner Kummel, T. Kroeger, and W. Eichhorn. 2001. "The need to reintegrate the natural sciences with economics." *BioScience* 51 (6): 663-673.

Hall, Charles A.S., Jessica G.Lambert, Stephen B. Balogh (2014) "EROI of different fuels and the implications for society." *Energy Policy*. 64,: 141–152.

Howard, Philip (2014) *The rule of nobody: saving America from dead laws and broken government*. W W Norton.

Kalambokidis, Laura (2014) "Creating public value with tax and spending policies: the view from public economics." *Public Admin. Rev*. 74(4): 519-526. doi: 10.1111/puar.12162

Kime, Patricia (2016) "Panel weighs closing all VA health care — vets' groups raise conflict of interest" *Military Times*, April 1, 2016. http://www.militarytimes.com/story/veterans/2016/04/01/panel-weighs-closing-all-va-health-care-vets-groups-raise-conflict-interest/82522000/

Krippner, Greta R. (2001) "The elusive market: Embeddedness and the paradigm of economic sociology. *Theory and Society* 30: 775-810, 2001.

Labonte, Marc (2010, June 14) "Size and role of government: economic issues." Congressional Research Service Report, 7-5700.

Losen, Daniel J., Michael A. Keith II, Cheri L. Hodson, Tia E. Martinez (2016) "Charter Schools, Civil Rights and School Discipline: A Comprehensive Review," Civil Rights Project UCLA, March 16, 2016. https://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/charter-schools-civil-rights-and-school-discipline-a-comprehensive-review

Luntz, Frank (2014, November 5) "The midterms were not a Republican revolution." *New York Times*, pA31 New York print edition. http://www.nytimes.com/2014/11/06/opinion/the-midterms-were-not-a-republican-revolution.html

MacKinnon, J. B. (2016) "The Great L.E.D. Quandary." *New Yorker*, July 14, 2016
http://www.newyorker.com/business/currency/the-l-e-d-quandary-why-theres-no-such-thing-as-built-to-last

Matchar, Emily (2016, July 13) "The fight for the 'Right to Repair'." smithsonian.com
http://www.smithsonianmag.com/innovation/fight-right-repair-180959764/?no-ist

Mayhew, Anne. (2016) "A Commentary on Asad Zaman's paper: 'The Methodology of Polanyi's *Great Transformation*'" *Economic Thought*, 5.1, pp. 64-69.
http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-5-1-Mayhew.pdf

MITRE (2015), *Independent Assessment of the Health Care Delivery Systems and Management Processes of the Department of Veterans Affairs*, Vol. 1, The MITRE Corporation, Sept. 1, 2015.

Moore, Mark H (2014) "Public Value accounting: establishing the philosophical basis." *Public Admin. Rev.* 74(4): 465-477. doi: 10.1111/puar.12198

Mundy, Alicia (2016) "The VA Isn't Broken, Yet" *Washington Monthly*, March/April/May 2016.
http://www.washingtonmonthly.com/magazine/marchaprilmay_2016/features/the_va_isnt_broken_yet059847.php?page=all

Olorunnipa, Toluse (2016) "Hurting for Cash, U.S. National Parks Turn to Companies", *Bloomberg News,* June 17, 2016, http://www.bloomberg.com/politics/articles/2016-06-17/hurting-for-cash-u-s-national-park-service-turns-to-companies

Paulson, Henry (2016, Sept. 20 "How to raise trillions for green investments", *New York Times*
http://www.nytimes.com/2016/09/20/opinion/how-to-raise-trillions-for-green-investments.html?_r=0

Persson, Jonas (2015) "CMD Publishes List of Closed Charter Schools", *PRWatch*, Sept. 22, 2015)
http://www.prwatch.org/news/2015/09/12936/cmd-publishes-full-list-2500-closed-charter-schools

Rein, Lisa (2016) "Yosemite, Sponsored by Starbucks? National Parks to start selling some naming rights." *Washington Post*, May 9, 2016
https://www.washingtonpost.com/news/powerpost/wp/2016/05/09/yosemite-national-park-brought-to-you-by-starbucks/

Schuck, Peter H (2014) *Why government fails so often, and how it can do better*. Princeton University, Princeton

Sekera, June (2016) *The Public Economy in Crisis; A Call for a New Public Economics*, Springer Briefs in Economics.

Shapiro, Joseph (2014, May 19) "As court fees rise, the poor are paying the price." National Public Radio, News Investigations, with update on May 23.
http://www.npr.org/2014/05/19/312158516/increasing-court-fees-punish-the-poor

Shionoya, Yuichi (2005) *The soul of the German Historical School; methodological essays on Schmoller, Weber, and Schumpeter*. Springer, Boston

Siltala, Juha (2013) "New public management: the evidence-based worst practice?" *Admin.* 45(4): 468-493. doi:10.1177/0095399713483385

Smith, Thomas A (2014) "Review of Why government fails so often, by Peter Schuck and Simeon E. Baldwin." *Yale Alumni Mag*. 77(6): 58 https://yalealumnimagazine.com/articles/3936/reviews

Stillman, Sarah (2014, June 23) "Get out of jail inc." *New Yorker*.
http://www.newyorker.com/magazine/2014/06/23/get-out-of-jail-inc

Stretton, Hugh, and Lionel Orchard (1994) *Public goods, public enterprise, public choice – the theoretical foundations for the contemporary attack on government*. St. Martin's, New York

Thiel, Peter (2015) "The New Atomic Age We Need", *New York Times*, Nov. 27, 2015.

Urry, Amelia (2016, Oct. 11) "Apple's recycling robot wants your old iPhone. Don't give it to him." Grist.org. http://grist.org/business-technology/apples-recycling-robot-wants-your-old-iphone-dont-give-it-to-him/

U.S. Department of Justice (March 2015) *Investigation of the Ferguson Police Department*.

Zaman, Asad. (2016) "The Methodology of Polanyi's *Great Transformation*." *Economic Thought*, 5.1. http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-5-1-Zaman.pdf

Zapotosky, Matt (2016) "Justice Department warns local courts about unlawful fines and fees", *Washington Post*, March 14, 2016

Author contact: junesekera@gmail.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

Why does capital flow from poor to rich countries? – The real puzzle

Michael Joffe   [Imperial College London, UK]

Abstract

Lucas' classic paper (1990) highlighted the paucity of capital flows from rich to poor countries, in contrast with predictions of standard theory – the Lucas puzzle (or paradox). Subsequently, abundant capital has flowed from certain low-income countries to rich ones, notably China to the USA. More generally, empirical research has shown that fast-growing developing countries rely less on foreign capital; and, international capital flows towards countries with lower productivity growth and lower investment – the "allocation puzzle". This paper considers post-reform China, finding that massive outflows are a *consequence* of growth that is readily understandable, and widely understood (except by some economists). Similar experiences also occurred in previous dynamic Asian economies. *The causal direction is different from what Lucas assumed*.

The academic literature seeks to explain the discrepancy from standard theory – not the phenomenon itself – by invoking financial-sector weakness or under-development and its impact on borrowing or saving, or accumulation of foreign reserves that affects the exchange rate. Mainstream economic accounts rely on what they see as "typical" behavior, which accords with standard theory. However, the empirical analyses strongly suggest that East Asia has a different dynamic, with different causal processes. This is not a minor exception, given these countries' population size, contribution to world economic growth in the past half-century, and influence over global capital imbalances. Their strong growth performance suggests that invoking financial "weakness" is misplaced. Rather, a better analysis is needed of the impact of these countries' channeling of capital for strategic purposes, and how this brings about a distinct pattern of causal processes.

JEL codes F21, F41, G20, O16

Key words Lucas puzzle, Lucas paradox, international capital flows, China capital outflow, allocation puzzle, financial-sector weakness

1. The Lucas puzzle – theory

In 1990, Robert Lucas published a classic paper, entitled "Why doesn't capital flow from rich to poor countries?" (Lucas 1990). He raised an issue which has since become known as the "Lucas puzzle" (or "Lucas paradox"): given that theory predicts that capital should flow from capital-rich high-income countries to capital-poor low-income ones, why do the data show that this does not happen?

The theoretical argument is based on the uncontroversial observation that rich economies have more capital than poor ones, and on the standard feature of neoclassical theory that capital is subject to diminishing returns. Poor countries, with little capital, should therefore have a higher rate of return than rich countries with abundant capital. Lucas' example was that India, with production per person about a fifteenth of that of America (according to estimates by Robert Summers and Alan Heston), should have a marginal product of capital 58 times larger – based on a Cobb-Douglas production function and a plausible capital share of 0.4. Investment in India should be highly attractive from an American viewpoint – "Indeed, one would expect *no* investment to occur in the wealthy countries". As he commented, "there is nothing at all delicate about this standard neoclassical prediction on capital flows. The

42

assumptions on technology and trade conditions that give rise to this example must be drastically wrong, but exactly what is wrong with them, and what assumptions should replace them? This is a central question for economic development."

Lucas presents possible reasons for this discrepancy. One is that the human capital of workers in the different societies means that investment in a country like India would be far less productive than the above calculation would suggest – although he points out that this theoretical addition would imply that there should also be no economic motive for labor flows either. A second is that there could be external benefits of human capital, the type of spillover proposed by Paul Romer, and calculates its magnitude using estimates from Denison. However, as Lucas says, this idea requires these external effects to be confined to their originating countries, whereas it seems plausible that at least some of them cross national borders. These two arguments depend on an implicit assumption that the capital would be invested in production rather than, say, government bonds or real estate. A further reason for the lack of capital flow from rich to poor countries is the existence of capital market imperfections, specifically the difficulty of enforcing the payment of interest payments or repatriated profits once an investment has been made – a form of political risk.

The arguments presented are intended to explain the absence of rich-to-poor country flows. If that were the dominant reality, Lucas could be regarded as having fulfilled his objective – he did not set out to account for the presence of poor-to-rich country capital exports. But as is well known, large-scale flows subsequently occurred, from relatively low-income countries like China to the rich world, including America. The first two explanations could account for such flows, if they were invested in the highly productive real economy within the recipient rich country, but lose their plausibility when it is realized that most of the investment has been in the financial sector (e.g. bonds) or in real estate. The third explanation is not directly relevant to such flows, but the corollary – more reliable institutions in the rich world – could well be part of the explanation for poor countries' decisions on where to place their money.

**2. The Lucas puzzle – evidence**

There is now abundant evidence of the scale as well as the direction of these capital movements, which are termed "uphill" because they flow in the direction opposite to that predicted by standard theory. For example, Prasad et al. (2007) divided countries in their sample into those having a surplus or a deficit in their current account, and calculated the purchasing power-adjusted per capita GDP for the two groups, weighting the estimates by each country's contribution to the surplus or deficit. In the early part of the period that they cover, the 1970s and early 1980s (and especially 1975-1981), surplus countries were richer than deficit countries, i.e. uphill flows did not exist. This is true also for most of the 1990s. From 1984 until 1990, the time that Lucas was writing, there is evidence of a small uphill flow. And from 1998 a wide gap develops, with the surplus countries now being clearly the poorer ones (figure 1). It is true that FDI flows downhill, but it represents only about 40 percent of private capital flows to developing countries.

Prasad et al. (2007) also provide a different perspective on this issue by examining the association of capital flows with growth rates rather than with level of prosperity. This is what Gourinchas and Jeanne call the allocation puzzle (see below). A focus on fast-growing countries should bypass what Prasad et al. call "a variety of problems – inadequate infrastructure, a poorly educated labor force, corruption, and a tendency to default on debt

from abroad, among other factors – that reduce the risk-adjusted returns to investment". This is because whilst the Lucas puzzle in principle can be explained away by such factors, on the grounds that they would impede profitable investment and therefore also growth, the same does not apply to fast-growing countries that have evidently largely overcome them. Their findings are that in 1970-2007, the net amount of foreign capital flowing to relatively high-growth developing countries was smaller than that flowing to the medium- and low-growth groups. A more extreme pattern was seen in 2000-2004, when out of all developing countries, only the low-growth ones received significant amounts of capital, with China and other high-growth countries exporting large amounts of capital, and with India and medium-growth countries exporting moderate amounts.

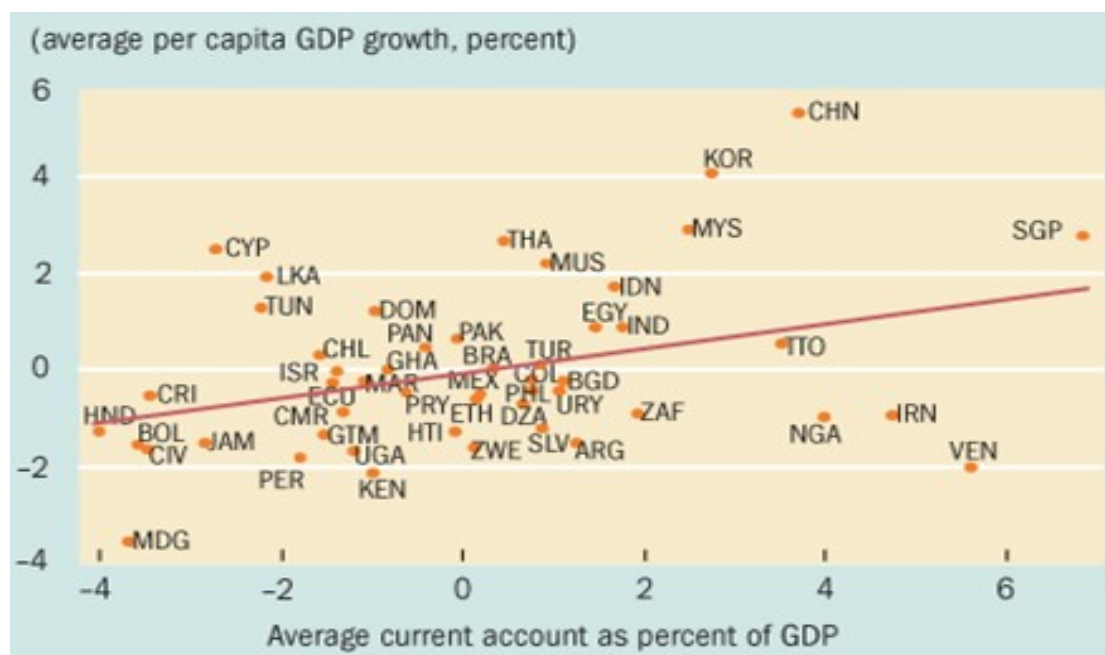**Figure 1** The relative income of capital surplus and deficit countries, 1970-2005
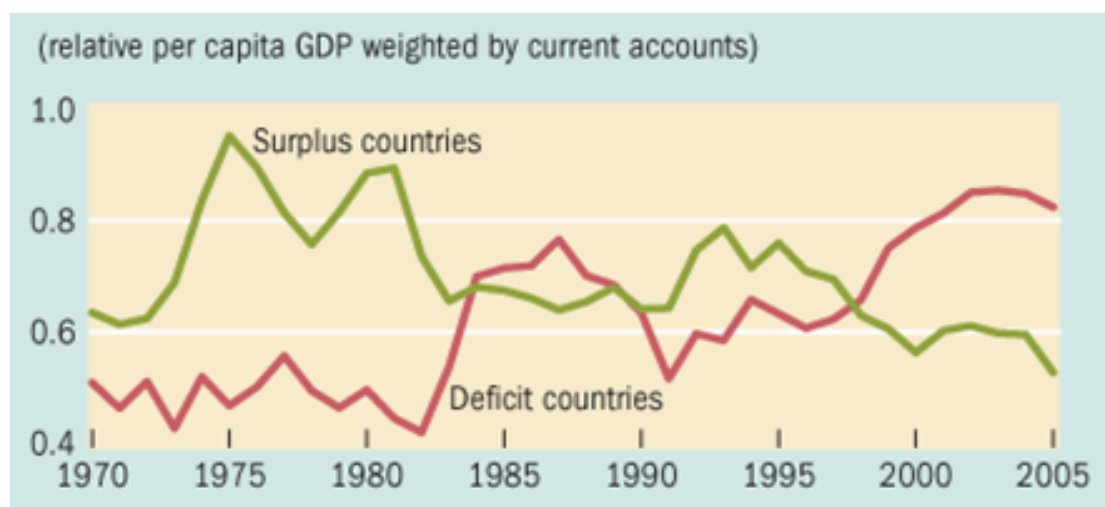
Taken with permission from Prasad E, Rajan R, Subramanian A. The paradox of capital. Finance & Development (IMF) 2007; 44(1), Chart 1. Available at
http://www.imf.org/external/pubs/ft/fandd/2007/03/prasad.htm

Note: their sample of countries was divided into two groups – those with current account surpluses and those with deficits in that year. They then computed a current account-weighted measure of the incomes of each group of countries, relative to that of the United States.

In addition, Prasad et al. (2007) carried out a cross-country analysis, plotting the average level of GDP growth for each country against the average current account as a percentage of GDP, for 1970-2004. Theory predicts a negative relationship – a downward-sloping curve. Instead, the scatterplot shows a rising regression line (figure 2). And in particular, the group of economies with a positive current account of more than 2 percent of GDP that also had strong growth contains China, South Korea, Singapore and Malaysia. The other countries in their sample with a positive current account of more than 2 percent of GDP were Venezuela, Iran, Nigeria and Trinidad & Tobago – large producers of oil and/or gas – an issue to which we will return.

Prasad et al. (2007) also state that "countries that had high investment ratios *and* lower reliance on foreign capital (lower current account deficits) grew faster – on average, by about 1 percent a year – than countries that had high investment but also a greater degree of reliance on foreign capital." This finding reinforces that of Aizenman et al. (2004), who observed that countries with high self-financing ratios grew faster.

Figure 2 Cross-country growth rates and average current account, 1970-2004

Taken with permission from Prasad E, Rajan R, Subramanian A. The paradox of capital. Finance & Development (IMF) 2007; 44(1), Chart 4. Available at
http://www.imf.org/external/pubs/ft/fandd/2007/03/prasad.htm

3. The economic emergence of China

In order to explore the forces that lie behind this apparently puzzling phenomenon, I will first examine the most dramatic case, that of China. I will then discuss the extent to which China's experience has been typical of the whole phenomenon.

In 1978, not long after the death of Mao Zedong, economic reforms began to be implemented. The main changes were: (i) rural households were now allowed to keep their own surpluses (the "household-responsibility system"); (ii) Township and Village Enterprises were allowed to operate in a manner similar to capitalist firms; (iii) Special Economic Zones such as Shenzhen were set up, based on foreign capital and the export market; and (iv) State-Owned Enterprises were increasingly required to operate according to market logic to improve their economic efficiency (Lin et al., 2008). The first two of these, peasant agriculture and Township and Village Enterprises, did not need large quantities of investment, as they were both low-cost activities; in very many instances they gradually expanded by ploughing their profits back into the business. The capital for investment in State-Owned Enterprises continued to be the responsibility of government, in continuity with the pre-reform era. In contrast, the Special Economic Zones did rely on new sources of funding, largely foreign direct investment – which had the additional advantage of bringing technology and knowhow with it – but also some portfolio investment. Much of this foreign capital was from neighbors that had already developed substantial modern industry, and that also had close cultural links, especially Hong Kong and Taiwan.

Replacement of Soviet-style centralized planning by organizations that operated more like capitalist firms had a dramatic impact on the economy. In particular, the manufacturing sector

developed on the basis of very low unit costs – low wages relative to the productivity level. This, together with an undervalued currency, enabled Chinese products to be marketed extremely cheaply, which became known as the "China price". The result was that Chinese manufactures conquered the world.

Within China, the large and ever-growing volume of exports led not only to unprecedented levels of sustained economic growth, and rising living standards for an increasing proportion of the population, but also to soaring quantities of capital. This largely consisted of corporate profits from export sales, predominantly in foreign hard currency. In addition, household saving rates were extremely high, due to increasing wages together with an important precautionary element because of low social security provision, plus very likely a strong cultural element as well. These household savings were channeled by state banks to State-Owned Enterprises, allowing massive capital investments to be made, albeit not always in the most efficient manner.

The saving rate, as a percentage of GDP, fluctuated between 35 and 43 percent – already high by international standards (especially if one excludes oil exporters) – until the early 2000s, when it rose to 50 percent or above (figure 3). The well-known near-exponential Chinese GDP growth was thus accompanied by equally strong growth in gross savings, with an even steeper increase during 2002-2006 (figure 3). It is plausible that the rise in percent savings in this latter period was at least partly due to the ever-increasing prosperity of industry and also of its employees, whose consumption level did not keep up with their increase in earnings.

Figure 3 Growth and savings in China, 1982-2012

Source: World Bank http://data.worldbank.org/indicator/
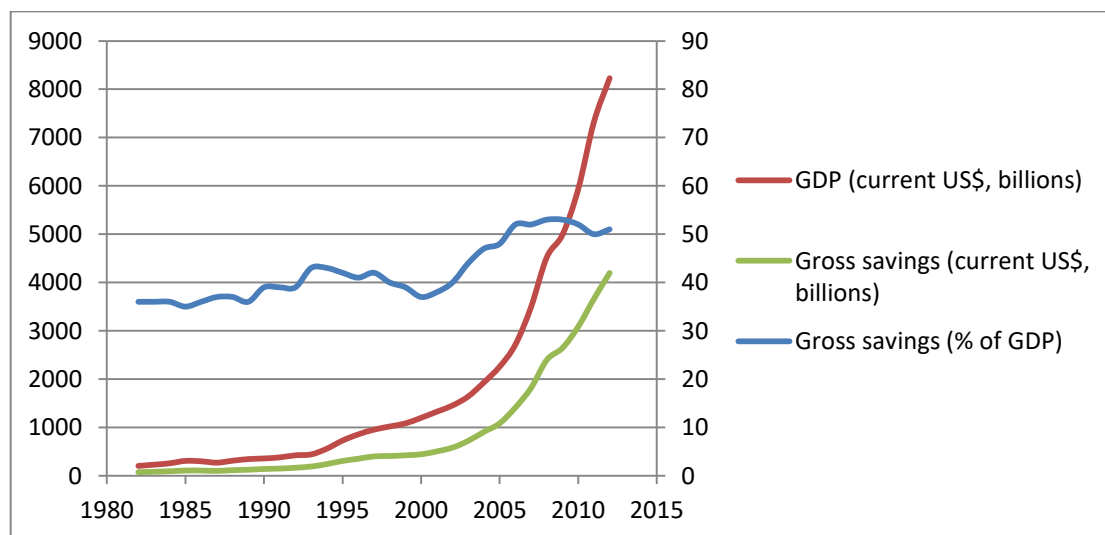
Much of this capital was ploughed back into domestic investment in industry and infrastructure. But not all of it – copious quantities flowed overseas. The destinations were diverse: some was used to purchase bonds, e.g. US Treasury bonds. Some went into buying existing infrastructure, or building new infrastructure (especially in Africa). Some went into productive investment in western industry, giving access to technology and brands. The Chinese current account rose from its previously positive but relatively moderate level close to

the range 20-40 billion US dollars annually in 1998-2003 to a peak of 420 billion in 2008, before falling back to approximately 150-250 billion since then (State Administration of Foreign Exchange, China; World Bank).

One factor that may have contributed to the export of capital from China was a precautionary motive, following the experience of many East Asian countries during the crisis of the late 1990s. However, the figures do not support this as an important factor, because the main rise in capital exports did not begin until 2004, several years after the East Asian Crisis.

In summary, international capital flows involving China showed a persistently positive current account starting in the late 1990s. In other words, capital was exported from this relatively poor country, e.g. in terms of GDP per capita, mainly to rich countries such as the USA. There is no puzzle about this, because the quantity of corporate profits and of domestic savings has been so enormous that it is unsurprising that some of it would flow abroad – especially as much of it was in hard currency, derived from exports.

4. How typical is China?

One response to the analysis so far could be, China is unique. There is some plausibility to this idea, particularly in the magnitude of the transformation of the Chinese economy and its impact on the rest of the world. But in fact it is only an extreme example of a more general, if not universal, phenomenon.

The East Asian economies that have previously experienced prolonged rapid growth have had highly profitable industry, and have also been major capital exporters at least for parts of their periods of growth. This was true of Japan, and later of the four smaller "tigers", Taiwan, South Korea, Hong Kong and Singapore. Later, they were joined as capital exporters by Malaysia, Thailand, Indonesia and subsequently India (Alfaro et al., 2014). Most of these countries also showed a rise in corporate saving during this period (Bacchetta and Benhima 2015). Another recurring pattern is that the capital has flowed from the early developers to later ones, e.g. Japan to Taiwan, then Taiwan to China, and subsequently China to elsewhere in the region (as well as outside it).

A similar outflow of capital is seen if it is generated from a different source. In figure 2, we noted that the countries in the sample of Prasad et al. (2007) which had a positive current account of more than 2 percent of GDP included not only Asian economies with strong growth records in 1970-2004 (China, South Korea, Singapore, Malaysia), but also four with poor or negative growth during this period: Venezuela, Iran, Nigeria and Trinidad & Tobago. This suggests that large-scale production of oil and/or gas is an alternative source of abundant exportable funds. Table 1 shows all the countries that exported more than ten billion US dollars' worth of capital in 2012. Economies with a strong track record in manufacturing exports, both Asian and European, are strongly represented. The other main category is a group of oil and/or gas producers, which may or may not have economies that perform well in broader terms, but which have large foreign-currency incomes from hydrocarbon exports (James 2014). These findings confirm the idea that copious quantities of capital tend to lead to a strongly positive current account.

Table 1 Countries that exported capital amounting to more than 10 billion US dollars in 2012

East Asian manufacturers	(billions of US$)	Oil and/or gas producers	(billions of US$)
China	215	Algeria	12
Japan	60	Angola	14
South Korea	51	Azerbaijan	15
Macao, China	18	Iraq	30
Singapore	52	Kuwait	79
		Libya	24
European manufacturers		Malaysia	16
Denmark	19	Nigeria	17
Germany	249	Norway	64
Netherlands	90	Qatar	62
Sweden	32	Russian Federation	97
Switzerland	69	Saudi Arabia	165

Source: World Bank
http://databank.worldbank.org/data/reports.aspx?source=2&series=BN.CAB.XOKA.CD&country

There is another parallel. Much of the poor-to-rich investment is done by sovereign wealth funds and central banks, both in the industrial and the oil/gas power houses. It puts a country in a strong position if it is able to "help" the rich world, e.g. by bailing out troubled financial institutions in a crisis, which can be useful in geopolitical terms. Furthermore, the hydrocarbon exporters are thus able to diversify their economies, so that reducing reliance on a finite resource is one motivation. But in addition, many of these countries contain extremely wealthy individuals who are able to buy assets in the West, including prestigious buildings and other assets such as football clubs. Non-state organizations may also be involved, e.g. private equity firms. The common thread is that *abundantly-generated capital creates pressure to find outlets, and this occurs through multiple pathways*.
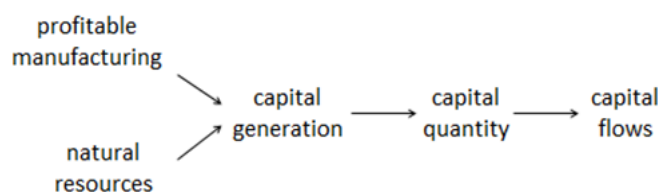
5. Causal direction

In figure 4, panel (a) shows the causal direction assumed by Lucas, and by more recent authors who have adopted the same theoretical framework. The starting point is the existing quantity of capital, which implies a certain marginal productivity of capital – given a negative sign in the diagram because of diminishing returns. This in turn leads to a predicted capital flow, which depends positively on the marginal productivity of capital as an incentive. This two-link causal model is applied to the real world by adding the uncontroversial assumption that in general, poor countries tend to have a lower capital *stock*.

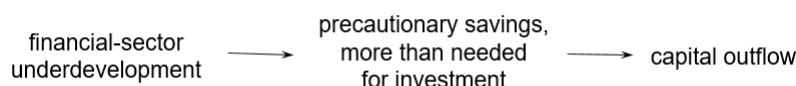**Figure 4** Comparison of causal directions

(a) causal direction assumed by Lucas (1990):

capital quantity $\xrightarrow{\ -\ }$ marginal productivity of capital $\longrightarrow$ capital flows
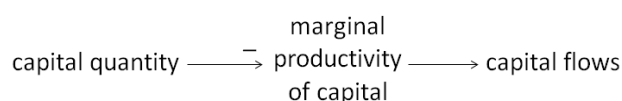
(b) causal direction suggested by the evidence on profit generation and capital flows:

(c) causal direction assumed by Sandri (2010):

In contrast, the evidence on profit generation and capital *flows* that we have just reviewed suggests that capital generation leads to its increase in quantity, and thence to its outflow – all of these being positive relationships, so that the predicted consequence has the reverse sign of that predicted by Lucas. The generation of capital can come from profitable manufacturing, or from a natural resource such as oil. This is shown in a schematic causal diagram in figure 4, panel (b).

The view expressed schematically in panel (b) of figure 4 implies that saving would be a *consequence* of economic growth. This contrasts with the assumption that is conventionally made that the causal direction is from savings to growth. Statistical evidence, mainly using Granger-causation techniques to investigate the time order, strongly favors the former (Blomström et al., 1996; Carroll et al., 2000; Rodrik 2000; Attanasio et al., 2000).

Another way of looking at the same issue is that panel (b) answers the question, "where does capital come from?", whereas panel (a) does not appear to be aware of it, except for the observation that the stock is lower in poor countries. In a world where most growth is catch-up growth – convergence, as predicted by the Solow neoclassical growth model – one would expect capital to be generated largely in those economies that are catching up. The Lucas conception is static, whereas the Solow model implies a dynamic process of convergence. Of course, it could be that capital generation in catch-up growth economies is merely incremental, adding gradually to that country's stock of capital. What the data show is that beyond a certain level of export-led growth, the quantity of the generated surplus exceeds the amount that can be profitably reinvested in the domestic economy.

There is a sense in which the pathways depicted in figure 4 panel (b) are not new discoveries – indeed, that "everybody knows" this. Informed commentators on the international economic scene are certainly familiar with these ideas, and they occur regularly in descriptions of current events. To take just two examples, from a single issue of *The Economist*, a high-quality publication that is generally well disposed to orthodox economics (with my emphasis):

> "China's financial repression … has contributed to China's remarkably high rate of saving, which reached over 50% of GDP in 2012. **This is more than China can invest at home, obliging it to export some of its savings** (typically 2-3% of GDP) abroad. This incurs the wrath of its trading partners" (*The Economist,* 29 March 2014, p. 65).

"… the real interest … rate … has been dragged down by long-term structural trends. A global savings glut is partly to blame: **export powerhouses like the OPEC countries and China buy vast quantities of rich-world debt**, depressing borrowing costs in the process" (*The Economist,* 29 March 2014, p. 75).

6. Existing explanations

My discussion of the literature attempting to explain the Lucas puzzle, and its modern version of poor-to-rich countries flows, will focus specifically on the mainly middle income capital-exporting countries that were relatively poor until quite recently, notably China and certain other Asian countries.[1] The studies reviewed here are based on datasets covering a much wider range of countries. This means that their focus is far broader, often global, and also highly heterogeneous: a negative correlation between capital inflow and growth could apply equally to a dynamic economy that exports capital and to a shrinking economy that relies on capital imports. Furthermore, in many of these papers, different patterns of growth dynamics are not separated out, e.g. Mexico versus the state-directed strategy of South Korea. This is one reason that my interpretation may differ from that of the authors of each study. Another point to bear in mind is that many studies are unweighted, giving equal weight to all the participating countries – unlike Prasad et al. (2007), whose weighting meant that countries were influential in proportion to their contribution to international capital flows.

Lucas' puzzle: rich-to-poor country flows

Much of the literature that has explored these issues has followed Lucas' own ideas on the likely explanation for the puzzle about the relative lack of rich-to-poor country flows. Alfaro et al. (2008) found that institutions in the form of government stability, bureaucratic quality, non-corruption, and law and order are the major factors. Similarly, Papaioannou (2009) accentuated property rights, legal efficiency and contractual institutions that guard against expropriation; low corruption and bureaucratic efficiency were also found to be important specifically for FDI. In addition, Göktan (2015) demonstrated that the tendency for the very poorest countries (*per capita* GDP less than $12,000) to have the lowest inflows of cross-border bank lending could be explained by institutional quality. These institutional accounts go a long way to explaining Lucas' original puzzle.[2]

Other interpretations, that focus on the financial system within the capital-exporting country or on international capital frictions, have a much more uncertain role. Reinhardt et al. (2013) demonstrated that in financially open economies the Lucas puzzle is not seen for the period 1980-2006, and also that there is no systematic relationship in countries with a closed capital account – thus "the 'failure' of the neoclassical model to predict international capital flows can also be explained by a violation of one of the model's key underlying assumptions: capital can flow freely across countries". They did not address Lucas' observation that *all* investment

---

[1] I use the shorthand of "poor" countries, following the common usage in this literature, stemming from Lucas' original paper. For brevity and clarity of presentation, I focus only on the contributions in this vast and rich literature that are most relevant to the central theme of the present paper. For a fuller treatment of this literature, see e.g. Chinn (2013) and Gourinchas and Rey (2014).

[2] Some poor-to-rich country flows could also occur for such reasons, e.g. because heads of government, rich households, etc. seek a safe haven for their money. However, the magnitude of these flows falls far short of the main phenomenon outlined above.

should be in poor countries. They predicted that the paradox will disappear as financial openness spreads over time, noting that capital account restrictions have been gradually lifted in most countries over the course of the past three decades – but did not consider this observation alongside the evidence that poor-to-rich country flows became large during the last third of the period studied. And in contrast to Reinhardt et al., Göktan (2015) found that capital market imperfections were unimportant in explaining countries' capital inflows.

An alternative interpretation is that what Reinhardt et al. classify as lack of financial openness is in fact a different developmental model in East Asia. As Chinn and Ito (2008, p.489) say, "More open financial markets do not appear to have any impact on current account balances for this group of countries… If anything, arguments based on this thesis have inappropriately extended a characterization applicable to industrialized countries to less developed countries".

In principle, international capital frictions could reduce the magnitude of rich-to-poor country flows. Their role is subject to conflicting evidence. Caselli and Feyrer (2007) showed that the marginal productivity of capital is similar across countries, implying that international credit frictions are unlikely to explain the Lucas puzzle. Rather, the major factors are endowments of complementary factors and efficiency – i.e. lower productivity for reasons other than lack of capital.

On the other hand, Kalemli-Ozcan et al. (2010) examined capital flows between the states of the USA, and found that they accord with theory, in the sense that "capital flows to fast growing states from slow growing states and as a result high growth states pay capital income to other states". This leads them to suggest that the Lucas puzzle is due to frictions associated with national borders – i.e. that international capital markets are de facto incomplete – although other interpretations are possible for the discrepancy between these findings and the international-flow context.

Poor-to-rich country flows

Caballero et al. (2008) approach financial issues from the viewpoint not of the potential borrower, but rather of the supply of store-of-value financial instruments. They attribute the sustained rise in the US current account deficit to the lack of capacity of other regions of the world to generate financial assets from real investments. They are doubtless correct that US financial assets are seen as desirable, and that most parts of the world are unable to produce anything comparable. Again, the magnitude of the effect would be small compared to the observed international capital flows; (see also Mendoza et al., 2009).

More nuanced analyses of international capital flows separate out the different types of capital flow. Alfaro et al. (2014), for the period 1970-2007, found that net *private* flows (including FDI as well as portfolio investment) went to growing countries, even if these countries were net exporters of *total* capital, highlighting the need to explain the puzzling direction of public capital flows. They note that East/Southeast Asia is atypical – their central banks buy reserves in developed countries, hence the outflow of capital (cf. also Krishnamurthy and Vissing-Jorgensen, 2012; Reinhart and Tashiro, 2013) – although they do not mention the *ability* of these central banks to do this, i.e. the source of the abundant capital. In other regions, the private sector conforms with theory in its direction – although the paper provides no quantitative estimates of the magnitude of such flows, to see whether they are comparable with theoretical predictions, as Lucas did in his original paper.

A somewhat different pattern was found by MacDonald (2015) for 1980-2010: portfolio investment outflows, which are privately held, exceeded FDI inflows in high-growth countries. This was attributed to a desire for international portfolio diversification in liquid assets. The observed pattern was found to be greater with liberalized capital accounts. Again, the source of the funds was not explicitly traced to the high profit generation and rising incomes in these countries.

Hypotheses concerning the reason for these observations fall into three main groups: a focus on the weakness of the financial systems in capital-exporting countries, an excess of saving over investment (also resulting from financial under-development), and the effect on exchange rates of foreign reserve accumulation.

The hypothesis of a weak financial system

Prasad et al. (2007), who we relied on above when presenting the evidence on the issue, do actually ask whether fast-growing countries may *need* less foreign capital, because higher growth generates higher domestic savings. But they reject this idea, because "typically, as countries grow (that is, when they experience a positive productivity shock), they should want to consume more (because they are richer) and invest more (because of the investment opportunities)". Their response to this conundrum is that it results from a weak financial system.

One result is that entrepreneurs are forced to rely on self-financing, because incomplete financial markets mean that they are unable to borrow: "Corporate investment could be limited to the funds firms generate internally from past investment". This idea is shared by other authors (e.g. Sandri, 2010; Song et al., 2011; Alfaro et al., 2014). Nevertheless it is odd, because there is abundant evidence that firms in a wide variety of different circumstances have generally relied on retained profits for their continued investment (e.g. O'Sullivan, 2007) – there is no need to invoke financial restrictions to make entrepreneurs do this, and in a high-profit economy such as post-reform China it would be even less necessary.

Another consequence of a weak financial system is that consumers could not borrow in anticipation of higher future income, as predicted by the permanent-income hypothesis. However, it is much more plausible that in the context of an expanding economy, where those who are participating in it receive rising incomes, their consumption would increase more slowly than wages – possibly influenced by past habits – rather than by future-orientated conjectural possibilities. It is likely that this is what happened in China, as noted above. The importance of past habits would accord with the formal non-stochastic *AK* growth model, with perfect foresight, proposed by Carroll et al. (2000).

Prasad et al. (2007) also state that a weak financial system might not be good at intermediating foreign capital, leading them to wonder, in that case "where are the productivity gains coming from?". Although not an easy question to answer in a causal sense, it is clear descriptively that the sequence of East Asian economic miracles has had at its root a competitive advantage based on low unit cost – i.e. low wages for the level of productivity. There is no compelling reason why capital imports (other than FDI) should be essential to this.

There is another still more compelling reason to reject the hypothesis that capital exports from East Asian countries are due to their having a weak financial system. Many countries across

the world have weak financial systems, especially in low- and middle-income economies. Most of these do not have a positive current account balance. Something else must be going on.

Excess saving over investment in poor countries

Another explanation focuses on investors' precautionary response to the risk they face, again with an emphasis on financial underdevelopment. Sandri (2010) suggests that the resulting high investment risk encourages entrepreneurs to undertake excess saving. There would then be an excess of saving over investment, which would lead to the export of capital. Such a process "can explain why growth accelerations in developing countries tend to be associated with current account improvements". This interpretation, if applied to China, would imply that its capital exports have been due to a relative lack of investment, which would be strange given that China's investment levels have been famously high (Bai et al., 2006).

Sandri's focus is thus on the observation that the increase in savings rate in fast-growing developing countries is even stronger than the increase in investment rates. His response is to try and explain why excess savings would occur, as a *decision* made by entrepreneurs in response to financial underdevelopment (figure 4 panel (c)). In his view, they also "increase saving to finance the investment which triggers higher growth". This is in line with the conventional assumption that the causal direction is from savings to growth, whereas the evidence has been clear for some time now that the reverse is true, as noted above. The problem is that this evidence has not displaced standard theory and the habits of thought that go with it. If one instead follows the evidence, the obvious interpretation is that growth generates increased wealth, which in turn facilitates increased saving, manifest as corporate profits, household saving and likely increased tax receipts. It is a question of *flow* rather than primarily of decision making.

In any case, Sandri's emphasis on the causal importance of a weak financial system on saving is contradicted by the evidence of Chinn and Ito (2008), who studied 19 industrial and 69 developing countries during 1971-2004. As they concluded, "our empirical findings are *not* consistent with the argument that the more developed financial markets are, the less saving a country undertakes. **Especially for most of the East Asian emerging market countries, we find that more financial development leads to** *higher* **saving**" (pp. 480-81; italics in the original, bold emphasis added). This conclusion applies to all of emerging South, Southeast and East Asia, including China, except for Hong Kong and Singapore (Chinn and Ito, 2008, Figure 3). In China specifically, financial development has led to an even greater increase in investment than in savings (p. 489). They conclude, "we found evidence that the oft-cited effects of financial and legal development are only applicable to industrial countries" (p. 493).

A related analysis is that of Buera and Shin (2009), who postulate that in countries that have undergone reforms which remove distortions but leave financial frictions intact, allocation improves, and consequently TFP increases. Saving rises immediately as a permanent-income effect, but investment only later, leading to a short-lived surplus of saving over investment, and therefore to capital exports during this period. The transient nature of this phenomenon would only be relevant to temporary outward capital flows, not to the longer-term ones seen in China and comparable countries.

Song et al. (2011) have a different interpretation: in China, the large trade surplus has resulted from high-productivity but financially-constrained private firms outgrowing low-

productivity State-Owned Enterprises with good access to credit markets. "The downsizing of financially integrated firms forces domestic savings to be invested abroad, generating a foreign surplus." However, their theory predicts falling investment rates, whereas the evidence shows no such decline and a possible increase starting in the mid-1990s (Bai et al., 2006). Moreover, their notion that the private firms "must rely heavily on retained earnings to finance investment" needs to be considered in the light of the very high rates of return. According to Bai et al. (2006), the return to capital in China has been at least 20 percent since 1979, and this must have been considerably higher for the high-productivity firms that are alleged to have been starved of capital. In any case, private sector firms have in fact received substantial bank loans, especially during the firms' start-up period (Allen et al., 2005).

The hypotheses of Buera and Shin and of Song et al. have been tested empirically by Fan and Kalemli-Ozcan (2016). They used firm-level data for several Asian countries undergoing financial reforms during the period 2002-11. Firms that were previously credit constrained did decrease their savings more (or increase their savings less) than those that were not, as predicted. However, overall corporate savings rose after the reforms, rather than falling as predicted, and the national current account surplus was not significantly affected.

Broner and Ventura (2016) present an ingenious model of the interaction between domestic and foreign debts in which domestic debts support foreign debts. A ''financial depth'' effect allows the country to sustain more foreign borrowing than in the representative-agent benchmark, and more domestic borrowing than in autarky. Another possible consequence is a ''capital flight'' effect with low domestic and foreign borrowing. Clearly, neither of these applies to the experience of China or of other fast- growing capital exporting countries, with low foreign borrowing and high domestic saving.

Foreign reserve accumulation and exchange rates

A further suggestion of Prasad et al. (2007) is that excess foreign capital can lead to currency appreciation, so avoiding that would be good for export-oriented manufacturing. This appears to be true for capital-exporting countries, albeit in mirror image – the export of capital leading to currency depreciation will have reinforced the competitive advantage based on low unit cost (Rodrik, 2008; McMillan et al., 2014). It is likely to have been an important contributory factor in a country like China, but not the main driving force, because a decision to export capital will only have this effect if the economy is already highly competitive internationally, and is generating capital. A low-income country with inefficient industry and capital scarcity would not succeed with such a policy.

