THE SHARE OF THE TOP ONE PERCENT: IS IT DUE TO THE MARGINAL PRODUCT OF LABOUR OR FINANCIALISATION?

John S.L. McCombie
University of Cambridge

&

Marta R.M. Spreafico
Catholic University of Milan

CAMBRIDGE CENTRE FOR ECONOMIC AND PUBLIC POLICY

CCEPP WP03-16

DEPARTMENT OF LAND ECONOMY

UNIVERSITY OF CAMBRIDGE

OCTOBER 2016
THE SHARE OF THE TOP ONE PERCENT: IS IT DUE TO THE MARGINAL PRODUCT OF LABOUR OR FINANCIALISATION?

Abstract

The last thirty years or so have seen the rapid increase in the share of income of the top one percent, especially in the United States. This has led to increasing concern about the consequences of the degree of income and wealth inequality and whether or not policies should be introduced further to reduce it. However, for a long time neoclassical economics has ignored the problem, generally because of its uncritical acceptance that individuals are paid their marginal products in largely competitive markets. It is a short step from this to John Bates Clark’s normative argument that this is what they should receive, a view espoused recently by Mankiw (2013) together with the ‘just deserts’ ethical argument. Nevertheless, it is shown that the marginal productivity theory is deeply flawed empirically and cannot, as a matter of logic, be substantiated theoretically. The rise in the share of the top one percent is largely due to the increase in chief executive officers’ salaries which was due to institutional factors, notably the role of remuneration committees and the widespread adoption of stock options in the 1990s. These were introduced in a mistaken belief that they would overcome the principal-agent problem. There was also a rapid change in the economic milieu with the rise of financialisation, defined broadly to include the increasing role of financial markets. The neoclassical approach is of limited use in explaining these phenomena.

Key-words: financialisation, inequality, income distribution, CEO pay, marginal productivity, aggregate production function.
1. Introduction

The degree of income inequality, in both the developed and developing countries, despite being neglected for many years, has now become of increasing importance and a major policy issue (OECD, 2011; Asian Development Bank; 2012; Oxfam, 2013). This has been due not only to considerations of the adverse social consequences, and issues of equity, that inequality raises, but also to the recent findings that greater inequality harms growth (Berg et al., 2011, Ostry et al., 2014, Cingano, 2014) and through rising household debt has short-term harmful consequences for the level of aggregate demand (Cynamon et al., 2013).

The questions posed by inequality have been highlighted in a number of recent books, inter alios, by Stiglitz (2012), Deaton (2013), Picketty (2014), and Atkinson (2015). Undoubtedly, Picketty (2014) has attracted the most attention and controversy. One of the reasons for impact of Capital in the Twenty First Century is that it focuses on the dramatic increase in the share of the top one percent in income and wealth over the last thirty years, or so, in the US, the UK, Australia and Canada. The increase has been much less in the other advanced European countries. This focus on the change in share of the top one percent presents a much more dramatic picture than, say, a substantial change in the Gini coefficient.

The figures on this for the US are remarkable. The labour compensation of the top one percent over the period 1979 to 2007 accounted for 60 percent of the growth of market-based incomes (38 percent of post-tax incomes) (Bivens and Mishel, 2013). The income of the top one percent is largely driven by the earnings of the chief executive officers (CEOs), not only because they comprise a substantial proportion of the top one percent, but because there is a comparability effect on the salaries of the other top earners. In the US, over the period 1965 to 2013, the remuneration of the average CEOs’ annual income increased from just over $800,000 to $15.3 million in 2013 prices (Mishel and Davis, 2014). The ratio of the pay of the average CEO to that of the average worker was 20:1 in 1965; peaking at 383:1 in 2000 and is nearly 300:1 in 2013. In the UK, the FTSE 100 senior executives today earn 150 times that of their average employees; in 1998 the figure was about 50.