Accumulation of foreign reserves is taken as the primary causal factor in the model of Benigno and Fornaro (2012). It induces a real exchange rate depreciation and a reallocation of production towards the tradable sector that boosts growth. They compare the optimal reserve policy with a laissez-faire equilibrium, and find that it "entails a fast rate of reserve accumulation, as well as higher growth and larger current account surpluses compared to the economy with no policy intervention." However, it is implausible to attribute any sizable growth performance to this mechanism; they estimate the gain as a 1 percent permanent increase in consumption. Typical East Asian growth rates have been far higher during this period.

Similarly, Korinek and Servén (2016) focus on "neo-mercantilist" reserve accumulation, and find that it entails a static cost in the reduced consumption of tradables, but a dynamic gain

due to an increase in growth. This model applies to countries with limited investment opportunities, not to China and similar countries.

The puzzle of productivity growth and lack of inflows

Gourinchas and Jeanne (2013) consider what they term the "allocation puzzle", which is related to that of Lucas. Rather than focusing on the pre-existing quantity of capital, they consider the rate of productivity growth: "the textbook neo-classical growth model … predicts that countries that enjoy higher productivity growth should receive more net capital inflows. We … find that this is not true. … The non-OECD countries that have grown at a higher rate over 1980-2000 have tended to export (not import) more capital." Contrary to expectations, the results are similar for financially open and closed countries. The findings reflect stagnant or contracting countries with capital inflows (e.g. Mozambique) as well as dynamic economies that exported capital (e.g. Taiwan).

The authors introduce a saving wedge that needs to be strongly negatively correlated with productivity growth – "the allocation puzzle is a *saving* puzzle".[3] Furthermore, they (like Sandri, 2010) note that "savings not only has to be positively correlated with productivity growth, but the correlation must be stronger than that between investment and productivity growth". In the context of China, their saving puzzle corresponds to the generation of copious savings – i.e. it indicates the need for theory to consider where the capital originates.

Further analysis shows that the negative correlation of capital inflows with productivity growth is due to the influence of six outliers, five from East Asia (China, Korea, Taiwan, Hong Kong and Singapore), together with Botswana; if these are excluded, the expected positive correlation is observed for the remaining 62 countries (Rothert, 2016).

Gourinchas and Jeanne conclude that the tendency for capital "to flow more toward countries with lower productivity growth and lower investment … is puzzling for neoclassical models of growth – in fact, this makes one wonder if the textbook neoclassical framework is the right model at all to think about the link between international financial integration and development." They then review some possible explanations:

(1) saving leads to growth, which would require a friction that prevents foreign savings substituting for domestic savings;
(2) growth leads to savings, as households' income rises faster than consumption, in accordance with Carroll et al. (2000) – but this only explains the allocation puzzle if savings rise faster than investment;
(3) domestic financial-sector frictions, because low financial development implies (a) less borrowing against future income, (b) less responsiveness of investment to productivity growth, and (c) possible precautionary savings, by households because they lack social insurance, and/or by governments to deal with the risk of crisis;
(4) productivity growth *in the tradeable sector* could lead to a surge in exports, and capital outflows.

Of these, (1) can be discarded (at least for China), because of the evidence of abundant domestic savings that would prevent competition from external finance, as well as of the

---

[3] An exploration of this idea emphasizing age differences in borrowing has been presented by Coeurdacier et al. (2015).

evidence on causal direction. (2) is in accordance with the reality of East Asian growth, but greatly understates the case because it refers only to households, omitting the huge profitability of industry during the rapid growth phases in Japan, the "little tigers" and China. (3) would not apply to very rapidly-growing economies such as China, at least for variants (a) and (b), because of rapidly rising living standards and high investment levels from retained profits; (c) is very likely to have occurred, but is quantitatively insufficient on its own as an explanation of the massive capital flows that were outlined above. (4) is an accurate description of the East Asian experience, as depicted schematically in panel (b) of figure 4.

7. Interpretation

Of the various suggested explanations, the accumulation of foreign reserves and its consequences for the exchange rate is likely to have been a contributory factor in China and similar countries. There is also likely to have been a precautionary element in savings behavior. But what was the primary driving force?

There are two major types of interpretation. One is that financial weakness or under-development, either directly or via the balance between saving and investment, can explain the export of capital from countries like China. The other emphasizes the level of productivity – a focus primarily on the real economy rather than the financial sector. This is the viewpoint suggested in this paper, as well as by Caselli and Feyrer (2007), and some of the possibilities put forward by Gourinchas and Jeanne (2013). Low unit costs lead to profitability, which in turn leads to an increase in corporate saving, and rising wages may also lead to increased household saving, so that large quantities of capital are generated. This can then lead to net capital outflow and a positive current account. The evidence briefly reviewed in this paper gives support to this second view for the historical experience of large parts of Asia.

One possibility is that there are (at least) two patterns to the relationship between development and capital flows, perhaps relating to two distinct models of development. As we have already seen, Chinn and Ito (2008) found that the traditional views about the consequences of a weak financial system and of financial openness, based on the experience of the industrial countries, do not apply to emerging Asia. This also could be connected with the observation made by Alfaro et al. (2014) that there are "typical emerging market countries" such as those of Eastern Europe which imported capital during their growth phase. And on the other hand, a pattern is seen in "a few Asian countries" that is "not typical of the average emerging market". These countries are China, South Korea, Malaysia, Singapore, and (less robustly) Indonesia, Thailand and India.[4] The Lucas puzzle in its modern reversed form, of large capital flows from relatively poor to rich countries, does in fact pick out these particular economies. It also picks out oil exporting countries, whose source of wealth is different but has similar consequences; these were however excluded by outlier tests in the main sample analyzed by Alfaro et al. (2014).

This raises the question, on what grounds is a particular pattern judged to be "typical" or not? When Prasad et al. (2007) explained the reason for their rejection of the idea that fast-growing countries need less foreign capital, they started with the word "typically" – see the

---

[4] This generalization is imperfect: e.g. South Korea received substantial foreign capital at key stages of its development (Studwell, 2014). In addition, Vietnam has been a high-growth country since 1990 (*The Economist*, 6 August 2016, pp.59-62) following the 1986 *doi moi* reforms, but had a negative current account until 2010 (World Bank Data).

quotation above. Alfaro et al. (2014) also used "typical" in both the above-cited quotations. But this may be an inappropriate term, given that the area of the globe covered by these Asian countries contains approximately half of the world's population. They have also been responsible for many of the largest international capital flows, and therefore are central to the global imbalances that built up in recent decades. They are not small or marginal exceptions.

It could be that "typical" here means that the observed pattern conforms well to standard theory. If so, this would merely mean that the theory is limited to explaining the experience of e.g. Latin America and Eastern Europe, but has failed to comprehend this group of Asian countries. It would imply that another theory is required for this purpose – or better, a more inclusive theory that can encompass both models of development: their conditions of existence, and their successes and drawbacks. As the growth records of these Asian economies have been far superior to the "typical" countries, a revisionist theory along these lines could be very important.

The question of economic growth

Indeed, this issue of economic growth raises an important question. If an atypical financial sector, characterized by weakness and/or under-development, does indeed lead to capital exports and a large positive current account, what then is its relationship with growth?

There are three possible patterns of causation in this apparent three-way association between an atypical financial system, capital exports and growth. Empirical research is needed to establish which best represents the real world. One is that the association between this atypical financial model and sustained growth is not directly causal – that it is coincidental or epiphenomenal (figure 5 panel (a)).

Figure 5 Possible causal relationships between atypical financial system, capital export, and growth
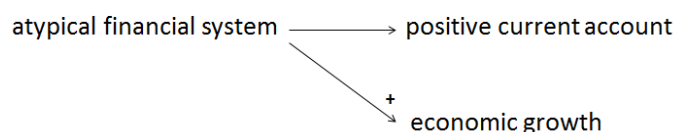
(a) no causal connection

atypical financial system ⟶ positive current account

economic growth

coincidence?
epiphenomenon?
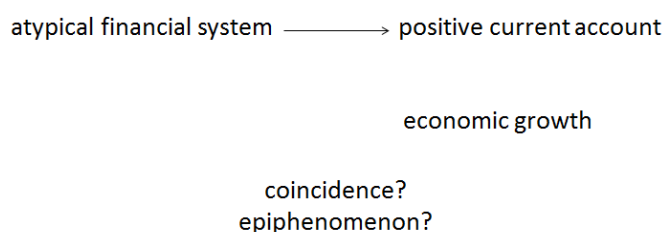
(b) causal relationship with negative sign (standard theory)

atypical financial system ⟶ positive current account

economic growth
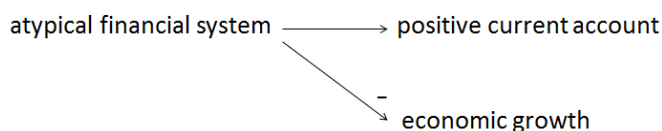
(c) causal relationship with positive sign (suggested by the evidence)

On the other hand, if there is a causal relationship, the immediate response from the viewpoint of standard theory would be that the atypical financial system must have been relatively inefficient – after all, it is characterized by weakness and/or under-development. It must therefore have inhibited growth in productivity and output (figure 5 panel (b)).

Alternatively, it could be that an atypical financial system of this kind in fact promotes productivity and output growth (figure 5 panel (c)). If one interprets this atypicality as it is presented in the literature, equating it with weakness or under-development, this would suggest the rather radical conclusion that certain Asian economies have become rich and powerful because their financial systems are weak and/or under-developed!

It is probably safe to assume that the association of an atypical financial system and a successful growth record is not merely a coincidence as in figure 5 panel (a). If the starting point is standard theory, and figure 5 panel (b) is correct, it would imply that the various economic miracles would have been even more miraculous if the financial systems had been brought up to standard earlier – in particular, Chinese growth would then have been even more stellar during the thirty years following the start of the reforms, which is hard to imagine. Conversely if figure 5 panel (c) is correct, and if atypicality implies weakness, it would be a "paradox of weakness puzzle": that something weak and/or under-developed is a source of strength.

The two views have different implications for the relationship between financial development and growth. Sandri (2010) argues in favour of the former:[5] "financial development can… improve welfare… risk-sharing instruments can considerably speed up the process of growth, limit the need for precautionary savings, and drastically change the implications for the current account by allowing for capital inflows… The welfare gains from financial development can be substantial". This position appears to be in clear contradiction with the observations: "typical" developing countries that import capital have a far inferior growth record compared with the capital exporters, as shown in the studies cited above – think Mexico vs. South Korea. It therefore proposes a course of action that is likely to have exactly the opposite effects from what is intended, in welfare terms.

7. Conclusion

The repeated empirical finding that certain emerging Asian economies behave differently from what is expected (or "typical") has two implications. First, a practical methodological one, that the mechanical application of econometric analysis to such heterogeneous samples may not always be a fruitful way to proceed – the implicit assumption of homogeneity in some of the empirical studies discussed above might not be justified.

---

[5] This passage is not present in Sandri (2014), which is otherwise similar.

Secondly, heterogeneity is a pervasive fact of economic life. Whilst many features are common to different economies, and persist over time, there is a limit to this homogeneity. The causal processes are not necessarily the same across space and across time. Economic theories need to take this into account, e.g. by an analysis of the way that institutions and culture impact the economy. Standard theory tends to treat different types of economy as if the same laws apply universally. But as we have seen, its mechanical application does not perform well in all situations.

A further general observation is on the nature of causation in economic life. Standard economic theory is generally framed in terms of the motivations and actions of agents – it is a form of decision theory. I have previously argued that the causal impact of flows should be given a more prominent place in economic theory; economic decisions are decisions about the destination and magnitude of flows (Joffe, 2017). In the present context, the causal direction in figure 4 (b) indicated the importance of identifying the *source* of the capital that has flowed copiously from China and other East Asian countries. Once one introduces flows, in the causal sense of something that has real-world effects, the Lucas puzzle and its associated problems disappear.

Related to this, as we have seen, in this literature there is a tendency to focus on the *decision* (Alfaro et al., 2014) or the *desire* (MacDonald, 2015) to export capital. The *ability* to do so, i.e. the presence of copious capital, has tended to be overlooked.

Many of the studies cited above propose a correction element, such as a weak financial system, that operates in addition to the process described by standard theory. The discrepancy between what is expected "typically" and the observed reality requires the authors to postulate a second process. The result is a two-stage interpretation comprised of (a) "what should happen" according to theory, and (b) the *deviation* from that assumed process. Neither of these actually takes place in the real world – in terms of the actual mechanisms by which things are brought about, it is a double error.

Substantively, the dynamic East Asian economies have not only exported capital. They also share the feature – despite significant differences in other policy respects – that they adopted policies of state channeling of capital for strategic purposes (Amsden, 1989; Wade, 1990; Studwell, 2014). This could well be relevant to their particular relationship between profitability, capital exports, the nature of the financial sector, and economic growth. State channeling of capital could produce a different set of causal relationships from those observed elsewhere, and assumed by neoclassical theory.[6] International capital flows and global imbalances cannot be fully understood unless this distinctiveness is given its due weight.

To conclude: the massive outflow of capital from China, and from similar countries, is not difficult to explain if we look at the evidence and draw the obvious conclusions. There is no puzzle – or at least, no puzzle about finding a good explanation for the real-life phenomena. A puzzle only arises if one starts from standard theory. This reinforces the suggestion of Gourinchas and Jeanne (2013), that the textbook neoclassical framework may not be the right model at all to think about the link between international financial integration and development. As Lucas said, "The assumptions … must be drastically wrong, but exactly

---

[6] For example, advanced financial systems increasingly show a preference for investing in existing assets rather than productive investment (Ryan-Collins et al., 2011).

what is wrong with them, and what assumptions should replace them? This is a central question for economic development." The remaining puzzle is the continuing reliance on a theory that is unable to explain the phenomenon, when its explanation is so obvious.

References

Aizenman, J., Pinto, B., and Radziwill, A. (2004). "Sources for financing domestic capital – is foreign saving a viable option for developing countries?" (NBER Working Paper 10624). http://www.nber.org/papers/w10624

Alfaro, L., Kalemli-Ozcan, S., and Volosovych, V. (2008). "Why doesn't capital flow from rich to poor countries? An empirical investigation." *Review of Economics and Statistics* 90 (2): 347-368. http://web.b.ebscohost.com.iclibezp1.cc.ic.ac.uk/ehost/pdfviewer/pdfviewer?vid=2&sid=6a425696-6360-4a34-8cf5-5d30c4da073e%40sessionmgr105&hid=118.

Alfaro, L.., Kalemli-Ozcan, S., and Volosovych, V. (2014). "Sovereigns, upstream capital flows and global imbalances." *Journal of the European Economic Association* 12 (5)**:** 1240-1284. http://web.b.ebscohost.com.iclibezp1.cc.ic.ac.uk/ehost/pdfviewer/pdfviewer?vid=6&sid=f9227ab6-5421-4eb6-b4c6-e944f0f08e77%40sessionmgr120&hid=118.

Allen, F., Qian, J., and Qian, M. (2005). "Law, finance, and economic growth in China." *Journal of Financial Economics* 77 (1): 57–116.
http://ac.els-cdn.com.iclibezp1.cc.ic.ac.uk/S0304405X0500036X/1-s2.0-S0304405X0500036X-main.pdf?_tid=184ab72e-6aec-11e6-9cbb-00000aab0f26&acdnat=1472147473_843e5bbf0bc11ffb06ff442a9e95db87.

Amsden, A.H. (1989). *Asia's next giant: South Korea and late industrialization*. New York: Oxford University Press.

Attanasio, O.P., Picci, L., and Scorcu, A.E. (2000). "Saving, growth, and investment: a macroeconomic analysis using a panel of countries." *Review of Economics and Statistics* 82 (2)**:** 182-211. http://discovery.ucl.ac.uk/15278/1/15278.pdf.

Bacchetta, P., and Benhima, K. (2015). "The demand for liquid assets, corporate saving, and international capital flows." *Journal of the European Economic Association* 13 (6): 1101-1135. http://hdl.handle.net/10.1111/jeea.12132.

Bai, C.-E., Hsieh C.-T., and Qian, Y. (2006). "The return to capital in China." *Brookings Papers on Economic Activity* 2: 61-101.
http://faculty.chicagobooth.edu/chang-tai.hsieh/research/brookings%20china%20paper%20final%20version.pdf.

Benigno, Gianluca and Luca Fornaro, 2012, "Reserve accumulation, growth and financial crises." CEP Discussion Paper No 1161 http://cep.lse.ac.uk/pubs/download/dp1161.pdf

Blomström, M., Lipsey, R. E., and Zejan, M. (1996). "Is fixed investment the key to economic growth?" *The Quarterly Journal of Economics* 111 (1): 269-276.
http://qje.oxfordjournals.org/content/111/1/269.full.pdf.

Broner, F., and Ventura, J. (2016). "Rethinking the effects of financial globalization." *The Quarterly Journal of Economics* 131 (3): 1497–1542. http://qje.oxfordjournals.org/content/131/3/1497.full.pdf+html.

Buera, F.J., and Shin, Y. (2009). "Productivity growth and capital flows: the dynamics of reforms." NBER Working Paper 15286. http://www.nber.org/papers/w15268.pdf

Caballero, R., J., Farhi, E., and Gourinchas, P.-O. (2008). "An equilibrium model of 'global imbalances' and low interest rates." *American Economic Review* 98 (1)**:** 358-393.
http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.98.1.358.

Carroll, C.D., Overland, J., and Weil, D.N. (2000). "Saving and growth with habit formation." *American Economic Review* 90 (3): 341-355. http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.90.3.341.

Caselli, F. and Feyrer, J. (2007). "The marginal product of capital." *The Quarterly Journal of Economics* 122 (2)**:** 535-568. http://qje.oxfordjournals.org/content/122/2/535.full.pdf+html.

Chinn, M.D. (2013). "Global imbalances." In: Caprio, G. (ed.) *The evidence and impact of financial globalization*, Volume 3, 67-79. Oxford: Elsevier Inc. http://www.ssc.wisc.edu/~mchinn/global_imbalances.pdf.

Chinn, M., Ito, H., 2008. "Global current account imbalances: American fiscal policy versus East Asian savings." *Review of International Economics* 16 (3), 479-498.

Coeurdacier, N., Guibaud, S., Jin, K. (2015). "Credit Constraints and Growth in a Global Economy." *American Economic Review* 105: 2838-881.

Fan, J., and Kalemli-Ozcan, S. (2016). "Emergence of Asia: reforms, corporate savings, and global imbalances." *IMF Economic Review* 64 (2): 239-267. http://link.springer.com.iclibezp1.cc.ic.ac.uk/article/10.1057/imfer.2015.33.

Göktan, M.G. **(**2015). "On the explanation of the Lucas Paradox." *Economics Letters* 137: 109-113 http://ac.els-cdn.com/S0165176515004504/1-s2.0-S0165176515004504-main.pdf?_tid=e38aadc2-6ad8-11e6-8095-00000aab0f26&acdnat=1472139224_a9a15a5959dfb4f0f11558e8efa9fd1d.

Gourinchas, P.-O. and Jeanne, O. (2013). "Capital flows to developing countries: the allocation puzzle." *Review of Economic Studies* 80 (4): 1484-1515. http://restud.oxfordjournals.org/content/early/2013/02/14/restud.rdt004.full.pdf+html.

Gourinchas, P.O., and Rey, H. (2014). "External adjustment, global imbalances, valuation effects." In Gopinath, G., Helpman, E., and Rogoff, K. (eds.) *Handbook of International Economics,* Volume 4, 585-645. http://www.sciencedirect.com/science/article/pii/B9780444543141000100.

James, H. (2014). "International capital movements and the global order." In Neal, L. and Williamson, J.G. (eds.) *The Cambridge History of Capitalism*, Vol. 2. Cambridge: Cambridge University Press, pp. 264-300, especially p. 291.

Joffe, M. (2017). "Causal theories, models and evidence in economics - some reflections from the natural sciences." *Cogent Economics & Finance* http://www.tandfonline.com/doi/pdf/10.1080/23322039.2017.1280983?needAccess=true

Kalemli-Ozcan, S., Reshef, A., Sørensen, B.E., and Yosha, O. (2010). "Why does capital flow to rich states?" *Review of Economics and Statistics* 92 (4): 769-783. http://www.parisschoolofeconomics.com/reshef-ariell/papers/CapFlowsStatesRESTATpublished.pdf.

Korinek, A., Servén L. (2016). "Undervaluation through foreign reserve accumulation: static losses, dynamic gains." *Journal of International Money and Finance* 64: 104-36.

Krishnamurthy, A. and Vissing-Jorgensen, A. (2012). *"*The aggregate demand for treasury debt." *Journal of Political Economy* 120 (2): 233-267. http://faculty.haas.berkeley.edu/vissing/demandtreas_jan6.pdf.

Lin, J.Y., Cai, F., and Li, Z. (2008). *The China miracle: development strategy and economic reform, 3rd edition*. Hong Kong: The Chinese University Press.

Lucas, R.E. Jnr. (1990). "Why doesn't capital flow from rich to poor countries?" *American Economic Review* 80 (2): 92–96. http://www.econ.nyu.edu/user/debraj/Courses/Readings/LucasParadox.pdf.

MacDonald, M, (2015). "Patterns of international capital flows and productivity growth: new evidence." *Review of International Economics* 23 (5): 846-872, Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2676967

McMillan, M., Rodrik, D., Verduzco-Gallo, Í. 2014 "Globalization, structural change, and productivity growth, with an update on Africa." *World Development* 63: 11-32.

Mendoza, E.G., Quadrini, V., and Ríos-Rull, J.V, (2009). "Financial integration, financial development, and global imbalances." *Journal of Political Economy* 117 (3): 371-416. http://www.sas.upenn.edu/~egme/pp/JPEreprint.pdf

O'Sullivan, M.A. (2007). "Funding new industries: a historical perspective on the financing role of the US stock market in the twentieth century." In Lamoreaux, N.R. and Sokoloff, K.L. (eds.) *Financing Innovation in the United States, 1870 to the Present*. Cambridge, MA: MIT Press.

Papaioannou, E. (2009). "What drives international financial flows? Politics, institutions and other determinants." *Journal of Development Economics* 88 (2): 269-281.

http://ac.els-cdn.com.iclibezp1.cc.ic.ac.uk/S0304387808000382/1-s2.0-S0304387808000382-main.pdf?_tid=113432a4-6aeb-11e6-babc-00000aacb35e&acdnat=1472147032_3e1e97dcbfa628fb4b9d90b61b3e007e.

Prasad, E., Rajan, R., and Subramaniam, A. (2007). "The paradox of capital." *Finance and Development (IMF)* 44 (1). http://www.imf.org/external/pubs/ft/fandd/2007/03/prasad.htm

Reinhardt, D., Ricci, L.A., and Tressel, T. (2013). "International capital flows and development: financial openness matters." IMF Working Paper WP/10/235. https://www.imf.org/external/pubs/ft/wp/2010/wp10235.pdf

Reinhart, C.M. and Tashiro, T. (2013). "Crowding out redefined: the role of reserve accumulation." NBER Working Paper No. 19652. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w19652

Rodrik, D. (2000). "Saving transitions." *World Bank Economic Review* 14 (3): 481-507. http://wber.oxfordjournals.org/content/14/3/481.full.pdf+html?hwshib2=authn%3A1470162982%3A20160801%253A9a825802-0de2-4543-9d90-3a9c439ccf8b%3A0%3A0%3A0%3AFr8jOjSV24HBP0igTbaPg%3D%3D.

Rodrik, D. (2008). "The real exchange rate and economic growth." *Brookings Papers on Economic Activity,* 365-439. https://www.brookings.edu/wp-content/uploads/2008/09/2008b_bpea_rodrik.pdf

Rothert, J. (2016). "On the savings wedge in international capital flows." *Economics Letters* 145: 126-29.

Ryan-Collins J, Greenham T, Werner R, Jackson A. (2011). *Where does money come from?* London: New Economics Foundation.

Sandri, D. (2010). "Growth and capital flows with risky entrepreneurship." IMF Working Paper WP/10/37. https://www.imf.org/external/pubs/ft/wp/2010/wp1037.pdf

Sandri, D. (2014). "Growth and capital flows with risky entrepreneurship." *American Economic Journal: Macroeconomics* 6 (3): 102-123. http://pubs.aeaweb.org/doi/pdfplus/10.1257/mac.6.3.102.

Song, Z., Storesletten, K., and Zilibotti, F. (2011). "Growing like China." *American Economic Review* 101 (1): 196–233. http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.101.1.196.

State Administration of Foreign Exchange, China, accessed through Trading Economics: http://www.tradingeconomics.com/china/current-account

Studwell, J. (2014). *How Asia works: success and failure in the world's most dynamic region*. London: Profile Books Ltd.

Wade, R. (1990). *Governing the market: economic theory and the role of government in East Asian industrialization*. Princeton: Princeton University Press.

World Bank Data: Current Account Balance. http://databank.worldbank.org/data/reports.aspx?source=2&series=BN.CAB.XOKA.CD&country

Author contact: m.joffe@imperial.ac.uk

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

The case for taxing interest[1]

Basil Al-Nakeeb[2]   [Economist and investment banker (retired)]

Abstract
The focus of this paper is Western economies (the West), primarily represented by the European Union and N. America, using the US – the largest Western economy – for illustrative purposes where appropriate. It notes the failure of past attempts, using a variety of measures, to prevent banking crises from reoccurring and to improve banking survivability. Regrettably, the economic literature has not classified interest bearing debt – the primary source of financial breakdowns – as a negative externality, despite potentially representing the largest and growing negative externality plaguing Western economies. Thus, the paper investigates this cancerous negative externality and its effect on cyclicality, households, corporations, economic uncertainty, national defense, and democracy. It ascribes the growth of the debt phenomenon to a biased tax system that favors debt over equity finance. It explains the rationale of imposing Pigovian (excise) taxes on demerit goods such as alcohol to curb their consumption and argues in favor of imposing a similar tax on interest bearing debt to restrict its use. It concludes with a brief discussion of the likely effects of such a tax on the economy, forms of financing, the structure of banking, and monetary policy. The paper uses qualitative analysis to arrive at its conclusions.

Keywords bank, debt, equity, externality, interest, tax

Introduction

This Time Is Different (Reinhart and Rogoff, 2009) is a fascinating book; it examines debt crises from as far back as 1258, reporting on a shocking frequency of financial fiascos worldwide: 268 crises between 1800 and 2008. Surprisingly, advanced economies – despite enjoying more advanced planning, better discipline, and greater financial sophistication – have suffered deeper crises, affecting more countries, and with greater frequency than developing countries, sustaining 19 banking crises since World War II. One plausible explanation is that the banking sector is more deeply entrenched in advanced economies and more tightly integrated into the legal, political, and economic fabric, particularly, the tax framework, facilitating more indebtedness relative to GDP and, consequently, deeper and more frequent financial breakdowns. Most illuminating is the book's ironic title, *This Time Is Different*, hinting that we presume every crisis is different instead of endless replays of variations on essentially the same theme.

Piles of books have been written on banking crises. *The Problem with Banks* (Rethel and Sinclair, 2012), coming on the heels of the global financial crisis, is a fitting example. It is an exposé of banks' chronic problems, requiring colossal support from the US government and the Federal Reserve to survive. A review of the book in Economic Record journal (Perumal 2016) accepts the book's conclusions that banks are "troublesome institutions"; however, it finds most of the material reiterates descriptions of the problems facing the banks without offering real solutions, the reforms proposed at the end of the book being brief and lacking in

---

[1] This paper entails no conflict of interest and the author has no business links with the industry concerned. It elaborates on and extends the analysis and concepts presented by the author in *Two Centuries of Parasitic Economics* (Al-Nakeeb, 2016).
[2] The author is deeply indebted to Imad A. Moosa for his invaluable comments and suggestions.

detail. Perhaps those scholars did not offer a comprehensive solution because so many of their peers have previously tried and failed.

Despite an abundance of solutions – proposed by celebrated scholars, governments, central banks, and international organizations – bank failures and crises have continued unabated. Typically, the solutions consist of tinkering, at the margin, with the banking model, its regulations, and policies, but without stepping outside the flawed banking box. The most ambitious and concerted international initiative to improve banking survivability and solvency has been the establishment in 1975 – in the wake of the 1974 crisis – of the prestigious Basel Committee on Banking Supervision (Basel), under the direction of the governors of the central banks of the ten largest economies. Its mandate states:

> The Basel Committee on Banking Supervision (BCBS) is the primary global standard setter for the prudential regulation of banks and provides a forum for cooperation on banking supervisory matters. Its mandate is to strengthen the regulation, supervision and practices of banks worldwide with the purpose of enhancing financial stability (Basel Committee on Banking Supervision).

So far, it has released three sets of banking guidelines and regulations: Basel I, Basel II, and, most recently, Basel III, in 2010-2011 – in the wake of the 2008 global financial crisis. By not succeeding in preventing bank failures and crises, those periodic Basel pronouncements have been, in reality, a series of illusions of solutions, which have kept hope alive that banking can yet be fixed. BCBS has been in a unique position to lead the indispensable structural reform that banking desperately needs, by declaring the present model defunct and proposing an alternative one. Instead, it has settled for the easier task of prolonging the demise of banking as we know it and with it the attendant crises that the world must endure, propagating a forlorn hope that next time is different – a grand disservice to the world and the banks it had set out to serve.

There is a dearth of effective ideas because banking does not take well to solutions that infringe on its sacrosanct banking box. Indeed, even lesser measures are not tolerated. Thus, in the aftermath of the Great Recession, Iceland's actions were condemned for allowing its failing banks to fail while the Greek government was showered with praise for following the instructions of its creditors. Today, Iceland's once sickly economy is healthy and vibrant while the Greek economy continues to linger in depression. Despite ample proof of what worked and what did not from the perspective of the citizens of those countries, there has been no formal international admission of reality and no retraction of any condemnation or praise.

Albert Einstein – a widely recognized genius – famously warned, "We cannot solve our problems with the same thinking we used when we created them." He also described futile repetitive behavior as: "Insanity: doing the same thing over and over again and expecting different results." It is most unwise to take Einstein's advice lightly; finding a durable solution to the banks' periodic and deepening calamities requires the intellectual courage to admit that past thinking has been amply and repeatedly demonstrated to be wrong and fruitless, making a fresh look from a different angle inescapable and absolutely necessary.

1.   The negative externality of interest bearing debt

A positive externality arises when the public (social) benefit from consuming a good (service) exceeds its private benefit, a spillover effect where additional benefits accrue to third parties who are not party to the transaction. Education is an often-cited example of a positive externality because besides benefiting the students concerned, it also benefits society as a whole.[3] By contrast, a negative externality arises when a transaction causes a negative spillover that affects third parties who are not party to the transaction, thereby raising the public cost of a good over and above its market price. Environmental pollution is a prime example of a negative externality, having a near-zero private cost to polluters and a large public cost to society.[4]

Doubtless, the full cost of debt is greatly underestimated, considering it may well be the largest negative externality facing advanced economies. For example, many borrowers unwittingly presume that nominal interest is the cost of debt,[5] until they default. Similarly, to the extent that indebtedness amplifies cyclicality it magnifies another negative externality because recessions entail enormous social costs such as loss of jobs, profits, increases in governmental outlays, and reductions in government revenue.[6]

Still, the following hardly attempts to quantify the full cost of interest bearing debt, but only the more modest endeavor of identifying some of its implicit incremental costs through its impact on cyclicality, households, corporations, etc.[7]

1.1 Cycle amplification

The best-publicized contractions during the past hundred years were all debt related, frequently accentuated by misguided monetary policy during the expansion, contraction or both, including the Great Depression, the twin Volker recessions in the early 1980s, the Internet bubble in 2000, and the global financial crisis in 2008. Thus, at a minimum, credit tends to amplify the business cycle, on the upside and downside.

Irving Fisher (1867-1947) was a neoclassical economist, but the Great Depression convinced him that debt creation induces economic expansion and fuels speculative asset bubbles, while credit contraction bursts them, followed by recession or depression. Sadly, his seminal

---

[3] Societal benefits of education include contributing to upgrading the skills of its work force, raising their standard of living, improving their employment prospects, reducing the need for income support, and increasing tax revenue. It also nurtures successful entrepreneurs that provide employment opportunities, develop new industries, and spur economic growth. Furthermore, a better educated society tends to enjoy a better quality of life and higher standards of living, makes better political and economic decisions, and suffers less crime. These positive spillover effects accruing to society makes education a merit good because, in addition to its private benefits, it has large public benefits; hence, pricing education based on its private benefits alone would result in less than optimal level of consumption from society's perspective, calling for public encouragement of education by subsidizing it or offering it free.

[4] More specifically, air, water, and soil pollution damage public health, increase medical bills, and harm buildings, forests, crops, fish, and animal herds. Hence, responsible governments enact laws, regulations, and impose penalties to protect the environment from all forms of pollution. In addition, governments impose taxes to restrict the magnitude of negative externalities.

[5] In what follows the negative externality of interest, interest bearing debt, and debt are used interchangeably to mean the same thing.

[6] For a through discussion of the negative externality of recessions, see Al-Nakeeb, 2016, "Economic Cycles as Externalities" in Chapter 11.

[7] Estimating the full cost of interest to Western societies is a huge task, requiring teams of skilled researchers and appropriate funding.

Debt-deflation theory (Fisher, 1933) did not receive the attention it deserved. Alas, Fisher stopped short of proposing a structural solution to curtail the pervasiveness of debt in the economy to dampen economic cycles. Interest in Fisher's theory has revived since the 1980s; post-Keynesian economists Hyman Minsky developed it further (Minsky, 1986), while Steve Keen modeled Minsky's financial instability mathematically (Keen, 1995).

During the expansion phase, optimism permeates the economy, inducing lenders to relax their credit standards, thereby accelerating the pace of expansion and, consequently, bringing forward its premature end. In contrast, during a recession, lenders become cautious and raise their credit standards, making a recession deeper than it would be otherwise. These pro-cyclical credit policies increase the amplitude of cycles. It is economically more efficient and less disruptive to have less violent fluctuations in asset prices (e.g., stocks, bonds, properties), gentler and longer lasting expansions, and shallower and shorter contractions.

The following reiterates and elaborates on certain aspects of how credit amplifies the expansion-contraction cycle:

- Easy credit facilitates excessive business investment during the expansion phase, resulting – during the contraction phase – in greater excess capacity, larger investment cutbacks, heavy indebtedness, and increased business failures.

- Excessive expansion of housing credit fuels higher house prices, attracting speculative demand that spurs housing construction and expands employment and consumption. Ultimately, an excessive housing inventory develops, causing house prices to drop followed by a contraction in housing credit and construction. The contraction spreads to the rest of the economy, causing the newly unemployed to default on their mortgages and the banks to repossess and sell their homes, crashing house prices and bringing new construction to a near halt, followed by another round of defaults. For a while, this self-feeding vicious cycle continues, contributing to a deeper contraction.

- Credit expansion also amplifies the cycle in stock prices. For example, a 50 percent credit margin policy permits the doubling of the size of stock portfolios, lifting stock prices higher in a bull market. Higher stock prices, in turn, increase the equity of stock accounts with brokers, providing more margin borrowing and purchasing power, which feeds the cycle of stock purchases followed by higher stock prices, further boosting margin availability. At some point, the market tops and begins to decline, inducing the cycle to work in reverse, this time as a price-credit contraction spiral. Lower stock prices diminish the market value of stock portfolios, triggering margin maintenance calls, which require investors to provide additional capital or else liquidate stocks to maintain credit margins, pushing prices lower. Thus, borrowing on margin first exaggerates the rise then fall in stock prices.

- Worse still, margin – by fueling leaps in stock prices – rouses a speculative, herd like, behavior in the investing public's psyche; as a result, the investing public's maximum portfolio positions tend to coincide with maximum credit availability near market tops, while their minimum stock positions tend to coincide with minimum credit availability near market bottoms. Record stock turnover around market peaks and bottoms supports this conclusion. The result is the worst investment strategy imaginable, running contrary to the cardinal rule of shrewd investing: buy low and sell high. It also affords astute investors opportunities to significantly increase their own return by acting as counterparties to the investing public, selling stock to the investing public at high prices then buying it back

from them at low prices. More generally, a credit cycle tends to have a similar negative influence on the investment performance of the investing public in other asset classes as well.

- Credit expansion, by inflating asset prices and investors' wealth, also induces incremental consumption due to the so-called wealth effect. When the bubble bursts, debt works in reverse, reducing wealth excessively and, with it, consumption. Thus, the credit cycle amplifies swings in consumption, adding to cyclical fluctuation.

- Monetary authorities have been known to misjudge the state of the economy. Moreover, the measures they take, with rare exceptions, are delayed reactions to economic events. Thus, during an upturn, they initially tend to raise interest rates modestly then raise them too high and – invariably – too late. During a downturn, they tend to delay cutting interest rates, then cut them too little and too late. Both effects amplify cyclical swings.[8]

- National debt can add to cyclicality. During an expansion, national debt facilitates more government spending than is available based on tax revenue alone, thereby exaggerating the expansion. During a recession, the deficit becomes too big, inducing banks to lobby governments to adopt austerity, inducing deeper recessions than otherwise. Thus, after decades of debt-financed overspending, excessive national debts have induced the governments of Greece, Italy, Portugal, Spain, Ireland, Hungary, Britain, France, and other nations to cut their budgets in the depth of contractions. As a result, they are facing widespread unemployment, unrest, deteriorating social services, and larger budget deficits, while their national debts continue to spiral out of control. In addition, heavy indebtedness accentuates the cyclicality in other countries through its effect on international trade.

1.2 Implicit costs to corporations

Aside from making timely interest and principal payments, indebted corporations must meet all the terms and conditions of their indentures at all times to avoid insolvency; these terms include the maintenance of liquidity and solvency ratios and cross-default[9] and negative pledge clauses.[10] Violating any of these terms risks triggering an event of default and, hence, these constitute incremental financial risks to corporations (especially during recessions). Moreover, violating these terms permits lenders to renegotiate loan conditions to gain greater advantage, yet another tacit cost to corporate borrowers.

Other things being equal, indebted corporations are more likely to fail than those that are debt free; furthermore, the more heavily indebted corporations are – under various measures of indebtedness such as debt to equity, debt to total assets, and so forth – the more likely they are to fail. Moreover, responsible corporations compensate for the financial risk of debt by reducing their business risk taking, resulting in slower growth for businesses and the economy.

---

[8] An economy substantially weaned off debt removes the interest handle from monetary authorities, eliminating a prime source of amplified cyclicality by leaving markets to determine rates of return instead of leaving it to a central bank, in effect, a socialist style central planning committee for capital markets in everything but name.

[9] A cross-default clause makes the default of a borrower on any loan an event of default on the present loan as well, thereby greatly magnifying the financial risk to the borrower.

[10] A negative pledge clause requires a borrower not to pledge any of its assets to a third party, thereby greatly restricting a borrower's room for maneuvering in a financial emergency.

Worse still, the effect of indebtedness is uneven across corporate sectors, with its negative impact falling most heavily on sectors with a natural tendency for wider cyclical fluctuations, particularly those with high operating leverage like capital-intensive industries. Several vital sectors fall in this category, including the automotive, airline, and steel industries. Even without borrowing, the presence of debt in the economy magnifies cyclical fluctuations for these industries and increases their business risk. Superimposed on this debt environment are monetary policies that can exaggerate cyclicality. Hence, many such companies go under during business downturns. For example, much of the US steel industry in Cleveland, Ohio, evaporated, permanently, in the wake of the 2000 recession; foreign industrialists bought the idle plants at a fraction of their replacement cost, dismantled them, and shipped them overseas, thereby providing fiercer competition to the unfortunate residual US steel producers.

1.3 Implicit costs to households

Defaulting mortgage borrowers suffer bank repossessions of their homes, typically losing their home equities accumulated over many years, an incremental cost that is over and above the nominal interest of the mortgage and a significant implicit cost of interest bearing debt. In addition, the personal traumas associated with defaults are grave. How deep is the pain of homelessness and broken families following mortgage defaults? What are the economic as well as the emotional costs of losing a family business to a bank? Can anyone imagine the desperation of farmers in India driven to suicide because they cannot withstand losing their farms and livelihoods for failing to meet a bank mortgage payment? Suicides also increase with the rise in poverty, which spreads wider as a recession deepens. What is the dollar equivalent of making tens of millions of people desperate and miserable? Those public costs are enormous and real but unaccounted for under current national statistics because present statistical and accounting methods are insufficiently sophisticated to do so, thus, seriously understating the true cost of debt. Fittingly, the original Latin meaning of "mortgage" is death.

1.4 Uncertainty and other economic effects

Keynes: The Return of the Master (Skidelsky, 2010) points to Keynes' analysis that uncertainty increases economic instability, raises the normal rate of unemployment, and slows economic growth.

More generally, we would expect high indebtedness to be associated with (1) greater uncertainty, (2) higher unemployment, (3) a fall in industrial investment, (4) higher interest rates, (5) currency overvaluation, (6) mounting budget and trade deficits, (7) industrial erosion, and (8) slower economic growth. In the case of the US economy, all of the above characteristics have progressively materialized over the past three-and-a-half decades, save high interest rates. US monetary policy has caused wild gyrations in interest rates, peaking at exceptionally high levels in the early 1980s then trending to abnormally low levels in recent years; Fed interventions have thwarted a market-determined interest rate, a serious violation of the conditions necessary for Pareto optimal resource allocation. Given the size of the US economy, its interest rate policy has been mirrored in other Western economies.

1.5 Threat to national security

Admiral Mike Mullen, a former chair of the US Joint Chiefs of Staff, is well qualified to assess threats to national security. In an interview with *Fortune* magazine, he commented on national debt by saying, "It's the single biggest threat to our national security" (*Fortune* Magazine, 2012). Niall Ferguson voiced a similar concern, predicting that by 2019 interest payments on the federal debt could potentially rise to 17 percent of the federal revenue, which would limit US military spending and US international power projection (Ferguson, 2010). Moreover, a significant downgrade of US credit would cause a sudden spike in the cost of federal debt, potentially constraining the financing of national defense.

1.6 Corruption of democracy

Indebtedness is addictive and, like all dependences, it is difficult to rein in. National debt permits nations to live beyond their means in the short to medium term at the price of loss of fiscal discipline, slower economic growth, and lower living standards in the medium term and beyond. In a properly functioning democracy, taxes are the price of public goods. Thus, voting for increased public goods implies accepting higher taxation. When national indebtedness is constitutionally permissible, democracies are prone to become increasingly indebted over time because cunning politicians can entice voters with a seemingly free lunch – increasing public goods without a corresponding increase in taxes.

Politicians who initiate big budget deficits know that biting tax increases must inevitably follow, but they hope they would have departed the political scene by then, leaving it to others to clean the fiscal mess they leave behind. They bequeath to future generations a debt burden, particularly if they used the debt to pay for current expenditures instead of financing long-lived infrastructure that benefits future generations. Taxing future generations without giving them corresponding benefits swindles future taxpayers. It represents an inter-generational inequity because it increases the benefits to and cuts the tax burden of mature citizens today, while increasing the tax without corresponding benefits on the young and yet-to-be-born. Moreover, deferring taxes increases their ultimate amount by the cumulative interest on the portion of national debt that financed the tax shortfall until its full repayment.[11] Consequently, national debt, by distorting public choice, undermines the foundations of democracy and sound public policy.

1.7 The escalating cost of debt

The problem of servicing national debts in the West is acute, but it promises to get worse still due to three factors. First, the interest rate has been at record lows, which is unsustainable over long periods; the inevitable rise in the interest rate will radically increase the debt burden. For example, Forbes magazine estimated the US federal debt to soon exceed $20 trillion (*Forbes* Magazine, 2016). Federal Reserve Bank of St. Louise data indicates that on December 29, 2016, the interest rate on a ten-year US Treasury bond was 2.49 percent compared to 6.31 percent twenty years earlier, on December 29, 1996 (Federal Reserve Bank of St. Louise). A return to the higher interest rate environment would dramatically increase the burden of the federal debt to about $1.3 trillion per year. Second, given the ongoing large budget deficits and slow economic growth, national debts in the West will

---

[11] Japan has one of the highest ratios of debt to GDP, but, so far, its consequences have been muted because of Japan's exceptionally low interest rate and because, as a nation, it is a net international creditor.

continue to rise much faster than the GDP. Third, longevity and an aging population will add to the growing financial burden, particularly if emigration rates fall.

These factors will affect most if not all Western countries. Thus, it is a matter of time before interest payments on Western national debts, without corresponding benefits, become unbearable for taxpayers and politically impossible to sustain. At that point, Western governments must choose from among three alternatives: higher taxes on the ultra-rich, inflation where the bulk of the debt is denominated in a domestic currency, or repudiation of the debt, or some combination of these.

2. Banking as a fiscal phenomenon

Unlike any other business, banking is based, since its inception, on a flawed business model that is prone to periodic bank runs and waves of mass failure. Instead of reconfiguring their business model to resolve this problem, banks have resorted to political pressure to obtain government appropriations in the form of massive treasury bailouts, ultra-cheap central bank funding for extended periods, and otherwise. This massive aid is in fact a thinly veiled public-sector subsidy to a sector that not only does not supply merit goods like the health care and education sectors, but rather supplies harmful demerit goods for which there is no economic justification, any more than subsidizing other producers of demerit goods like tobacco, alcohol, or pollution. Moreover, the size of the subsidies to sustain banking has reached colossus proportions, as the 2008 Great Recession demonstrated, and it will continue to grow as the debt grows, relentlessly.[12] At some point, probably in the not too distant future, the public will reject those exuberant subsidies for the benefit of ultra-rich bank owners; consequently, failing banks will be taken over by the state at considerable loss to the bank owners, as happened in Iceland. Therefore, it is in the interest of bank shareholders to reconfigure their business model while they still can, before events overtake them.

3. Borrowing as a tax phenomenon

The principle of tax neutrality is important for maintaining economic efficiency. In general, taxes should not distort economic behavior by influencing, for example, the choice of financing, unless there is a compelling economic reason to do so. Unfortunately, several taxes violate this principle by providing incentives for using debt instead of equity financing, such as:

  1. the corporate income tax

  2. the personal income tax

  3. tax-exempt municipal bonds

  4. the dividend withholding tax.

---

[12] Direct fiscal aid to the banks is not economically the most rational in other respects as well. For example, following the sub-prime crisis in 2008, the Federal government extended some $700 billion in emergency funds to bailout the big banks instead of the millions of delinquent homeowners; shortly thereafter the lending banks foreclosed on those properties, selling them at deep discounts, crashing the property market, and bringing construction to a dead halt. The moral economic policy was to bailout the delinquent homeowners who would have repaid the banks, thereby bailing out two parties instead of one, stabilizing the housing market, bringing about a faster recovery, saving the banks and the homeowners hefty losses, and avoiding clogging the courts with a tidal wave of court cases to boot.

Yet, economists have rarely voiced concern about the lack of tax neutrality with respect to interest. For example, the prominent economic scholar Friedrich von Hayek was an avid supporter of Pareto optimality and a keen student of economic cycles, yet he did not criticize the debt bias of the tax system or its role in economic instability. His silence on the subject is particularly puzzling given that he spent a lifetime studying economic cycles and advocating economic efficiency.

A wise saying states, "There are always good reasons for doing the wrong thing." The familiar justification for the corporate income tax interest deductibility is that it is a business expense, the cost of providing capital. However, applying this logic uniformly would require a similar deduction for dividends, the cost of providing equity capital. Such an even-handed treatment would provide neutral taxation that eliminates the tax incentive for corporate indebtedness.

The degree of interest bias of the corporate income tax is directly proportional to its tax rate; the higher the rate, the greater the tax advantage of debt over equity financing. Around the turn of the 20th century the top corporate income tax rate in many industrialized countries was no more than 15 percent. However, the enormous costs of two world wars compelled many Western countries to raise their corporate income tax rate to 50 percent or higher, making the after-tax cost of equity financing punitive, while the tax deductibility of interest provided an extraordinary tax advantage for debt financing. Many corporations adapted to the higher tax rate by increasing their use of debt to partially recover their previously higher after-tax rates of return on equity.

At a minimum, tax neutrality requires comparable treatment of interest and dividends. However, tax neutrality between equity and debt financing is not an economically sensible policy because debt is a negative externality but not equity. Indeed, improving economic efficiency requires reversing the present corporate income tax discrimination favoring debt by making dividends tax deductible but not interest. This would lower the after-tax cost of equity while raising that of debt, thereby encouraging equity and discouraging debt financing. Consequently, corporate financial risk would fall, resulting in a fall in corporate bankruptcies and job losses, shallower recessions, greater capacity for business risk-taking, and faster economic growth, potentially saving Western governments tens of billions of dollars during recessions.

The present tax bias also extends to the household sector. Thus, the personal income tax code has an interest deduction provision that lowers the after-tax cost of mortgage interest but without a corresponding tax deduction for rent. This encourages indebtedness, a major negative externality. Such tax treatment is also unfair and regressive because the incomes of those who rent are likely to be lower than those who have mortgages. Hence, removing this tax bias by giving a comparable treatment to rent or replacing the interest deduction with a housing deduction for all is not only economically sound, but also fairer and less regressive.

At this time, we can only speculate by how much indebtedness would fall if all the tax favoritism of interest is eliminated, or better still reversed, but it would likely be a significant economic adjustment in the right direction.

4. Pigovian taxes

As stated previously, externalities cause misallocation of resources, inefficiency, and market failure. They arise because market forces fail to internalize all costs and benefits of a good or service, resulting in a divergence between private and public benefits or private and public costs because prices are too high or too low, respectively. Governments can alleviate this condition by providing subsidies or imposing taxes to encourage beneficial or discourage harmful consumption, respectively.

Professor Arthur Pigou, the eminent English economist, formalized the concept of externalities and was the first to recognize that the taxation of demerit goods not only raises revenue for the treasury but also, by restricting the consumption of demerit goods, improves public welfare (Pigou, 1920). Nowadays, such taxes are labelled Pigovian. Pigovian taxes raise the private cost of demerit goods to levels closer to their public cost, thereby restricting their consumption and improving welfare. The additional tax revenue they generate also helps pay for the additional public cost associated with the consumption of demerit goods. For example, the spillover effects of excessive alcohol consumption include death, injury and damage to property due to car accidents, absenteeism from work, health costs caused by impairment of the brain, liver, heart, and other organs, shorter life expectancy, addiction, social problems, and harm to the family fabric. These considerable social costs are not included in the price of alcohol. Hence, many governments impose an excise tax on alcohol to restrict its consumption. A similar case on health grounds is made to justify imposing an excise tax on tobacco.

For a long time, coal and petrol were viewed as fuels without regard to their polluting effects. In recent decades, however, public awareness has improved, calling for restricting the pollution caused by these hydrocarbons. As a result, a range of measures have been adopted to curb their use, such as heavier taxation of petrol, particularly in Europe, and encouraging substitutes such as electricity from cleaner sources like wind power and solar energy. It is high time that a similar awareness emerges regarding the use of debt financing with comparable measures to restrict its negative effects.

Inexplicably, the literature continues to sidestep a profound discussion of interest. For example, *The Theory of Externalities, Public Goods, and Club Goods* (Crones and Sandler, 1996) presents an extensive discussion of externalities and Pigovian taxes but – as with other books on the subject – makes no mention of the negative externality of interest bearing debt or the need to tax it.

4.1 Taxing interest

As stated previously, the full cost of interest is not just its nominal amount but also the phenomenal cost of the crises it seeds, including lost jobs, profits, homes, and tax revenue to say nothing of the escalating cost of bank bailouts. Indeed, most major crises in the past two centuries can be traced to excessive interest bearing debt. Clearly, the public cost of interest far exceeds its private cost (the nominal interest), precisely the definition of a negative externality such as pollution, albeit, an imperceptible financial pollution. Pollution of the natural environment is physical while financial pollution is intangible, but just as real. Regrettably, while there are now a variety of restraints to curb physical pollution, there are still no comparable restraints on financial pollution; it continues to surge in parallel with

indebtedness. It is testimony to the power of banks that most goods and services are subject to tax but not lending, despite its huge negative externality.

Hence, a financial-pollution excise tax is essential, if long overdue. A suitable tax base would be the value of all new debt instruments such as loans, bonds, deposits, inter-bank lending, and central bank lending. It should also apply to short selling and derivatives in an appropriate fashion.[13]

What is the proper rate for taxing financial pollutants? The social cost of interest-bearing debt runs into the trillions of dollars, making any tax on debt, however huge, too small. Hence, banning interest altogether is economically logical, but it should not be implemented given the present pervasiveness of debt. Weaning the economy from debt needs to be a gradual process to avoid jolting the economy because economies need time to adjust and adapt.

Accordingly, the initial tax rate needs to be a small fraction of the social cost of debt, at say 0.5 percent per annum applied to new debt; however, the tax-rate on new debt would need to be increased annually by a similar amount until the economy is substantially weaned off debt. This would progressively discourage new borrowing without eliminating it outright. A lower rate would slow the adjustment process while a higher rate would make it faster. Moreover, in an inflationary environment, where interest rates are very high, the tax rate ought to be further linked to the prevailing nominal interest to achieve its objectives. One potential complication is that as the tax rate progressively increases over time, lenders would have a growing incentive to present their lending activities as trade transactions or equity financing to evade the tax; however, an effective tax code could easily pinpoint debt financing.

Administratively, the treasury could require the originators of interest bearing debt and financial brokers to collect the tax on its behalf in the same way that merchants collect sales taxes on behalf of the government.

5. Forms of alternative finance

The distinguishing feature of debt versus equity financing is that equity entails the assumption of business risk and therefore non-payment of dividends or principal does not constitute an event of default. There are several equity alternatives to debt. Foremost among them is common stock. In addition, certain types of preferred shares can be properly classified as equity rather than debt, provided they do not have clauses pertaining to cumulative dividends, renegotiation of terms in the event of interruption of dividends, capital prepayment triggered by financial difficulties, or priority in the net assets of a company in the event of liquidation. On the other hand, inclusion of these terms gives the preferred stock debt features with corresponding negative externalities.

The tax advantages enjoyed by debt have inhibited equity financing and the development of equity alternatives. Thus, if dividends become tax deductible under a revised corporate income tax regime, then preferred shares could become a significant source of corporate

---

[13] A short seller borrowers the security that he (she) sells short, resulting in a contingent loan to the owner of the security that is shorted, with the loan amount fluctuating in line with the market price of the security. On the other hand, derivatives, which Warren Buffet calls weapons of mass financial destruction, are best banned altogether because of the extraordinary risk they entail or, at a minimum, heavily taxed.

financing. Furthermore, a variety of tailored equity instruments for specific purposes could be developed; one such instrument for housing finance, *Equity Participation Certificates*,[14] has been proposed, with interesting economic and financial implications for banks and homebuyers, alike. The imposition of an excise tax on debt would radically accelerate this process.

5.1 Effects of alternative finance

Since the replacement of debt by various forms of equity finance should be a gradual process so as not to jolt the economy, the benefits from replacing debt with equity finance would also be realized gradually. Such benefits are expected to reverse the negative effects of indebtedness; thus, we can expect, among other effects, shallower recessions, reduced household losses caused by bank repossessions of their homes, a decline in corporate bankruptcies, a reduction in the cost of bank bailouts, a fall in economic uncertainty, lower unemployment, faster economic growth, and improved Pareto optimality and resource allocation.

The proposed changes to the tax code are also expected to affect the business of banking and monetary policy. It would likely induce banks to gradually change their business model, evolving into giant providers of equity instead of debt finance, by acting as a new form of fund managers; furthermore – in the absence of loan defaults – bank profitability would likely improve and the severity of periodic banking crises and their associated costs to the banks would likely subside.

The evolution to a substantially equity financed economy would also significantly reduce the role of monetary policy and, with it, the role of central banks in arbitrarily setting the rates of return in an economy. At the same time, the role of fiscal policy in managing the economy would grow in importance. In addition, as new debt shrinks so does the banks' role in creating new money, thereby the treasury would reap a financial windfall in the form of a monetary dividend through the provision of a growing share of the increase in money supply that is required by the economy.

6. Concluding remarks

The problem of interest bearing debt has echoes of a famous children's tale by Hans Christian Andersen, *The Emperor's New Clothes*. Briefly, the story is about two tailors who assured the king they would produce for him the best suit ever made but it would be invisible to idiots. Thus, the king's entourage and the king himself hardly dared to speak out that the suit is invisible lest it indicates they are idiots; as a result, the king ended up parading in the nude before his subjects. Similarly, the bankers have convinced the world that without interest bearing debt lubricating capitalism the capitalist economic engine would seize. Economists, financial analysts, and political leaders have observed the repeated banking meltdowns without daring to declare banking defunct. Thus, financial fiascos have persisted. Rescuing the banks and the rest of the world from this absurd situation must begin by facing the economic reality: interest bearing debt is a large negative externality. Still, one cannot underestimate the banks' instinctive resistance to this conclusion.

---

[14] For a full description of the mechanics of Equity Participation Certificates see Al-Nakeeb, 2016, Chapter 15.

In any case, establishing that interest-bearing debt is a major negative externality is of no consequence without satisfactory remedial measures. The solution requires government intervention to improve Pareto optimality by (1) ending the implicit giant subsidies offered to banks, (2) removing the tax favoritism of debt over equity finance, and (3) treating debt as a demerit good that requires a Pigovian tax to curb its pervasiveness. These measures are certain to encourage equity finance, discourage debt, and encourage the evolution to a more robust banking model that increasingly provides equity instead of debt financing. Still, a precondition for their adoption is a properly functioning democracy that can withstand the political clout of big banks.

References

Al-Nakeeb, B. (2016) *Two Centuries of Parasitic Economics*, USA: Smashwords.

Basel Committee on Banking Supervision: http://www.bis.org/bcbs/about.htm

Crones, R. and Sandler, T. (1996) *The Theory of Externalities, Public Goods, and Club Goods*, 2nd ed., UK: Cambridge: The Press Syndicate of the University of Cambridge.

Einstein, A., BrainyQuote.com, available at: https://www.brainyquote.com/quotes/quotes/a/alberteins133991.html

Federal Reserve Bank of St. Louise, Economic Research, 10-Year Treasury Constant Maturity Rate, USA: FRBS, available at: https://fred.stlouisfed.org/series/DGS10

Ferguson, N. March/April (2010) "Complexity and Collapse." *Foreign Affairs*.

Fisher, I. (1933) "The Debt-Deflation Theory of Great Depressions." *Econometrica*, 1(4), pp. 337-357.

Forbes Magazine (2016) U.S. Debt Is Heading Toward $20 Trillion: Where It's Been, Where It's Going And Why, March 28, 2016. *Forbes*, Internet edition.

Fortune Magazine (2012) interview of Admiral Mike Mullen, June 7, *Fortune* Magazine, Europe Edition, pp. 44.

Keen, S. (1995) Finance and Economic Breakdown: Modelling Minsky's Financial Instability Hypothesis, *Journal of Post Keynesian Economics*, Vol. 17, No. 4, pp. 607-635.

Minsky, H. (1986) *Stabilizing an Unstable Economy*, USA: Yale University Press.

Perumal, M. (2016) "Review-The Problem with Banks." *Economic Record*, Volume 92, Issue 299, pp. 673-674.

Pigou, A. (1920) *The Economics of Welfare,* UK: London: Macmillan.

Reinhart, C. M. and Rogoff, K. S. (2009) *This Time is Different: Eight Centuries of Financial Folly,* USA: Princeton University Press.

Rethel, L. and Sinclair, T. J. (2012) *The Problem with Banks*, UK: Zed Books.

Skidelsky, R. (2010) *Keynes: The Return of the Master*, Great Britain: Penguin, pp. xix.

Author contact: parasitic.economics@gmail.com

Why consumers are not sovereign: the socio-economic causes of market failure

John F. Tomer   [Manhattan College, Riverdale, NY]

"We should be awake to the consequence, far and near, of the way we earn our living. So many modern industries are harmful to humans and nature" (Thich Nhat Hanh, *The Heart of the Buddha's Teaching,* 1999).

"To gather one's profit out of the need of another is condemned by all laws, human and devine" (*Rerum Novarum*, Enitcyclical of Pope Leo XIII on Capital and Labor, May 1891).

"I play to people's fantasies … That's why a little hyperbole never hurts. People want to believe that something is the biggest and the greatest and the most spectacular" (*Trump: The Art of the Deal*, 1987).

"The relationship between empathy and profit is necessarily a fragile one… It should be possible to have constructive competition. The key factor is the motivation of those engaged in it. When the intention is to exploit or destroy others, then clearly the outcome will not be positive. But when competition is conducted with a spirit of generosity and good intention the outcome… will at least not be too harmful" (Dalai Lama, *Ethics for the New Millennium*, 1999).

Introduction

The purpose of this paper is to explain why competitive markets, for important noneconomic reasons, too often fail to serve the best interests of society. Before getting to these reasons, it is important to provide a brief review of the essential tenets of mainstream economics concerning the role that competitive markets play in maximizing the welfare of society.

Consumer sovereignty: mainstream and alternative views

According to mainstream economics, it is the self-interest motive (the invisible hand in the words of Adam Smith) that leads firms in competitive markets to pursue profits and avoid losses, and thereby, to maximize the net benefits to society in the long run (Frank, 1991, p. 350; Smith, 1776). This favorable economic outcome occurs, as Vilfredo Pareto recognized, when advantageous trades continue to be made until it is not possible to make another trade that will increase the well-being of any one person without reducing the well-being of any other person (Hyman, 2011, p. 58). When the latter point is reached, assuming that the values of the goods traded reflect the true values of these goods to the people involved, markets can be said to be efficient and fully functional for society. Economists, of course, recognize that there are a few important economic reasons why markets could still fail to achieve the optimum, maximizing the net benefits to society. They understand that society

will fail to achieve full economic efficiency if 1) some markets suffer from a lack of competition, i.e., have monopoly power; 2) some traded goods have positive or negative external effects (externalities) on others who are not buyers or sellers; 3) some valuable goods (public goods) cannot be traded because their benefits can be obtained even if people do not pay for them; and 4) would-be traders of some goods cannot obtain information about these goods' benefits and costs. In these cases of market failure in which the market activity by itself does not lead to an efficient outcome, most economists acknowledge that there is a strong case for government action to intervene in order to correct the market failure.

When competitive markets are efficient and maximizing the net benefits of society, consumers will in theory be sovereign in the sense that consumers' wishes will determine what is produced and offered for sale. "Most economists believe that consumer sovereignty reigns," that in the real world people as buyers really do control business (Colander, 2010, p. 65). On the other hand, quite a few people and some economists such as David Colander have doubts about consumers' control of business. For example, some people believe that their consumption choices are strongly influenced by businesses who fool them and control their choices using advertising and other means (p. 65). If that is the case, businesses can be said to control consumers rather than vice versa. To understand the argument against the existence of consumer sovereignty, it is important to make explicit a number of important assumptions. For consumer sovereignty to exist 1) consumers must really know what they want and what is best for themselves, 2) consumers must communicate this to businesses via their market behavior, and 3) consumers must not allow themselves to be influenced contrary to what is really best for them by businesses' communications and market behavior. Later in the paper it will be argued that for consumer sovereignty, it is also important that a fourth assumption must apply. Businesses must not just be motivated by profitability considerations; they must be motivated as well to supply goods that are really in the best interests of their consumers.

Colander (2010, p. 534) recognizes that "individuals sometimes do not do and pursue what is in their best interest"; in this sense they often do not behave rationally. In the absence of rational behavior on the part of consumers, market economic activity will not be efficient, and consumers will not be sovereign. And it is unlikely that consumers will get what they truly want. Although Colander, in the above quote from his introductory text, does not provide a full explanation, he clearly has some sense of the noneconomic reasons for market failure. Colander refers to this as a "rationality failure of individuals" (p. 534). A later section of this paper will develop this and related ideas further.

Akerlof and Shiller (A&S) (2015) have given considerable thought to why too often the free market system fails. In their view, it fails when the functioning of the market spawns manipulation and deception (p. vii), notably when business people behave in a purely self-serving way. According to A&S, businesses' manipulation, deception, and trickery causes the failure of markets because consumers wind up paying too much for products they do not need. A&S have a dual view of the workings of free market economies. On the one hand, businesses' profit incentives enable us to be supplied with many great products. On the other hand, free market economies lead to selfish business practices "that are analogous to biological cancers" in the human body (p. x). This happens when buyers in markets have weaknesses in knowing what they want, and such weaknesses give businesses the incentive to take advantage of buyers by learning about them, priming them, and then setting the trap for them (p. x).

Consumer preferences and business orientations

To fully explain why competitive markets may fail and consumers might not be sovereign, it is necessary to explain carefully about several types of consumer preferences and several different business orientations.

First consider consumer preferences. It is useful to distinguish between actual preferences, metapreferences, and true preferences (Tomer, 2008). Actual preferences reflect our ordinary wants and desires. They are the preferences that a consumer applies when he/she makes a choice among alternatives (p. 1706). Metapreferences are a person's preferences about one's actual preferences. Metapreferences reflect a person's capacity to critically stand in judgment of his/her actual preferences and contemplate their worth. True preferences are the preferences that are a person's ultimate ideal; they represent the unique truth about what is really the right and best for that person. True preferences are the preferences that a person would have if she were perfectly informed about not only herself but the desirability of goods that she might consume (p. 1706). Choosing in line with true preferences will lead the person to the highest possible health and vitality. To illustrate, consider two classes of foods, healthy (H) and junky (J), the latter being less healthy and potentially damaging to one's health (p. 1707). Suppose that her actual preferences are for J goods, and accordingly she will choose J food. Suppose, however, that her true preferences are for H goods as H food really is better for her health. If true preferences represent what is really right and best for the person, not only in the short-term but in the long-term, then the ultimate of rationality, true rationality, means choosing in line with true preferences (p. 1707).

Second, it is necessary to understand the different orientations of businesses. We are concerned here with businesses' orientation to act responsibly and ethically. It is useful to think of the orientations of firms as falling into three categories: 1) socially responsible, 2) market oriented, and 3) negatively opportunistic (see Tomer and Sadler, 2006; Tomer, 2013, pp. 95-97). "Socially responsible companies seek out strategies and behaviors that are simultaneously good for society, all their stakeholders, and themselves, and they avoid actions imposing costs on others" (p. 96). "Firms that are negatively opportunistic have a low ethical orientation and tend to behave not just self-interestedly but opportunistically in the sense of seeking opportunities in which they can gain at the expense of others" (p. 96). Market oriented firms are in between the other two. They respond in self-interested ways to economic and social incentives. However, these companies "are oriented to following conventional social norms; [they] are generally unwilling to cause [significant] harm to others, but are also unwilling to go out of their way to be helpful to others" (p. 96).

Phishing for phools: the essence

Why don't free competitive markets serve us well? In their book, *Phishing for Phools: The Economics of Manipulation and Deception*, Akerlof and Shiller (A&S) (2015) develop a theory that explains why competitive markets are often dysfunctional in the sense of not serving the interests of market participants. In their explanations, they use two words (phish and phool) associated with the internet. Originally, the word phish meant using misleading internet communications in order to induce individual computer users to provide valuable personal information, information that presumably could be used to make a profit at that individual's expense. A&S use the word phish in a broader sense. According to A&S, people who phish are trying to get other people, the target people, to do things that are in the interests of the

phishers or phishermen, but not in the interest of the target people (p. xi). A phool is a target person who is successfully taken advantage of, i.e., phished. "It is about angling, about dropping an artificial lure into the water and sitting and waiting as wary fish swim by, make an error, and get caught" (p. xi).

According to A&S, in the context of markets, there are two general kinds of phools: psychological and informational (Akerlof and Shiller, 2015, p. xi). And there are two types of psychological phools. In the first type, a person's "emotions override the dictates of his common sense, thereby making him vulnerable" (p. xi). In the other type, an individual's cognitive biases, which are like optical illusions, lead a person to misinterpret reality. When an individual acts on the basis of one or more of those misinterpretations, he becomes vulnerable (p. xi). Among these cognitive biases are availability, anchoring, representativeness, status quo, endowment effect, and loss aversion. Lastly, an individual becomes an information phool with its associated vulnerability when he "acts on information that is intentionally crafted [by businesses] to mislead him" (xi). It should be noted more generally that psychologists of all stripes ranging from Sigmund Freud to Daniel Kahneman (in contrast to mainstream economists) "understand that people frequently make decisions that are not in their best interest. Put bluntly, [people] do not do what is really good for them; they do not choose what they really want. Such bad decisions [and their corresponding vulnerabilities] make it possible for them to be phished for phools" (p. 1).

In the view of Akerlof and Shiller (2015, pp. xi-xii) businesses operating in competitive markets commonly phish, but it is not because they have an unethical motivation or orientation. As A&S understand it, the basic problem is that the competitive markets in which companies do much of their business put a lot of pressure on those companies (or incentivize them) to be less than scrupulous. In other words, it is the competitive market environment, not the ethical orientation of businesses, that A&S believe provides the crucial influence that leads business people to behave as phishermen with respect to their customers (p. xi). Further, according to A&S, these markets "rarely reward [either in terms of profits or social kudos]… those [business people] who restrain themselves from taking advantage of customers' psychological or informational weaknesses" (p. xii). A&S emphasize that whenever there is an opportunity for above equilibrium (or above normal) profits to be made by phishing, some would-be phishing businesses will act to take advantage of the opportunity and others will join in.

Although A&S do not mention anything about businesses' orientation, it is clear in their view that phishing businesses have a negatively opportunistic orientation. These businesses are actively seeking to discover biases and other weaknesses that make the buyer vulnerable and, thus, make it possible to take advantage of them. Their orientation is to make a profit by exploiting the consumer if that is possible. The flip side of this is that such businesses do not have a socially responsible orientation. That is, they do not have an orientation toward or convictions about doing good for their customers. These businesses are not trying to figure out what is really good for the consumer and then trying to supply that. Accordingly, they are not interested in their customers' true preferences. These businesses are only interested in their customers' actual preferences. And if their customers are found to have unhealthy or otherwise unwise actual preferences, they are not interested in helping these customers have different preferences. Further, these businesses are also not attempting to figure out how they

can avoid causing negative consequences for their customers. For example, they are not trying to find out about and lower the risks associated with the use of their products.[1]

The story of Cinnabon is recognized by A&S to be a very good example of a company that has been successful in phishing for phools. Cinnabon is a company that makes a very tasty cinnamon roll that has an attractive, strong smell, has 880 calories, and is slathered in frosting (Akerlof and Shiller, 2015, pp. 2-3). It is very tempting to buy Cinnabons, but they are not a healthy food. The success of Cinnabon is an example of a company that has phished for a long time and found many phools to buy its product. Moreover, it is an example of how "the free market system exploits our weaknesses" (p. 3). There are many other such companies phishing in the same waters. One suspects that if Cinnabon had not introduced this kind of tempting cinnamon roll, some other company surely would have (p. 3). There is good reason to believe that the choices made by Cinnabon's customers are not choices that are in their best interests, certainly not from a health perspective. Of course, buying a Cinnabon roll is not the worst mistake you can make in your life. To get a better idea of the full human cost of phishing for phools, Akerlof and Shiller provide many other examples where the stakes are much higher. These examples include rip-offs in cars, houses, credit cards, politics, food, pharmaceuticals, innovative products, tobacco, alcohol, and finance. Additional examples will be provided later in the paper.

Dual motive theory: the core underlying human motivations

While Akerlof and Shiller's phishing for phools model captures the opportunistic, exploitative aspect of all too much business behavior, their analysis fails to capture the essential core underlying human motivations that are at play. They fail to appreciate that not all business behavior has a negatively opportunistic orientation. They do not appreciate that businesses can aspire to and achieve behavior that is socially responsible. The point here is not that socially responsible behavior is attained every day by all businesses. The point is that socially responsible behavior does happen and is happening increasingly. To appreciate that businesses can have motivational orientations other than negative opportunism, it is necessary to consider dual motive theory which has its roots in brain physiology.

The two core human motivations are self-interest and other interest (or empathy). The starting point for understanding these two motivations is Paul MacLean's (1990) research on brain physiology. He conceived of the human brain as having three interconnected modular levels. The first part of the brain, the earliest in evolutionary terms, is the innermost core of the brain, the reptilian complex, which governs fundamental physiological operations and is concerned with self-preservation (Tomer, 2012, p. 78). It is associated with self-interest motivation. The second brain module, the paleomammalian brain, is located on top of the reptilian brain. It provides for the distinctively mammalian features of humans such as maternal care, parental

---

[1] Although the main thrust of Akerlof and Shiller's (2015) book is developing a theoretical perspective that explains the omnipresence of phishing in markets, the authors do mention that there are at least a few individual leaders who have resisted phishing activity (p. 136). A&S refer to these leaders as heroes. Some heroes have taken responsibility to reduce phishing by making good product information available to consumers. Others have been involved in measurement of quality standards and in enforcing these standards. Also, they have been involved with product standardization, grading, and certification, while working for governments and other organizations (pp. 137-139). A number of these heroes have been businessmen of conscience who have developed organizations concerned with product quality and ethical business practice such as the Better Business Bureau (pp. 140-141). Heroes working in governments have been concerned with developing laws, legal standards, contracts, and regulations that protect consumers from being victimized by information phishes.

responsibility, family life, and social bonding. This part of the brain is associated with caring, other interest, and empathic motivation. The third brain is the neo-mammalian brain or neocortex which envelops the other two brains. This brain provides the human capacities for problem solving, learning, memory, language, thinking, and related functions. According to MacLean, the neocortex is involved in determining how the two strong core motivations, empathy deriving from the paleomammalian brain and self-interest deriving from the reptilian brain, are interrelated and expressed. This dual human motivation view deriving from MacLean's research is in sharp contrast to mainstream economics' view of humans as motivated solely by self-interest. The dual motivation view also contrasts sharply with A&S's view that business decision makers almost inevitably behave in a negatively opportunistic fashion, that business decision makers too often behave in an aggressively negative manner with respect to their customers and others. Further, in A&S's perspective, there is relatively little hint of business being motivated by empathy or regard for others' interests.

Based on the brain physiology research of MacLean, Gerald Cory (1999) developed a model explaining how the two core human motivations tend to be balanced. In Cory's view, it is the executive functioning of the brain's neocortex that attempts to bring about a balance between the self-interest and empathy motivations which frequently are in conflict with each other (Tomer, 2012, pp. 78-9). This theoretical development has been labeled the dual motive theory (DMT).

DMT has important implications for understanding socio-economic behavior (see Cory 2006). It, of course, implies that in making decisions people have two dominant motivations, ego or self-interest and empathy or other-interest. Further, it implies that the meaning of rationality in DMT is different from its meaning in the mainstream economic model (Tomer, 2012, p. 80). Rationality in the DMT model does not involve simply maximizing the self's utility. Rationality a la DMT involves attempting to do well for oneself as well as attempting to do well by others. In the DMTs broad conception of rationality, the essence of rational behavior is attempting to live a well-balanced life in which one's own interests are integrated with others. Clearly, this behavioral perspective does not accord with the A&S view.

Tomer (2012) has proposed a revised DMT model that incorporates recent insights from brain science. In the revised model, an individual's empathic capacity is determined not just by genetics but also by brain changes that happen as a consequence of the individual's life experience. The latter phenomenon is known as brain plasticity; it is the ability of the brain to change structurally and functionally as a result of input from the environment (p. 81). Although every human has basically the same brain physiology, a particular person's brain functioning is shaped as well by that individual's unique path through life. That is, an individual's capacities such as his/her empathic capacity is shaped by every sustained activity of the person, that is, all the person's physical activities, sensory activities, cultural activities, learning, thinking, imagining, etc. (Doidge, 2007, pp. 287-91). In the revised DMT model of the human brain's functioning, people still have two dominant motivations (ego and empathy), but the strength and character of an individual's empathic motivation depends very much on the individual's life experience and whether the individual has made efforts to develop his/her empathic capacity. This revised DMT model implies that businesses and their decision makers could be motivated to develop their empathic capacities, and thereby, improve their relationships with the customers and other stakeholders of their business. In other words, businesses might for instance seek more balanced relationships with customers, not just exploitive ones.

According to Lynne et al (2016), the relationship between humans' self-interest and empathy motivations is somewhat different from the relationship articulated in the original and revised DMT models. As Lynne et al understand it, while the two motivations (ego and empathy) could be directly in conflict, generally the self-interest motive is primal, and a person's empathic capacity plays a restraining or conditioning role with respect to self-interest. This is particularly so when human self-interest is excessive. Excessive self-interest may manifest as egoistic, selfish, hedonistic, or greedy behavior. When this excess is present, humans are arguably in need of greater self-control as well as greater understanding and appreciation of the people and organizations they interact with. That is, it would be desirable if people in businesses would do more to temper, restrain, or condition their self-interest motivation. Otherwise, economic activity will not be sustainable because the economy will be dominated by the actions of business decision makers who do not consider the harm they impose on the public as a whole nor consider the interests of others they interact with (Lynne et al., 2016). It is important to realize that in a "good capitalism," humans' empathy along with other human virtues play a very positive role, ideally leading businesses to more balanced behavior that involves an integration of self-interest with the interests of others (McCloskey, 2006). Interestingly, as Lynne et al. (2016, pp. 244-246) note, taken together Adam Smith's two books, *Wealth of Nations* (1776) and *The Theory of Moral Sentiments* (1759), articulate the essence of this view.

Integrating the phishing for phools and dual motive theory perspectives

Although the phishing for phools (PfP) and the dual motive theory (DMT) perspectives are quite different, both theoretical perspectives arguably contain considerable truth. The purpose of this section is to suggest how these two theoretical perspectives can be integrated. The key to the integration is the recognition that the two perspectives apply in quite different situations. As A&S suggest, the PfP model is likely to apply where 1) businesses are strongly competing for customers to improve their profitability, 2) consumers in the market have significant weaknesses and/or lack of knowledge that make them vulnerable (in other words, these consumers are not sophisticated), 3) societal and community norms regarding proper business behavior are weak, 4) government regulations and enforcement with respect to businesses' market activity and products are weak, and 5) there is a strong tradition of self-interested, opportunistic behavior among ambitious business people. On the other hand, the DMT model is more likely to apply where 1) market competition is less cutthroat in character, 2) consumers are more sophisticated in the sense that they are not so easily deceived because they are more knowledgeable, have less weaknesses, and overall are less vulnerable, 3) societal norms strongly oppose businesses taking advantage of customers, 4) strong regulations and enforcement with respect to products and market behavior discourage businesses from exploiting customers, 5) there is a strong tradition of businesses serving society and community, 6) there exists a strong tradition of ethical business behavior that has roots in religious, spiritual, and humanistic practices, 7) business education is oriented to supporting socially responsible business behavior, and 8) business leaders have strong nonpecuniary motivations such as for prestige, good reputation, conformity, and desire to improve society.

As explained above, the PfP model has better explanatory power in certain kinds of situations, and the DMT model seems to apply better in alternative situations. In effect, these two models are complements. The PfP perspective explains better in situations where the ethical or social responsibilities of business decision makers are not likely to be a key factor.

PfP provides too little recognition that business decisions are often made in an excessively self-interested manner and that businesses are generally too quick to take advantage of customer weaknesses. On the other hand, the DMT is likely to have greater explanatory power in situations where the ethical and social responsibilities of business decision makers are an important factor. The DMT perspective is most useful for explanations that are related to humans' inherent motivations and to what could be if humans lived up to their highest social and ethical potentials. The DMT model can be faulted for insufficiently recognizing customers' weaknesses, and for not recognizing customers' inability to understand or appreciate what is really good for themselves. The upshot is that understanding the interactions of buyers and sellers in markets, particularly when sellers are likely to be taking advantage (perhaps unfairly) of buyers, requires both of these perspectives.

Let's consider an important example that illustrates both phishing for phools and the dual motive theory perspectives. This is the example of predatory mortgage lending practices. These practices reached their peak during the years 2002 to 2007, the period just preceding the advent of the financial crisis and the beginning of the great recession (2008). During this period, mortgage lending grew at an excessively rapid rate. This is indicated by the fact that in 1994, the value of subprime mortgage originations equaled $35 billion. Whereas in 2005, this value reached $625 billion, an almost eighteen fold increase in eleven years (Blinder 2013, p. 70). It is very clear that this huge increase in subprime mortgage lending was not due to a correspondingly huge increase in the number of creditworthy subprime borrowers. This lending increase was due to a large drop in lenders' standards for making these loans. Banks and nonbank lenders during this time were making far too many mortgage loans to people who had little or no ability to repay them. Lenders were unscrupulously taking advantage of unsophisticated subprime borrowers who wanted to buy houses but did not have the financial capacity to repay the loans. Lenders also were taking advantage of significant changes in the mortgage market. The fall in lending standards was indicated by the fact that many of these mortgage loans were low-doc, no-doc, liar loans, and Ninja loans, i.e., mortgage loans with little documentation, no documentation, false documentation, or loans to people with no income, no jobs, and no assets (p. 70). Many of these loans to financially unsophisticated borrowers were "designed to default" (pp. 68-71).

Why were lenders willing to make such loans? During this time, many lenders came to the realization that by making such mortgage loans, they could profit by earning the loan commission, and then quickly sell the loan, usually to a securitizer, thereby letting the securitizer, or some investor further down the line, worry about the frequently negative consequences (p. 69). Ultimately, of course, the securitizers as well as the investors (presumed to be sophisticated), who bought a great variety of mortgage backed securities from the securitizers, lost considerable money. But the biggest losers were the subprime mortgage borrowers (the home owners) many of whom reached a point where they could not pay what they owed, and therefore, were forced to default on their loans. Many of these people not only lost their houses but were forced to declare bankruptcy. These mortgage market events are consistent with the PfP perspective insofar as the borrowers' were vulnerable because of their significant weaknesses and lack of information and because government regulations and enforcement at that time were particularly weak. There is no doubt that during this period subprime mortgage lenders were phishing for subprime mortgage phools. It is noteworthy that mortgage loan company decision makers, not to mention many other decision makers in the financial sector of the economy, have lacked a strong tradition of ethical business behavior, and positive social norms have been lacking. Besides PfP, the DMT perspective also helps to explain the failure of the mortgage market.

Lenders were too often behaving in an excessively self-interested, greedy manner. They were single-mindedly pursuing profit, and their self-interest motivation apparently was not being sufficiently tempered or constrained by their empathic, other regarding motivation. Blinder's view on this is very clear. "Making loans that are 'designed to default' to financially unsophisticated borrowers who likely do not know what they are getting themselves into violates every principle of sound banking—not to mention of human decency" (Blinder, 2013, p. 71).

Once the causes of the financial crisis were coming to be widely understood and people realized how far from "normalcy" that things had gotten, it became very clear that mortgage lending practices had to change, that there had to be a very significant move toward sound banking practices. Now much of this has occurred. At least the worst practices have presumably stopped. Despite this, one suspects that some degree of phishing for phools still occurs and that, although it may be largely constrained, some degree of excessive self-interest survives and could conceivably under certain conditions emerge again in a full-blown fashion. For a more complete, lasting solution to the problem, people, at least those who are business decision makers, would have to develop their dual motivations in a much more balanced and integrated way, achieving a more desirable mix of the two core motivations, a mix with much more empathy and a mix in which self-interest is restrained by other regarding interest.

Manipulation

Manipulation occurs when a person uses influence to get other people (the choosers) to behave in a way that is to the benefit of the person who is manipulating. It is often done in an unfair or fraudulent way, but not always. When business people are said to phish, generally that means they are manipulating. Manipulation is not the same as persuasion. "With (non-manipulative) persuasion, people are given facts and reasons, presented in a sufficiently fair and neutral way" (Sunstein, 2016, p. 4). Not so with manipulation. An effort to influence a person's choices is "manipulative to the extent that it does not sufficiently engage or appeal to their capacity for reflection and deliberation" (p. 6). Manipulation is likely to have occurred "when choosers justly complain that because of the actions of a manipulator, they have not… had a fair chance to make a decision on their own" (p. 6). Some types of manipulation seek to influence by getting people to forgo deliberation altogether. Other types of manipulation attempt to influence by triggering automatic forms of mental processing (p. 7). Manipulation undermines people's ability to decide by engaging in rational deliberation. People who have been manipulated often feel that they have been tricked or fooled because they have not had the opportunity to make a decision using their own deliberative capacities (p. 11). Thus, when they are manipulated, people's choices may not promote their own welfare (p. 31).

In Hanson and Kysar's 1999 article entitled "Taking Behaviorism Seriously: Some Evidence of Market Manipulation," they state that their most important finding is that "individuals' perceptions and preferences are highly manipulable" (p. 1422). They also state their belief that to survive in a competitive market, manufacturers and marketers must manipulate their customers' perceptions. Based on their research, Hanson and Kysar conclude that businesses' manipulation is a very important reason why markets fail in the sense that consumers pay too much and do not get what is truly best for them. Their case studies of a variety of consumer markets "reveal sustained and deliberate efforts by manufacturers and retailers to manipulate consumer product perceptions" (pp. 1427-1428). This is particularly so

with respect to product risks. Based on their findings, they conclude not only that "manufacturers attempt to manipulate consumer risk perceptions… [but] that their manipulation succeeds" (p. 1428). Because of this manipulation, "many consumers are likely purchasing too many risky products" at too high a cost (p. 1428). Note also that as part of businesses' manipulative efforts, companies spend a huge amount of money studying consumers' behavior and psychology in order to understand consumers' behavior patterns and to determine the effectiveness of various kinds of marketing efforts (pp. 1429-1439). "The resulting studies and analyses provide the tools that manufacturers need 1) to shape consumer perceptions of their products, 2) to alter consumer behavior in the purchasing context, and 3) to influence consumer-safety risk assessments" (p. 1439).

In thinking about manipulation, it is useful to distinguish between soft manipulation and hard manipulation. Soft manipulation occurs when the seller's practices, while designed to influence consumer decision making, are not considered to be unfair, unethical, or illegal. In soft manipulation, the seller is not waiting for the consumer to make an error or to display a weakness so that the seller can unfairly gain an advantage. Nevertheless there is an element of manipulation, because the seller is taking an action that perhaps in a very subtle fashion is designed to induce the buyer to consider the seller's product(s) in a more favorable light than otherwise. Hard manipulation, on the other hand, involves a more definite, less subtle action by the seller to gain an advantage at the expense of the buyer. Hard manipulation would generally be considered unfair, unethical, or illegal. Hard manipulation is certainly phishing; soft manipulation is probably not. It is useful to think of a manipulation spectrum with soft manipulation on one end and hard manipulation on the other end. The idea of a manipulation spectrum indicates that there are many shades of manipulation involving different degrees of softness and hardness (Sunstein, 2016). Note that it is conceivable that even very soft manipulation could contribute to an element of market failure if the sellers' actions were somehow able to influence consumers to buy more (or less) of the seller's product than what is really best for them.