Consequently, two questions arise. What accounts for the rapid increase in the remuneration of the salaries of the CEOs? And secondly, is this a cause for concern? For a long time, there has been little interest amongst neoclassical economists about issues of inequality. As Milanovic (2013) dramatically put it: “Before the global crisis, income inequality was relegated to the underworld of economics. The motives of those who studied it were impugned. According to Martin Feldstein, the former head of Reagan’s Council of Economic Advisors, such people have been motivated by envy. Robert Lucas, a Nobel Prize winner, thought that ‘nothing [is] as poisonous to sound economics as ‘to focus on questions of distribution’ ‘. 
As far as we are aware, there is no discussion of inequality in any of the well-known microeconomic or macroeconomic textbooks. This is instructive, as, to paraphrase Kuhn (1970), if it is not mentioned in the textbooks, it cannot be important.

This neglect is largely because of the widespread acceptance of the proposition from the neoclassical theory of production and distribution (Ferguson, 1972) that individuals are paid their marginal products in markets that are generally competitive. Hence, both the incomes of individuals and the share of income going to labour are determined by technological factors underlying the production function. This has the great advantage that this is amenable to neoclassical mathematical modelling. There is no need to consider institutional factors (such as how, in reality, salaries are actually determined) or sociological explanations. Wages can simply be explained by a simple partial derivative. From here it is a short step to arguing normatively that their marginal product is what individuals deserve to be paid, as it represents their contribution to output.

In this paper, we cover some well-travelled, and some not so well-travelled, arguments as to why the marginal productivity theory of distribution is deeply flawed. We then consider the questions, if this is the case, how are the salaries of CEOs actually determined and how do we explain their rapid increase in recent years?

We begin with a consideration of the neoclassical argument that there is no need to be concerned about inequality because their remuneration is largely determined by market forces. This has recently been forcefully argued by Mankiw (2013). Picketty (2014), however, is skeptical of the relevance of the marginal productivity theory for determining the salaries of the top one percent. Nevertheless, he still considers that it is applicable to the determination of the pay of those undertaking “replicable” work. We show, by means of a simple example, that this still concedes too much to the marginal productivity theory and which confounds his argument. We show that problem of testing that an individual is paid his or her marginal product faces insurmountable problems. Indeed, even a few moments thought would suggest that is difficult to get out of the circularity of the neoclassical argument that there is severe competition for the best CEOs, who are paid according to the contribution they make to output. This is none other than their marginal products, which in turn are reflected in the salaries they receive.

However, because of the problems of determining whether an individual is paid his or her marginal product, neoclassical economics, instead, uses an indirect test. Production theory shows that if markets are perfectly competitive, then a test for whether or not a homogeneous factor of production is paid its marginal product is whether or not, when an aggregate production functions is statistically estimated, the estimated relevant output elasticity equals its factor share. As in many cases this proves to be the case, it is inferred that the payments to all categories of labour, including individuals, are paid their marginal products.
As we show, this argument founders on two problems.

The first is the theoretical one that, as shown by the Cambridge capital theory controversies and the more general aggregate problem, the well-behaved aggregate production function does not exist, even as an approximation. These are the well-established results of (mathematical) logic, yet the neoclassical tradition has been simply to ignore them. One reason that suggests itself is the adherence of neoclassical economists to Friedman’s (1953) methodology of positive economics. This considers that the realism, or otherwise, of the assumptions does not matter. The fact that aggregate production functions usually give good statistical fits to the data with their output elasticities equal to the factor shares is interpreted as evidence that labour in aggregate is paid its marginal productivity, despite the theoretical objections. From here, the implicit assumption is made that it shows that individuals are paid their marginal products. However, we show the problem that the empirical estimations of aggregate productions, because the data used is constant price value data and not physical measures as theory dictates, means that as a matter of logic, such estimates can never be taken as reflecting the underlying technology of the economy.

If this explanation is not tenable, what then determines the incomes of the top one percent and, in particular, their rapid increase over the last three decades or so? We next turn to a consideration of this starting with how CEOs pay is actually determined, namely through executive remuneration committees. This and the neoclassical explanation could not be more different. We examine the way the attempt to solve the principal-agent problem, caused by the divorce of ownership and control of corporations, through the use of stock options, has had completely the opposite effect. This combined with financialisation and the introduction of “share-holder value”, with changing social norms, explains much of the increase in compensation of the top one percent.

The last section concludes.