Consider the following business practices that Hanson and Kysar consider to be manipulation. In my view, the following practices should be considered types of soft manipulation. First is the used car salesman who never shows the potential customer just one car; he shows the customer many cars including irrelevant options (decoys). The idea is that the car that initially interested the buyer becomes more attractive as additional cars are shown (Hanson and Kysar, 1999, p. 1440). Also, typically the used car salesman prominently displays a price on the car; this sticker price then usually becomes an anchor in negotiating the selling price. Second is the seller who sets the price (say of gasoline) a little below some round number, for example, $1.99, $2.99, $4.95. Arguably, this causes consumers to "think they are getting a better deal than they really are" (pp. 1441-1442). Third, in a similar vein, are the typical pricing strategies of home sellers. Home sellers typically ask for an amount below some 'round' number (for instance $195,000). This takes advantage of the typical buyers' mental accounting related to their house buying budget (their budgets are usually expressed in round numbers such as $200,000) (p. 1442). Fourth is the design and atmosphere of the modern supermarket. These supermarkets are "marketing marvels [with] a shopping climate scientifically calibrated to induce as many unplanned purchases as can possibly be wrought from the 'sovereign' consumer" (p. 1444). Such supermarkets are designed to induce in the consumer a particular state of relaxation and positive mood most conducive to consumption (p. 1444). Fifth, is the supermarket's product placement. For example, "staples such as milk, bread, and eggs are placed at opposite extremes of the supermarket to force the shoppers to cover as much store real estate as possible" (p. 1447). Other store design features include

aisles designed as mazes to encourage meandering and wide aisles to encourage browsing, especially among the highest margin products. Sixth is the supermarket's pricing strategy. One example is that staples are typically priced "very competitively, counting on consumers to use the low milk or egg price as a proxy for other items in the store" (p. 1449).

There is no shortage of examples of hard manipulation. Let's consider just a few, particularly ones in which sellers take advantage of buyer biases or weaknesses. First is an example related to products marketed to thrillseekers, consumers who typically underestimate a product's risk. Fast sports cars are a product in this category. Advertisements for these sports cars typically do not adequately account for the risks of driving them very fast. One example of this is the advertisements for the Chevrolet Corvette, a well-known fast sports car. It turns out that the Corvette has "the worst death rate of any automobile in the United States," a fact not reflected in their advertisements (Hanson and Kysar, 1999, p. 1462). Second are products marketed to people who have many fears and tend to overestimate the negative consequences associated with particular products or activities. Such overestimation can become a company's selling opportunity (p. 1462). The anxiety that many people feel over possibly losing their jobs should they be forced to be absent from their jobs due to a medical problem became the angle used by the manufacturer of a cold medicine. The advertisement sought to persuade readers/listeners/viewers that the cold medicine was needed so that the cold sufferer would be able to work despite his/her cold: "No work, no pay" (p. 1463).

One interesting and noteworthy example of manipulation relates to health clubs. These clubs supply opportunities to use fitness facilities and receive health related instruction to people who want to become more fit and healthy. Their new customers are typically people who have recently resolved to improve their physical fitness and health. To use the club's facilities, the new members must choose from three different payment plans: 1) pay by visit, 2) sign a contract to pay monthly by credit card with automatic monthly renewal unless cancelled, or 3) sign an annual contract (Akerlof and Shiller, 2015, pp. 3, 167, 169; see also DellaVigna and Malmendier, 2006). These new members typically suffer from "present bias." That is, they are inclined to put off until "tomorrow" some of their fitness and health efforts. They always aspire to do more tomorrow. Most club customers choose the monthly contract with automatic renewal. This is understandable in light of their plans to do more tomorrow. The clubs understand their customers' behavior pattern and motivation and take advantage of this by making it difficult for them to cancel the contract (Akerlof and Shiller, 2015, p. 3). As a consequence, these members almost always wind up doing less physical exercise than they had hoped to and paying more on an annual basis ($600 per year more) than they would have if they paid per visit (p. 3). This is clearly a manipulation (a hard one) as the clubs are taking advantage of many new and not so new members' bias and/or lack of self-control, and profiting at their expense.

Consider another interesting way that sellers can take advantage of buyers. This is the example of goods with "shrouded" attributes (Akerlof and Shiller, 2015, pp. 167-169; Gabaix and Laibson, 2006). Any good or service in this category will include a base good and an add-on good(s). In the case of printing, the printer is the base good and ink (ink cartridges) are the add-on. The printer company typically will advertise the printer including its price and compete aggressively to sell it. On the other hand, in the case of the add-on, ink cartridges, the company typically will hide or shroud its price and will not compete to sell these cartridges, which represent a very significant part of the customer's cost of printing as well as the company's profits (p. 506). As a consequence, many printing customers (the unsophisticated ones) will not know about and do not think about the ink cartridge add-ons when buying the

base good. They only focus on the price of the base good. Because of these customers' myopia or unawareness concerning a very significant part of the cost, the sellers can manipulate or take advantage of the buyers of this product. Not surprisingly, firms typically decide to shroud attributes like add-on costs when a relatively high percentage of buyers are unsophisticated (p. 506). Another example of a good with shrouded attributes is the case in which a bank advertises the merits of one of its accounts, but the marketing materials hide or do not disclose add-on account costs such as ATM usage fees, bounced check fees, minimum balance fees, etc." (p. 506). One other example is the hotel room whose price is listed and known by customers but there are many other attributes whose prices are hidden. The latter attributes or add-ons might include parking, telecommunications, room service, etc. (p. 507) The upshot is that in a market where there are unsophisticated customers and shrouded attributes, there are sure to be 1) customers who are being taken advantage of and 2) a market that is to some degree failing.

The above examples of manipulation have only scratched the surface of the universe of possible manipulations. Note that there are plenty of ways to exploit people that have much more serious negative consequences than the ones mentioned above. But even a short listing is useful to get a better idea of the scope of the phishing/manipulation problem and its associated market failure. As Akerlof and Shiller (2015, pp. 103- 116) explain in an important chapter, the consequences of phishing are probably worst in the case of the four great addictions: tobacco, alcohol, drugs, and gambling. In these areas, the phools are not simply losing money or failing to receive benefits, they are often losing essential elements of their physical and mental health and well-being. Among the negative consequences of tobacco use are lung cancer and early death. For people who become alcohol dependent, the negative consequences include chronic, debilitating physical and mental illness, notably personality changes such as the loss of one's capacity for intimacy (pp. 109-115). Addiction to drugs and gambling carry with them negative consequences that can be similarly dire. With regard to tobacco and alcohol, the most basic fact "is that they are easily available… This easy availability of tobacco through the market, in and of itself, is the basic phish of the smokers, likewise, the easy availability of alcohol is the basic phish of those who end up drinking too much" (p. 116).

Changes in patterns of manipulation

Not surprisingly, the patterns of manipulation in the economy change over time. The "free-market system brings [into being] ever more sophisticated manipulations and deceptions" (Akerlof and Shiller, 2015, p. 136). Over time manufacturers learn more about the existing pattern of biases and shortcomings of consumers as well as learning about new patterns of biases and weaknesses of consumers. What sellers learn about buyers naturally leads them to develop new manipulations (Hanson and Kysar, 1999, p. 1467). According to Hanson and Kysar (p. 1555), it may very well be that "the most successful sellers will be those who, wittingly or not, are the most successful manipulators" or phishers.

Capuchin monkeys' dysfunctional market behavior

To get an idea of how dysfunctional markets might become when phishing is omnipresent, Akerlof and Shiller use the example of Capuchin monkeys. Research has shown that these monkeys can learn how to use money and acquire goods in markets (Akerlof and Shiller,

2015, p. 4). Capuchins can understand the prices of goods and anticipate these goods' payoffs. Imagine an experiment in which these monkeys were given incomes and encouraged to buy goods in markets without any regulatory safeguards. In these laboratory experiments with Capuchin monkeys, the experimenters can easily phish the monkeys with different food offerings, many of which the Capuchins find good tasting, but which are relatively unhealthy. Note that "Capuchins have limited ability to resist temptation" (p. 4). These monkeys will be able to get whatever they choose in the market. How would these monkeys behave? Obviously, they will buy whatever they find good tasting. The problem is that those choices are likely to be very different from what makes them happy. We know, for example, that Capuchins love Marshmallow Fluff-filled Fruit Roll-Ups (p. 4). It is unlikely that they would resist the Roll-Ups. Moreover, there is good reason to believe that the Capuchins would not only choose the Roll-Ups but make many similar choices. And as a result of their many poor choices, "they would become anxious, malnourished, exhausted, addicted, quarrelsome, and sickened" (p. 4). In other words, Capuchins' free market choices are likely to be phoolish ones; their choices are not likely to make them better off. No doubt, humans are smarter, have more self-control, and are generally more capable of making good choices than Capuchin monkeys. But as Akerlof and Shiller (2015, p. 4) explain, humans in many contexts too often make choices that are not much better than the Capuchins (for more on Capuchin monkeys' trading behavior, see Chen, Lakshminarayanan and Santos, 2006).

Behavioral market failure

As indicated earlier, mainstream economists use the term market failure for the purely economic reasons why markets fail to achieve optimum efficiency (for example, monopoly power, externalities, public goods, and lack of information). In addition to the standard market failure concept, behavioral economists refer to behavioral market failures in which the failure of the market occurs for noneconomic reasons, in particular due to the human propensity to err (Sunstein, 2014, p. 16). These behavioral failures typically stem from the all too common opportunistic orientation of businesses to take advantage of the weaknesses of their customers. "Businesses know… that consumers are easily manipulated by sellers into making bad choices—choices they would never make if they knew better" (pp. 8-9). And "free markets … reward sellers who attempt to exploit human errors" (p. 10). When businesses manipulate buyers, people's genuine, real, true needs (the needs that reflect their true preferences) are not satisfied. If that is the case, it implies a lack of consumer sovereignty because businesses are not serving the best or true interests of their customers. These behavioral market failures reflect both the weaknesses of consumers and the negatively opportunistic orientations of businesses. Such businesses are acting with an excess of self-interest and too little empathy.

Degree of socio-economic dysfunction

The socio-economic dysfunction associated with behavioral market failure has a resemblance to the phenomenon that Harvey Leibenstein's concept of X-inefficiency refers to. According to Leibenstein (1978, p. 17), "When an input is not used effectively, the difference between the actual output and the maximum output attributable to that input is a measure of the degree of X-inefficiency." In the analysis presented in this paper, when sellers act to gain at the expense of buyers using manipulation, deception, and trickery, the outcome is that customers' true needs are to some degree not being satisfied. In concept at least, one could construct a

measure of the degree to which customers true needs are not being satisfied because of the relationship of sellers to their customers. This measure could be called the degree of customer exploitation or the degree of phishing for phools or, perhaps better, the degree of socio-economic dysfunction. The degree of socio-economic dysfunction has a similarity to the degree of X-inefficiency in the sense that it helps us understand how the economy is not working as well as it could. If one could devise such a measure, it would provide important new insight into how our economy is performing or not performing.

Conclusion

Consumer sovereignty is an ideal. In the view of mainstream economists, that ideal is to a great extent realized as businesses' efforts to self-interestedly pursue profits leads them to supply what consumers want. It is important to note that the mainstream economic analysis is purely economic in nature; noneconomic considerations are left out. Further, that analysis assumes that the products that people buy reflect their actual preferences. There, is, however, good reason to believe that what people really desire are satisfaction of their true preferences, the preferences a person would have if he/she were perfectly informed and mindful. The mainstream economic analysis does not consider the possibility that businesses' pursuit of their self-interest might not lead them to produce what satisfies people's true needs. The mainstream economic analysis also does not consider the possibility that firms might seek to gain at the expense of their customers. This paper attempts to develop a more accurate conception of both the consumer sovereignty ideal and how businesses and consumers actually behave in their market interactions. Part of the story is that businesses often behave in a negative opportunistic fashion. They learn about consumers' lack of information, weaknesses, biases, and emotionality, and then they learn how to take advantage of this knowledge at the expense of the customer. The model in Akerlof and Shiller's *Phishing for Phools* helps to understand this part of the story. The dual motive theory model (self-interest and other interest (empathy)) provides another important part of the story. It indicates how businesses need to be oriented and to behave if they are to produce goods and services that truly satisfy their customers. Integrating these two perspectives enables us to understand why consumers are often not sovereign.

References

Akerlof, George A. and Shiller, Robert J. (2015) *Phishing for Phools: The Economics of Manipulation and Deception*. Princeton: Princeton University Press.

Blinder, Alan S. (2013) *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*, New York: Penguin.

Chen, M. Keith; Lakshminarayanan, Vendat; and Santos, Laurie R. (2006) "How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior," *Journal of Political Economy*, 114(3), 517-537.

Colander, David (2010) *Microeconomics*. 8th Edition. New York: McGraw-Hill Irwin.

Cory Jr, Gerald A. (1999) *The Reciprocal Modular Brain in Economics and Politics: Shaping the Rational and Moral Basis of Organization, Exchange and Choice*. New York: Kluwer Academic.

Cory Jr., Gerald A. (2006) "Physiology and Behavioral Economics," in Altman, Morris. Ed. *Handbook of Contemporary Behavioral Economics: Foundations and Development*. Armonk, New York: M. E. Sharpe.

DellaVigna, Stefano and Malmendier, Ulrike (2006) "Paying Not to Go to the Gym," *American Economic Review*, 96(3), June, 694-719.

Doidge, Norman (2007) *The Brain That Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. London: Penguin Group.

Frank, Robert H. (1991) *Microeconomics and Behavior*. New York: McGraw-Hill.

Gabaix, Xavier and Laibson, David (2006) "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121(2), May, 505-540.

Hanson, Jon D. and Kysar, Douglas A. (1999) "Taking Behavioralism Seriously: Some Evidence of Market Manipulation," *Harvard Law Review*, 112(7), May, 1420-1572.

Hyman, David N. (2011) *Public Finance: A Contemporary Application of Theory to Policy*. Mason, Ohio: South-Western.

MacLean, Paul (1990) *The Triune Brain in Evolution: Role in Paleocerebral Functions*. New York: Plenum.

McCloskey, Deirdre (2006) *The Bourgeois Virtues: Ethics for an Age of Commerce*. Chicago: The University of Chicago Press.

Leibenstein, Harvey (1978) *General X-Efficiency Theory and Economic Development*. New York: Oxford University Press.

Lynne, Gary D.; Czsp, Natalia V.; Czap, Hans J.; and Burbach, Mark E. (2016) "A Theoretical Foundation for Empathy Conservation: Toward Avoiding the Tragedy of the Commons," *Review of Behavioral Economics*, 3, 243-279.

Smith, Adam (1982 [1759]) *The Theory of Moral Sentiments*. 6th ed. Edited by D. D. Raphael and A. L. Macfie. Indianapolis, IN: Liberty Classics..

Smith, Adam (1937 [1776]) *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York: Modern Library.

Sunstein, Cass R. 2014. *Why Nudge? The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

Sunstein, Cass R. (2016) "Fifty Shades of Manipulation," *Journal of Marketing Behavior*, 213, 1-32.

Tomer, John F. and Sadler, Thomas R. (2007) "Why We Need a Commitment Approach to Environmental Policy," *Ecological Economics*, 62(3-4), 627-636.

Tomer, John F. (2008) "Beyond the Rationality of Economic Man, Toward the True Rationality of Human Man," *Journal of Socio-Economics*, 37(5), 1703-1712.

Tomer, John F. (2012) "Brain Physiology, Egoistic and Empathic Motivation, and Brain Plasticity: Toward a More Human Economics," *World Economic Review*, 1(1), 76-90.

Tomer, John F. (2013) "Stemming the Tide of Obesity: What Needs to Happen," *Journal of Socio-Economics*, 42, 88-98.

Author contact: johnftomer@gmail.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

Economics as a science: understanding its procedures and the irrelevance of prediction

Adam Fforde[1]   [Victoria University, Australia]

Abstract
The paper clarifies economics' status as a science, using as an empirical base the most-cited textbooks in microeconomics and macroeconomics (Varian, 2010 and Blanchard and Fischer, 1989). To avoid the now sterile "positivist debate", it focusses on issues of method, citing two alternative accounts of scientific method – those of Crombie and Nisbet – and exploring which fits better the evidence implied by the two textbooks. It concludes that Nisbet, reporting a very long Western tradition requiring that accounts of social change be "natural histories" (empirically-founded metaphors), fits well the views found in the textbooks. Crombie's view, arguing that science requires management of scepticism by framing procedure in terms of inductive and deductive phases, with requirement for comparison between theories through use of a predictive criterion, fits badly. This suggests that decisions about which economic accounts are deemed correct are not defined by economists' methods, but rather outside economics. It concludes by suggesting that this supports arguments for a "right to scepticism" in both the creation and consumption of policy advice, because this allows judgements to better engage with forces attempting to deem certain accounts as "correct".

Keywords policy rationality, scepticism, economists' methodology, prediction, philosophy of science

Introduction

It is self-evident that economics – what economists do – is both important in the creation of policy advice, and also that, as a procedurally-governed science, consumers of economists' accounts of the world should place trust in the validity and nature of economists' scientific procedures or methods as guiding what is deemed to be correct and so what good policy is. Yet, it is not as easy as it could be to establish precisely the methods that govern it.[2] This paper discusses these methods and argues that an examination of economists' scientific procedures suggests that, in the absence of a criterion within economics requiring exhaustive testing of accounts (such as predictive power), selection of the account deemed correct must, logically, occur *outside* economic method. It takes as exemplars of economists' normative views on procedure Varian 2010 and Blanchard and Fischer 1989, which are reportedly the most widely-cited microeconomic and macroeconomic textbooks respectively. Whilst on one metric these textbooks are the most-cited, of course there are other statements about economists' normative views, some of which are far harder and assertive in their prescriptions.

What is deemed to be correct policy, this paper argues, is better seen as not decided upon by economics, specifically through the scientific method of economics, but by something else. In

---

[1] *Victoria Institute for Strategic Economic Studies*, Victoria University: PO Box 14428, Melbourne, Victoria 8001, Australia.
[2] The author comes to the issue here as an applied economist with considerable experience in policy advice and a significant publications record. This paper is therefore in part a "reflective excursion" into matters of method and their relevance to action.

this, the paper argues, economics is far better understood as sharing, in its method, characteristics of other social sciences than how natural sciences are usually understood. In a time of Trump and Brexit, this perhaps helps explain low public trust in economists' assertions, and indeed in government based upon policy (Fforde, 2013; 2017).

Economics is a powerful presence in discussions of policy and governance, and I think it self-evident that it asserts that important parts of change processes are predictively knowable. Discussions, for example, of the pros and cons of austerity policies after the global financial crisis included forecasts of growth, tax revenues, state spending and fiscal positions. Yet, it is also self-evident that the predictive power of such accounts is extremely low, if not spurious, and examination of confirmation bias alerts us to the need for far better management of belief and scepticism alike (Fforde, 2016; 2017). This also means that students of economics and consumers of economists' ideas need, though they often do not get, some assistance in how they judge economics "as a science": what is meant by "as a science" and how can they form judgements about alternative answers? What method do economists use and what can be made of answers to this?

The paper throws light on this. For reasons of space and hoped-for utility, it focusses on presenting its own argument and therefore ignores much of the very large existing literature on the nature of scientific methodology in general and economic methodology in particular; this seems appropriate here and does not intend to suggest that this literature is unworthy, merely that the argument of the paper seems valid as it stands, and that it can be wise to be economical with words.[3] Its focus is upon method and statements of method.

The paper also offers a novel and useful interpretation of the meaning of prediction as a possible element of scientific procedure, of special significance for a highly policy-relevant "real world" science such as economics, but of more general potential value. This comes down to an explanation of why there is a tendency for forces or factors outside social science in procedural terms to be what determines "the truth of the matter". This is of great relevance to understanding how knowledge becomes policy, and here economics is a very useful example of wider and more general trends.

It first presents two statements, chosen for their relative simplicity and convenience, laying down which criteria are required to be met for theories or accounts within a science to be acceptable. They are quite different and clearly refer to distinct and alternative sets of criteria that may be used to judge a practice as scientific or not; in effect, they give two alternative "rules of the game". They may be, if one wants, labelled "natural science" and the other "social science", though this is unnecessary and perhaps confusing, and they draw upon the work of two scholars working in quite different fields who both share, however, a focus upon scientific method understood in terms of procedural *criteria*. I contrast these two statements in terms of their different lists of acceptability criteria, intending then to use these lists as tools

---

[3] A search for cited titles containing both "economics" and "method" using Harzing's *Publish or Perish* (which uses Google Scholar) "maxed-out" after returning 6527 citations to 1000 works (17th May 2015). Most of the highly-cited works are relatively old, and come from before 2000: 1st is Latsis 1976 (530 citations), 2nd comes Knight 1956 with 223 and 3rd Katouzian 1980 with 222. These arguably predate shifts in the centre of gravity of various ways of thinking about science epitomised by scholars such as Escobar 1995 and Said 1978. This "social epistemological", or relativistic, or linguistic, turn has of course deep roots, such as in Lakatos' stress on observation theory (Lakatos, 1970), not to mention Goedel's work on logical systems in the early 1930s. See also, however, Arndt 1981 for an early discussion by an economist of tensions inherent in the term "economic development" due to a frequent lack of clear distinction between transitive and intransitive uses of the verb "develop", which seems to me to be close to the nub of the matter (Fforde, 2013, Chapter 5).

with which to examine economic science. I do so, therefore, in order to prepare a ground for what is to come, the main point of the paper.

I look for empirically-important discussions of method that can be argued to be particularly relevant to students and teachers of economics. I therefore identify and examine the most popular textbooks as defined by citations data to establish significant views of the criteria said by them used by economists to define their science. Examining these with the tools established in the first part of the paper then permits assessment of what we find. This shows that economics is best viewed as following the criteria loosely defined as "social science", with some important implications – above all that it focuses upon providing insights and understandings, and in terms of method does not apply a predictive or indeed a comparative criterion. The problem the paper then turns to is to explicate what this means, and here the paper offers a novel insight. This is to suggest a re-interpretation of the nature of prediction, as a criterion with which accounts – theories – may be compared and judged, that is intended to help both economists and those who use the knowledge they create. This re-interpretation is that predictive power is usefully understood, not primarily as the ability of a theory to predict, but rather as a very particular potential member of the list of criteria applied to gauge and accept theories that would require their comparison and how it should be done. Awareness of the significance of the absence of such a criterion helps, I argue, better understand economics as a science.

There is of course a large literature on methodology. Beed 1991 attempts a summary of ongoing changes in natural science and concludes:

> "… that the question of whether or not economics is a science, or makes progress, is indeterminate because of a widespread uncertainty about what science is" (p. 488).

This denies any sense that economics as a knowledge production practice exhibits patterns and as such cannot be itself researched, to analyse and present arguments as to what methods are explicitly or implicitly followed. This is denied by the presence of fascinating studies of "what economists do", such as Yonay, 1998 and Yonay and Breslau, 2006. Such studies allow us to reflect on what their results suggest in comparison with representative studies of scientific methods. My focus here is upon method, as a core analytical focus, and I look for clear statements of method that I can use when examining the two textbooks. Here I deploy two. I avoid arguments as to just how correct or representative they are.

Scientific method # 1- Crombie and Grosseteste

If we search for an accessible investigation of scientific method, a good idea is to look for an account of its historical origins, and a convenient one can be found in Crombie 1953. Crombie looks at a scholar called Grosseteste (c. 1168–1253) who taught Roger Bacon (1214-1294) to whom many histories of science refer. I take Crombie thus as a useful entry point to discussion rather than an established and accepted statement of the truth of the matter. Crombie himself, in the introduction to the second impression, expresses self-criticism in that his particular focus (upon the 12th-century scholar Grosseteste) led to his "writings {being} credited with too much influence on science, as distinct from logical and epistemological theory associated with science" (Crombie, 1953, p. v).

We can learn much from Crombie, and he offers the advantage of both historical distance and clear definition. The emphasis upon method is what I stress here.

Crombie argues that the most important aspect of what Grosseteste formulated was *procedural* (Crombie, 1953, p.1). Based upon a belief that science was about the finding of truth, grappling with "the conception of rational explanation contained in scientific texts recently translated from Greek and Arabic" (Crombie, 1953, p.1), what was done, Crombie argues, was to add to an Aristotelian view of procedure a requirement that deductions from theory be tested empirically. Aristotelian thought, it was believed, as a part of Greek science:

> "…was dominated by the desire to discover the enduring and intelligible reality behind the constant changes perceived through the senses… and was brought into the realm of logical discourse through the idea of… demonstration or proof, the great methodological discovery of the Greeks which has occupied an essential place in all ideas of scientific explanation ever since. It meant, broadly speaking, that a particular fact was explained when it could be deduced from general principles which related it to other facts" (Crombie, 1953, p. 3).

This meant that, before Grosseteste (in Crombie's account, viewed in terms of method and focussing upon Aristotle) "scientific investigation and explanation was a twofold process, the first inductive and the second deductive" (Crombie, 1953, p. 25). Regarding the first aspect of the process, the inductive one, Aristotle, according to Crombie:

> "… gave a clear psychological account. The final stage in the process was the sudden act by which … intuitive reason[4]… after a number of experiences of facts, grasped the universal theory explaining them, or penetrated to knowledge of the substance causing and connecting them" (Crombie, 1953, p. 27).

As an explanation, this was both positive about the power of "intuitive reason" and stressed the possibility of science apprehending links between the world of thought and the essential and natural aspects of reality, which are clearly considered knowable through and in this inductive stage. Deduction was then secondary and, in the main, simply showed-off the acquired knowledge. Thus:

> "The investigator must begin with what was prior in the order of knowing, that is, with facts observed through the senses, and he must ascend by induction to generalizations of universal forms or causes which were most remote from sensory experience, yet causing that experience and therefore prior in the order of nature. The second process in science was to descend again by deduction from these universal forms to the observed facts, which were thus explained by being demonstrated from prior and more general principles which were their cause" (Crombie, 1953, p. 25).

Crombie then argues that the advances he reports, which he deems crucial, added experimentation to this duality, which implied that whilst the inductive aspect could lead the theorist to believe their theory was true, it was then necessary to relinquish this belief in some

---

[4] That is, *nous*.

way and, now sceptical, assess their theory. Deduction then served empirical testing and the relationship between the two moments – induction and deduction – changed, with the latter given greater importance.

Inductive work would be seen as involving suspension of disbelief, a phrase fitting well with the language of theatre and metaphor, where what is obviously just theatre and metaphor can, through suspension of disbelief, be treated as real. We agree to pretend. What is crucial here, and why Crombie stresses *method*, is how belief and disbelief are managed and how they are treated as part of a social epistemology – whether what is done is deemed to be an example of good application of method or not; compliance with method validates what is and was done.

However, Grosseteste was a priest and Christian, who argued that in the process of induction "the mind was assisted by Divine illumination (Crombie, 1953, p. 57)."[5] Thus:

> "The special merit of Grosseteste's theory of science was that he recognized clearly that although causal theories of this kind could not be inferred from the facts they served to explain but could only be suggested by them, nevertheless they could be tested by deducing from them consequences not included in the original generalizations and then carrying out observations of experiments to see if these consequences did in fact happen" (Crombie, 1953, p. 72).

The reasons for this shift away from Aristotle's position were, it appears, linked closely to Grosseteste's Christianity and his belief that human reasoning could not, without reengaging with Divine order, find truth. This implies that in the inductive phase the theorist was seen as relatively distant from the Divine, and this needed reversal, hopefully through the deduction of empirically-testable predictions. Mediation – the relationship between theory and empirics – is here, as is surely the case throughout most Christian thought, linked to Christ's presence in the world, as divine and human — both God and man. Theory therefore had to be tested for it to get closer to truth. Yet, believing that Divine illumination played a crucial role in theorization, in contrast to but not so different from Aristotle's psychological metaphor (the power of *nous*), Grosseteste had confidence in the ideas he generated inductively. Theorising about optics, he did not bother to test his own theories experimentally. Thus, if Crombie's account is to be believed, at the very historical origin of modern scientific method, we find the key contributor deciding that their theory "must be true":

> "Very simple experiments could have shown Grosseteste that his quantitative law of refraction was not correct. He was, in fact, a primarily a methodologist rather than an experimentalist… it was one of the basic principles of his theory of science that theories must be put to the test of experiment and that if they were contradicted by experiment then they had to be abandoned. In the next generation such natural philosophers as Roger Bacon and Petrus Peregrinus … were to use this principle as the basis of some really thorough and elegant pieces of experimental research" (Crombie, 1953, p. 124).

---

[5] Quoting Grosseteste "For in the Divine Mind all knowledge exists from eternity, and not only is there in it certain knowledge of universals but also of all singulars…. Intelligences receiving irradiation from the primary light see all knowable things" (Crombie, 1953, p. 73).

This perhaps evokes for a contemporary observer, in world far more secular, the powerful general attraction of theorisation, as a task and practice.

We can then view, using Crombie's account, prediction as a criterion that may or may not be present within a scientific procedure. It appears as a requirement that theory, having been created through a suspension of scepticism in an inductive phase, be confronted with a resumption of scepticism as deductions from theory are confronted with empirical testing. This framing means that a predictive criterion can be seen as essentially *procedural*, seeking to manage the relationship between theory and what it is meant to be about, rather than about prediction *per se*. This is in part because theorisation requires a belief that a theory being created "matters", let us say empirically, and this in turn requires some protection of the process of theorisation, which is removed when the theory is then deemed testable. Theorisation, as the quote above states, "must begin with what was prior in the order of knowing, that is, with facts observed through the senses" (Crombie, 1953, p. 25). One can reflect that what was "prior in the order of knowing" for Grosseteste, in other words possibly "what he saw around him", was thus procedurally deemed to be an inadequate empirical foundation for accepting a theory, and more was needed.

I now turn to a second and also powerful statement of scientific method, which offers a very different set of procedural criteria.

Scientific method # 2 – Nisbet and metaphor

If Crombie's account goes back to the twelfth century, Nisbet's goes back to well before the start of the first millennium (Nisbet, 1969). His focus is upon the rules governing accounts of social change in the West, and he argues that analysis of these takes a long historical perspective. The key points to take from him are three.

First, much can be learnt from a historical discussion of accounts of social change. As Nisbet puts it in his Preface:

> "Whatever novelty or originality may lie in the book comes from my having brought into single perspective ideas and themes which are ordinarily considered in isolation from one another. … Nowhere to my knowledge are all of them united within a single frame of reference that is formed by their common assumptions in the history of Western social thought. This I have tried to do" (Nisbet, 1969, pp. vii, viii).

What Nisbet sees as underpinned by "common assumptions" is the "Western idea of social development" (ibid., vii). Like Crombie, he is examining the shared criteria applied to judge knowledge production. He argues that much can be learnt from digging deep into history to elucidate and map these assumptions, and he concludes that there is a shared pattern. His book goes back to the classical Greeks and forward to the contemporary (the 1960s).

His second point is that beliefs about social development have, over time, usually contained two distinct sets of ideas that are in mutual tension.

Third, that these two sets of ideas are, on the one hand, that social change is particular, contextual and real, and, on the other, that social change is best treated through *metaphor*.

His discussion of the second is a discussion of the rules applied to determine whether accounts are acceptable, that is, scientific, and is therefore a discussion of scientific procedure, equivalent to Crombie's but quite different.

> "It is, however, the principal argument of this book that the metaphor … {is} much more than adornments of thought and language. {It is} quite inseparable from some of the profoundest currents in Western thought on society and change. They were inseparable in ancient Greek thought and in the thought of the centuries which followed the Greeks; and they remain closely involved in premises and preconceptions regarding the nature of change which we find in contemporary social theory" (Nisbit, 1969, pp. 8, 9).

Nisbet stressed how standard accounts of social change in what he calls The West occurred in two different forms: first, detailed "histories" that offered contextual and contingent accounts of what happened; second, "abstract realities" that provided an understanding of essential common patterns in social change, which were, in the main, self-consciously quite different from the first form – natural histories – histories of the *nature* of change. These natural histories presented accounts of what were believed to be true and essential patterns of change. In the long period Nisbet considers (two and half millennia) most scholars understood that such accounts were essentially different from detailed contextualised historical accounts, with a sense quite different from that given to natural history nowadays. Nisbet argues, I think convincingly, that natural histories in Nisbet's sense have retained certain characteristics over this long period and are powerful, because their characteristics meet the criteria of foundational beliefs about what makes an account valid.

Nisbet calls these accounts of abstract reality – theories - *natural* histories. They are histories about the nature of things, for focussing on their nature is the main task for metaphorical accounts. He concludes that, in the broad cultural field he is studying (for him, The West), such accounts share specific attributes:

> "For twenty-five hundred years a single metaphoric conception of change has dominated Western thought. Drawn from the analogy between society and the organism, more specifically between social change and the life-cycle of the organism, this metaphor very early introduced into Western European philosophy assumptions and preconceptions regarding change in society that have at no time been without profound influence on Western man's contemplation of past, present and future" (Nisbit, 1969, p. 211).

Nisbet lists the requisite characteristics of such metaphors (the acceptability criteria used to assess the validity of theories: their method) as follows:

> "From the metaphor came the notion of change as natural to each and every living entity, social as well as biological, as something as much a part of its nature as structure and process. Second, social change – that is, natural change, was regarded as immanent, as proceeding from forces or provisions within the entity. Third, change, under this view is continuous, which is to say that change may be conceived as manifesting itself in sequential stages which have genetic relation to one another; they are cumulative. Fourth, change is directional; it can be seen as a single process moving cumulatively from a given point in time to another point. Fifth, change is necessary; it is

necessary because it is natural, because it is as much an attribute of a living thing as is form or substance. Sixth, change in society corresponds to differentiation; its characteristic pattern is from the homogenous to the heterogeneous. Seventh, the change that is natural to an entity is the result of uniform processes; processes which inhere in the very structure of the institution of culture, and which may be assumed to have been the same yesterday as they are today" (Nisbit, 1969, p. 212).

Such a list is deeply instructive. Consider the following, from a much-cited book in the field of international political economy [Held et al 1999] where the question is asked – "What is globalisation and how should it be conceptualised?",[6] and they offer a list of criteria as follows:

"…any satisfactory account of globalization has to offer: a coherent conceptualisation; a justified account of causal logic; some clear propositions about historical periodization; a robust specification of impacts; and some sound reflections about the trajectory of the process itself" (Nisbit, 1969, p. 14).

Like Nisbet's list, but unlike Crombie's, this says nothing about how accounts or theories should be compared. What they focus on in the main is the (logical) form of the account, almost taking for granted that there is some empirical support for it. This is however very muted in both lists. Let us now consider significant statements about economic method.

Statements on economic method

Statements

As a science, a producer of knowledge, to be coherent economics must be governed by, and so explicitly or implicitly contain, rules that give scientists assessable criteria for judging candidates for knowledge, including the procedures that should be followed. The quote from Held et al above is an example. It is hard to imagine an economic account that was deemed illogical that would be accepted by economists as valid. There is thus an empirical question, which is what these rules are.

Study of such rules, how they change and how they are viewed, is familiar to many economists from the works of scholars such as Kuhn, 1962; Popper, 1959 and Lakatos, 1970. They may be less familiar with other scholars, such as Said 1978, Escobar 1995 and Foucault. One difference between these two groups is that the former tend to maintain a focus upon understanding scientific practices as in some sense progressive, in that they may be read as implying that science creates, on the whole, better knowledge over time, whilst the latter are more focussed upon issues such as the power implications of knowledges. What they share is an epistemological interest – in studying aspects of knowledge rather than knowledge itself: they are reflective. However, if we look at canonical texts in economics, we tend to find that matters of method are treated *ex cathedra*: that is, they are treated as given – perhaps to be stated, perhaps not, but not something meriting much reflection.

---

[6] Harzing's *Publish or Perish,* based upon Google Scholar, gives 7909 citations as of March 30th 2015, far more than either Varian or Blanchard and Fischer.

Here I treat textbooks as canonical: the place to look for normative statements of scientific method. Using as a citations metric the data from Google Scholar, the most highly cited microeconomics and macroeconomics textbooks are Varian 2010[7] and Blanchard and Fischer 1989.[8] I consider them in turn.

Varian

Varian advances various scientific criteria for the validity of what he teaches. He states that his:

> "… aim … was to present a treatment of the methods of microeconomics" (Varian, 2010, xix).

That:

> An analytical approach to economics is one that uses rigorous, logical reasoning (Varian, 2010, xix).

And that:

> "The conventional first chapter of a microeconomics book is a discussion of the 'scope and methods' of economics. Although this material can be very interesting, it hardly seems appropriate to **begin** your study of economics with such material. It is hard to appreciate such a discussion until you have seen some examples of economic analysis in action … Economics proceeds by developing **models** of social phenomena. By a model we mean a simplified representation of reality" (Varian, 2010, p. 1).

This is the only place in his text where the phrase "scope and methods" can be found. He does not return at the end of the book to discuss it - the final chapter is, like the others, about theory.

The book exposits the well-known body of microeconomic theory, and deploys powerful and elegant metaphorical argument. Thus:

> The great virtue of a competitive market is that each individual and each firm only has to worry about its own maximization problem. The only facts that need to be communicated among the firms and the consumers are the prices of the goods (Varian, 2010, p. 627).

Searching on "facts" shows that this means for him the facts of theory (e.g. pp xix, 90, 162, 279 (where the phrase "mathematical facts" is used), 370, 398 and 479. Footnote 5 on p. 532, however, cites a Wall Street Journal article to support the assertion that "threat of retaliation then serves to keep all prices high". There is but one mention of empirics (search under "empiric"), in a discussion of what the standard models say about the effects upon work of changes in wages:

---

[7] Using Harzing's *Publish or Perish* (8th April 2015) this work, dated 2010 – the 8th edition - but including citations of earlier editions when Varian had co-authors - had 3357 citations.
[8] Using the same source as before, Blanchard and Fischer had 4929 citations; running a close second and third were Romer 2011 with 4912 and Obstfeld and Rogoff 1996 with 4619.

> "As the wage rate increases, people may work more or less. … Why does this ambiguity arise? When the wage rate increases, the substitution effect says work more in order to substitute consumption for leisure. But when the wage rate increases, the value of the endowment goes up as well. This is just like extra income, which may very well be consumed in taking extra leisure. Which is the larger effect is an empirical matter and cannot be decided by theory alone" (Varian, 2010, p. 176).

His last sentence is a clear metaphor for the relationship between theory and reality. Theory captures the essence of reality, and beyond that empirical investigation is needed.

Varian is clearly following a scientific procedure. As he says, the core of this procedure is the construction of models that are "a simplified representation of reality". Therefore, as for the accounts of natural histories Nisbet reports and analyses, empirical aspects of method are far less important than exposition of theory – he therefore, consistently, does not need to elucidate, for example, how economists should judge whether a representation of reality is a good one, other than that it be "logical".[9]

Blanchard and Fischer

Turning to Blanchard and Fischer, we find again belief in the presence of shared and coherent procedure - that economics is a science. Thus they argue that the existence of "multiple truths" in macroeconomics does not mean that it is not a science:

> On the surface, macroeconomics appears to be a field divided among schools, Keynesians, monetarists, new classical, new Keynesian, and no doubt others. Their disagreements … leave outsiders bewildered and skeptical ... This is not our assessment … We believe that macroeconomics exists as a science, an admittedly young, hesitant, and difficult one. Its inherent difficulties stem from the need to draw from all branches of microeconomics, deal with aggregation, make contact with data, and eventually make policy recommendations (Blanchard and Fischer, 1989, xi).

---

[9] He has little to say, an issue shared by Crombie and Nisbet (and Held et al), about what exactly it means to be logical. Compare Kline 1980, arguing, for example, that as a believing Christian it was quite natural for Newton to believe what we could now call his intuition, but is better called his belief in a revealed or revealable natural order, leaving proofs of important steps in his formal argument until later as he pressed on to his conclusions. Also Priest, who, in a provocative and heterodox book argues for the possibility of true contradictions (Priest, 2002). For him, true contradictions are illustrated by the proposition that, standing in a doorway, somebody can be both in, and not in, the room. Varian would presumably disagree with Priest, asserting that Priest was illogical. This suggests that such statements and their acceptability would depend on what one means (in part, who one is) and how in that context meaning is interpreted, so that a procedural requirement that "one be logical" should also state what that means – what logic should be followed and how disputes about being illogical be resolved. Winch 1958 restates Lewis Carroll's paper *What the tortoise said to Achilles* to conclude that "The moral of this, if I may be boring enough to point it, is that the actual process of drawing an inference, which is after all at the heart of logic, is something which cannot be represented as a logical formula … Learning to infer is not just a matter of being taught about explicit logical relations between propositions; it is learning to do something" (p. 57). I conclude from all this that it would be better to describe such accounts as metaphors rather than theories.

And:

> "We have written this book to … present the common heritage, the conceptual framework and the set of models that are used and agreed upon by the large majority of macroeconomists … {and to} present life at the frontier, showing the various directions in which researchers are currently working" (Blanchard and Fischer, 1989, xi).

Their stance regarding empirics is somewhat different from Varian. They start "with the basic facts that need to be explained, the existence and persistence of economic fluctuations and their characteristics" (Blanchard and Fischer, 1989, xi) and then exposit standard models used to explain them. But they point out up front that whilst these building blocks "shed light on the fundamental issues" (ibid*.,* xi), they are essentially "equilibrium economics" (ibid*.,* xii) and at this point there is disagreement amongst economists, who divide into the various schools they mentioned at the start. At this point, though, as strikingly as Varian's remark that "Which is the larger effect is an empirical matter and cannot be decided by theory alone", they argue that "Working economists, like doctors treating cancer, cannot wait for all the answers to analyze events and help policy. They have to take guesses and rely on a battery of models that … have repeatedly proved useful" (ibid*.,* xii).

The point here is that the gauge of a model, of an explanation, is for them linked strongly to the ability to use it to give policy advice. Yet, there is no discussion of the extent to which macroeconomics contains, in its method, either of two things: first, ways of comparing, procedurally, different models; second, whether it would or could be better, given the way in which models are empirically founded, to "do nothing". Here, then, we find empirically-founded metaphorical accounts, expressed in terms of sophisticated models with varying degrees of econometric support, asserted to possess predictive power.

Discussion

The power of economic ideas, especially in policy debates, clearly draws upon many things, and just how exactly they gain authority is far from certain, though what does seem clear from my exposition so far is that, like social science more generally, what is deemed to underpin a knowledge-based policy cannot, if we follow Nisbet, be understood solely in terms of scientific procedure.

Part of the story, however, surely is that audiences expect economists to be, in some sense, scientists, seeing economics as rule-governed. The question therefore examined here is, "what rules"? Audiences view economists, amongst other things, as producers of knowledge, and they expect economists to follow rules in doing so. Thus, as we have seen, these textbooks provide for students and others rules defining what is acceptable as microeconomic or macroeconomic theory, and these allow economists to refer to shared criteria that make theories acceptable, for without this discussion and debate would not only be chaotic, but lack the potential to gain audiences and so support for policy proposals. But of course many other factors come into play.

The three examples I have given (Held et al, Varian and Blanchard and Fischer, 1989) clearly all follow and share the same basic rule, which is that they offer accounts of what *should* be done to produce valid accounts, that is, scientific knowledge. These lay down the procedural rules that should be followed, and this is what we should expect. These are all, in equivalent

ways, statements about what a theory should be (for Varian, as for Blanchard and Fischer, acceptable models; for Held et al a satisfactory account). By implication, theories that do not meet their criteria are unacceptable. They are writing for audiences, so if you read or are taught any of these three books (Held et al, Varian and Blanchard and Fischer, 1989), then you learn that a theory or account that does not follow, in its production, the particular given criteria, is wrong, and should be rejected. If somebody uses different criteria they are wrong because they are not following the procedure that defines what science is.[10]

So, I ask what these criteria may be. It is clear from the way these texts are written (see the quotes above) that these are to a considerable extent taken by the authors as obvious, clear enough and not worth (at least in the texts) much deliberation. None of them, for example, give any citations as to the origins of their methodological statements, nor discuss alternatives, nor use citations to support their positions. This is striking, for, as I have discussed, there are choices being made because there are identifiable alternatives.

It is clear that the canonical economics texts discussed do not suggest following anything like the procedure reported by Crombie. No distinction, for example, is made between the empirics of, on the one hand, theorisation, and on the other use of deductions from theory to create assessable predictions. Theory is essentially metaphorical, showing the essence of what is happening, with deviations from it to do with the particular circumstance.

Microeconomics, *pace* Varian, is a statement of theory. Readers are therefore offered almost no discussion of empirics or facts, and data is referred to in order to provide passing support, in a manner reminiscent of what Crombie has to say about induction, to theory. There is no sense of a managed movement between a suspension of scepticism and then its resumption. There is no distinction between "things reliably known and things less reliably known".

Searching through the text for references to "data" is illuminating. On pp. 83-84 data is presented to show how a utility function can be derived from data describing consumer behaviour. This is no more than a demonstration that a particular functional form, selected ad hoc, "fits" the data presented. The particular functional form used for this exercise is not theoretically justified (as, for example, an inverse square law is justified in Newtonian theories of gravitation). Of itself this suggests strongly that we dealing with a science of metaphor – that is with a production of "natural histories" that grasp and explain what is said to be essential. The etymology of the word metaphor appear to be "to carry beyond, or over", which points to the status of theories and accounts as being related but somehow "beyond" something else, what is often called "reality". The discussion here, drawing upon Nisbet, perhaps suggests that social science knowledge production is usefully seen as *essentially* theorisation, a rule-based production of theories that are usefully seen as metaphors, and only at great risk seen as truthful expressions of reality (and so reliable guides to prediction). Inductive methods that produce such metaphors, or theorisations, are not divorced from reality, they are empirically-based metaphors, *but that is all*. Nobody would trust (or be able to insure) an aeroplane whose design was based upon theory alone; what gets insurance is a judgement that risks are acceptable.

---

[10] Obviously, and this is abundantly clear from practice, there is and can be extensive debate about the particular meanings of terms such as "accepted", "procedure" and so on; but the point stands, as it is about social norms, not truth (or rather it is about the implications of the idea that the truth of a matter can be decided – sometimes a big ask). See the quote from Winch above.

As Varian states, however:

> "We can estimate a utility function that describes their consumption patterns and then use this estimated utility function to forecast demand and evaluate policy proposal" (Varian, 2010, p. 85).

Data is also presented on p.126, again to show how demand varies, theoretically, with price. This data is constructed for this purpose only, to illustrate theory, for example, leading to the conclusion that it:

> "… could not be generated by a consumer with stable preferences who was always choosing the best things he or she could afford" (Varian, 2010, p. 128).

Data here is used to see whether theory works, in terms of matching the data.

> "Think, for example, of a household consisting of several people. Will its consumption choices maximize "household utility"? If we have some data on household consumption choices, we can use the Strong Axiom of Revealed Preference to see" (Varian, 2010, p. 130).

This is the empirics of inductive reasoning, in Crombie's sense. It seeks to manage empirical aspects of theorisation, not by deduction and prediction, but by continuing to believe in the theory. At root, *it seeks to defend the theory*. Data is used to support the theory; scepticism is suspended, disbelief is too. The scientific method applied is thus very different from that described by Crombie.

Similar considerations apply to Blanchard and Fischer. It is clear that, for them, macroeconomics is mainly to be defined as what macroeconomists do, and this is, essentially, to use a shared "conceptual framework and … set of models that are used and agreed upon by the large majority of macroeconomists" (Blanchard and Fischer, 1989, xi). The main thrust is to do with "with the basic facts that need to be explained" (ibid., xi]. The word prediction is not to be found in their Index, nor is there an entry for forecasts or forecasting. Whilst some may argue that it is self-evident that macroeconomic modelling is not predictively powerful, more importantly, prediction is not important to its method.

What is important for Blanchard and Fischer is very similar to what Nisbet is reporting, and is the idea that an economic theory should offer an insight into the economic logic of what is observed. Like Varian, what we find here is a science of metaphor.

Consider the basic stance of microeconomics as Varian exposits it, and the role within it, well-known to any trained economist, of competition as modelled through comparative statics. Competition is seen as natural and the primary force of change, coming from forces within the economy, as theorised. In terms of his third criterion, change is cumulative, as competition pushes the economy to changes in levels of output and consumption. Change is directional, as competition pushes towards optimal outcomes, *unless inhibited by market failure*. What comes through particularly clearly is conveyed well by Nisbet's very particular use of the term "natural history": economists' theories offer us accounts of an essential nature of social change, for example in microeconomics put in terms of deviations from competitive outcomes. After all, rents are "a gain or advantage that cannot be competed away" (Levy, 1995, p. 96).

Finally, reconsider the clear statement from Held et al about the criteria any account should meet to be deemed acceptable. Again, this omits the criterion of prediction, and indeed any criterion requiring comparison of competing theories or accounts; it states the rules that allow scientists to validate theorisation. In all three examples, therefore, we find a science of metaphor, close to what Nisbet reports and very distant from Crombie.

The remainder of the paper first draws together the discussion of the nature of economic science, and then offers a novel account, that greatly clarifies the situation, of how we should best view prediction as a criterion.

Economic science

The discussion above relied for its empirical basis for discussing economic science upon textbooks. Although there is not much research that examines what economists do when they choose what to model, we can examine Breslau and Yonay, 2006.[11] They conclude:

> "The truth of economic statements is … the product of economists' success in enlisting the support of other economists, data, whole economies, mathematics, and other agents, rather than adherence to an established and rule-based method" (Breslau and Yonay, 2006, p. 5).

Breslau and Yonay point out that whilst a model with the approved building blocks (statements about agents' preferences, etc.) and an analytic solution may be challenged on the grounds of empirical plausibility, this is not a *predictive* criterion:

> Referees and editors often cite implausibility as a reason for rejecting articles. They use their sense and knowledge of the economy to assess whether a model offers an important explanation of an economic phenomenon. Thus, an article can handle an important subject, be rigorously constructed, and still be rejected if the referees and the editor believe that it fails to address a main mechanism behind the phenomenon in question (Breslau and Yonay, 2006, p. 28).

This is clearly interpretable through Nisbet's lens, as a deliberation on whether the theory captures empirically essential ("natural" in Nisbet's word) processes that exist in reality.

I conclude that economic science, not following a methodology that includes something equivalent to a predictive criterion, is best seen as empirically-based theorisation that focusses upon the generation of models deemed to improve understanding. The absence of a comparative criterion from procedure is striking, and, as Nisbet suggests, this corresponds to scientific regulation that permits – has no formal criterion to prevent – the co-existence of "multiple truths", any criterion for choice between which, if it happens, exist outside the rules scientists are following. Choice between theories – for example as part of debates about

---

[11] They state – "[W]e want to ask how neoclassical economists themselves make the connection between their models and economic realities. . . . [O]ur goal is to elucidate the 'epistemic culture' of economists that guides their own routine work of model-building and their evaluation of their colleagues' models. Such empirical studies of economics are strikingly missing, despite economics' allegedly huge influence on economic policymaking, and consequently on the lives of us all" (Breslau and Yonay, 2006, p. 6).

public policy - must be made *outside* the rules followed by economists' that govern their endeavours as scientists.[12]

Viewed as an element of scientific method, are tests for predictive power best seen as tests for the ability of a theory to predict?

Whilst it may superficially appear clear, an alleged ability of a theory to predict is easily shown to depend upon a host of tangled factors, so things are not clear at all.

At an extreme, to start with, a theory that is right 51% of the time could feasibly be described as predictive, but is not likely to be. Yet if the point is to win bets placed very many times, then it could be thought of as predictive. Theories from physics, such a Newton's laws of motion, are widely felt to be predictive, but this is within certain bounds, about which quite a lot is known. On the one hand, for example, as velocities approach the speed of light, so mass, assumed constant, is thought to vary. Again, just as Newtonian space is conceptually made up of lines, with no presence outside one dimension, and points, with no presence at all, so mass is assumed to be something that can be situated at a single point, a centre of gravity. All of this can be understood to mean that the apparent clarity of Newtonian physics is not what makes it acceptable under some circumstance as a guide to action. The extent to which it matters that observables necessarily seen to flout the scientific metaphor involved – lines as measured have width, points in time have duration, forces cannot be directly observed - and are therefore associated with an ability to insure the resulting object (say, an aeroplane) depends on the local and social context. To develop this argument, if gun-laying was being done for "extremely inaccurate" riflemen in a war of accepted extreme levels of attrition (consider if the guns were aimed by cloned animals), then prediction that entailed a 51% accuracy rate could be, one can imagine, accepted, as it would arguably "win the war". There is no escape from the social context in which beautiful theory like Newton's might – or might not – be used.[13]

Further, as Lakatos 1970 pointed out, to make sense of data requires observation theories, and the accuracy of observation – whatever that means – likely has some bearing on the way in which terms within theory map to observables. Thus, whilst predictive power may seem clear, it is not. One is tempted to conclude that predictive power exists when it is said to exist; this is done by some community, with reference to all the complex tangles human communities generally seem to be able to manage. They will therefore likely often argue about it. If this conclusion is reasonable then what can be said about predictive power?

What comes from my discussion of the contrast between the different criteria defining the acceptability of theory that we find in Crombie and Nisbet is that prediction is most important

---

[12] Such processes can be researched. Two studies that are striking for me are Yonay, 1998 and Rodgers and Cooley, 1999.

[13] As McCloskey 1985 puts it: "The numbers are necessary material. But they are not sufficient to bring the matter to a scientific conclusion. Only the scientists can do that, because "conclusion" is a human idea, not Nature's. It is a property of human minds, not of the statistics." (p. 112). And: "It is not true, as most economists think, that . . . statistical significance is a preliminary screen, a necessary condition, through which empirical estimates should be put. Economists will say, "Well, I want to know if the coefficient exists, don't I?" Yes, but statistical significance can't tell you. Only the magnitude of the coefficient, *on the scale of what counts in practical, engineering terms* as nonzero, tells you. It is not the case that statistically insignificant coefficients are in effect zero" (stress added p. 118). Quoted in Fforde, 2013.

in that it requires two things, and neither are to do with prediction *per se*, as it is generally understood (e.g. "getting a rocket to the moon").

First is the requirement for comparison between theories *as a matter of procedure*. If, however, this is not part of scientific procedure and a single truth is required, then this choice is logically done outside of scientific procedure.

Second is explicit management of the shift between suspension of scepticism in theorisation (Crombie's inductive phase, when theory is empirically-founded) and its resumption when theory can be, if the empirics suggest, abandoned. Following such norms, theory has to be protected, but not for ever, and it has also to be killable.

This view of the nature of predictability seems to me to be novel, and also to allow us to get away from somewhat fruitless debates. Economics as a science is about providing insights and improved understandings, and this is shown by its method.

Discussion

The idea that, because social change is unpredictable, the notion of development, of intentional social change itself (based upon statements that policy X will lead to change Y), is particularly problematic in international development. Fforde, 2005 reported citations of the application of robustness-testing methods to studies of the causes of variations in economic growth globally (Levine and Zervos, 1993). Levine and Zervos concluded that in the data there were almost no robust relationships, in other words that the articles in large literature asserting that the causes of growth were known, and reporting statistical analyses to support this, were spurious. Citations examined in Fforde, 2005 showed that most economists dealt with this anomaly by ignoring it, though a minority did not. Kenny and Williams, 2001 suggested that these spurious statistical results stemmed from assumptions of ontological and epistemological universalism, in other words that the world was far more varied than economic theory and its language suggested. Fforde, 2017 points out that, in international development practice, this set of scientific assumptions, as development workers well know, leads to denial of voice and a well-publicised series of "horror stories" as, totally unsurprisingly, outcomes are unexpected and often perverse [e.g. Ferguson, 1997]. The tension between viewing intentional social change, such as the deployment of a given economic policy, as something that is both done and also happens, as intentional and also part of some predictively knowable process, was clarified by the work of Cowen and Shenton 1996, who argued that historically two apparent solutions had been deployed. Both were answers to the question: what is correct policy? Like Levine and Zervos, they imply that social change is unpredictable. Given this, they argue that both solutions preserved the stance that change was predictively knowable. One asserted that correct policy was simply policy that fitted with the logic of change (this they term the Marxist solution); the other that correct policy was simply what those in authority said was correct policy.

Conclusions

The focus of this paper has been upon method. Economics is probably the single most important policy-relevant discipline in the social sciences. It is therefore important to understand matters of procedure – method – as they apply to economists' knowledge

production. It is also important for students of economics and consumers of economists' ideas to have a clear understanding of the rules, sometimes implicit, that validate economists' judgements in that they "follow the rules" – this is what makes them acceptable to other scientists.

Based upon an examination of three important texts we find that all have much to say about method and procedure. This is to be expected. What we find, though, is that it is impossible to link their statements about procedure to an arguably canonical exposition, following Crombie, of a science procedure that entails empirical assessments of theories derived from deduction using inductively-derived theorisation that use a predictive criterion or an equivalent. This is not what they are doing nor is it what they think they should do. Rather, Nisbet's arguments about the criteria required for accounts of social change seem far more appropriate, and lead to an understanding of economic science as the production of empirically-founded metaphors. Nisbet's arguments elucidate what these economic texts say they do and what they think they should do, as economists.

This helps explain just why and how economics exists as a powerful "real world" source of policy-relevant knowledge and popular beliefs about social realities. To carry weight in such areas, where, if we agree with Nisbet the competition is between empirically-founded metaphors (rather than Crombian prediction), that is the type of knowledge that has to be deployed. Arguments about the value of competition and free markets, supported and informed by economic theory, sit well within what Nisbet has to tell us about the particular and deep-rooted beliefs he reports governing what is required for accounts of reality to be accepted – to be given a "seat at the table". For me, this very much helps explain the power of economics as a science.

But this was not linked, in my argument, to some notion that economics is "not a predictive science". Rather, prediction, I have argued in what I think is a novel contribution, is more usefully seen as a criterion present in some scientific procedures, but not in others. It is usefully seen as acting, I have argued, as a requirement that theories be procedurally compared, with the implication that if it or an equivalent is absent, and a single truth required, theory selection will be done by something *outside* scientists' procedure. From this point of view economic science (understood in the terms here, that is, a science that is following Nisbettian rather than Crombian procedural rules) is, not being so protected by its procedures from outsiders' influence, usefully seen as required to manage that influence, in ways this paper has not addressed, partly for reasons of space, as the literature is vast, but also as the point I am making does not require it.

Further, the analysis showed that, whilst the key point to grasp about prediction is not that "it tells you whether theory is right", but that it is absent or present in the different criteria adopted by different types of science, the key point about science method, as Crombie presented it, was the prescribed management of belief/disbelief during and after theorisation. Empirically-founded induction, or in a modern language theorisation, requires a suspension of disbelief in theory: a suspension of scepticism for the theorist to theorise and believe that theory, a metaphor, has an acceptable relationship to reality from which it is separated and to which it has somehow to be mediated. Predictive power, *in terms of its scientific method alone*, is therefore clearly not important to economic science. What is important, in terms of how we may interpret Nisbet, is the ability to generate understanding of economic aspects of social reality that makes sense in that it offers powerful metaphors about the nature of economic phenomenon, in a complex and confusing world.

Various obvious and important implications follow for many who sit within the policy – process. For those seeking to use policy to guide action, it follows that a right to scepticism is vital. People should be free to assert that in a particular context change is predictively unknown, and organise accordingly (Fforde, 2017). For those seeking to gauge policy analyses, such as politicians and their political advisers, my argument suggests that they wisely be keenly aware that it is they, not the procedures shared by the array of knowledge producers confronting them, that decide the "truth of the matter" – that is, which amongst competing theories will be used. This has important implications for accountability, as some modern democratic electorates may have realised, or be realising, in one interpretation of Trump and Brexit. Economics, in this framing, has to be sure - if it is progressive - just how its positions are deemed correct: that is, for and by whom, and in whose name.

References

Arndt, H.W. (1981) "Economic development: a semantic history", *Economic Development and Cultural Change* 29(3) pp. 457-466.

Beed, Clive (1991) "Philosophy of Science and Contemporary Economics: An Overview", *Journal of Post Keynesian Economics*, Vol. 13, No. 4 (Summer, 1991), pp. 459-494.

Blanchard, O.J and Fischer S. (1993) *Lectures on Macroeconomics*, Cambridge MA: The MIT Press.

Cowen, Michael, and Robert Shenton (1996) *Doctrines of Development*, London: Routledge.

Crombie, A.C. (1953) *Robert Grosseteste and the origins of experimental science 1100-1700*, Oxford: Clarendon Press.

Escobar, Arturo (1995) *Encountering Development: The Making and Unmaking of the Third World*, Princeton, NJ: Princeton University Press, Princeton Studies in Culture/Power/History.

Ferguson, James (1997) "Development and Bureaucratic Power in Lesotho", in Ed Majid Rahnema and Victoria Bawtree, *The Post-Development Reader*, London: Zed Books.

Fforde, Adam (2005) "Persuasion: Reflections on Economics, Data and the 'Homogeneity Assumption'", *Journal of Economic Methodology*, 12:1 pp. 63-91 March.

Fforde, Adam (2009) *Coping with facts – a sceptic's guide to the problem of development*, Bloomfield, CT: Kumarian Press.

Fforde, Adam (2011) "Policy recommendations as spurious predictions: toward a theory of economists' ignorance", *Critical Review,* Vol. 23, nos. 1-2, pp. 105-115.

Fforde, Adam (2013) *Understanding development economics: its challenge to development studies*, London: Routledge.

Fforde, Adam, 2016, Confirmation Bias: Methodological Causes and a Palliative Response, *Quality and Quantity*, July, DOI:10.1007/s11135-016-0389-z.

Fforde, Adam (2017) *The sceptical change agent: reinventing development*. New York: Springer (Palgrave/MacMillan).

Fforde, Adam (2015) "What might international development assistance be able to tell us about contemporary 'policy government' in developed countries?", *Administration and Society,* May doi:10.1177/0095399715583891.

Fforde, Adam and Katrin Seidel (2015) "Cambodia - donor playground? Defeat and doctrinal dysfunction in a hoped-for client state", Mar, *South East Asian Research* 23 1 pp. 79-99.

Katouzian, H. (1981) *Ideology and method in economics*, New York: New York University Press.

Kenny, Charles and David Williams (2001) "What Do We Know About Economic Growth? Or, Why Don't We Know Very Much?" *World Development*, 29 (1):1–22.

Kline, Morris (1980) *Mathematics: The loss of certainty*, Oxford: Oxford University Press.

Knight, F.H. (1956) *On the History and Method of Economics - Selected Essays*, Chicago: University of Chicago Press.

Kuhn, Thomas (1962) *The structure of scientific revolutions,* Chicago: Chicago University Press.

Lakatos, Imre (1970) "Falsification and the methodology of scientific research programmes", *in*, ED Imre Lakatos and Alan Musgrave, *Criticism and the growth of knowledge*, Cambridge: Cambridge University Press.

Latsis, S. (ed.) (1976) *Method and Appraisal in Economics*, Cambridge: Cambridge University Press.

Levine, Ross, and Sara J. Zervos (1993) "What Have We Learnt About Policy and Growth From Cross-Country Regressions?", *American Economic Review*, Papers and Proceedings, 82 (2) (May): 426–430.

Levy, J.M. (1995) *Essential Microeconomics for Public Policy Analysis*, London: Praeger.

Nisbet, R.A. (1969) *Social Change and History: Aspects of the Western Theory of Development*, Oxford: Oxford University Press.

Obstfeld, M. and Rogoff, K. (1996) *Foundations of International Macroeconomics*, Cambridge MA: The MIT Press.

Priest, Graham 2(002) *Beyond the limits of thought*, Oxford: Oxford University Press.

Romer, D. (2011) *Advanced macroeconomics,* McGraw-Hill.

Said, E.W. (1978) *Orientalism*, New York: Pantheon Books.

Varian, H.R. (2010) *Intermediate Microeconomics – a modern approach*, 8th edition, New York: W.W. Norton.

Winch, Peter (1958) *The idea of a social science and its relation to philosophy*, London: Routledge.

Yonay, Yuval and Daniel Breslau (2006) "Marketing Models: The Culture of Mathematical Economics", *Sociological Forum.*

Yonay, Yuval P. (1998) *The struggle over the soul of economics*, Princeton: Princeton University Press 1998.

Author contact: adam@aduki.com.au

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

An ontology for the digital age?
David Elder-Vass, *Profit and Gift in the Digital Economy.* Cambridge:
Cambridge University Press, 2016, 257pp. (ppk) ISBN: 978 1 316 50938 8

Jamie Morgan [Leeds Beckett University, UK]

Though he studied economics in Cambridge in the 1970s, David Elder-Vass is not an economist in the professional/academic sense. He is a well-known social theorist and sociologist working within the broad approach of critical realism that readers may be familiar with from the work of Tony Lawson. In *Profit and Gift* he sets out to argue that we need a new conceptual framework for the economy. To be clear, by conceptual framework he literally means *conceptual* framework: a fundamental ordering set of concepts that presupposes the need for and then seeks to supply a generalised framing of what it means for something to be designated as economic, and so, by extension of an economy. His point of departure is a combination of empirical claim and theoretical critique: real economies are manifestly diverse, current approaches, though heterodox economics is a source of important critique and insight, do not adequately conceptualise this diversity – in so far as they are able to encompass *all* of that diversity as in some sense economic. More specifically, in order to apply a more encompassing set of concepts Elder-Vass begins from a "clearing" exercise focused in particular:

> "The best-established ways of understanding our economy are the neoclassical tradition that dominates mainstream academic economics and the Marxist tradition that dominates critical politics. For both, despite individual dissenters and substantial differences in the details, the contemporary economy is a monolith: a capitalist monolith, characterised more-or-less universally by the production of commodities for sale at a profit… The real economy, however, is far more diverse. It is neither overwhelmingly capitalist as most Marxists assume nor overwhelmingly a market economy as most mainstream economists assume. Both traditions tend to ignore vast swathes of the economy that do not fit with their stylised models, but because their models have thoroughly shaped our thinking they have largely succeeded in obscuring these diverse economic forms from view" (2016, pp. 3-4).

In what follows I focus first on the core positive aspect of Elder-Vass's argument before making some comment on the material used to provide the point of departure for that core (including the way, neoclassical economics, Marxism and monolithic capitalism are positioned).

The positive argument

Elder-Vass's main point is that thinking in terms of markets tends to peripheralise already existent economic forms that do not fit the concept of market capitalism (at least the concept he then sets out). Economic forms that are *not* based on commodity production for profit are widespread (according to Elder-Vass, "the non-market economy as a whole, far from being

marginal, is at least similar in its size to the market economy in contemporary global society, and arguably larger" 2016, p. 7, also p. 35). However, the concept of a gift economy also does not quite capture the full range of ways this is so. For Elder-Vass, there are many "hybrid" forms. Importantly, it is in recognising this that one understands the need for a different more encompassing account of an economy.

According to Elder-Vass, it is only by providing this more encompassing account that the range becomes properly visible and this is important because the rethink provides a way to support, nurture and "think creatively" about economic forms that already exist and for a constructive future. For Elder-Vass there is a current danger of viewing capitalism as overwhelmingly dominant and solutions to its problems as all or nothing outcomes. So, the point is to develop an approach that is less zero-sum and which recognizes the changes and potentials around us (inviting a constructive optimism, which finds a place for capitalism, markets etc rather than simply dichotomises). So, Elder-Vass approaches economics critically and with purpose, motivated to address:[1]

1. A social system that fails to prioritize the needs, let alone the flourishing, of the vast majority of the population;
2. An utterly distorted system of economic values shaping policy and perceptions of social issues;
3. An obsession with the idea that the market is the only significant (and either necessarily good or necessarily evil) form of the contemporary economy;
4. A discourse of the economy that sees it as essentially distinct from other aspects of society;
5. And reflecting that perception, a wall dividing economics from the rest of the social sciences;
6. A thoroughly formalist attitude to economics, at least in the mainstream, both in the sense of its mathematical form and in Polanyi's sense of being narrowly focused on optimising activities in markets;
7. A dominant tradition of economics that is flawed at a number of different (though thoroughly interconnected) levels – empirical, methodological, theoretical, epistemological and ontological.

Elder-Vass's solution is what he terms a *political economy of practices* (Chapter 5). That is, a concept of the economy as an integral constituent of social reality based on a common ontology. Events in the world are caused by the interacting causal powers of entities. *Social* entities involve people in contexts, typically organized through relations involving roles, and where the form of organization makes possible powers and activities that could not otherwise be possessed. Practices enacted can vary by organization, and given one can categorise as similar some kinds of organization, which Elder-Vass terms "complexes", one can designate economic activity as a social variant of complexes. He follows the heterodox inspired approach of designating social activity as "economic" (and so exploring this is to explore the "economy") if focused on "provisioning" (and emphasises an activity may have an economic "dimension" rather than is restrictively "economic"). Activities oriented on provisioning are termed by Elder-Vass "appropriative practices". The complexes can be explored at small scales (the causal powers operative, conditioned and expressed, for some category of economic entity, such as a business, or within a sector) and at large scales (price determination in markets, obligations in gift transfers etc). So:

---

[1] Personal communication 18th April 2017.

> "The unit we can characterise as belonging clearly to a specific economic form is neither the social formation [the entire society, since economic activity can be of varying kinds within it] nor the social entity [since there can be variation here too], but rather the *practice*. A *practice* is a tendency to act in a certain way, usually a tendency that is both reinforced for the individual and standardised across individuals by normative social expectations, although other factors can also contribute to the standardisation of practices… Practices are primarily the product of social norms: standardised expectations of how we should behave…." (2016, p. 99).

The underlying concept of the "economic" Elder-Vass wants to articulate draws initially on Allan Gruchy, Karl Polanyi, Andrew Sayer, Julie Nelson and various others. That is, the substantive component of social activity aimed at provisioning, or the interchange with the natural and social environment meeting the needs of human beings (2016, pp. 28-32). Elder-Vass acknowledges that there is a danger in extending the range of coverage of the economic through contestation of a market focus. In enabling a focus on what is "economic", opting for "provisioning" might become elastic to the point of meaninglessness (involving dispute regarding real versus false needs, varieties of immaterial needs of one kind or another etc). This may undermine the constructive potential of using the term provisioning that initially seems to be created because the term does not involve an *a priori* focus on markets. However, he states and suggests:

> "Our definitional strategy must be guided by our purposes. My purpose here is critical. Driven by the need to recognize that non-market activities can often meet our needs just as well as those activities that are considered *economic* under the market-oriented definitions of the economy. It seems more viable to persuade people that our familiar concept of the economic needs to be expanded in this way than to invent a completely new non-economic terminology to encompass the market and social alternatives to it. This is, of course, a strategic not an ontological argument… unlike the usual concepts of provisioning or the substantive economy [it] suggests a practical way of delimiting what activities we will call economic: the provision of goods and services through commodity exchange, plus the provision of equivalent goods and services through other social practise… [this] is not a definition of an objective social category… But the task we face today is not to be fully free of market thinking; it is to re-establish belief in the possibility of alternative social practise in a context where being fully free of market thinking is literally unthinkable" (2016, pp. 31-32).

Thereafter in Chapter 5 he sets out what appropriative practices are:

> "Using economic in the sense discussed in Chapter 2, that is, those practices concerned with provisioning our needs, either in the form of commodities or through goods and services that could have been supplied as commodities. Appropriative practices are related to – and ultimately dependent on – the production of goods or services (defined broadly, to correspond to this broad definition of economic) but their defining feature is that they are concerned with the allocation of the benefits (and indeed harms) that arise from production to individuals or social groups. In a sense this is a functional

definition, as it picks out a specific set of practices on the basis of their tendential effects rather than, say, some common feature of their structure, mechanisms or the actors involved: it includes all those *practices that significantly, systematically and more or less directly influence the allocation of the benefits of production*… [but] Let me be clear, then, that my use of the term *appropriative* neither refers to the initial creation of property nor carries any implication as to who is in charge of the practice [it involves no ethical judgement, focus on unowned to owned or simply ordinary language sense of taking]" (2016, p. 102).

Having defined appropriative practices in the context of his social theory position Elder-Vass then develops the way different complexes of appropriative practices can be stated and explored. For example, he terms "the pursuit of capital accumulation by the employment of wage labour to produce commodities" *canonical capitalism* (2016, p. 106). In the rest of the book he attempts to apply his approach to practices in order to illustrate that there are in fact a wide range of economic forms (and that his account can encompass the diversity of real economies). He provides five extended examples drawing on the digital economy: Apple, Wikipedia, Google Search, Facebook and YouTube (Part III: Chapters 6-9). Wikipedia illustrates an effective non-market economic model in operation, Apple a form of market power, but one based on practices that do not conform to mainstream economic accounts, and the other 3 different hybrid economic forms.

Profit and Gift concludes by both drawing together key aspects of the argument whilst also stating the ultimate purpose served by a broader provisioning inspired approach to economics:

> "Capitalism is not one such form but a range of different complexes of appropriative practices that share one central feature: activity in capitalist forms is ultimately driven by the need to accumulate ever-increasing amounts of capital. Our economy, however, is not only the site of capitalism but also of a wide range of other complexes of appropriative practices, including gift economy and hybrid forms…" (2016, pp. 216-217).

> "To make it possible to think beyond the economic systems we have today – and even just to understand the economy we already have – we need a new kind of political economy… We need a political economy that evaluates actions and practices against explicit ethical standards that reflect the actual needs of people" (2016, p. 216).

> "How, then, might we go about changing the economy to better meet human needs?... Part of the problem we confront is what David Harvey calls a 'double blockage': 'the lack of an alternative vision prevents the formation of an oppositional movement, while the absence of such a movement precludes the articulation of an alternative' [what is required is a spiral of critical experimentation with forms but also conceptual clearing of the kind Elder-Vass undertakes to facilitate this]… We will only be able to engage productively in such a process by abandoning monolithic visions of nirvana and working instead towards multiple partial real utopias. This is not a step backwards but a step forwards for progressive politics: we must reject the

dogmas of both the old political economies and instead engage creatively with our diverse economy and its open future" (2016, pp. 231-232).

Critical comments

It should be clear from the above that *Profit and Gift* does not emerge, depend on or develop an existing school of thought in economics as (traditionally conceived) substantive economic theory. Also, though drawing on the same tradition of philosophy and social theory work as Tony Lawson, the Cambridge Social Ontology Group and various other similar approaches (which also draw inspiration from the methodological and philosophical elements within the history of economic thought – original institutionalism, Keynes, Marx and so forth; Pratten, 2015; Dow, 2012), it is quite different. The intent is to clarify what the "economic" is and what an economy is, in order also to clear the ground for a progressive economy of the future, recognizing more explicitly the actual range now. This of course, in argumentation terms, presupposes that there is a need for a clarification of the economic and the economy, and requires that Elder-Vass's account is adequate as that clarification. Moreover, given the actual argument uses as its point of departure the claim that both a neoclassical and Marxist framework are dominant (albeit differently) and that both obscure real diversity, the initial presupposition seems to rest on the claims made regarding these two. However, one might argue that it only *seems* to. Elder-Vass's accounts of both neoclassical and Marxist economics can be misleading or ill-founded and it can still be the case that we *do* in fact tend not to pay due attention to the diversity of economies beyond (in his terms) archetypal market forms and processes, and this is worth highlighting and discussing. I would suggest this is the case.

Elder-Vass's account of neoclassical economics and Marxism is reductive in ways that serve his argument rather than represent the nuance of these points of departure as subjects (see Milonakis and Fine, 2009; Martins, 2012; Morgan, 2015; Stillwell, 2011). However, this merely makes the first part of the book something a reader is liable to read critically rather than just read as *critique.* The purpose of the book still remains interesting, and though liable to be provocative one should not neglect that being provoked is an important exercise to prevent complacency. Provocation is the flipside of the *practice* of pluralism (Fullbrook, 2008). As Fourcade et al (2015) reminds us, in terms of received practice economists more than any other branch of human inquiry need this kind of provocation. Oleg Komlik, Jakob Kapeller and Wolfram Elsner (amongst others) have done much to encourage genuine inter-disciplinary engagement with sociology.[2] Still, as political economists who work in sociology tend to note, sociology has its own problems and has become more introverted (for counter-tendencies see Jessop, 2008; Sayer, 1995; 2014).

There is also something inherently interesting about a focus on provisioning as a way to order our concerns regarding an economy. It does open up a wholly different way of thinking about ultimate concerns and the role of an economy (see e.g. Jo et al., 2017; Lee and Cronin, 2016). However, it does not follow that a focus on provisioning requires one to endorse Elder-Vass's way of expressing the constitution of that ordering. The concept of appropriative practices in complexes is required to do a great deal of work. One can always ask what kind of work is it doing and is it doing that work effectively? Arguably it is a theory of generalised

---

[2] See for example: https://economicsociology.org/oleg-komlik/ and
http://www.heterodoxnews.com/HEN/current%20issue.html

mutual definitions that *different* elements of an economy are then *re-described* in terms of. As a general purpose framework it integrates with Elder-Vass's prior social theory work, and there are competing varieties and critiques of this (see Archer and Elder-Vass, 2012; Porpora, 2015). Still, Elder-Vass is clear that his approach is more than just an exercise in categorisation, it allows one to clarify the range of causal mechanisms and real complexity of economic complexes. The task set is immense:

> "It seems, therefore, that we need concepts for at least three levels of economic form: i) the character of the whole economy or world system, which is clearly not a purely capitalist system but rather one that also encompasses non-capitalist economic forms; ii) capitalism as a rather varied form of economy that shares the core characteristics of being driven by the accumulation of capital; and iii) lower level (more concrete) forms, whether capitalist or not, that can be characterised by particular forms of organization of the productive practices (these are the forms that I propose to theorise using the concept of *appropriative practices*)" (2016, p. 60).

It is an open question whether re-description (if that is what it is) adds any insight to the different foci to which it is applied. One needs first to read the book and absorb the way of thinking before one can decide whether such statements as "x is an appropriative practice", "y is an appropriative practice" creates clarity and triggers lines of thought that would not otherwise exist for that practice, whilst also creating consistent or coherent thinking in regard of all such practices, *as an economy*. For example, what would recent post-Keynesian work on financialisation, and variants on credit theory-money creation through the banking system as unstable *real* tendencies in contemporary economies, and alternative wage-led aggregate demand approaches, look like in *appropriative practice-complex* terms (see e.g. Keen, 2017; Wray, 2015; Palley, 2013; Hudson, 2015; Lavoie and Stockhammer, 2013)?

More generally, there is a great deal more to say regarding the *mutuality* of diversity, once recognized. Not least because (as Elder-Vass as a proponent of emergence is aware) the character of a system acts down on as well as emerges from the complexity of (however described) the activity and organization of parts (see Elder-Vass, 2010). If a system contains capitalist forms but also non-capitalist ones then that system, *if it is a system*, still has operative conditioning aspects that make a difference to its parts – otherwise they are not *parts* of a *system*. They are just juxtaposed entities not components. What are those conditioning aspects and what is that "system"? *Non-monolithic* capitalism?[3] Alternatively, in what sense could a non-capitalist total system condition (serve to reproduce) capitalism within it? I have no answer here.[4] It merely struck me that Elder-Vass has implicitly set himself

---

[3] I was never quite certain what status Elder-Vass afforded to capitalism and what *he* meant by it: system (within which *not all* activity is commodity based extraction of surplus labour, since it is inherent to Marx's general theory that not all activity is of this kind – the household, the finance architecture in Volume 3 of *Capital* and relations to M-C-M, some of the activity in any given economic sector, political institutions etc.?) or specific operative form within a system etc? He uses the term in several senses based on different sources (e.g. 2016, p. 35) and acknowledges in each case that it is used so (but what is Elder-Vass's eventual positon on the relations of system and specifics could be usefully clearer, his attempts at clarification notwithstanding). Moreover, many of the sources of inspiration for the progressive conclusions to the positive argument derive from socialists and Marxists (Harvey, Olin Wright, Nelson etc.); these thinkers do not seem to have been hampered in thinking creatively in terms of recognized diversity by Marxism – and are part of the always existent critical tendency in Marxism and related thought.
[4] Elder-Vass argues that Marxist ways of thinking tend to render alternatives impossible – the flipside is Marx's argument that the reality of Capitalism is to spread and also to co-opt or subordinate any

a larger task by virtue of the nature of his critique, and this is not just about the names we apply.

Conclusion

One should remember that *Profit and Gift* is also a book about the digital economy and so its interest also stems from the many things written regarding trends in this aspect of the economy.[5] However, the abiding question that the book begs is *does one understand (and can one explain) a/the economic system any better based on the tools of the book, or has one merely acquired tools to explore specific parts in a new language of recognition?* One might be tempted to ask the same question of some of Tony Lawson's work – but it is important also to note that Lawson's initial contribution was to *bring out* the ontological and to draw attention to the problem of justifying the realism of economics (rather than necessarily what is real; see Lawson, 2003; 2015). In the spirit of pluralism I would recommend one read *Profit and Gift* and one do so with an open mind. Elder-Vass is an original thinker, one prepared to do the difficult work of building frameworks of thought, this is rare, and a quality that always offers the prospect of new insight, if at the risk of great error. Still, the prospect of progress without error is idealistic, and if one thinks that way, one may as well just go back to endorsing models of rational economic man where neither error nor insight is possible.

References

Archer, M. and Elder Vass, D. (2012) "Cultural system or norm circles? An exchange," *European Journal of Social Theory* 15 (1): 93-115.

Dow, S. (2012) *Foundations for new economic thinking: a collection of essays.* Basingstoke: Palgrave Macmillan.

Elder-Vass, D. (2010) *The Causal Powers of Social Structures.* Cambridge: Cambridge University Press.

Fourcade, M. Ollion, E. and Algan, Y. (2015) "The superiority of economists." *Journal of Economic Perspectives* 29, no. 1: 89-114.

Fullbrook, E. (ed.) (2008) *Pluralist Economics.* London: Zed Books.

Hudson, M. (2015) *Killing the Host.* Dresden: ISLET-Verleg.

Jessop, B. (2008) *State Power: A Strategic Relational Approach.* Cambridge: Polity.

Jo, T. H., Chester, L. and D'ippolita, C. (eds) (2017 [2018]) *The Routledge Handbook of Heterodox Economics: Theorizing, Analyzing and Transforming Capitalism.* London: Routledge.

Keen, S. (2017) *Can We Avoid Another Financial Crisis?* Cambridge: Polity.

Lavoie, M and Stockhammer E. (2013) *Wage-Led Growth: An Equitable Strategy for Economic Recovery.* Basingstoke: Palgrave Macmillan.

Lawson, T. (2003) *Reorienting Economics.* London: Routledge.

Lawson, T. (2015) *Essays on the State and Nature of Modern Economics.* London: Routledge.

---

alternatives beyond or emerging within it. Though, of course, Marx also notes the contradictions that create the grounds for people to organize and overcome this – so for Marx there seems to be no inconsistency, but for Elder-Vass there is (albeit one of Marxists).
[5] For another and different take on issues arising from the new world of information etc. see Moore, 2017.

Lee, F. and Cronin, B. (eds) (2016) *Handbook of Research Methods and Applications in Heterodox Economics.* Cheltenham: Edward Elgar

Martins, N. (2014) *The Cambridge Revival of Political Economy.* London: Routledge.

Milonakis, D. and Fine, B. (2009) *From Political Economy to Economics.* London: Routledge

Moore, P. (2017) *The Quantified Self in Precarity: Work, Technology and What Counts.* London: Routledge.

Morgan, J. (ed.) (2015) *What is Neoclassical Economics? Debating the origins, meaning and significance.* London: Routledge.

Palley, T. (2013) *Financialization*. Basingstoke: Palgrave.

Porpora, D. (2015) *Reconstructing Sociology.* Cambridge: Cambridge University Press.

Pratten, S. (ed.) (2015) *Social Ontology and Modern Economics.* London: Routledge.

Sayer, A. (1995) *Radical Political Economy: A Critique.* Oxford: Blackwell.

Sayer, A. (2014) Why we can't afford the rich. London: Policy Press.

Stilwell, F. (2011) *Political Economy: The Contest of Economic Ideas.* Oxford: Oxford University Press.

Wray, R. (2015) *Modern Money Theory.* Basingstoke: Palgrave Macmillan.

Author contact: jamiea.morgan@hotmail.co.uk

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

The resource curse reloaded: revisiting the Dutch disease with economic complexity analysis

Paulo Gala, Jhean Camargo and Guilherme Magacho [Fundação Getulio Vargas, Brazil]

Abstract
This paper shows that the Dutch disease can be more formally characterised as low economic complexity using ECI-type indicators; there is a solid and robust inverse relationship between exports concentrating on natural resources and economic complexity as measured by complexity indicators for a database of 122 countries from 1963 to 2013. In a large majority of cases, oil answers for shares in excess of 50% of exports. In addition to empirical panel analysis, we address case studies concerned with Indonesia and Nigeria and introduce a brief review of the theoretical literature on the topic. Indonesia is considered in the literature as a good example in avoiding the negative effects of the Dutch disease, whereas Nigeria is taken as a bad example in terms of institutions and policies adopted during the seventies and eighties. The empirical results show that complexity analysis and Big Data may offer significant contributions to the still-current debate surrounding the Dutch disease.

Introduction

The term "Dutch disease" was coined to describe the problems that emerged in the Netherlands in the 1960s and 1970s after the discovery of gas reserves. The sudden increase in exports of this product caused important changes for the Dutch economy. The excessive currency appreciation arising from the income that the new discovery generated implied a retraction for the Dutch manufactured goods industry, which ultimately led to unemployment and lower growth rates. The country's economic situation worsened after the reserves' discovery, in a paradox that became known as the "Dutch disease" problem, or the resource curse. The literature analyses many more such cases. Some authors showed how the discovery of gold in Australia in the 19th century caused similar problems for manufacturing in that country, or how the strong influx of gold into Spain in the 1500s as a result of discoveries in the Americas was also harmful. More generally, the Dutch disease paradoxically connects with the negative effects of the economic rents generated by great discoveries or the abundance of natural resources such as gold, oil, and gas.

This paper shows that the Dutch disease can be more formally characterised as low economic complexity with the use of ECI-type indicators; a solid and robust inverse relationship exists between exports that concentrate natural resources and economic complexity as measured by ECI indicators. In a large majority of cases, oil accounts for more than 50% of exports. The article is divided into five sections. Section one reviews the importance of manufacturing as an economy's productivity generator and analyses how the Dutch disease literature addresses this issue. Section two introduces the concept of economic complexity and analyses the main characteristics of the manufacturing sector and of the production of natural resources based on this new conceptual view. The third section is concerned with case studies for Nigeria and Indonesia. The fourth section uses panel data to

empirically analyse 122 countries from 1963 to 2013. The paper's fifth section summarises our conclusions.

1.   Theoretical aspects of the Dutch disease

The theoretical arguments that explain the channels by means of which the Dutch disease may affect an economy essentially relate to the reduction of the manufacturing sector. The currency appreciation that arises from the export of natural resources hampers the production of agricultural tradables and, in particular, manufactured ones that carry greater potential for technological innovations and productivity gains. The natural resources sector crowds out agricultural and manufacturing production. Capital and labour shift to natural-resource extraction and non-tradables production. The country's manufacturing sector turns inwards, specialising in the production of non-tradables with better returns because of the appreciated currency. Depending on the intensity of the process, the economy becomes "inward-looking", which also harms its efficiency level because of the absence of the competition it would face in the world market (Bresser-Pereira, 2007).