2. Why Should the Inequality of Income be a Matter of Concern? Aren’t CEOs Paid their Marginal Products?

The neoclassical standard explanation of how factors of production are rewarded developed from Ricardo’s model of distribution by applying the marginal principle to all factors of production and not just to land (Kaldor, 1955-6). Although, the early models concerned themselves with homogeneous labour, it is a small step to apply this at the microeconomic level to individuals. Consequently, in a nutshell, those workers with higher productivities earn higher incomes that reflect their greater contribution to society, and this is determined solely by the technical conditions of production and factors affecting the supply of labour. As John Bates Clark (1899, p.v) memorably wrote many years ago, “[i]t is the purpose of this work to show that the distribution of income to society is controlled by a natural law, and that this law, if it worked without friction, would give to every agent of production the amount of wealth which that agent
creates”. While John Bates Clark’s statement does not imply that this is what every agent ought to get, it is often implicitly assumed that this is the case (Mankiw, 2013). Moreover, the implication is that any attempt to alter the free market distribution of earnings will lead to a “great contradiction” as Okun (1977) termed it, namely a trade-off between equity and efficiency. As altering the distribution of income is likely to reduce the efficiency of the allocation of resources, it, therefore, comes at an economic cost.

A recent statement defending the present distribution along these lines, albeit with some minor qualifications, is that of Mankiw (2013). Mankiw clearly believes that in a competitive economy individuals are paid their marginal products. For example, in outlining what he sees as the criticism of what he describes as the “left”, he writes as follows “In the standard competitive labor market, a person’s earnings equals the value of his or her marginal product.” The normative implications of this are made explicit when he attempts to defend the earnings of the top one percent along the following lines of the ethical argument of “just deserts”. “If the economy were described by a classical competitive equilibrium without any externalities or public goods, then every individual would earn the value of his or her marginal product, and there would be no need for government to alter the resulting income distribution” (p.32).

Consequently, this is may be taken as the neoclassical benchmark. The key, Mankiw continues, is whether the earnings of the top one percent reflect their higher (marginal) productivity or represent the extraction of rents. Indeed, he concedes that if the increase in the share of the top one percent were attributable to successful rent seeking, then he would deplore it. (It is interesting that the example Mankiw gives is not the rise in the share of the one percent due to the change in the way CEOs are remunerated, but to the inequities of government policies in creating monopolies.) He asserts that on his own reading of the evidence the earnings of the top one percent, and their rapid growth over the last thirty years is due to their increased productivity.

The evidence Mankiw offers in support of this is not compelling. He invokes the superstar theory that “changes in technology have allowed a small number of highly educated and exceptionally talented individuals to command superstar incomes in ways that were not possible a generation ago” (p.13), citing Steve Jobs of Apple and the authoress J.K.Rowling. But as Dean Baker, inter alios, has pointed out, their large incomes are heavily dependent on institutions set up by governments in the form of patents, copyright monopolies and state expenditure on R&D (Mazzucato, 2013), all of which are the antithesis of the free market. Moreover, such huge salaries are not necessary to persuade individuals to make substantial contributions to society. Just think of the unsung heroes who developed the internet and indeed the role of the US government in facilitating it. Then there is the Genotype project which makes the results freely available to all, compared with the smaller project of the Celera Corporation, whose aim was to appropriate the private rents from advances in this area. One could go on almost indefinitely. Finally, the share of the top one percent is dominated by CEOs and the finance sector, not talented innovators.
The second line of reasoning is that Mankiw argues that the increase in the share is due to the “race between education and technology” (Goldin and Katz, 2008). This is the hypothesis that skill-biased technical change has increased the demand for skilled relative to unskilled labour and has led to a college premium. This, according to the hypothesis and which is Mankiw’s view, has led to rising income inequality, which is nothing to do with rent-seeking, but is simply the operation of supply and demand for labour. He argues that, while Goldin and Katz (2008) concentrate on the full distribution of income rather than the top one percent, “it is natural to suspect that similar forces are at work” and that they follow a similar U-shaped pattern. However, unfortunately for this thesis, the college premium flattened out in the 1990s, while the growth of the top one percent was very much faster and bears little resemblance to the path of the college premium. Moreover, the skill-based explanation cannot explain the fact that there has also been a rapid increase in the share of the top one percent in capital income (Mishel, 2013). We are not going to discuss the merits of this hypothesis of skill-biased technical change here, save to say that it is predicated upon the existence of a well-behaved CES production function, and the indirect measure of different types of technical change.¹