The models describing Dutch disease problems assume an economy with three sectors: non-tradables (NT), tradable natural resources such as oil (RT), and agricultural and manufactured tradables (MT). Corden (1984) analyses the issue in terms of a non-tradables sector (NT), a "booming sector" (B), and a "lagging sector" (L), where B represents the tradable natural resources sector and L stands for the manufactured exports sector. According to these models' traditional rationale, expansion of the natural resources sector (RT) causes exchange appreciation via two possible channels: increased flow of funds as a result of rising exports and increased prices of non-tradables because of the rising domestic demand stemming from the income gains in the natural resources sector. The manufactured tradables sector (MT) suffers because it loses capital and labour which shift to the non-tradables (NT) and natural resources (RT) sectors.

These consequences may also be described in terms of two effects: a resources shift, where capital and labour transfer to the "booming" sector because of increased returns, and the expenditure effect, according to which the factors used in the "non-booming" tradables sector also transfer to non-tradables because of increased demand. The latter effect's highlight lies in the increased demand for non-tradables as a result of the rising income that the booming sector generates. Agricultural and manufactured tradables then start being imported. By the end of a process with these characteristics, the non-tradables and natural resources sectors will have expanded, and the non-traditional tradables sector, manufacturing in particular, will have shrunk (Gelb, 1988).

On the problem of deindustrialisation in association with the Dutch disease, Palma (2003) points out the unemployment present in such cases. A pattern of manufacturing job decreases emerges in connection with a development process that naturally shifts workers from manufacturing to services industries. The negative effects of excessive currency appreciation stemming from the funds inflow generated by the commodities and tourism services sectors cause a premature eviction of workers from manufacturing. This splits countries into two categories: those that pursue an external surplus in the manufacturing sector because of a need to import other types of goods and those that pursue a strategy of industrialisation and manufactured goods exports despite being rich in natural resources. The author analyses the path of several countries that navigated the problem of the Dutch disease, particular note

being due to Southeast Asian ones. Malaysia, Indonesia, and Thailand are illustrative of what Palma (2003) calls "swimming against the tide" of deindustrialisation. Despite the presence of vast natural resources, they have been able to develop significant manufacturing industries.

The negative effect of the Dutch disease therefore relates to blocking the development of a non-traditional agricultural or manufactured tradables goods sector as a result of exchangev rate appreciation (Bresser-Pereira, 2007). Assuming positive externalities, "learning by doing" (Wijnbergen, 1984), or forward and backward linkages *à la* Hirshman in the production of such goods, the absence or retraction of such a sector has severe consequences in terms of technological dynamics and productivity gains. The agricultural and manufacture tradables sector's retraction causes – possibly irreversible – losses of knowhow, local capabilities, and production plants. The literature also refers to this outcome as the effect of the loss of technological spillovers due to the Dutch disease (see also Williamson 2005, 2003 and Sachs and Warner, 1995). As van Wijnbergen (1984) argues, economic development success cases after World War Two, without exception, involve aggressive promotion of the tradables sector, where technological progress is faster.

2. The Dutch disease defined as a loss of economic complexity

Empirically, it has never been easy to determine whether a country is experiencing or has experienced the Dutch disease. Great steps in this direction are possible with the use of the *Atlas of Economic Complexity*. Based on the *Atlas's* measures, one may more accurately define the Dutch disease in empirical terms: loss of economic complexity or a drop in the ECI indicator. The benefits of using the complexity index (ECI) to measure the Dutch disease lie in comparability and the availability of data for the past 50 years. Hausmann *et al* (2011) use computing, networks, and complexity techniques to create an extraordinarily simple method capable of measuring countries' productive sophistication, or "economic complexity". Based on the analysis of a country's exports, they can indirectly measure the technological sophistication of its productive tissue. The method created to build economic complexity indices culminated in an *Atlas* (http://atlas.media.mit.edu) that collects extensive material on countless products and countries since 1963. How does one measure an economy's "complexity"?

The two basic concepts for measuring a country's economic complexity, or sophistication, are the ubiquity and diversity of the products it exports. If a certain economy is capable of producing non-ubiquitous goods, this indicates a sophisticated productive tissue. Naturally, a problem of relative scarcity exists here, particularly for natural goods such as diamonds and uranium, for example. Non-ubiquitous goods must be divided into those with high technology content, which are therefore difficult to produce (airplanes, for example) and those that are highly scarce in nature (niobium, for example), which are therefore naturally non-ubiquitous. To control for the problem that scarce natural resources pose for measuring complexity, Hausmann *et al* (2011) use an ingenious technique: they compare the ubiquity of a good made in a certain country with the diversity of goods that other countries that also make the same good can export. For example, Botswana and Sierra Leone produce and export something that is rare and, therefore, non-ubiquitous: uncut diamonds. However, their exports are extremely limited and undiversified. These, then, are cases of non-ubiquity without complexity.

At the opposite end lie, for example, products such as medical imaging equipment, which practically only Japan, Germany, and the United States can manufacture; they are clearly non-ubiquitous. In this case, however, the exports of Japan, the United States, and Germany are extremely diversified, indicating that these countries are capable of making many different things. That is, non-ubiquity with diversity means "economic complexity". In contrast, a country with very diversified but ubiquitous exports (fish, fabrics, meat, ores, etc.) does not show great economic complexity; the country makes what every country makes. Diversity with ubiquity means a lack of economic complexity. The authors' trick with their complexity measures is to use diversity to control for ubiquity and vice-versa. The Netherlands, for example, is considered complex because its exports are diversified and non-ubiquitous; it is one of the few countries in the world to export X-ray machines. Ghana, in contrast, is a non-complex country because its exports lack diversity and cover ubiquitous products: fish and agricultural goods. Argentina lies at an intermediate position, with more diversified and less ubiquitous exports than Ghana but less diversified and more ubiquitous ones than the Netherlands'; it is therefore regarded as of intermediate complexity. The *Atlas*'s complexity calculation routine turns these important differences into a number called economic complexity that applies to both countries (ECI) and products (PCI). For example, in 2014, Pakistan's economic complexity was -.75 and Singapore's was 1.40, meaning that the latter was more complex than the former in that year.

One of the main virtues of these complexity indicators is that they rely on quantitative measures based on linear algebra calculations to arrive at results. They do not consider qualitative issues relevant to producing and exporting such goods. That is, they do not pass judgment on the complexity or non-complexity of products. Another interesting benefit lies in the measures' power to coherently capture huge changes in production technologies over time. A TV set from the 1970s is entirely different from another made in 2014. A car, airplane, or motorcycle from the 1980s is far from what we now call a car, motorcycle, or airplane. But, even so, the *Atlas*' methodology can capture the relative difficulty of producing each good at any point in time. A country now capable of producing a motorcycle may have been unable to do so in 1980 due simply to the fact that with today's technologies and trade integration it is far easier to make a motorcycle. But, for the purposes of the *Atlas*, a motorcycle is probably far less sophisticated today than it was in 1980. The concept of complexity remains over time a relative measure across countries and goods. Along these lines, Hausmann *et al* (2011) classify countries to arrive at impressive correlations between per-capita income levels, growth and economic complexity; this indicator may be used as proxy for the relative economic development across countries. It is no accident that Japan, Germany, the United States, the United Kingdom, and Sweden have always been among the world's most complex countries over the past 20 years.

The *Atlas of Economic Complexity* makes a very interesting contribution to the discussion of the Dutch disease; from the angle of a strictly empirical analysis based on the *Atlas*'s algorithm, manufactured goods are clearly characterised as more complex goods and commodities as less complex ones. By calculating the probability of products being co-exported by several countries, the *Atlas* creates an interesting measure of the productive knowhow contained in the products and of the local capabilities needed for their production: the "productive space" (Hidalgo *et al,* 2007). The greater the probability of two products being co-exported, the greater the indication that they contain similar characteristics and, therefore, require similar productive capabilities for production; they are "sibling" or "cousin" products. The co-exportation indicator ultimately serves as a measure of the productive knowhow linkage between products. That is, it indicates the productive "connections" between several

goods that emerge from their shared production requirements. Goods with high "connectivity" are therefore loaded with potential for technological knowhow. This makes them "knowledge hubs", whereas goods with low connectivity require simple productive capabilities, with low potential for knowhow multiplication.

For example, countries that make advanced combustion engines probably have engineers and knowledge that enable them to make a series of other similar and sophisticated products. Countries that only produce bananas or other fruits have limited knowhow and are probably incapable of producing more complex goods. It is important to emphasise that the difficulty observing this stems from the inability to directly measure and capture these local productive skills. International trade shows products, not the countries' ability to produce them. Empirically, the *Atlas* shows that manufactured goods are generally characterised as more "connected" and complex goods, and commodities emerge as less complex and "connected". Out of the 34 main communities of products calculated based on a compression algorithm from the *Atlas* (Rosvall and Bergstrom, 2007), machinery, chemicals, airplanes, ships, and electronics stand out as more complex and "connected" goods (that is, knowledge hubs). In contrast, precious stones, oil, minerals, fish and shellfish, fruits, flowers, and tropical agriculture show very low complexity and connectivity. Grains, textiles, construction equipment, and processed foods lie at an intermediate position between more and less complex and connected goods.

As for criticism of and potential problems with the complexity analysis methodology, its main fault may perhaps be to rely solely on export data as proxies for the productive structures of the various countries. This is indeed a weakness, as it is a known fact that, for a number of reasons, many countries produce but do not export some goods. The entire analysis is based on what can be "seen" from world trade data: a broad, disaggregated, standardised base that extends from the 1960s. The main benefit of these trade databases lies precisely in the standardisation, capillarity, and longevity of the data; its disadvantage lies in not capturing each country's domestic idiosyncrasies. In contrast, the national accounts databases that might contain such data do not, at this time, capture the same type of information at the level of granularity needed for the kind of analysis done here; they usually have few productive disaggregation layers. Another issue with the methodology is that it does not identify "maquila" countries: those that merely import and then export complex products, Mexico being the notorious case. Regarding this issue, Schteingart (2014) does an interesting job qualifying countries' "genuine" complexity by taking into account the number of patents filed and R&D expenditures as a percentage of GDP.

3. Cases of Dutch disease and complexity dynamics: Indonesia and Nigeria

In one of the most relevant references in the literature, Gelb (1988) discusses the case of countries that experienced the disease in the 1980s because of large oil reserves. He analyses the results of the oil shock for six economies with an abundance of this resource: Indonesia, Algeria, Ecuador, Nigeria, Trinidad and Tobago, and Venezuela. The author builds a single index capable of measuring the effects of the Dutch disease in each of these economies based on the evolution of their non-oil tradables sectors after the shocks. Nigeria and Trinidad and Tobago show the worst outcomes, with high currency appreciation in 1974-1978 through to 1984. Algeria, Venezuela, Indonesia, and Ecuador show better outcomes. The former three were able to keep their non-oil tradables sectors relatively unscathed during the period, although its representativeness in Algeria and Venezuela had always been very

small. The highlight in the sample is Indonesia, which was able to use currency devaluations to maintain the dynamism of its non-oil tradables sector (Gelb 1988).

In the Nigerian case, funds from the vast oil reserves were misused and ended up harming the country's growth path. Oil was found in 1956 and exports for the world market began in 1958. By the 1970s, 50% of Nigeria's exports came from oil, although agriculture remained the main activity, at 40% of non-oil GDP and 70% of the workforce. By the late 1970s, non-oil exports had all but vanished because of the price shock. The oil sector grew from 1% of GDP in the 1960s to 25% in the 1970s, and by 1979, oil accounted for 90% of the country's total exports. The agriculture sector retracted markedly, while the government concentrated funds from oil revenues in the non-tradables sector, which was extremely poor at the time (Gelb, 1988).

Over the course of the two oil shocks, the Nigerian real exchange rate appreciated intensely. The increased domestic demand generated by the rising revenues from oil exports led to a persistent rise in inflation during the 1970s. Because of import restrictions and inelastic food product supply, prices rose rapidly. The nominal exchange rate, pegged as it was to a basket of currencies of trade partners, together with annual inflation rates of around 20% from 1973 and 1978, led to an extremely appreciated real exchange rate. According to Gelb's (1988) calculations, the real exchange rate went from an average 100 in 1970-1972 to 287 in 1984. The Nigerian government's response was primarily to increase import tariffs in an effort to protect domestic manufacturing. By the mid-1970s, non-oil exports were practically insignificant, dropping in 1982 to one-seventh of their 1973 value. Nigeria became a net importer of agricultural goods in 1975. The country's total imports rose sharply in 1975-1978 and 1980-1982, with a massive loss of technological sophistication compared to the 1960s, as illustrated by the economic complexity index (ECI) and the export maps below.

Figure 1 Economic complexity in Nigeria
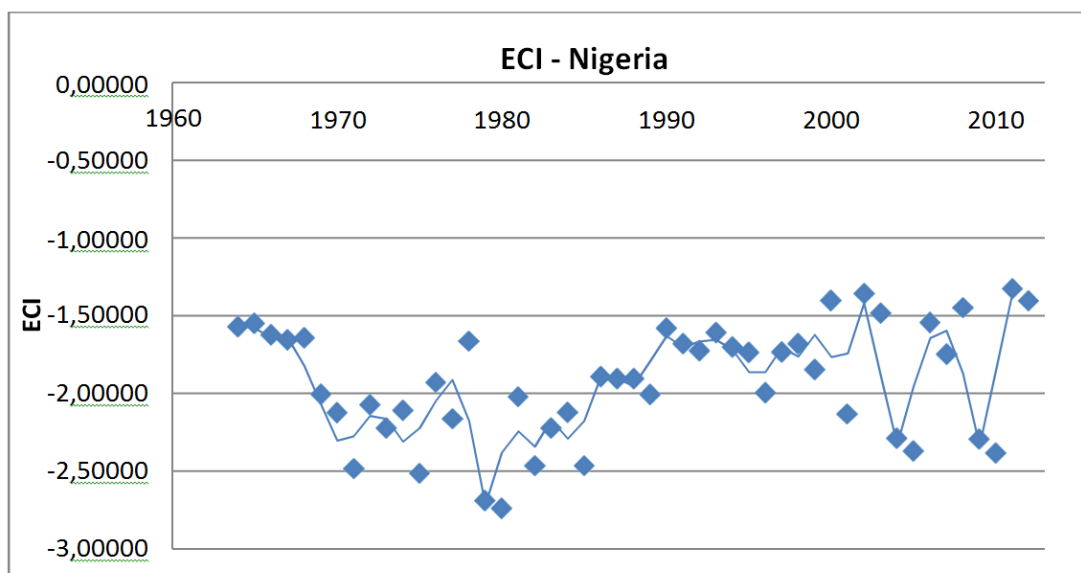
Source: elaborated by authors with data from the Atlas of Economic complexity:
http://atlas.media.mit.edu/en/

Figure 2 Exports of Nigeria
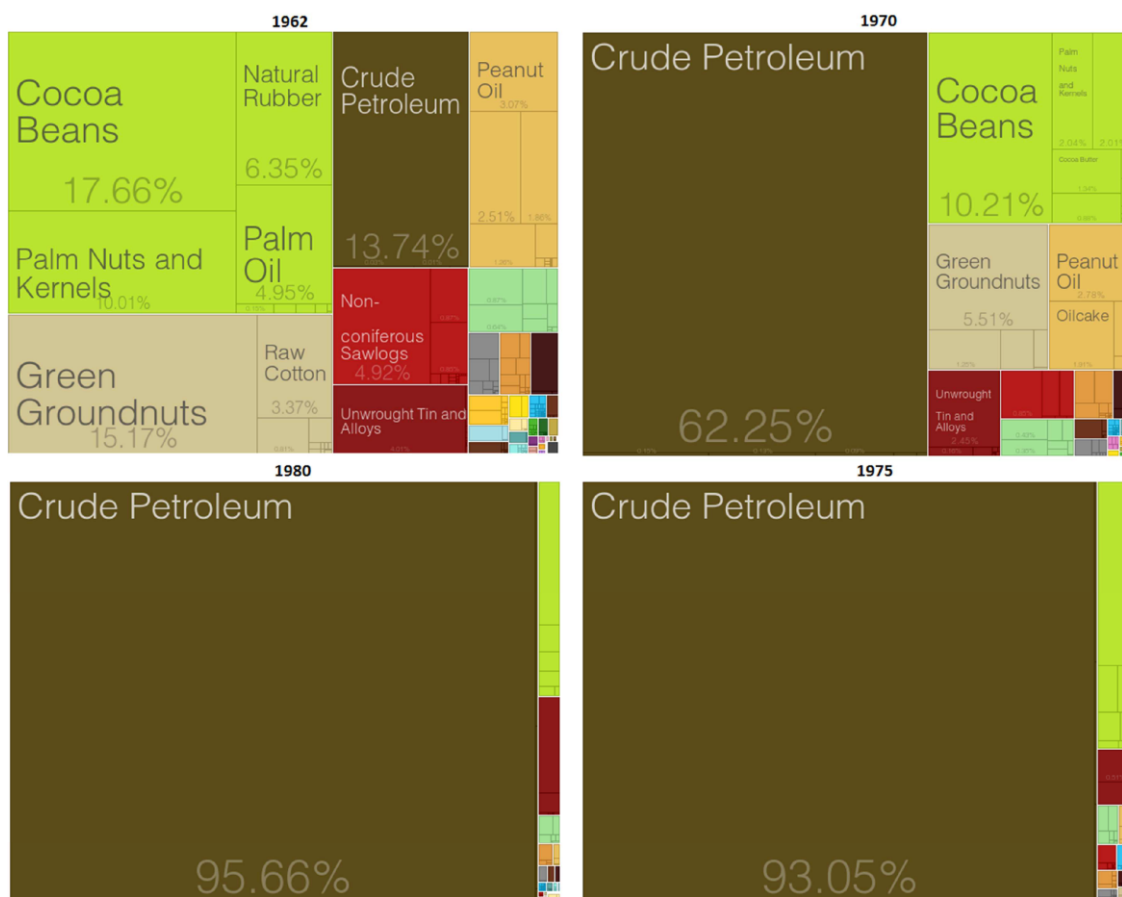
Source: elaborated by authors with data from the Atlas of Economic complexity: http://atlas.media.mit.edu/en/

An important contrasting case relative to Nigeria's is Indonesia, which is equally rich in oil but was able to manage its resources more rationally. It developed a dynamic agricultural and manufactured export tradables sector in parallel with the oil industry. Unlike Indonesia, where oil funds were also used to invest in agriculture, which prospered during and after the shock, the agricultural sector collapsed in Nigeria. The exchange rate appreciation caused by the rising domestic prices was intense, and the government took no steps to realign the real exchange rate. Productive activities were mostly redirected to the production of oil instead of tradables.

Despite the abundance of natural resources such as oil, rubber, and lumber, Indonesia was one of the poorest and most populated countries in the world in the 1970s, with a per-capita GDP of U$715 (1985 US dollars) and 129 million inhabitants. The team of bureaucrats that took over with Suharto's coup launched a successful economic action plan with the following guidelines: stabilising the economy, strengthening agriculture, manufacturing, and infrastructure, fostering foreign trade and exports, fiscal austerity, and reforming the financial industry. Among the specific measures taken, we emphasise the 1970 unification of the foreign exchange market, with the subsequent devaluation of the rupiah and the adoption of a pegged exchange regime starting in 1971. By the early 1970s, certain indices already

reflected the plan's positive effects: inflation below 10% annually, rising agricultural and manufacturing production, and an increase in currency reserves. Tax revenues rose from 4% of GDP in 965 to 10% in 1970.

Figure 3 Complexity and per capita income in Nigeria and Indonesia
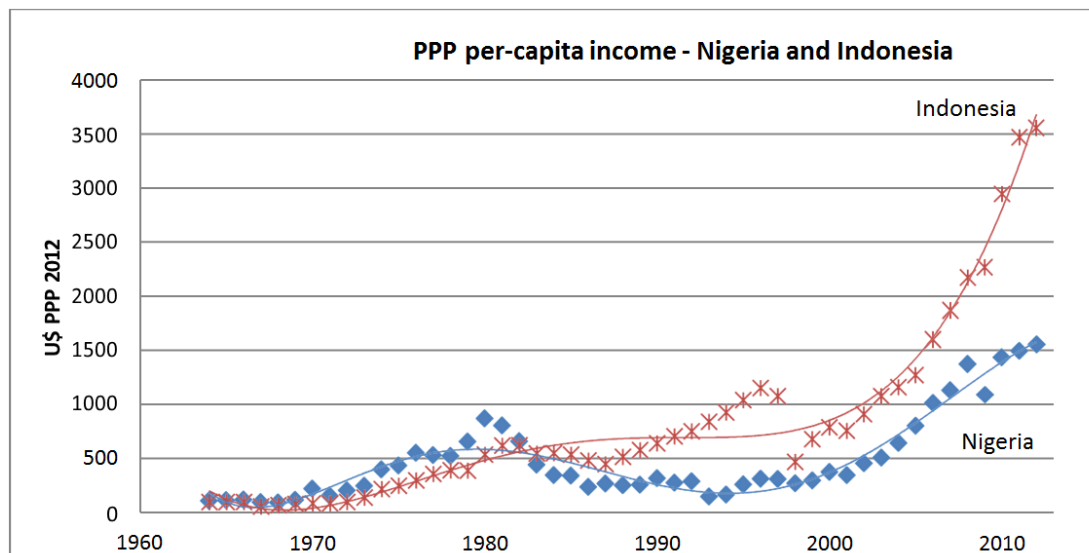
Source: elaborated by authors with data from the Atlas of Economic complexity: http://atlas.media.mit.edu/en/

Before the first oil shock, a huge crisis abated over Pertamina, Indonesia's biggest state oil company. Because of excessive debt with no central-government control, the company was unable to honour its obligations in 1974. To navigate the crisis, the government was forced to redirect two-fifths of all annual national oil revenues to pay the company's debt. The massive negative consequences of the event for Indonesia's economy taught the government a lesson about the potential risks of excessive economic dependence on oil. The Pertamina case strongly influenced the country's response to the oil shocks. In the md-1970s, the fixed exchange rate regime, associated with an inflation spike stemming from a failed food product harvest and the rising income and demand because of the positive oil-price shock led to a marked appreciation of Indonesia's real exchange rate. According to Gelb's (1988) calculations, the real exchange rate rose from a 100-base in 1970-1972 to 133 in 1978. The government responded with fiscal controls and monetary restrictions. That same year, authorities devalued the rupiah from 415 to 625 per US dollar. Many analysts (World Bank, 1993; Edwards, 1991; and Hill 2000, for example) emphasised that this devaluation was not associated with balance-of-payments problems or trouble paying for imports as oil revenues continued to flow into the country. Gelb (1988) argued that the authorities' purpose was to protect the non-oil tradables sector, in particular rubber, coffee, and the nascent manufactured goods industry. The public revenue gains stemming from the rise in oil prices in 1973 and 1974 were largely used to invest in infrastructure, agriculture, and manufacturing.

For the second oil shock, the authorities' response was also quite positive. They maintained a reasonably balanced fiscal position, even achieving surpluses at times. In terms of foreign accounts, Indonesia went from a current-account deficit of 1.6% of GDP (except mining activities) to a 2.3% surplus in the late 1970s. Currency reserves rose from 2.6 billion US dollarsin 1978 to 5 billion in 1980, despite high foreign debt. Compared with five similar cases analysed in Gelb (1988) – Ecuador, Nigeria, Venezuela, Algeria, and Trinidad and Tobago – Indonesia's real exchange rate remained practically stable in 1978-1982, showing almost no real appreciation (Gelb, 1988). With the drop in oil prices seen in the early 1980s and markedly deteriorating foreign accounts, the government again devalued the rupiah to 970 on the US dollar in 1983.

Throughout the 1980s, the exchange policy strategy continued, with a new devaluation of the rupiah in 1986 in response to the dropping oil prices and the adoption of a crawling peg system until the early 1990s. At first, the exchange rate followed the path of the US dollar, which devalued after the Plaza Accord, particularly against the yen. With the stabilisation of the US currency and moves toward appreciation in the early 1990s, Indonesia's authorities introduced a wider crawling peg, letting the rupee again devalue against a basket of currencies. Indonesia's performance was most remarkable in non-traditional exports. From 1982 to 1992, non-oil exports rose by 300%. In 1979, after the first devaluation, manufactured goods exports increased by 260%. A similar behaviour occurred after the 1983 devaluation (Gelb, 1988). The marked increase in manufactured goods exports was among the main factors responsible for keeping the economy on a positive path after the marked drop in oil prices in the mid-1980s. One of the sectors that prospered the most then was manufactured goods. Exports-oriented foreign direct investment started to flow into the country vigorously in 1986. Some authors mention an Indonesian exports boom after 1985 (see Hill, 2000). In terms of exports breakdown, the change is quite remarkable. In 1980, manufactured goods accounted for 3% of Indonesia's exports. The share rose to 7% in 1983 and reached 50% in 1992. However, despite the importance of exports in transforming the economy's structure, one cannot properly say that the country followed an export-led growth path like its Asian neighbours. Until the early 1990s, Indonesia's exports had never exceeded 31% of GDP. A series of preventive measures, including proper foreign exchange rate policy, made Indonesia an important case of overcoming the Dutch disease, as seen in the exports and economic complexity (ECI) mapsbelow.

Figure 4 Exports of Indonesia

Source: elaborated by authors with data from the Atlas of Economic complexity: http://atlas.media.mit.edu/en/

Figure 5 Economic complexity in Indonesia

Source: elaborated by authors with data from the Atlas of Economic complexity: http://atlas.media.mit.edu/en/

4. Panel analysis

In this section, we set up a panel of 12 four-year periods (1965 to 2013) and 122 countries with distinct export structures to evaluate the impact of concentrating exports on the change in the Economic Complexity Index (ECI). A high share of a good in total exports is understood as a symptom of the Dutch disease and reduced complexity is one of its harmful consequences for economic development, as discussed previously. Our analysis here attempts to evaluate these ideas in broader empirical terms. Our database has 5,108,305 lines covering the years 1963 to 2013, and the information includes the SITC-4 codes for the exported products, the origin and complexity of the exports, and the amount of exports in US dollars for each country in each year. The complete data bank covers 250 countries and 986 different products. To calculate the panel regressions, the data were treated and filtered using a routine to capture the top two products out of all exports. The graph next summarises the 2013 data with economic complexity on the X-axis and the share of the main exported product on the Y-axis.

Figure 6 Main exports of selected countries in 2013

Source: elaborated by authors with data from the Atlas of Economic complexity: http://atlas.media.mit.edu/en/

The economic complexity indexes for 1965-2013 have been obtained from the *Atlas of Economic Complexity* (Hausmann et al, 2011), and the control variables used in the regressions were drawn from the Penn World Table 9.0 (Fenestra et al., 2015). The selected econometric method was System GMM (Blundel and Bond, 1998). The estimator was developed based on the Arellano-Bond (1991) GMM estimator, which takes into consideration two sources of persistence over time: autocorrelation arising from the inclusion of lagged variables and individual effects to control for heterogeneity among individuals. With these estimators, the orthogonality between the time-lagged variables and the disturbances generate additional instruments. The difference between the Arellano-Bond GMM estimator and System GMM is that the latter enables causal analysis without the need for additional

exogenous instrumental variables, as it relies on lagged variables as instruments for the level equations and level variables as instruments for the lagged variables.

The impact of export concentration on the change in economic complexity is estimated based on the following equation:

$$dECI_{i,t} = o\ dECI_{i,t-1} + ß\ top\_pr_{i,t} + y\ Z_{i,t} + a_i + \mu_{i,t} \qquad (1),$$

where $dECI$ is the change in the economic complexity index, top_pr is the share of the country's top export product, and Z is the set of control variables. The long-term impact of export concentration in economic complexity increase is therefore given by

$$b = f3\ /\ (1 - o) \qquad (2).$$

Because the analysis focuses on the long-term impact of concentration in complexity increase instead of having *t* represent continuous years, the analysis chose to use four-year windows. In this sense, in the equation for period t, the period $t - 1$ does not represent the previous year but he previous analytical period, that is, the previous four-year period.

To prevent distorting the analysis by including secondary effects in the coefficient that measures the long-term impact, a few controls have been added. Initially, to prevent the economies' natural convergence and divergence processes from impacting the measurement of coefficient b, the (previous period's) economic complexity index itself was added as a control. In addition, some control variables were selected that are usually employed in studies that seek to evaluate the determinants of countries' income and productivity, such as government spending, population, and economic openness degree. Government spending, gov, is generally used to measure its weight, which may have a positive or negative influence on growth. For openness degree, we chose exports as a share of GDP, exp. Inclusion of this variable is particularly relevant to this study because it reduces the bias generated by the use of export complexity instead of the productive structure as a measure of economic complexity. In addition, to avoid ignoring the effect of population growth on the change in economic complexity, we included the variable $gPOP$ to represent the population's geometric growth rate in the previous four years.

Table 1 shows the estimations' results. The first model has no controls; the second includes only the complexity index to control for the impact of convergence or divergence, and the third model includes all other control variables mentioned.

Table 1 Impact of concentration on the change in economic complexity

	(1) $dECI_t$	(2) $dECI_t$	(3) $dECI_t$
$dECI_{t-1}$	-0.00804	-0.0160	0.00957
	(0.0348)	(0.0375)	(0.0383)
$top\_pr_t\ ECI_{t-1}$	-0.00123**	-0.00376***	-0.00450***
	(0.000600)	(0.000778)	(0.000931)
exp_t		-0.102***	-0.107***
		(0.0168)	(0.0200)
			0.00117***
			(0.000439)
gov_t			-0.00241*
			(0.00137)
$gPOP_t$			0.0150
			(0.0154)
Constant	-0.00709	0.0812	0.102
	(0.0649)	(0.0714)	(0.0781)
Long-term impact (b)	-0.00122**	-0.00370***	-0,00454***
	(0.000600)	(0.000778)	(0.000931)
Observations	963	963	953
Number of code	122	122	122
Hansen test	3.528	7.637	8.607
Hansen p-value	0.966	0.664	0.570

Standard errors in parentheses; ***: p<0.01, **: p<0.05, *: p<0.1.

In every model tested, the long-term impact of concentrating imports in a single product on the change in economic complexity is negative at a significance of at least 5%. This means that, although benefits may ensue for income in the short run, production concentration is harmful to technological development in the long run. In the first model, which lacks controls, the coefficient that measures this long-term impact is a negative .00123. Analysis of the Hansen test, however, indicates that the model is underidentified. The p-value of the Hansen test is expected not to be very high (close to unity), as this indicates a lack of sufficient variables to ensure the exogeneity of the instruments. The second and third models, in their turn, show appropriate Hansen test values. The tests' p-values neither approach unity, which would indicate underidentification, nor lie below 5%, which would indicate overidentificaiton.

In the latter two models, the long-term impact is measured by the coefficient b, which derives from the coefficients connected with the short-term impact and the autoregressive term, again indicates a negative relationship between export concentration and increased economic complexity. This indicates, as noted, that the Dutch disease, which occurs when exports concentrate on a specific product, may reduce countries' economic complexity in the long run and, therefore, reduce their technological development. The value of the coefficients is apparently small (although statistically significant at 1%). However, because this is about an impact of the **level** of concentration on the **change** in complexity, the consequences can be quite significant. A coefficient of .00454, as in the case of the third model, indicates that, for

every additional 1 p.p. of the main export, economic complexity is reduced by .00454 every four years. However, given that a country afflicted by the Dutch disease may have a single product accounting for 50% of its exports, as is often the case with oil exporters, the impact would be a reduction in complexity of .227 every four years. If such a country were to maintain this structure for 20 years, its complexity would decrease by 1.135, meaning that it could go from a situation such as Denmark's, with a complexity index of 1.284, to a level close to India's .102.

Despite the significant results that this analysis provides, we have not so far taken into consideration the complexity of the product that accounts for the largest share of the countries' exports. As discussed before, the Dutch disease does not stem exclusively from concentrating exports in a single product but from concentration in a low-complexity product. For this second stage in our analysis, we use heterogeneous regressions techniques (Agung 2014, Woodridge 2002) where the degree of exports concentration interacts with the product's complexity in such a manner that, instead of having the long-term impact b as a parameter, it works as function of the product's complexity. We begin by estimating the following model:

$$dECI_{i,t} = o \, dECI_{i,t-1} + \textit{ß}_1 \, top\_pr_{i,t} + \textit{ß}_2 \, top\_pr_{i,t} \cdot pci\_top_{i,t} + y \, Z_{i,t} + a_i + \mu_{i,t} \quad (3),$$

where pci is the complexity index of the main export.

Unlike equation 1, equation 3 includes an interaction term in association with the coefficient $f3_2$.

This term, which is the product of the share of the main product in exports times its complexity index, allows analysis of the impact of concentration in low- or high-complexity products. In this case, the long-term impact coefficient, b, is given by equation 4 and, therefore, is a function of the product' complexity rather than a parameter as in equation 2.

$$b = \frac{f3_1 + f3_2 \, pci\_top}{(1-o)} \quad \quad (4)$$

As in the analysis that did not take product complexity into account, three models were estimated: one without controls; another controlling for the potential impact of natural convergence or divergence, and yet another that includes, in addition to the latter control, certain variables generally employed to prevent the relevant coefficient from absorbing secondary effects. Table 2 shows our results:

Table 2 Impact of concentration on the change in economic complexity considering the complexity of the main export product

	(1) $dECI_t$	(2) $dECI_t$	(3) $dECI_t$
$dECI_{t-1}$	-0.00684	-0.0252	0.000278
	(0.0311)	(0.0330)	(0.0328)
top_pr_t	0.00330**	0.00122	0.000898
	(0.00141)	(0.00140)	(0.00141)
$top\_pr_{i,t} \cdot pci\_top_{i,t}$	0.00157***	0.00198***	0.00225***
	(0.000486)	(0.000470)	(0.000490)
ECI_{t-1}		-0.118***	-0.126***
		(0.0157)	(0.0181)
exp_t			0.00128***
			(0.000396)
gov_t			-0.00197
			(0.00130)
$gPOP_t$			0.0178
			(0.0145)
Constant	-0.0443	0.0458	0.103*
	(0.0639)	(0.0690)	(0.0566)
Observations	963	963	953
Number of code	122	122	122
Hansen test	3.777	7.216	7.701
Hansen p-value	0.957	0.705	0.658

Standard errors in parentheses; ***: p<0.01, **: p<0.05, *: p<0.1.

In every model, the coefficient associated with the interaction term is statistically significant at 1% and shows the expected sign, indicating that a positive product complexity index has a positive impact on the change in complexity and that the lower this index and the greater the concentration in the product are, the greater the complexity reduction will be. However, since we are now analysing a function rather than a parameter, quantitative analysis of this impact becomes more concrete considering different product complexity indices. Table 3 shows some product complexity indices and the long-term impact of concentration in the respective products.

As the table describing the impact of each product shows, although concentration in a specific product may negatively impact countries' complexity growth, as in the case of oil, minerals, and soy beans (considering model 3, which is the most comprehensive), products exist that positively affect the change in complexity, such as medicaments, cars, personal computers, and microcircuits. Concentrating exports in oil, as do Nigeria, Colombia, Egypt, Oman, Yemen, Kuwait, and others, negatively impacts the country's complexity change by .00483 per percentage point. As previously analysed, if oil accounts for 50% of exports, the country's economic complexity will decrease by .242 every four years, meaning that complexity will drop by 1.208 over a period of 20 years. Similar results, albeit at a lesser scale, apply to iron ore, the main export product for Brazil and Australia, and soy beans, the main export for Uruguay and Paraguay. In contrast, concentration in more complex products has no negative effect on complexity change. To the contrary, the effect of concentration in microcircuits, which are the main exports for Malaysia, the Philippines, and Singapore, and in personal

computers, which is the case of China, positively impacts complexity. Based on this analysis, one may infer that the Dutch disease, one of whose main symptoms is a high concentration of exports in low-complexity sectors, is an important factor in explaining the low technological development level of certain countries and, therefore, of their lower long-term economic growth.

Table 3 Impact of concentration on the change in economic complexity, examples of main export products

	ECI_{2013}	(1)	(2)	(3)	Countries
Liquid Petroleum Gases	-2.67383	-0.00089	-0.00397	-0.00512	Qatar
Crude Petroleum	-2.54636	-0.00069	-0.00373	-0.00483	Nigeria, Colombia, Egypt, Oman, Yemen, Kuwait
Iron Ore	-2.2224	-0.00019	-0.0031	-0.00410	Brazil, Australia
Soy Beans	-1.81141	0.000453	-0.00231	-0.00318	Uruguay, Paraguay
Petroleum Gases	-1.52749	0.000896	-0.00176	-0.00254	Bolivia, Netherlands
Copper	-0.43187	0.002604	0.000356	-0.00007	Chile, Bulgaria, Zambia
Medicaments	0.500071	0.004057	0.002156	0.002024	Ireland, Switzerland, Greece, Slovenia
Cars	0.587872	0.004194	0.002325	0.002221	Austria, Turkey, USA, Spain, Germany
Personal Computers	0.908917	0.004695	0.002945	0.002944	China
Electronic Microcircuits	1.4119	0.005479	0.003917	0.004076	Malaysia, Philippines, Singapore

5. Conclusion

To a certain extent, the results of the complexity and export concentration analysis shown here were expected; empirical investigation in the *Atlas of Economic Complexity* pointed in this direction. The complexity measures already use exports diversity as a criterion in their ECI calculations. The measures shown here are somewhat redundant. In contrast, the formal association between the excessive concentration of exports in products of low complexity and the general level of productive complexity of the various countries adds interesting analytical and empirical gains to the discussion of the Dutch disease. The intersection of the economic complexity literature with studies of the resource curse may provide new insights, such as the idea that productive specialisation in products such as oil or iron ore may not be fruitful from the angle of generating knowhow and innovation externalities; these are low economic complexity products. An interesting way to understand economic development is to think in terms of productive sophistication. Rich and developed countries are those capable of producing complex goods and selling them in world markets. The poor ones are those that can only produce and sell simple and rudimentary goods. This is why economic development can also be understood as a society's ability to know and control productive techniques, particularly in the more relevant world markets. This article attempts to contribute to this literature through economic complexity analysis applied to discussions of the Dutch disease.

References

Agung, I.G.N. (2014) 'Panel Data Analysis using EViews', Chichester, UK: John Wiley and Sons, Ltd.

Arellano, M., and Bond, S. (1991) 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations'. Review of Economic Studies, vol. 58, pp. 277–297.

Blundell, R., and Bond, S. (1998) 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, vol. 87, pp. 115–143.

Bresser-Pereira, L.,C., (2007) "The Dutch disease and its neutralization, a Ricardian approach", Brazilian Journal of Political Economy, São Paulo, v. 28, n. 1, Jan./Mar.

Corden, W., M., (1984) "Booming Sector and Dutch Disease Economics: survey and consolidation", *Oxford Economic Papers*, vol.36, n.3, November.

Edwards, S., (1991) Real exchange rates, devaluation and adjustment. *Exchange rate policies in developing countries*, MIT Press, Cambridge, Massachusetts.

Gelb, A., H., (1988) *Oil windfalls, blessing or curse?* World Bank Research publication.

Hausmann, R., Hidalgo, C., Bustos, S., Simões, A., Yildirim, M. A. (2011) *The Atlas of Economics Complexity – Mapping Paths to prosperity*. Puritan Press

Hidalgo, C. A., Klinger, B, Barabasi, A., L., and Hausmann, R., (2007) "The product space conditions the development of nations", *Science 27* July: 317 (5837), 482-487. Doi:10.1126/science.1144581

Hill, H., (2000), *The Indonesian Economy*, Cambridge University press, Second edition, Cambridge.

Palma, G., (2003) "Four sources of de-industrialization and a new concept of the Dutch Disease", in Ocampo, J. A. (ed.) *New Challenges for Latin American Development*, ECLAC-World Bank.

Rosvall, M. and Bergstrom, C. (2008) "Maps of random walks on complex networks reveal community structure," Proceedings of the National Academy of Sciences 105, 1118

Sachs, J., and Warner, A., *(*1995) "Natural resource abundance and economic growth", NBER 5398, December.

Schteingart, D. (2014) "Estructura productivo-tecnológica, inserción internacional y desarrollo", *Tesis de Maestría en sociología económica*, Idaes-Unsam, Buenos Aires.

Williamson, J., (2005) *Curbing the boom-bust cycle: stabilizing capital flows to emerging markets*, Institute for International Economics, Policy Analyses in International Economics, n.75, July.

Williamson, J., (2003) "Exchange rate policy and development", presented in Initiative for Policy Dialogue Task Force on Macroeconomics, New York.

Wijnbergen, S., (1984) "The Dutch Disease: a disease after all?", *Economic Journal*, vol.94 n.373, March.

Woodridge, J. M. (2002) *'Econometric Analysis of Cross Section and Panel Data', London: The MIT Press.*

World Bank (1993) *The East Asian Miracle, economic growth and public policy*, Oxford University Press, NY.

Author contact: paulo.gala@fgv.br, Jhean.camargo@gmail.com, guilherme.magacho@gmail.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

Transient development
Frank M. Salter

Abstract
A physically rigorous first principles quantitative assessment is made of the transient development of manufacturing projects as tools are made and applied to create final products. Output quantities and production rates are compared for different development histories applying existing technologies. Technical progress is excluded from projects, but no limit is set on the technology available. The effects of the division of labour are examined and the conditions for maximising output determined. Predictions and empirical facts are compared, from which it is concluded abductively that transient solutions provide quantitative descriptions of the development histories of manufacturing projects and industries.

1. Introduction

From the earliest times, members of the species *homo sapiens* or perhaps more pertinently *homo faber*[1] have used tools to improve their ability to survive. The improvement of existing and the invention of new tools continues to this day. Limits to the process are not apparent; simple extrapolation into the future suggests the possibility of unlimited expansion.[2] Since the industrial revolution, increase in the range and scale of tools has facilitated development from simple workshops to major industrial complexes. The application of new and existing tool designs to create output is the origin of economic development and growth. Economic analysis may then be seen as man's attempts to understand and describe the mechanisms involved in this continuing process as human behaviour and social structures respond to the pressures created by the deployment of new and existing technologies.

Despite the effort expended in attempting to understand this process, the outcome can only be described as piecemeal. Posing the question, 'What do economists really know?', Blaug (1998) concludes that the formalism in use is the underlying problem and that:

> "Economics has increasingly become an intellectual game played for its own sake and not for its practical consequences. Economists have gradually converted the subject into a sort of social mathematics in which analytical rigor as understood in math departments is everything and empirical relevance (as understood in physics departments) is nothing. If a topic cannot

---

[1] Byrne (2004, p.31) states 'Tool use is an important aspect of being human that has assumed a central place in accounts of the evolutionary origins of human intelligence'.

[2] Resource depletion will ultimately force limits upon some technologies. However as human understanding is incomplete and new technologies unpredictable, future development may proceed in unexpected directions. In 'The Tragedy of the Commons', Hardin (1968) discusses implications of living on a finite planet.

be tackled by formal modeling, it is simply consigned to the intellectual underworld" (Blaug, 1998, p.12–13).

In a paper first presented in 2006, before the economic upheavals which followed, Lavoie (2008) concluded that neither neoclassical nor heterodox analyses provide valid theories of production. Subsequently the many inadequacies of established economic thinking have been demonstrated: by its failure to predict; to establish control over; or even to explain the mechanisms allowing the extreme swings in global economic performance experienced in 2007 and the years following. The inappropriate hypotheses of the many schools of economic thought were laid bare.

Setterfield and Suresh (2014, p.812), in examining the dichotomy between micro and macro levels of economic analysis, conclude a necessity "for the prosecution of successful macroeconomic analysis by appeal to first principles". Bigo and Negru (2014, p. 329) state "Since the start of the global financial crisis, economists have increasingly acknowledged failures in their discipline" and (p.341) conclude "many economists across the board have tended to reaffirm their position... advocating the development of newer, better mathematical models that this time, allegedly, achieve greater realisticness (i.e. achieve a closer match to reality), promising a greater ability to successfully predict".

Clearly a mathematically-sound, physically-valid analysis from first principles is being called for; the subject of this paper.

1.1. Time in economic theory

Despite economic cycles being the norm from the beginnings of the industrial revolution, major areas of economic thought present equilibrium as an appropriate basis for analysis. Blaug (1998, p.23) comments "indeed real business cycle theory is, like new classical macroeconomics, a species of the genus of equilibrium explanations of the business cycle (which would yesteryear have been considered an oxymoron)." The formal treatment of time is eschewed.

In their articles, *The Production Function and the Theory of Capital*, both Robinson (1953) and Solow (1955) express their concerns about the use of time in economic analysis. Robinson points out that time is unidirectional in the real world, and that some mathematical descriptions fail to reflect the fact. Solow (1955, p.102) expresses his concerns, "But the real difficulty of the subject comes... from the intertwining of past, present and future."

Robinson (1980) continues to voice her critical assessment of the treatment of time in economic analysis. Later, no longer maintaining the concerns of his earlier insight, Solow (1994, p.47) states "Substitution along isoquants is routine stuff."

Setterfield (1995, p.23–24) concludes "it is not easy to introduce historical time into economic models in a manner that is at once meaningful and tractable. This is surely explained by the very nature of historical time, which must, by definition, defy deterministic, structural modelling"[3] and notes "the unwillingness of economic theorists to confront issues relating to historical time in the context of economic models".

---

[3] The present analysis is deterministic. Appendix A provides a different explanation of the intractability.

Boland (2005) in reviewing a three volume collection of articles, *Time in Economic Analysis*, by Zamagni and Agliardi (2004), concludes, "Namely, how can we build economic models where time matters because it is irreversible?" expressing his belief that this collection demonstrates the failure of economic analysis to offer any meaningful answer.

Clearly a physically valid analysis of economic development through time is required.

1.2. Physical dimensions

The International System of Units (SI) published by the Bureau International des Poids et Mesures (2006: updated 2014), is the English translation of the authoritative French text which defines physical units. It is a statement of the best scientific understanding of the nature of units of measurement. It states that physical quantities are expressed as products of a numerical value and a physical unit. Standard physical units can be one of a number of base units or units derived from them. *All* derived units are formed from the base units in combinations which have *negative, zero or positive small integer powers*. Dimensional analysis provides techniques by which to apply these requirements and thereby confirm the theoretical validity of the mathematical equations being applied to describe physical phenomena. It is a critical tool ensuring that symbolism conforms to reality but is little used in economic analysis. Barnett (2004) and Mayumi and Giampietro (2010) point out that the requirements of dimensional analysis are frequently neglected in economic publications.

Attributing theoretical significance to arbitrary equations containing physical units with fractional powers, violates the requirements of dimensional analysis. That the equations are good descriptions of the numerical data is irrelevant. That is an argument by analogy without first proving the analogy to be valid. A valid analysis using appropriate physical units is required.

1.3. Production and aggregate production functions

Economic analysts have introduced a bewildering range of mathematical expressions by which the production process might be described. Humphrey (1997) describes production functions before Cobb-Douglas and acknowledges that his description is incomplete as he has excluded others. S. K. Mishra (2010) presents a more recent overview of production functions. Zellner and Ryu (1998) seek to develop even more forms of *generalised production functions*; a process appropriate for curve fitting rather than developing theoretical understanding.

Georgescu-Roegen (1970) provides a rigorous but abstract mathematical description of the processes he perceived relevant to the quantitative description of the production process. He stresses the basic requirement that scientific symbolism should correspond to the real world. Amongst the wide range of factors he recognises as significant are time, utilisation factor and maintenance; parameters frequently ignored in conventional analysis. From an examination of the manner in which economic theorists formulate production functions, he (1970, p.2) infers that an "analytical imbroglio" exists in that "returns to scale must be constant in absolutely every production process".

Robinson (1953, p.82) states "To treat capital as a quantity of labour time expended in the past is congenial to the production-function point of view, for it corresponds to the essential nature of capital regarded as a factor of production". Solow (1955, p.101) dismisses her

suggestion by asserting that "This has a faintly archaic flavour"; though a few sentences earlier he asserts that labour and capital are measured in "unambiguous physical units", not appreciating that, for this assertion to be true, he requires capital to be labour-time.[4] Their different positions led into the *Cambridge Controversies*.[5]

In perceiving capital to be the appropriate parameter of analysis, time *per se* is excluded from production theory. *Calculus: One and Several Variables*, states

> "If the path of an object is given in terms of a time parameter t and we eliminate the parameter to obtain an equation in x and y, it may be that we obtain a clearer view of the path, but we do so at considerable expense. The equation in x and y does not tell us where the particle is at any time t. The parametric equations do" (Salas, Hille and Etgen, 2007, p.499).

The elimination of time from production functions should make their overall shape more apparent. However the range of empirical data is insufficiently wide to make proper de-termination without further theoretical justification. Invalid hypotheses well established in conventional literature introduce distortions into how reality is interpreted.

Allen (2012) provides the widest range of empirical data and notes (p.6) that "Neutral technical change is detected especially between 1880 and 1965" and that "The regressions show that the rate of productivity growth increased with the capital-labor ratio". However technology-in-use is introduced into production functions in different ways; depending on how the technology is conjectured to affect the performance of the production equipment. Three particular hypotheses are Hicks-neutral, Solow-neutral and Harrod-neutral which are respectively

$$y \ = \ Af\,(k,l), \qquad\qquad y \ = \ f\,(Ak,l), \qquad\qquad y \ = \ f\,(k,Al), \qquad\qquad (1)$$

where y is the output rate being evaluated, A, the technology-in-use factor,[6] k, the capital value and l, the labour applied.

1.3.1. The intercept of the aggregate production function

The value of the intercept of the aggregate production function is not to the fore of economic analysis. S. K. Mishra (2010, p.8) notes "It is surprising, however, that modern economists never formulate a production function in which labor alone can produce something." Von Thünen clearly understood the significance of the intercept. Jensen (2016), in examining von Thünen's contributions to the theory of production functions (von Thünen, J. H. *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, (Hamburg 1826)), identifies (2016, p.7), in English and the original German, the mathematical term (2016, p.7, equation(8)) which corresponds to von Thünen's "product of a man without capital (Arbeitsprodukt eines Mannes ohne Kapital)."

---

[4] Appendix A ends with the corollary that by representing production functions as $y \ = \ f\,(a,k,l)$, dimensional validity requires the units of capital to be labour-time.
[5] The Cambridge Controversies are discussed in general by Cohen and Harcourt (2003) and Lazzarini (2011) amongst others.
[6] The *factor* A is applied to different quantities in these general equations of production functions. Therefore, it must be a scalar quantity and as such can have no theoretical significance despite the appellation. This implies that the equations themselves are merely curve fitting devices.

That the value of the intercept is greater than zero is readily proved. At some point in human development, man had no tools. Man now has tools. Therefore the first tools were created without the use of tools. Labour is therefore able to create some output without tools.

The above-zero value presents empirically. All twelve figures shown by Allen (2012, pp.3–12) have intercepts above zero. His figure 7 (2012, p.8), plotting output rate against a low-value range of capital per worker, plots a relationship which he comments (2012, p.6) has continued for over two centuries. Badunenko and Zelenyuk (2004, p.469) show above zero values for the intercept in their figure 1.

Others presume the intercept to occur at the origin. Kumar and Russell (2002, p.534) base their tripartite decomposition on this presumption and introduce the origin as a point on the production frontier in figures 5, 6 and 7 (2002, pp.538–539). Makieła (2014) uses Cobb-Douglas and translog models which force the equations through the origin. The approaches taken by Koop, Osiewalski and Steel (1999) and (2000) produce equations passing through the origin but with very different slopes at the intercept: figure 1 (1999, p.478) shows the intercept being reached by lines with slopes of approximately one; figures 2 to 7 (1999, pp.481–492) with slopes of zero; figures 1 to 5 (2000, pp.294–295) with slopes of infinity.

It is shown that the intercept is determined by equation (14) and its significance is discussed in section 4.5, "The aggregate production function" and in section 4.7, "The Verdoorn coefficient and the intercept of the production function".

1.4. The present analysis

In his 1993 Nobel Prize lecture, *Economic Performance Through Time* (North, 1994) introduces a class of economic analysis as

> "A theory of economic dynamics comparable in precision to general equilibrium theory would be the ideal tool of analysis. In the absence of such a theory we can describe the characteristics of past economies, examine the performance of economies at various times, and engage in comparative static analysis; but missing is an analytical understanding of the way economies evolve through time.
>
> A theory of economic dynamics is also crucial for the field of economic development. There is no mystery why the field of development has failed to develop during the five decades since the end of the second World War. Neo-classical theory is simply an inappropriate tool to analyze and prescribe policies that will induce development.... The very methods employed by neo-classical economists have dictated the subject matter and militated against such a development.... In the analysis of economic performance through time it contained two erroneous assumptions: (i) that institutions do not matter and (ii) that time does not matter.
>
> This essay is about institutions and time. It does not provide a theory of economic dynamics comparable to general equilibrium theory. We do not have such a theory.[1] [The footnote states: In fact such a theory is unlikely. I refer the reader to Frank Hahn's prediction about the future of economic theory (Hahn, 1991).]" (North, 1994, p. 359).

The present analysis is a member[7] of the set North expected to remain unrealised and is presented with the intent of meeting the proposition:

> "Economic hypotheses can be judged by their logical coherence, their explanatory power, their generality, their fecundity, and, ultimately, their ability to predict. Economists, like all scientists, are concerned with predictability because it is the ultimate test of whether our theories are true and really capture the workings of the economic system independent of our wishes and intellectual preferences" (Blaug, 1998, p. 29).

The analysis is organised in the following manner. A physically valid mathematical representation of the production process is developed using both algebraic and differential equations. Solutions, limiting values and optimal production paths are determined. Output rates are shown to be aggregative. The solutions provide the theoretical rationale explaining Kaldor's stylised facts. Predictions of previously unknown relationships and their significance are revealed by the solutions. They are tested against empirical evidence which shows them to be consistent with that evidence and thus leads inexorably to a single conclusion.

2. Mathematical representation

Transient Development is a precise description of the scope of the present analysis. From first principles, it seeks a quantitative description of manufacturing projects' *development in and through time*.[8] *Development* is applied as an operant definition to describe projects which create tools by applying existing knowledge and techniques, to produce a final output. Projects do not experience technical progress.

This is quantitatively equivalent to the qualitative approach discussed by Robinson (1971, p.255) where she describes, "A book of blueprints representing a 'given state of technical knowledge'... Time, so to say, is at right angles to the blackboard on which the curve is drawn. At each point, an economy is conceived to be moving from the past into the future."

An Inquiry into the Nature and Causes of the Wealth of Nations begins

> "The greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgement with which it is any where directed, or applied, seem to have been the effects of the division of labour" (Adam Smith, 1776).

From this insight, he develops a qualitative analysis, perceiving the division of labour as specialisation into separate manufacturing processes which may be further divided into subsidiary processes.

---

[7] To the best of the author's knowledge, no other published analyses are available which apply physically rigorous analytical techniques to quantify the production process.
[8] In and through time is derived from titles of papers by Robinson (1980), "Time in Economic Theory" Boland (2005), "Economics in time vs time in economics: Building models so that time matters", and North (1994), "Economic performance through time".

The present analysis begins at exactly the same point as Adam Smith but proceeds quite differently. The initial ideas he presents are fundamental truths and provide basic abstractions for quantitative assessment of economic projects. In current terminology, the *productive powers of labour* may be regarded as productivity; the *division of labour* remains unchanged; the *greater part of the skill, dexterity, and judgement with which it is any where directed*, *or applied* may be seen to be the strategic choices about how human effort is allocated.

One further abstraction is required for a quantitative assessment of production; labour is divided into two parts, the first makes and maintains the production tools, and the second uses them to make a final product. Tools are created before they can be used and those in use require maintaining. The basic parameters quantifying the production process are:

1.   The numbers of people making, maintaining and using tools.
2.   The periods of time over which tools are being made, maintained and used.
3.   The number and effectiveness of the tools in use.
4.   The quantity of output.

2.1. The defining relationships

The present analysis is made on a *per capita* basis. The variables used are: h – the number of people; q – total quantity of output; t – time. Suffixes are: d – identifying tool-making and maintenance; p – identifying production of the final output. The physical units are the *natural units* of the problem space; established by the definition of the unit of quantity and initial values.

Idealising assumptions, definitions and relationships are:

1.   Quantitative relationships are linear; returns are to scale.

2.   Human effort is defined as the product of the number of people working and the period of time worked. Two parameterisations are possible:
     a)   retain the separation of h and t.
     b)   introduce labour-time as a single variable, $H \equiv ht$ – this, however, creates significant difficulties in solving the resulting equations – the complications are examined in Appendix A, "Labour-time".

3.   Productivity p, is defined as the rate of change of output with respect to the effort applied, $h_p t$;

$$p = \frac{\partial^2 q}{\partial h_p \partial t}.$$ (2)

4.   The level of technical progress is defined by the constant, a, its value determines the effectiveness of the tools being made. a is the rate of change of the productivity of the tools being made with respect to the effort expended in making them, $h_d t$;

$$a = \frac{\partial^2 p}{\partial h_d \partial t}.$$ (3)

5.   Maintaining and replacing worn equipment is a necessary part of manufacturing processes. Experience shows that new equipment requires less repair and replacement of worn parts than equipment which has been extensively used. Routine maintenance is generally carried out at predetermined intervals. Replacement of worn parts tends to be on a less precisely specified but at statistically determinable intervals. Attempting to follow such specific patterns might be appropriate for simulating the detailed performance of particular projects but is inappropriate in the search for general understanding. The regularities noted suggest that the maintenance requirement is best idealised as being in direct proportion to the machine-time expended.

        A maintenance constant, *m*, is defined to be the ratio of the effort required to maintain the tools to the machine-time used in operation.

6.   The natural unit of output is the quantity produced by one person, working without tools, for one unit of time.

7.   As the analysis is made on *per capita* basis, $h_d + h_p = h = 1$. In principle there is no implied restriction in the proportions of effort expended, except that the highest priority is given to maintaining the tools as they are used. Within this constraint all valid allocations of effort are possible but most of these will have no formal solution.

8.   Initial values for projects are $t = 0$, $p_0 = 1$ and $q_0 = 0$. The suffix $_0$ indicates values at time zero. These values and the natural unit of output establish the natural system of units for the problem space.

In science and engineering, it is a *necessary condition* that, *when equations are used to represent physical reality, they are dimensionally correct*. Failure to satisfy this condition is considered *proof* that any such representation is *wrong*. The variables relevant to the present analysis are a, h, q, t and m; m acts only to modify the effort available for tool making. So from Buckingham's theorem, systems represented by four physical properties are quantifiable by single dimensionless groups. Dimensionless groups appropriate to the development process are examined in appendix B, "Dimensional analysis".

As an analysis from first principles, no form is imposed upon the nature of manufacturing projects. All descriptions are provided by solutions to the equations.

3. Solutions

No general analytical solution exists for partial differential equations (2) and (3) when reified to describe projects in which *ad hoc* decisions change the proportions of effort being applied. However, two idealised development patterns with formal solutions are applicable.

An algebraic solution is available for projects consisting of two phases. An initial tool-making phase, during which all the production tools are made, followed by a second phase during which the tools are used and maintained whilst creating the final output. This is a typical development pattern in which factories are constructed, commissioned and operated to produce final products.

A solution from calculus is available when the proportions of effort, allocated to production and to tool-making and maintenance, are fixed. This allows the partial differential equations to be expressed as ordinary differential equations which have formal solutions. This pattern is an idealised description of projects for which production continues while tools are made and brought into use. Early artisans are likely to have followed such a pattern of development: making their own tools whilst continuing to make the products providing their livelihood. The successful would have expanded, training apprentices, building bigger workshops and introducing improvements which ultimately allowed craft workshops to become factories. Setting h_p constant allows the productivity definition, equation (2), to be written as the ordinary differential equation

$$\dot{q} = p h_p \tag{4}$$

3.1. Boundaries

The definitions adopted imply boundaries.
The no-development boundary occurs when production continues without tool making: output rates remain unchanged; $h_p = 1$; $p = 1$. Integrating equation (4) over the time interval $[0, t]$ gives

$$q = h_p t = 1 \cdot t, \qquad \therefore \frac{q}{t} = h_p = 1. \tag{5}$$

The upper output rate boundary occurs when all the effort available for tool making and maintenance is fully committed to maintaining the tools already in use; no further effort is available to make more. While maintenance levels may be reduced through technological improvement, maintenance free tools are impossible. The physically achievable boundary lies within that defined by the no-maintenance boundary. By setting the division of labour constant; partial differential equation (3) becomes the ordinary differential equation

$$\dot{p} = a h_d \tag{6}$$

which integrated over the time interval $[0, t]$ gives

$$p = 1 + a h_d t.$$

Substituting p from equation (7) into equation (4) gives

$$d_q = h_p (1 + a h_d t) dt \qquad \dot{q} = h_p (1 + a t h_d) \tag{8}$$

which integrated over the time interval $[0, t]$ gives the total output and the overall mean output rate respectively as

$$q = h_p t + \frac{1}{2} a h_d h_p t^2, \qquad \therefore \frac{q}{t} = h_p \left(1 + \frac{1}{2} a h_d t \right). \tag{9}$$

3.2. Solving the differential equations

By letting the function $\eta(t)$ represent the total effort applied to tool making, as a function of time, the total effort expended in tool making and maintenance, at time t, is

$$h_d t = \eta(t) + m \int_0^t \eta(t)\mathrm{d}t \tag{10}$$

which when differentiated with respect to time and rearranged gives

$$\dot{\eta}(t) + m\eta(t) = h_d \tag{11}$$

which multiplied by the integrating factor e^{mt} has the solution

$$\eta(t) = \frac{1}{m}h_d(1 - e^{-mt}). \tag{12}$$

Analogous to equation (7), productivity is

$$p = 1 + a\eta(t). \tag{13}$$

Substituting productivity from equation (13) into equation (4) gives

$$\dot{q} = h_p[1 + a\eta(t)] = h_p\left[1 + \frac{a}{m}h_d(1 - e^{-mt})\right] \tag{14}$$

which integrated over the time interval $[0, t]$ gives

$$q = h_p\left\{t + \frac{a}{m}h_d\left[t - \frac{1}{m}(1 - e^{-mt})\right]\right\} \tag{15}$$

and the overall mean output rate is

$$\frac{q}{t} = h_p\left\{1 + \frac{a}{m}h_d\left[1 - \frac{1}{mt}(1 - e^{-mt})\right]\right\}. \tag{16}$$

3.2.1. Development front

By substituting $(1 - h_d)$ for h_p in equation (14), the output rate development front as a function of h_d and t is

$$\dot{q} = (1 - h_d)\left[1 + \frac{a}{m}h_d(1 - e^{-mt})\right] \tag{17}$$

with output rate limit,

$$\lim_{t \to \infty} \dot{q} = 1 + (\frac{a}{m} - 1)\,h_d - \frac{a}{m}h_d^2. \tag{18}$$

3.2.2. Optimal output rates

Equation (15) maps the total output as a function of time. The allocation of effort which maximises the output quantity produced for the application of minimum effort is determined by the following procedure. Substituting $1 - h_d$ for h_p in equation (15) and rewriting gives

$$q = (1 - h_d)\left\{t + \frac{a}{m}\left[t - \frac{(1 - e^{-mt})}{m}\right]h_d\right\}$$

for which the first derivative with respect to h_d is

$$\frac{dq}{dh_d} = \frac{a}{m}\left[t - \frac{(1 - e^{-mt})}{m}\right] - t - 2\frac{a}{m}\left[t - \frac{(1 - e^{-mt})}{m}\right]h_d \qquad (19)$$

and the second derivative is

$$\frac{d^2 q}{dh_d^2} = -2\frac{a}{m}\left[t - \frac{(1 - e^{-mt})}{m}\right].$$

The sign of the second derivative is determined by $t - (1 - e^{-mt})m^{-1} \geq 0$ for all $0 < m < 1$ and $t \geq 0$. Therefore $\frac{d^2 q}{dh_d^2} \leq 0$ and so maximum output rates occur at values of h_d determined by setting $\frac{dq}{dh_d} = 0$ in equation (19) which gives optimum values for h_d as

$$h_d = \frac{1}{2}\left\{1 - \frac{m}{a\left[1 - \frac{(1 - e^{-mt})}{mt}\right]}\right\} \qquad (20)$$

with limiting values for h_d and h_p:

$$\lim_{t \to \infty} h_d = \frac{1}{2}\left\{1 - \frac{m}{a}\right\} \qquad\qquad \lim_{t \to \infty, a \to \infty} h_d = \frac{1}{2} \qquad (21)$$

$$\lim_{t \to \infty} h_p = \frac{1}{2}\left\{1 + \frac{m}{a}\right\} \qquad\qquad \lim_{t \to \infty, a \to \infty} h_p = \frac{1}{2} \qquad (22)$$

3.3. Algebraic solutions

Equation (7) describes the productivity reached along the no maintenance boundary, which corresponds to the situation before tools are brought into use. By making tools over the time interval $[0, t_k]$, utilising all available manpower, $h_d = 1$, the effort expended is $h_d t_k$, productivity of the tools produced is $p_k = 1 + at_k$. When these tools are used to produce the final output, they require a maintenance effort of mt_k which leaves $(1 - mt_k)$ units of manpower to create the final output. Thus the *per capita* idealisation implies an upper limit to

the useful quantity of tools created and the total output is

$$q = \begin{cases} 0, & 0 \leq t \leq t_k \\ (1 - mt_k)(1 + at_k)(t - t_k), & t \geq t_k \end{cases} \qquad (23)$$

and the overall mean output rate is

$$\frac{q}{t} = \begin{cases} 0, & 0 \leq t \leq t_k \\ (1 - mt_k)(1 + at_k)\left(1 - \dfrac{t_k}{t}\right), & t \geq t_k \end{cases} \qquad (24)$$

with the limiting overall mean output rate of

$$\lim_{t \to \infty} \frac{q}{t} = (1 - mt_k)(1 + at_k).$$

4. Discussion

Wide ranging relationships, based rigorously on physical principles, have been derived. Proven mathematically, they are necessarily true. As appropriate descriptions of the development process, their validity is determinable by comparison of prediction with empirical fact.

The assumption that the coefficient, a, remains constant, excludes technical progress from individual project histories. Actual values associated with projects, however, are established by the types and quantities of equipment installed. If mathematical relationships are reified by reference to specific empirical data, then the level of technical progress and associated limiting values are those established by the equipment in use. With no constraint on the value of a, the analysis is applicable at every level of technical progress.

Comparison of economic theory with empirical evidence, in academic publication, is generally presented as an assessment of how closely some hypothetical mathematical relationship or model fits empirical data; frequently alternate hypotheses are advanced. Coefficients for each are determined by curve fitting. The closest fit is then selected as an appropriate description of reality. While this process provides relationships suitable for interpolation within the range of the data, over-interpretation implying theoretical validity allowing extrapolation beyond that range, is likely to provide distorted views of reality.

Appendix B, "Dimensional analysis", shows that several dimensionless groups are available to represent different views of reality. Each provides a theoretically valid description of an aspect of the production process. Conventional analysis interprets some relationships detected in the empirical data as being paradoxical. The existence of several valid dimensionless groups provides a ready explanation. All forms should be detectable within the empirical data.

In the following discussion, it will be seen that some hypotheses of conventional analysis are falsified. Falsification *per se* is not the purpose of this discussion. Comment, on this, is restricted to an essential minimum. However, in order to use some published empirical data, an assessment is necessary to separate actual facts from views of those facts which have been distorted by inappropriate hypotheses.

4.1. Aggregation

The total output relationships of equations (5), (15) and (23) are aggregative; algebraic totals are valid mathematically. All relationships developed for individual projects are, therefore, equally applicable to manufacturing industries. Aggregate descriptions of production are based on a theoretically valid summation process.

By the mean value theorem, a description, within the mathematical forms presented, will exist for aggregated empirical data. This conclusion will not be explicitly reiterated; reference will be made, without further comment, to projects or industries whichever is appropriate in context.

The many arguments against the use of aggregate production functions, based on heterogeneity (Felipe and Fisher, 2003; Felipe and McCombie, 2014; Felipe and McCombie, 2013), must therefore be seen in context. If the numeraire (generally money) presents an affine transformation from the theoretically valid labour-time measurements then the resulting equations will be representative of the underlying reality and therefore economically useful. This is sufficient to explain the widespread and successful use of aggregate production relationships in macroeconomic analysis.

4.2. Stylised facts

Economic analysis accepts that the complexities of reality make difficult the comparison and full understanding of theory and fact. Without analyses from first principles, there are few generally agreed interpretations by which to compare or guide further analysis. Economic models are tested against stylised facts that are accepted, in general if not in detail, as demonstrably true.

Beginning "As regards the process of economic change and development in capitalist societies", Kaldor (1961, pp.178–179) suggested six stylised facts to which economic models should conform. Jones and Romer (2010, p.2) state that after nearly fifty years the first five of Kaldor's stylised "facts have moved from research papers to textbooks" and that currently "researchers are grappling with Kaldor's sixth fact".

All the facts, except Kaldor's statement of the capital:output ratio, are qualitative. The present analysis is quantitative. Simple direct comparison is inappropriate, but if the axioms of this analysis are empirically valid then its solutions should provide the theoretical explanation for Kaldor's observations.

Fact 1   "The continued growth in the aggregate volume of production and in the productivity of labour at a steady trend rate; no recorded tendency for a falling rate of growth of productivity." *And*

Fact 2   "A continued increase in the amount of capital per worker, whatever statistical measure of 'capital' is chosen in this connection."

Both the output rate equation (14) and the productivity equation (13) contain $\eta(t)$. The final two paragraphs of appendix A, "Labour-time", demonstrate that the general production function implies capital to be the labour-time expended in tool making, and that it is the monotonically increasing function $\eta(t)$, equation (12). Therefore, both of Kaldor's stylised facts are realised by the monotonically increasing capital relationship, $\eta(t)$.

Fact 4   "Steady capital-output ratios over long periods; at least there are no clear long-term trends, either rising or falling, if differences in the degree of utilization of capacity are allowed for. This implies, or reflects, the near-identity in the percentage rates of growth of production and of the capital stock – i.e. that for the economy as a whole, and over longer periods, income and capital tend to grow at the same rate."

This *fact* may be summarised as *the output:capital ratio is stable over time*.

4.2.1. The stability of the output:capital ratio

Since Kaldor (1961, p.178), various other studies have provided a range of differing conclusions about the stability of this ratio.

D'Adda and Scorcu (2003, pp.1180–1181) present values for seven industrialised economies for years in the range 1890 to 1990. They interpret the values for the USA as matching Kaldor's stylised fact 4, although the values they show graphically are about 0.3 before World War II and are generally above 0.4 afterwards. While the data for the other countries exhibits various excursions before, they all show falling values after the second world war, from which they conclude that, except for the USA, "output–capital ratios, show a tendency to reduce progressively over time" (2003, p.1189).

Madsen, V. Mishra and Smyth (2012, p.214) plot output:capital ratios for sixteen OECD countries from 1870 to 2004. The values they present for the USA follow a similar pattern to, but with numerical values apparently much greater than, those presented by D'Adda and Scorcu. However, whilst citing the D'Adda and Scorcu paper, Madsen, V. Mishra and Smyth do not comment on the numerical differences. They conclude that depending upon differing hypotheses of structural breaks, conclusions as to whether the data represents trend or mean reversion changes but that neither determination is totally sustainable. However they lean towards the interpretation of the ratio being constant. This may be for the pragmatic reason that this "ease[s] the interpretation of factors that determine the balanced growth path" (Madsen, V. Mishra and Smyth, 2012, p.233).

Allen (2012, p. 6) examines production data between 1820 to 1990 and considers whether the capital:output ratio is constant and concludes that it is not. He determines a value of 0.59 for one constant which, if the ratio was actually constant, would have the value 1.0. The value determined, 0.59, is fifteen standard deviations below that of the 1.0 required to indicate the ratio to be constant.

Quite clearly, the empirical data, in both confirming stability and movement, present a reality which is more complex than simply testing whether the stability of the ratio provides a binary

result. Transient analysis provides quantitative descriptions of the output rate, equation (14), and of capital, equation (12). The precise evaluation of the output:capital ratio is determined as follows.

Output:capital ratios for projects, expressed as the fraction $\frac{y}{k}$, are able to take on any value in the interval $[0, \infty]$. When output is being produced without the benefit of tools, capital is zero; the ratio is infinity. Initially when tools are made before being brought into use, no output has been produced; the ratio is then zero. In all economies, projects will be present at every stage of development. Stability can only be explained through aggregate values approaching the limiting values of the relevant relationships.

Appendix A concludes with the corollary that the concept of capital in production functions is the labour-time expended in producing the tools used. The output:capital ratio, $\frac{y}{k}$, is therefore

$$\frac{y}{k} \equiv \frac{\dot{q}}{\eta(t)} = \frac{h_p(1 + a\eta(t))}{\eta(t)} = h_p\left(\frac{1}{\eta(t)} + a\right) = h_p\left[\frac{m}{h_d(1 - e^{-mt})} + a\right] \tag{25}$$

$$\therefore \lim_{t\to\infty} \frac{y}{k} = h_p\left(\frac{m}{h_d} + a\right) \tag{26}$$

The limiting values represented by equation (26) remain in the interval $[0, \infty]$. However, as projects respond to competitive pressure the limiting values for h_d and h_p of equations (21) and (22), will be approached. Substituting these values into equation (26) gives

$$\lim_{t\to\infty} \frac{y}{k} = \frac{1}{2}\left(1 + \frac{m}{a}\right)\left[\frac{m}{\frac{1}{2}(1 - \frac{m}{a})} + a\right] = \frac{1}{2}\left(1 + \frac{m}{a}\right)\frac{a}{a}\left[\frac{2m + a - m}{\left(1 - \frac{m}{a}\right)}\right] = \frac{1}{2}\frac{(a + m)^2}{(a - m)} \tag{27}$$

With continuing technical development, a will increase, so

$$\lim_{t\to\infty, a \gg m} \frac{y}{k} = \frac{a}{2} \tag{28}$$

The limiting values for the ratio will therefore demonstrate apparent stability over shorter time periods[9] and with increasing technical competence, a ratio increasing over generationally long time periods.

---

[9] Equation 31 shows the value of technical progress obtained from Table 1 to be 0.49. The corresponding limiting value for the output:capital ratio for US manufacturing industries from the Solow (1957) data is therefore 0.25. For the whole US economy, D'Adda and Scorcu (2003, p.1180), in their Figure 1, show comparable values of between 0.2 and just over 0.3.

Figure 1 Output:capital ratio as a function of time – equation (25)
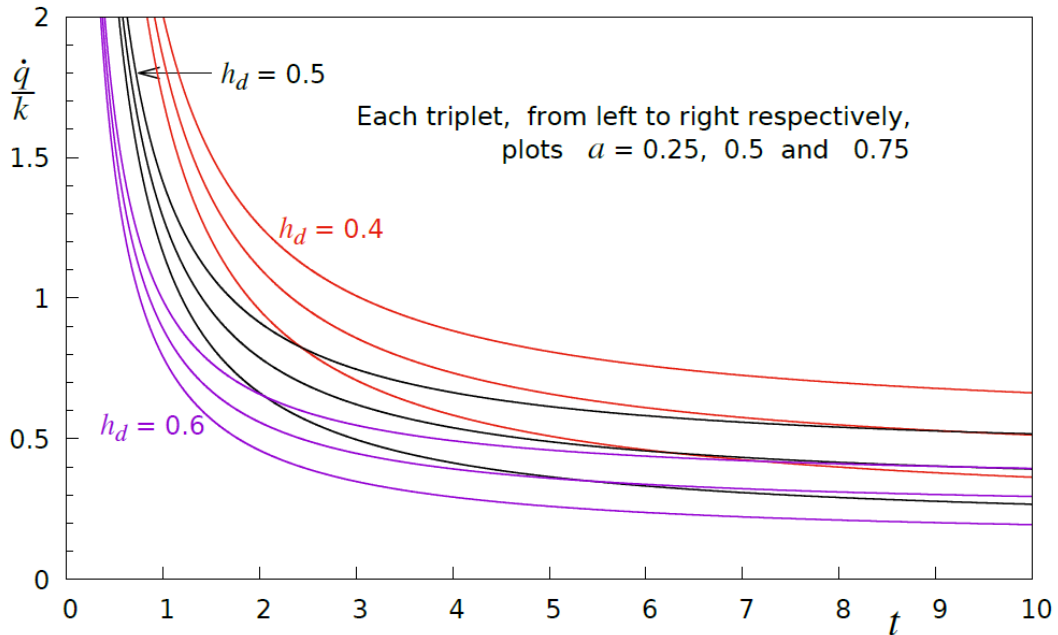
However, empirical evidence is restricted to shorter periods of time, collected over many countries and industries, and with the effects of many independent decisions, both good and bad. The complexities of the output:capital ratio progression may be seen in Figure 1, which plots equation (25) for some possible combinations of the fraction of effort allocated to tool making and maintenance, $h_d = 0.4$, 0.5 and 0.6 and for technology levels of $a = 0.25$, 0.5 and 0.75 natural units. Comparing the patterns shown in Figure 1 with those presented by Madsen, V. Mishra and Smyth' sub-figures of their Figure 1 (2012, p.214), shows that they all present similarities in their complexity, their decay with time and their convergence.

Equation (25) and the limiting value equations (27) and (28) provide a quantitative explanation of how the apparently inconsistent conclusions of the papers cited above arise: Kaldor (1961) (the ratio is effectively unchanging – true with no technical progress – consistent with equation (27)); D'Adda and Scorcu (2003, pp.1180–1181) (slowly increasing values for the USA – consistent with equation (28), the reducing values for the other countries – consistent with equation (25)); Madsen, V. Mishra and Smyth (mean reversion (stable ratio) – consistent with equation (27), trend reversion (increasing ratio) – consistent with equation (28)) and Allen (2012, p.6) (non-constant output-capital ratio – consistent with equation (28)).

Fact 6   expresses differences in the rate of growth of labour productivity and of total output in different societies whilst the other facts are maintained.

Conventional economic analysis, postulating equilibria, suggests unconditional convergence. The present analysis demonstrates that the performance of manufacturing industries depends upon decisions made about the quantity and the quality of the tools being created and applied. Inappropriate choices will produce less competitive industries; results consistent with Fact 6. A discussion of more recent research is considered in section 4.4, "Competition in manufacturing".

The third and fifth facts express relationships involving monetary values and as such they are directly outside the scope of the present analysis. However, with consistently distributed shares, both facts are in accordance with the steady capital-output ratio of Fact 4.

4.3. The development hypersurface

Equations (16) and (24) are the loci of all possible overall mean output rate histories of manufacturing projects and industries; the *development possibilities hypersurface*. Each strategy creates its own *world-line* traversing its own distinctive path over the hypersurface. Project-lines are the output rate histories of individual projects. Industry world-lines describe the movement of an industry's mean aggregated value over the hypersurface.

Figure 2 Overall mean output rate ($\bar{q}$) hypersurface views
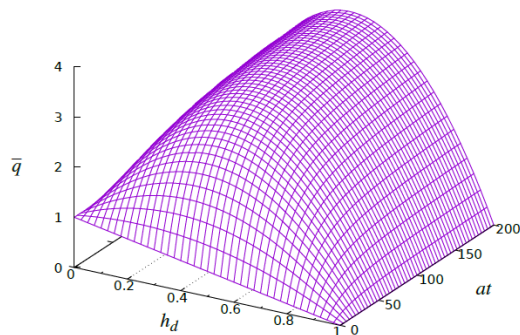
(a) Differential Equation Solutions
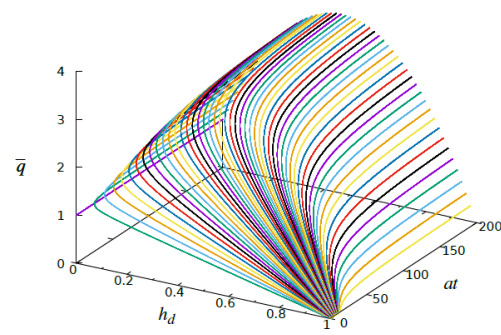
(b) Algebraic Solutions
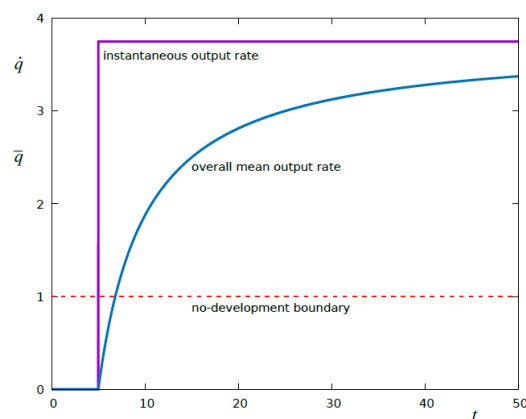
(c) Output Rates–Algebraic Solution
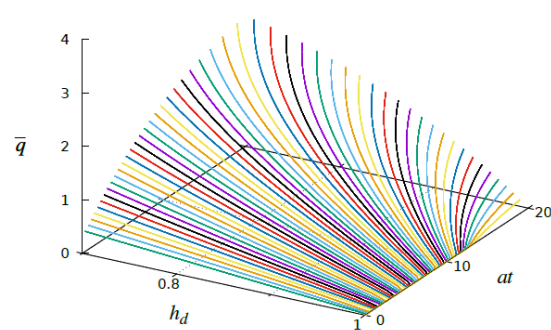
(d) Early Development–Algebraic Solutions

Projections of the development possibility hypersurface as functions of h_d and at[10] are shown in Figure 2. The values of the technical progress constant and the maintenance requirement are 1.0 and 0.075 respectively for each figure.

Project-lines for equation (16) are shown in Figure 2(a). The lines begin on the straight line joining points $(0,0,1)$ and $(0,0,0)$ and are at intervals of $\Delta h_d = 0.05$. Lines of constant at serve only to delineate the hypersurface.

The project-line trajectories of equation (24) are shown in Figures 2(b) and 2(d). Figure 2(d) enlarges a portion of Figure 2(b) to clarify the early development of the trajectories. A greater investment in tools, with a consequently later start of production, brings about very different project-line trajectories.

Equations (16) and (24) solve the same partial differential equations with very different development patterns, so while individual *project-line's* trajectories differ, they traverse the same hypersurface. This is clearly visible in Figures 2(a) and 1(b).

Figure 2(c) plots instantaneous and overall mean output rates along project-lines with production starting at $t_k = 5$. The dotted line is the output rate at the no-development boundary. For the initial tool investment case, at time t_k, the instantaneous output rate rises from zero and remains constant until production ends, whereas the overall mean output rate remains at zero until production starts and then converges asymptotically towards the value of the instantaneous output rate. Parameterisation into capital and labour implicitly introduces the instantaneous output rate view of reality and obscures the transient nature of the development process.

Contour plots, corresponding to Figures 2, are shown in Figure 3(a) for the development front equation (17), as functions of h_d and at, and for the limit equation (18), as a function of h_d. The development of the no-maintenance boundary is shown in 3(b).

The no-development boundary, where no tools are being made, is the point $(0,1)$. At the point $(1,0)$, tools are being made but not used and so it represents the overall mean output rate of the algebraic equations before production begins, $t \leq t_k$.

The contours are points of equal effort expended. Assuming the efforts of tool makers and users have the same unit cost, it follows that these contours are the *isocost* lines of conventional analysis. Horizontal lines between the outermost points record the same overall mean output rate. So by definition these lines are *isoquants*.[11]

The monotonically increasing dashed curves, originating at point $(0, 1)$, are the locus of points of maximum output rate with minimum effort. That of Figure 3(a) plots equation (20). It is the competitive growth path which competitive industries would have followed and it may be conjectured that this is the basis for the conventionally hypothesised *balanced growth path*.

---

[10] Equation (9) is the no-maintenance boundary but it is also the straight line from which equation (16) diverges under the effects of the maintenance decay term – at is therefore an appropriate parameter against which to map output rates.
[11] Isoquants have no meaningful existence in transient analysis. To provide a *realistic* description requires the introduction of concepts found in science fiction. Movement along an isoquant would be akin to travel in a time-machine for which movement through time requires simultaneous movement through alternate realities.

Figure 3 Overall mean output rate $(\bar{q})$ as a function of h_d and at

(a) With maintenance   (b) The no-maintenance boundary
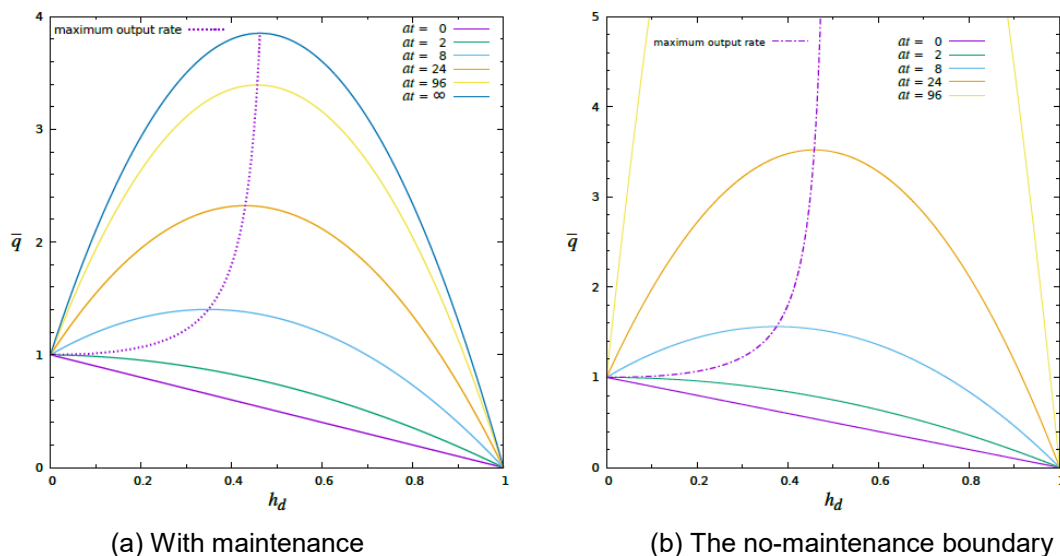
4.4. Competition in manufacturing

In examining British productivity performance from the late-nineteenth to the early 21st century, Crafts (2012, p.18) notes "Economic theory gives somewhat ambiguous messages about the impact of competition on productivity performance" and that competition and good productivity performance moved together throughout the period. He (p.27) concludes "Applied economists in the UK are now generally agreed that strengthening competition in product markets is good for productivity performance."

Similar tendencies are reported on a worldwide basis:

> "I show in this article that unconditional convergence does exist, but it occurs in the modern parts of the economy rather than the economy as a whole. In particular, I document a highly robust tendency toward convergence in labor productivity in manufacturing activities, regardless of geography, policies, or other country-level influences" (Rodrik, 2013, p. 166).

Figure 3(a) provides the explanation for these observations. The most competitive industries would have followed the maximum output rate curve of equation (20). New techniques, embodying the latest technological advances, implying increasing values of a, would have been introduced as new projects were completed and brought into production. The mean value of a for the industry would have continued to increase, and with a the limiting values of equations (21) and (22) are changed.

The introduction of more productive equipment would have allowed greater output to be achieved more quickly. By a process of Darwinian selection those following this path most closely survived. Others fell by the wayside. At any level of technical development with appropriate apportioning of tool making and productive efforts, a single point of maximum output will be achieved. Through competition, unconditional convergence occurs.

At all levels of technical progress, the best engineering produces maximum productivity. Vergeer and Kleinknecht (2007, p. 20) confirm that low wages do not equate to competitiveness, by their conclusion, "Our panel data analysis shows that a causal link indeed exists between wage growth and labour productivity growth".

Bénétrix, O'Rourke and Williamson (2015) examine the spread of manufacturing to the poor periphery between 1870 to 2007. They note that the industrial catching up on which they are reporting is different from the manufacturing productivity convergence reported by Rodrik (2013). They describe the catching-up process to be present throughout the period and that it reached a high point between 1950 and 1973. Again this is completely consistent with the predictions of the present analysis. There is no reason that the most appropriate engineering decisions will be made and thereby achieve competitiveness. Even valid choices may be undermined by other factors; inadequate infrastructure, civil unrest, corruption etc. There is no preferred direction of development; only better or worse choices.

Adam Smith's metaphor of the *invisible hand* may be seen as an anthropomorphic interpretation of the pressure exerted by competitors' increasing productivity. Good engineering increases competitive advantage. Competition enforces good engineering and allows output to be increased with no overall increase of effort. Economic competence and good engineering are the same. The invisible hand is the pressure created by others as they successfully introduce new and improved techniques into the production process.

The algebraic solution provides a description of a single manufacturing project which is built, brought into service, continues in production for a number of years and finally closes. This pattern may be aggregated to represent multiple projects, on single or many production sites, starting and finishing at different times and introducing the latest techniques; a scenario very familiar in advanced economies.

The solution through calculus provides a different scenario describing the earliest development patterns from household production to workshops operated by single artisans making their own tools to produce their wares. With the immediate feedback present at this scale the introduction of new effective techniques occurs organically by the selection of the best and discarding the less good. Again aggregation might be used to describe this process but it is unnecessary as calculus provides a direct description of the introduction of improving techniques, equation (20) shown in Figure 3(a). The *per capita* representation encompasses the increasing scale of production from workshop to factory. By the mean value theorem this solution is also the description of the development of competitive industries over extended periods of time. As the scale of industry increases even large plants are effectively infinitesimal in the expanding world of manufacturing.

4.5. The aggregate production function

Equation (14) is the transient relationship expressing the instantaneous output rate which is also evaluated by the production functions of conventional economic analysis. Through aggregation, it is comparable to the aggregate production function.

Solow (1957) analyses United States GNP data of private non-farm economic activity from 1909 to 1949 to obtain an assessment of the aggregate production function. He separates the form of the function from the effects of technical change. In commenting on this paper,

McCombie (2000, p.271) observes that Solow "adopted a novel estimation procedure that has not been followed since".

Solow (1957, p.312) warns, "It will be seen that I am using the phrase 'technical change' as a shorthand expression for *any kind of shift* in the production function" and (p.320, footnote 18) "I have left Kendrick's GNP data in 1939 prices and Goldsmith's capital stock figures in 1929 prices". He explicitly acknowledges (p.314) that in using GNP, "the share of capital has to include depreciation". Solow's Table 1 contains numerical errors which were corrected by Hogan (1958, p.407). The corrected data are used for testing the present analysis.

Solow hypothesises that the general statement of the aggregate production function is $Q = A(t) f(K, L)$ and that by evaluating $A^{-1}(t) \Delta Q \Delta^{-1} t$, the effects of technical change may be separated so as to isolate the effects of the other factors. By applying this procedure he creates a numerical mapping of the function $\dot{q} = f(k, 1)$. In order to provide a mathematical description of the mapping, he (1957, p.318) observes, "As for fitting a curve to the scatter...I can't help feeling that little or nothing hangs on the choice of functional form, but I have experimented with several. In general I limited myself to two-parameter families of curves, linear in the parameters (for computational convenience), and at least capable of exhibiting diminishing returns (except for the straight line, which on this account proved inferior to all others)." The curves he fitted are plotted in Figure 4, over the capital range of zero to four, together with the matching transient relationship. The perimeter of Solow's chart 4, is indicated so that the range of the data may be seen in relation to the functional domain necessary for theoretical validity.

Despite starting (1957, p.319) "Thus any conclusions extending beyond the range actually observed in Chart 4 are necessarily treacherous", he expresses the full extent of the functional mapping verbally, describing the curves shown explicitly in Figure 4 with:

> "The particular possibilities tried were the following... Of these, (4d) is the Cobb-Douglas case; (4c and e) have upper asymptotes; the semi-logarithmic (4b) and the hyperbolic (4c) must cross the horizontal axis at a positive value of k and continue ever more steeply but irrelevantly downward (which means only that some positive k must be achieved before any output is forthcoming, but this is far outside the range of observation); (4e) begins at the origin with a phase of increasing returns and ends with a phase of diminishing returns— the point of inflection occurs at $k = \beta/2$ and needless to say all our observed points come well to the right of this" (Solow, 1957, p. 318).

The most notable omission from his list of fitted equations is equation (4a). It is the *linear relationship*, the only one having a *positive intercept* with $k = 0$ and therefore the *only possible contender* which *might have theoretical validity*.

The most notable assertion in this list is "*means only that some positive k must be achieved before any output is forthcoming*", appealing to *transient concepts in order to justify behaviour impossible in his universe of discourse!*

Figure 4 Solow's Chart 4 with the fitted curves and the transient approximation
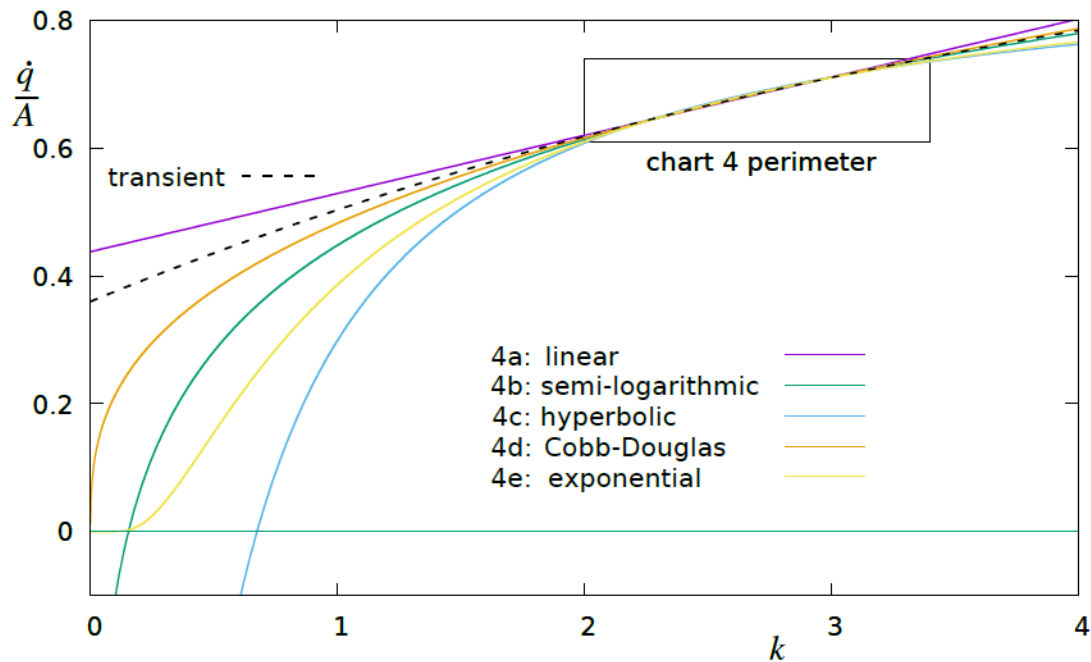
In responding to Hogan (1958), Solow (1958, p.412) notes "I can always choose q-units so that $A(t) = 1$ in some specified initial year" but he does not expand further nor discuss the implication that this assertion reduces his presentation of *the aggregate production function* to be merely a single possibility, arbitrarily selected, from the *forty-one alternatives* available!

The implications of this multiplicity are to be seen in Figure 5. $A(t)$ is set equal to one and applied to each of four years across the full range of the data. The years selected are: 1909, the first year of the data and the year Solow presents as *the aggregate production function*; 1929, the capital stock price basis of the data and the year for which he notes an increase in the rate of change of technical progress; 1939, the price basis for the GNP data and the start of World War II; 1949, the final year of the data. The mapped data points and linear fits to them are plotted over an extended *per capita* capital range, $-5 \leq k \leq 4$. The extension beyond the point of convergence allows recognition that the transformations are essentially affine. The mappings demonstrate that these manipulations create *four very different views* of the aggregate production function.

In replying to Hogan, Solow (1958, p.412) discusses the possibility that "a production function net of technical change... *wiggles and all*... will pass through the right points with the right outputs and with the right slope" but he does not appear to have carried out such an exercise. However, all is not lost. Transient analysis provides a method by which the mean aggregate production function for the data may be ascertained.

Figure 5 Solow's data – linear fits – $A(t) = 1$ for the years shown
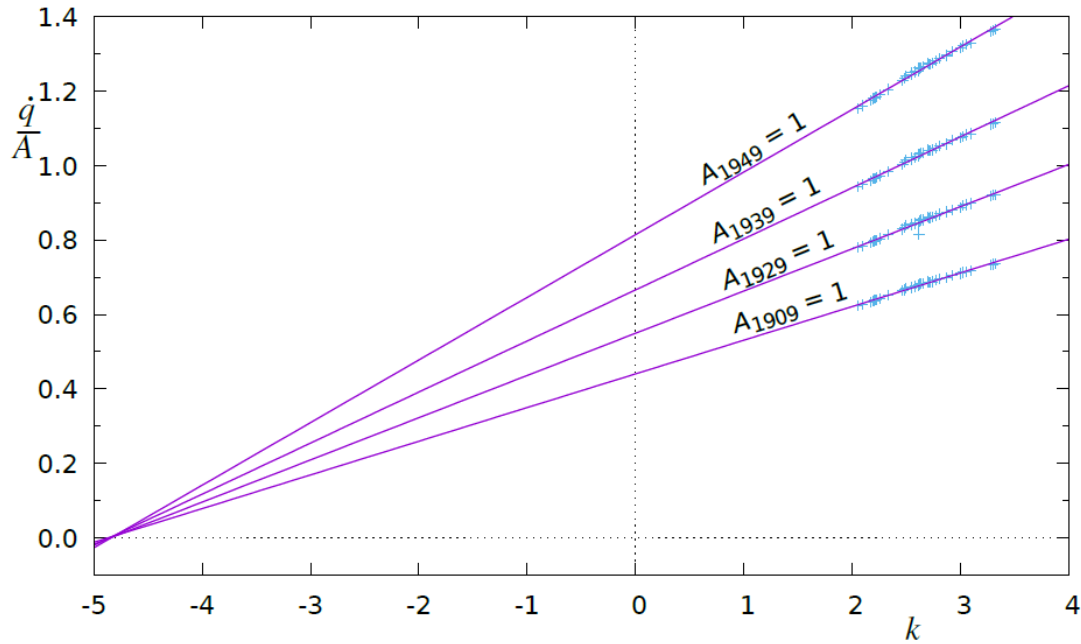
The output rate relationship used by Solow is $\dot{q} = A(t) f(k, l)$. Equation (29) expresses this relationship without technical change and equation (30) is the equivalent transient relationship.

$$\dot{q} = f(k, l) \qquad\qquad \dot{q} = f(0,1); k = 0 \qquad\qquad (29)$$

$$\dot{q} = (1 + a\eta(t))h_p \qquad\qquad \dot{q} = h_p; \quad t = 0 \qquad\qquad (30)$$

The value of the intercept with $k = 0$ is determined by the simultaneous solution of these relationships: the mappings of the transformed numerical data representing equation (29); and the particular forms of equation (30) which correspond to the mappings.

For each base year mapping, an initial estimate of h_p is obtained from the intersection of the linear fit with $k = 0$ and of a from the slope of the fitted line.[12] An initial value of m is set to some likely fraction of effort; say $m = 0.1$. These three parameters effect the position, slope and curvature of the transient function being mapped: h_p is the point of intersection with the y axes; a determines the slope at the intercept of the transient relationship and of the no-maintenance boundary; m determines the curvature and hence how the curve separates from the no-maintenance boundary. The parameters are adjusted iteratively so that the resulting curve matches the transformed data points.

Figure 6 shows the transient relationships, the no-maintenance boundary and the data points transformed from capital into time by implementing the above procedure for the four years plotted in Figure 5.

---

[12] Differentiating equation (8) with respect to time gives $\ddot{q} = ah_d h_p$ which rearranged is $a = \ddot{q}[1 - h_p)h_p]^{-1}$.

Figure 6 Solow data – instantaneous output rates as functions of time
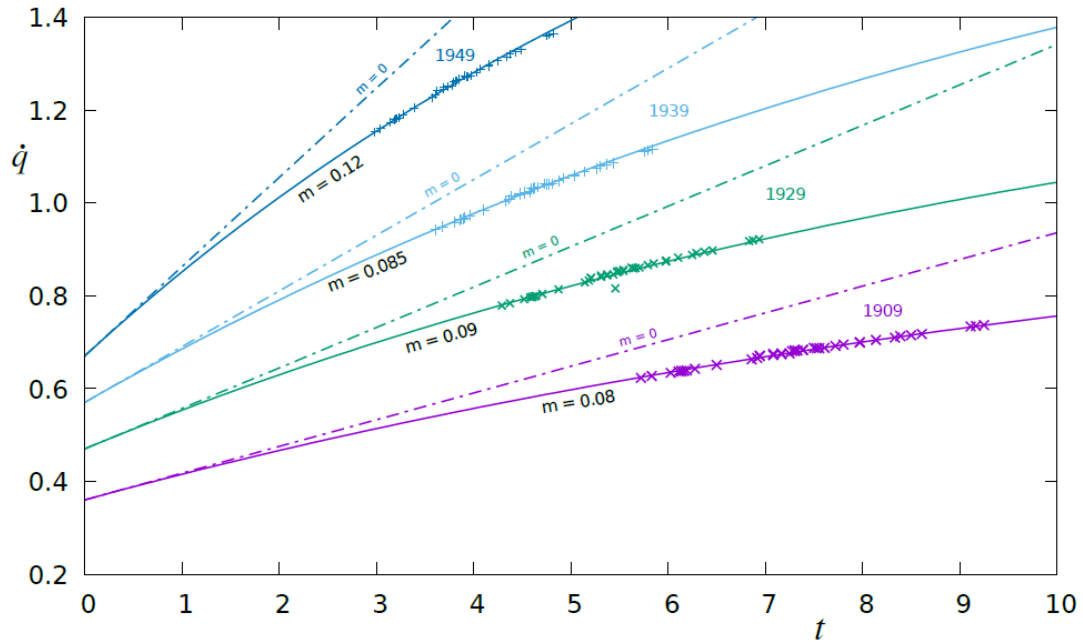
An estimate of the mean transient function describing the Solow data may now be obtained. Historical events determined the actual proportions of effort applied to tool-making and tool use. That manufacturing was competitive, over the period of the data, means that limiting values of h_d and h_p were achieved. Thus the mean relationship is to be found as the equation presenting consistent values for h_p, a and m. The best estimate would be achieved by: determining all forty-one sets of transformed data; finding, for each, the values of h_p, at the intercept and that of the limiting value evaluated from the estimates of a and m; fitting the two sets of estimates as functions of time; determining the intersection of the two which is the point with the same values for the intercept and the limit.

This procedure is demonstrated in principle using the four base years' data plotted in Figure (6) by reference to Table 1 which shows the values of the coefficients of equation (14) for the four years and the limiting values of h_p determined from a and m.

Table 1 Coefficients of equation (14) for the Figure 6 mappings

base year	intercept h_p	a	$\dfrac{a_{\text{year}}}{a_{1909}}$	m	limiting value $h_p = \frac{1}{2} + \frac{m}{2a}$
1909	0.36	0.25	1.00	0.08	0.66
1929	0.47	0.35	1.40	0.09	0.63
1939	0.57	0.49	1.96	0.085	0.59
1949	0.67	0.87	3.48	0.12	0.57

Inspection of Table 1 suggests that the transient equation providing an appropriate description of Solow's production data is

$$q = 0.58 \left[1 + \frac{0.49}{0.09}\overset{.}{(1 - 0.58)}(1 - e^{-0.09t}) \right] \tag{31}$$

4.6. First derivative of the productivity definition

Differentiating the productivity definition of equation (4), $\dot{q} = h_p p$, with respect to time gives

$$\ddot{q} = h_p \dot{p} \tag{32}$$

As a relationship derived solely from first principles, it has universal validity. Output rates increase as tools are brought into use (capital deepening) and productivity is increased. The relationship describes the changes occurring along the lines of constant h_d in Figures 2 and 3.

For competitive industries at any given level of technical progress, the limiting value of h_p will be approached; the $t \to \infty$ limit of equation (22), tends to a constant $\ddot{q} : \dot{p}$ ratio. For industries with advanced technical capability, the ratio will move closer towards its ultimate limiting value of one half. Different industries with their own levels of technical progress and maintenance, will, for this ratio, exhibit distinctive values.

Empirical measurements will demonstrate this simple relationship with extremely high levels of statistical significance. Non-competitive industries and those with little use for tools will not exhibit such a constancy of ratio.

In the introduction to the conference proceedings, *Productivity Growth and Economic Performance: Essays on Verdoorn's Law*, the editors state

> "Verdoorn's law, in its simplest form, refers to a statistical relationship between the long-run rate of growth of labour productivity and the rate of growth of output, usually for the industrial sector. The term 'Verdoorn coefficient' denotes the regression coefficient between the two variables" (McCombie, Pugno and Soro, 2003, p. 1).

Taking this assertion as a *de facto* definition, then the *Verdoorn coefficient* to which they refer must be either h_p or h_p^{-1}. Since Verdoorn (2002) first noted the empirical relationship, many efforts have been made to explain the underlying mechanisms. Since attempts to express transient physical behaviour in terms of equilibria, are doomed to inevitable failure, specific difficulties arise in comparing published empirical data with the present predictions. The more significant of these are examined in appendix C, "The Verdoorn relationship", where conventional hypotheses and empirical determinations appear to conflate the two interpretations of the Verdoorn coefficient.

Angeriz, McCombie and Roberts (2008, p.64) note that "the dynamic law can be derived directly from the static law by differentiating with respect to time", thereby identifying the static law as the productivity definition and the dynamic law as its first derivative. Traditional

reporting is frequently in terms of *returns to scale*.[13] Measurements detecting the productivity defining equation (4) are interpreted as *constant returns to scale* and those detecting equation (32) as *increasing returns to scale*. Paradox is perceived when analysis of the data detects both relationships and the expectation is of there being only one

.

For 118 firms between 1983 and 2002, Hartigh, Langerak and Zegveld (2009) examine the "Verdoorn Law" relationship for a number of industries. They find that, for the majority of the firms and industries examined, the relationships have statistical significance levels better than 0.001. Wide variations in the values of the coefficient are observed and are consistent with the predictions of the present analysis.

4.7. The Verdoorn coefficient and the intercept of the production function

The proportion of effort directly committed to producing the final output is h_p. It occurs in all quantitative descriptions of the production process and their derivatives. Its ubiquity provides a critical test by which transient analysis might be falsified. Values present empirically should, for competitive industries, sharing the same technologies, be close to the limiting values. For industries in which productivity increase is not from the use of tools and for non-competitive industries, the values may differ, without the competitive pressures driving them to convergence.

The value derived from the Solow data is $h_p = 0.58$ (Table 1 and equation (31)). Table 2 presents published values for the $\ddot{q} : \dot{p}$ ratio. All of these lie close to the 0.58 value. In his Table 1, Verdoorn (2002) presents, for the USA, three different values for the $\ddot{q} : \dot{p}$ ratio: from 1869 to 1899 as 0.42; from 1899 to 1939 as 0.57; from 1924 to 1939 as 1.67. The value for the USA between 1924 and 1939 lies within the period of the Solow data. The values (Verdoorn, 2002, p.29. Table 2.1) of 0.6 for the percentage increases in annual production and 1.0 for the percentage increases in productivity represent $h_p = 0.6$.

For competitive industries, while data scatter and the many equation forms used for estimates of the $\ddot{q} : \dot{p}$ ratio demonstrate a range of values, central values lie close to that obtained, from the Solow data. For non-competitive industries the reported values (Hartigh, Langerak and Zegveld, 2009) for the ratio are not driven towards limiting values of h_p and do not match them. That the $\ddot{q} : \dot{p}$ ratio and the proportion of effort dedicated to production, h_p, are the same variable, is not falsified by the empirical evidence.

---

[13] After proving that general statements of the production function imply constant returns to scale and perceiving it as an *imbroglio*, Georgescu-Roegen (1970, p.9) concludes "once we have untangled the imbroglio hatched by blind symbolism. The economics of production, its elementary nature notwithstanding, is not a domain where one runs no risk of committing some respectable errors. In fact, the history of every science, including that of economics, teaches us that the elementary is the hotbed of the errors that count most".

Kaldor (1972) discusses the scale effects of equipment dimension on performance and the difficulties in establishing appropriate mathematical representation of physical systems. While scale effects are important in equipment design, they are subsumed into mathematical representation as technical progress. This is not how scale is understood in current economic analysis. Kaldor (1972, p.1255) ends "The problem then becomes not just one of 'solving the mathematical difficulties' resulting from discontinuities but the much broader one of replacing the 'equilibrium approach' with some, as yet unexplored, alternative that makes use of a different conceptual framework".

Transient analysis, demonstrating returns to scale are constant, resolves the imbroglio and provides such a "different conceptual framework".

Table 2 Published values of the ratio $\ddot{q} : \dot{p}$

Author(s)	Area	Period	$\ddot{q} : \dot{p}$
Verdoorn (2002, p.29)	USA[‡]	1924 to 1939[‡]	0.6[‡]
Fingleton and McCombie (1998, p.92)	Europe	1979 to 1989	0.575
Vergeer and Kleinknecht (2007, p.35)	OECD	1960 to 2004	0.55
Hartigh, Langerak and Zegveld (2009, p.363)	Holland	1983 to 2002	0.672
Angeriz, McCombie and Roberts (2008, p.78)	Europe	1986 to 2002	0.579
Pons-Novell and Viladecans-Marsal (1999, p.448)	Europe	1984 to 1992	0.560, 0.587, 0.628

[‡] Within the period of the Solow data

To summarise, the predictions of transient analysis are consistent with published empirical data. Critical tests of limiting values are confirmed. They are consistent with the underlying identity of the coefficients.

5. Conclusion

Transient mathematical relationships, derived from *first principles*, provide a *parsimonious quantitative description of the development histories of manufacturing projects and industries*.

Appendices

A. Labour-time

The use of an alternate representation for human effort is introduced in sub-section 2.1, "The defining relationships", item 2b. By rewriting equations (2) and (3) in terms of a single labour-time variable, H, the two equations and equation (35) become respectively

$$\frac{dq}{dH_p} = p, \qquad \frac{dp}{dH_d} = a, \qquad \frac{\partial^2 q}{\partial H_d \partial H_p} = a.$$

Integrating the central relationship and substituting the result into the leftmost form, produces a relationship analogous to equations (7) and (13),

$$p = 1 + aH_d$$

and therefore the total quantity of output is

$$q = H_p(1 + aH_d). \tag{33}$$

The total quantity of output produced over time period, Δt, is $\sum \Delta q_t$. The overall mean output rate at time t is therefore $(\sum \Delta q_t) \div \Delta t$. Empirical data is available in this form for extended periods of time, providing empirical estimates of the instantaneous output rate found by differentiating equation (33) with respect to time,

$$\dot{q} = \frac{\mathrm{d}}{\mathrm{d}t}[H_p(1 + aH_d)]. \tag{34}$$

Evaluation of equation (34) requires the use of the product rule. The derivative of H_p is h_p. Only if H_d can be expressed as a function of time, is it possible to find its derivative. Without knowing this relationship, the problem is intractable. The relationship *is known* from equation (12), *derived from an analysis separating* the h and t parameters, and thereby rendering parameterisation into H_d and H_p *irrelevant*.

An interesting corollary presents itself. By comparing equation (34) with $\dot{q} = f(a, l, k)$, the ubiquitous general statement of production functions in conventional analysis, leads to the singular conclusion that H_d is equivalent to k.

Conventional economic analysis implicitly asserts the equality $k \equiv H_d$ and thereby declares *capital to be labour-time*! The neoclassical production function and the labour theory of value are tautologies.

B. Dimensional analysis

The units of the present analysis are the *natural units* of the problem space and so subsume dimensional analysis. Equations (2) and (3) are axioms of the analysis. Substituting p from equation (2) into equation (3) gives

$$a = \frac{\partial^2 p}{\partial h_d \partial t} = \frac{\partial^2}{\partial h_d \partial t}\left(\frac{\partial^2 q}{\partial h_p \partial t}\right) = \frac{\partial^4 q}{\partial h_d \partial h_p \partial t^2}. \tag{35}$$

Dividing equation (35) by a establishes

$$\frac{\partial^4 q}{a \partial h_d \partial h_p \partial t^2} = 1 \tag{36}$$

with units $[QA^{-1}H^{-2}T^{-2}]=1$; 1 is the identity element of dimensionless groups.

The general equation $f(q, h_d, h_p, t, a) = 0$ may be used to specify a project's transient development history; as $h_d + h_p = 1$, h_d and h_p represent a single variable. With four independent variables Buckingham's Π theorem states that the relationship may be represented by a single dimensionless number Π; such that $F(\Pi) = 0$ where $\Pi = q^{i_q} h^{i_d} h^{i_p} t^{i_t} a^{i_a}$ and i_q, i_d, i_p, i_t, i_a are integer. By inspection of equation (36), dimensionless groups of the form $\left[\frac{q}{ah_d h_p t^2}\right] = c$ are appropriate, where c is an arbitrary constant whose value is determined from actual data. Buckingham's Π theorem does not provide a single dimensionless parameter, but a set from which suitable forms may be selected.

Equations (4), (14) and (23) are derived rigorously from first principles and so provide alternate appropriate representations of dimensionless groupings. Dividing each equation by their right hand sides gives three groups[14]

$$\left(\frac{\dot{q}}{h_p p}\right) = 1, \qquad \left[\frac{\dot{q}}{h_p(p_0 + a\eta(t))}\right] = 1, \qquad \frac{a}{(1 - mt_k)(p_0 + at_k)(t - t_k)} = 1. \qquad (37)$$

The initial value of productivity, p_0, is presented to emphasise the unit rather than specific values. In the right hand group $(1 - mt_k) \equiv h_p; (h_d + h_p = 1, \ h_d \equiv mt_k)$.

The leftmost statement of equations (37) is the productivity definition of equation (4); the central is a statement of the instantaneous relationship of equation (14); the rightmost is the overall mean output relationship of equation (24).

C. The Verdoorn relationship

McCombie, Pugno and Soro (2003, p.5) observe "The story of the theoretical explanations of Verdoorn's Law is even more complicated, and far more open to debate". Fingleton (2001, p.7) describes a number of possible explanations of the relationship and states "Verdoorn's Law appears to be consistent with different theoretical positions or with different underlying technical relationships".

Without valid theoretical understanding, many alternative relationships, frequently without precise definition of the concepts being examined, hypothesise mechanisms to explain the empirical data. One common and the simplest statement (Kaldor (1975, p.891), McCombie and Ridder (1984, p.269), Fingleton and McCombie (1998, p.80), Angeriz, McCombie and Roberts (2008, p.65) etc.) is of the form

$$\dot{p} = \alpha + \beta\ddot{q} \qquad (38)$$

where α is described as the exogenous rate of productivity increase and β as the Verdoorn coefficient.

Two possible interpretations of equation (38) and their implications are:

1. if the statement is a functional definition, then it is incomplete. Specifications of the domain and codomain are required.
2. if the statement is a differential equation, then integrating over the time interval [0, t], gives

$$p_t - p_0 = \alpha t + \beta(\dot{q}_t - \dot{q}_0).$$

Setting $\dot{q}_t = \dot{q}_0$ gives $p_t = p_0 + \alpha t$. If $\alpha \neq 0$ then, for no obvious technical reason, productivity increases indefinitely with time (implied by the description exogenous rate of productivity increase). Only if $\alpha = 0$ can equation (38) have any theoretical validity.

Both interpretations lead to *reductio ad absurdum* positions which invalidate any claim to theoretical validity for equation (38).

---

[14] A further dimensionless group may also be derived from equation (16).

Setting $\alpha = 0$ in equation (38) and comparing the result with equation (32) demonstrates that $\beta \equiv h_p^{-1}$, thereby revealing a further problem. By definition $0 < h_p \leq 1$, therefore $\beta \geq 1$ but some publications present theoretical relationships incorporating equation (38) but report values in the range $0 < \beta \leq 1$ (Kaldor (1975, p.891) requires only that $\beta > 0$). Thirlwall (1983, p.350) observes "*Productivity growth has exceeded output growth in every country*" (no emphasis in the original). This statement and values, for the Verdoorn coefficient being less than one, are *mutually exclusive*, which implies a mismatch between theoretical description and empirical observation. Suggesting that empirical values, reported as being less than one, are arrived at by the following. Productivity and output rates are determined separately (frequently using regression techniques to provide an averaging procedure). The ratio is then reported conventionally, presenting fractional values rather than values greater than one. Conformance to the mathematical relationships presented requires $\beta > 1$.

References

Allen, R. C. (2012). 'Technology and the great divergence: global economic development since 1820'. In: *Explorations in Economic History* 49(1), pp. 1–16. url:
http://www.sciencedirect.com/science/article/pii/S0014498311000416

Angeriz, A., J. S. L. McCombie and M. Roberts (2008). 'New estimates of returns to scaleand spatial spillovers for EU Regional manufacturing, 1986–2002'. In: *International Regional Science Review* 31(1), pp. 62–87.

Badunenko, O. and V. Zelenyuk (2004). *Technological Change, Technological Catch-up, and Capital Deepening: Relative Contributions to Growth and Convergence During 90's; a Comment*. Univ. of Leuven, Department of Economics.

Barnett, W. (2004). 'Dimensions and economics: some problems'. In: *Quarterly Journal of Austrian Economics* 7(1), pp. 95–104.

Bénétrix, A. S., K. H. O'Rourke and J. G. Williamson (2015). 'The spread of manufacturing to the poor periphery 1870–2007'. In: *Open Economies Review* 26(1), pp. 1–37. url:
http://dx.doi.org/10.1007/s11079-014-9324-x.

Bigo, V. and I. Negru (2014). 'Mathematical modelling in the wake of the crisis: a blessing or a curse? What does the economics profession say?' In: *Cambridge Journal of Economics* 38(2), pp. 329–347.

Blaug, M. (1998). 'Disturbing Currents in Modern Economics'. In: *Challenge* 41(3), pp. 11–34. eprint:
http://dx.doi.org/10.1080/05775132.1998.11472030
url: http://dx.doi.org/10.1080/05775132.1998.11472030

Boland, L. A. (2005). 'Economics in time vs time in economics: building models so that time matters'. In: *History of Economic Ideas* 13(1), p. 121.

Bureau International des Poids et Mesures (2006: updated 2014). *The International System of Units (SI)*. 8th. International Committee for Weights and Measures.

Byrne, R. W. (2004). 'The manual skills and cognition that lie behind hominid tool use'. In: *The Evolution of Thought. Evolutionary Origins of Great Ape Intelligence*. Ed. by A. E. Russon and D. R. Begun. Cambridge University Press, pp. 31–44.

Cohen, A. J. and G. C. Harcourt (2003). 'Retrospectives: Whatever happened to the Cam bridge capital theory controversies?' In: *The Journal of Economic Perspectives* 17(1), pp. 199–214.

Crafts, N. (2012). 'British relative economic decline revisited: the role of competition'. In: *Explorations in Economic History* 49(1), pp. 17–29.
url: http://www.sciencedirect.com/science/article/pii/S0014498311000295

D'Adda, C. and A. E. Scorcu (2003). 'On the time stability of the output-capital ratio'. In: *Economic Modelling* 20(6), pp. 1175–1189.
url: http://www.sciencedirect.com/science/article/pii/S0264999302000810.

Felipe, J. and J. S. L. McCombie (2014). 'The Aggregate Production Function: "Not Even Wrong"'. In: *Review of Political Economy* 26(1), pp. 60–84.

Felipe, J. and F. M. Fisher (2003). 'Aggregation in production functions: what applied economists should know'. In: *Metroeconomica* 54(2-3), pp. 208–262.

Felipe, J. and J. S. L. McCombie (2013). *The Aggregate Production Function and the Measurement of Technical Change*. Edward Elgar Publishing.

Fingleton, B. and J. S. L. McCombie (1998). 'Increasing returns and economic growth: some evidence for manufacturing from the European Union regions'. In: *Oxford Economic Papers* 50(1), pp. 89–105.

Fingleton, B. (2001). 'Equilibrium and economic growth: spatial econometric models and simulations'. In: *Journal of Regional Science* 41(1), pp. 117–147.

Georgescu-Roegen, N. (1970). 'The economics of production'. In: *American Economic Review* 60(2), pp. 1–9.

Hahn, F. (1991). 'The Next Hundred Years'. In: The Economic Journal 101(404), pp. 47–50.
url: http://www.jstor.org/stable/2233836

Hardin, G. (1968). 'The tragedy of the commons'. In: *Science* 162(3859), pp. 1243–1248.

Hartigh, E. den, F. Langerak and M. Zegveld (2009). 'The Verdoorn law, firm strategy and, firm performance'. In: *Value-added Partnering and Innovation in a Changing World*. Ed. by M. van Geeuhaizen et al. Purdue University Press. Chap. 21, pp. 358–372.

Hogan, W. P. (1958). 'Technical progress and production functions'. In: *The Review of Economics and Statistics* 40, pp. 407–411.

Humphrey, T. M. (1997). 'Algebraic production functions and their uses before Cobb-Douglas'. In: *Economic Quarterly-Federal Reserve Bank of Richmond* 83, pp. 51–83.

Jensen, B. (2016). 'von Thünen: Capital, Production Functions, Marginal Productivity Wages, and the Natural Wage'. In: *German Economic Review*, pp. 1–30.

Jones, C. I. and P. M. Romer (2010). 'The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital'. In: *American Economic Journal: Macroeconomics* 2(1).
url: http://www.ingentaconnect.com/content/aea/aejma/2010/00000002/00000001/art00009

Kaldor, N. (1961). 'Capital accumulation and economic growth'. In: *The Theory of Capital: Proceedings of a Conference held by the International Economic Association*. Ed. by D. C. Hague. London: Macmillan & Co Ltd. Chap. 10, pp. 177–222.

Kaldor, N. (1972). 'The irrelevance of equilibrium economics'. In: *The Economic Journal* 82(328), pp. 1237–1255. url: http://www.jstor.org/stable/2231304

Kaldor, N. (1975). 'Economic growth and the Verdoorn law–a comment on Mr Rowthorn's article'. In: *The Economic Journal* 85(340), pp. 891–896.

Koop, G., J. Osiewalski and M. F. J. Steel (1999). 'The Components of Output Growth: A Stochastic Frontier Analysis'. In: *Oxford Bulletin of Economics and Statistics* 61(4), pp. 455–487. url: http://dx.doi.org/10.1111/1468-0084.00139.

Koop, G., J. Osiewalski and M. F. J. Steel (2000). 'Modeling the Sources of Output Growth in a Panel of Countries'. In: *Journal of Business & Economic Statistics* 18(3), pp. 284–299.
eprint: http://www.tandfonline.com/doi/pdf/10.1080/07350015.2000.10524871
url: http://www.tandfonline.com/doi/abs/10.1080/07350015.2000.10524871

Kumar, S. and R. R. Russell (2002). 'Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence'. In: *American Economic Review*, pp. 527–548.

Lavoie, M. (2008). 'Neoclassical empirical evidence on employment and production laws as artefact'. In: *Economía Informa* 351, pp. 9–36.

Lazzarini, A. (2011). *Revisiting the Cambridge Capital Theory Controversies: A Historical and Analytical Study.* Pavia University Press.

Madsen, J. B., V. Mishra and R. Smyth (2012). 'Is the output-capital ratio constant in the very long run?' In: *The Manchester School* 80(2), pp. 210–236.

Makieła, K. (2014). 'Bayesian Stochastic Frontier Analysis of Economic Growth and Productivity Change in the EU, USA, Japan and Switzerland'. In: *Central European Journal of Economic Modelling and Econometrics* 6(3), pp. 193–216.

Mayumi, K. and M. Giampietro (2010). 'Dimensions and logarithmic function in economics: a short critical analysis'. In: *Ecological Economics* 69(8), pp. 1604–1609.

McCombie, J. S. L. and J. R. de Ridder (1984). '"The Verdoorn law controversy": some new empirical evidence using U.S. state data'. In: *Oxford Economic Papers* 36(2), pp. 268–284.
eprint: http://oep.oxfordjournals.org/content/36/2/268.full.pdf+html
url: http://oep.oxfordjournals.org/content/36/2/268.short

McCombie, J. S. L. (2000). 'The Solow residual, technical change, and aggregate production functions'. In: *Journal of Post Keynesian Economics*, pp. 267–297.

McCombie, J. S. L., M. Pugno and B. Soro (2003). *Productivity Growth and Economic Performance: Essays on Verdoorn's Law*. Palgrave Macmillan.

Mishra, S. K. (2010). 'A brief history of production functions'. In: *The IUP Journal of Managerial Economics* VIII(4), pp. 6–34.

North, D. C. (1994). 'Economic Performance Through Time'. In: *The American Economic Review* 84(3), pp. 359–368. url: http://www.jstor.org/stable/2118057

Pons-Novell, J. and E. Viladecans-Marsal (1999). 'Kaldor's laws and spatial dependence: evidence for the European regions'. In: *Regional Studies* 33(5), pp. 443–451.

Robinson, J. (1953). 'The production function and the theory of capital'. In: *The Review of Economic Studies* 21(2), pp. 81–106.

Robinson, J. (1971). 'Capital theory up to date: a reply'. In: *The Canadian Journal of Economics/Revue Canadienne d'Economique* 4(2), pp. 254–256.

Robinson, J. (1980). 'Time in economic theory'. In: *Kyklos* 33(2), pp. 219–229.

Rodrik, D. (2013). 'Unconditional convergence in manufacturing'. In: The *Quarterly Journal of Economics* 128(1), pp. 165–204.
eprint: http://qje.oxfordjournals.org/content/early/2013/01/21/qje.qjs047.full.pdf+html.
url: http://qje.oxfordjournals.org/content/early/2013/01/21/qje.qjs047.abstract

Salas, S. L., E. Hille and G. J. Etgen (2007). *Calculus: One and Several Variables*. 10th. Wiley New York.

Setterfield, M. (1995). 'Historical time and economic theory'. In: *Review of Political Economy* 7(1), pp. 1–27.

Setterfield, M. and S. G. Suresh (2014). 'Aggregate structural macroeconomic analysis: a reconsideration and defence'. In: *Cambridge Journal of Economics*, pp. 797–815. Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Project

Gutenberg (2013). url: https://www.gutenberg.org/ebooks/3300

Solow, R. M. (1955). 'The production function and the theory of capital'. In: *The Review of Economic Studies* 23(2), pp. 101–108.

Solow, R. M. (1957). 'Technical change and the aggregate production function'. In: *The Review of Economics and Statistics* 39(3), pp. 312–320.

Solow, R. M. (1958). 'Technical progress and production functions: reply'. In: *The Review of Economics and Statistics* 40(4), pp. 411–413. url: http://www.jstor.org/stable/1926346.

Solow, R. M. (1994). 'Perspectives on growth theory'. In: *The Journal of Economic Perspectives* 8(1), pp. 45–54. url: http://www.jstor.org/stable/2138150.

Thirlwall, A. (1983). 'A plain man's guide to Kaldor's growth laws'. In: *Journal of Post Keynesian Economics* 5(3), pp. 345–358. url: http://www.jstor.org/stable/4537750

Verdoorn, P. J. (2002). 'Factors that determine the growth of labour productivity'. In: *Productivity Growth and Economic Performance*. Ed. by J. S. McCombie, M. Pugno and B. Soro. Basingstoke: Palgrave Macmillan, pp. 28–36.

Vergeer, R. and A. Kleinknecht (2007). 'Jobs versus productivity? The causal link from wages to labour productivity growth'. In: *TU Delft Innovation Systems Discussion Papers*, IS 1.

Zamagni, S. and E. Agliardi (2004). *Time in Economic Theory*. 3 vols. Cheltenham: Edward Elgar Publishing Limited.

Zellner, A. and H. Ryu (1998). 'Alternative functional forms for production, cost and returns to scale functions'. In: *Journal of Applied Econometrics* 13(2), pp. 101–127.

Author contact: fms-econ@virginmedia.com

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-81/

_____

Colignatus, M. Shahid Alam, Bryant Chen, Judea Pearl, John Pullen, Tom Mayer, Thomas Oechsle, Emmanuel Saez, Joseph Noko, Joseph Huber, Hubert Buch-Hansen, Brendan Sheehan, C P Chandrasekhar, Heikki Patomäki, Romar Correa, Piet-Hein van Eeghen, Max Koch, John Robinson, Oscar Ugarteche, Taddese Mezgebo, Donald Katzner, Crelis F. Rammelt, Phillip Crisp,John B. Benedetto, Alicia Puyana Mutis, Leon Podkaminer, Michael Kowalik, Mohammad Muaz Jalil, José A. Tapia, Robert W. Parenteau, Alan Harvey, C. T. Kurien, Ib Ravn, Tijo Salverda. Holger Apel, John Jeffrey Zink, Severin Reissl, Christian Flamant, Rainer Kattel, Amit Bhaduri, Kaustav Banerjee, Zahra Karimi Moughari, David R Richardson, Emil Urhammer, Michel Gueldry, Rüya Gökhan Koçer, Hee-Young Shin, Kevin Albertson, John Simister, Tony Syme, Geoff Tily, Ali Abdalla Ali, Alejandro Nadal, Steven Klees, Gary Flomenhoft, Bernard C. Beaudreau, William R. Neil, Ricardo Restrepo Echavarría, Carlos Vazquez, Karen Garzón Sherdek, Paul Spicker, Mouvement des étudiants pour la réforme de l'enseignement en économie, Suzanne Helburn, Martin Zerner, Tanweer Akram, Nelly P. Stromquist, Sashi Sivramkrishna, Ewa Anna Witkowska, Ken Zimmerman, Mariano Torras, C.P. Chandrasekhar, Thanos Skouras, Diego Weisman, Philip George, Stephanie Kelton, Luke Petach, Jørgen Nørgård, Jin Xue, Tim Di Muzio, Leonie Noble, Kazimierz Poznanski, Muhammad Iqbal Anjum, Pasquale Michael Sgro, June Sekera, Michael Joffe, Basil Al-Nakeeb, John F. Tomer, Adam Fforde, Paulo Gala, Jhean Camargo, Guilherme Magacho, Frank M. Salter

_____

Articles, comments on and proposals for should be sent to the editor at pae_news@btinternet.com
_____