However, for neoclassical economists, the concept of the marginal product of labour and the necessary adjunct of the (aggregate) production function is taken as axiomatic. In the language of Lakatos, the latter is part of the “hard core” or, in Kuhnian terms, it is a paradigmatic heuristic or paradigmatic pseudo-assumption.² Its existence is taken for granted and is deemed untestable by fiat. Consequently, the mainstream view has been that income inequality is not a major issue. It merely reflects differences in the marginal productivities of labour. Moreover, the decline in labour’s aggregate share, which has been observed in many advanced countries, is explained solely in terms of the aggregate production function and the value of the elasticity of substitution, together with changes in the capital-output ratio, over which there has been a recent extensive debate.

3. On Piketty’s “Illusion of Marginal Productivity”

It is difficult to discuss wealth or income inequality without mentioning Picketty’s (2014) influential *Capital in the Twenty-First Century*, which emphasized the rapid growth in the share of the top one percent over the last thirty years or so, especially in the US. Piketty (2014a, pp. 330-

---

¹ See Solow (2014) for a criticism of some of the other arguments Mankiw makes.
² These are termed pseudo-assumptions because, in the natural sciences, they are a hybrid of analytic-synthetic, or quasi-analytic, statements. They are analytic because they are not falsifiable by fiat. They are taken as self-evident and demarcate the paradigm. But they are synthetic in that they may initially have been part of the empirical basis of the paradigm, but “they are by no means the product of arbitrary definitional stipulations. They are rather in part the products of painstaking empirical and theoretical research” (Hoyningen-Huene, 1993, p.210). While neoclassical economics views the production function as this, we shall show the interpretation of the empirical support for the aggregate production function is always erroneous.
333) is rightly extremely skeptical of the concept of marginal productivity as an explanation for the determination of wages and salaries of the top one percent. The hedge fund manager, for example, Paulson earned $3.7 billion dollars in 2007 (Rajan, 2010, p. 80). Was this his marginal product? How do we test this proposition? Should the marginal products of a handful of CEOs of the banks that precipitated the Great Recession be regarded as substantially negative over this period?

It is worth citing Piketty (2014a, pp. 330-331) at some length on these points.

To my mind, the most convincing explanation for the explosion of the very top US incomes is the following. As noted, the vast majority of top earners are senior managers of large firms. It is rather naïve to seek an objective basis for their high salaries in individual “productivity”. When a job is replicable, as in the case of an assembly-line worker or fast food server, we can give an approximate estimate of the “marginal product” that would be realized by adding one additional worker or waiter (albeit with a considerable margin of error in our estimate). But when an individual’s job functions are unique, or nearly so, then the margin of error is much greater. Indeed, once we introduce the hypothesis of imperfect competition into standard economic models (eminently justifiable in this context), the very “individual marginal productivity” becomes hard to define. In fact, it becomes something close to a pure ideological construct on the basis of which justification for higher status can be elaborated. (Emphasis added)

What is interesting here is that although Piketty dismisses the concept of marginal productivity for senior managers and executives, he still seems to consider that theoretically it can be measured for those doing “replicable” jobs, albeit imprecisely. This seems a somewhat contradictory position. As the top one percent took the vast majority of the increase in income over the last thirty years in the US and this had nothing to do with their marginal productivity (which, as Piketty notes, cannot be independently measured), how could the remainder of the labor force be paid their marginal products?

Nevertheless, it is a short step from Piketty’s statement to assuming that for these employees with replicable jobs, competitive markets will ensure that they are paid the contribution they make to the economy. However, while the evidence discussed below provides support for Piketty’s arguments regarding CEOs’ pay, we shall argue that even for replicable jobs, the marginal productivity theory, qua a theory, is logically problematical.

To show what, in retrospect, may be seen to be a straightforward point, let us, following Piketty, take the simple example of a small restaurant managed by the owner. The manager has no idea of the elasticity of demand for his meals, and so undertakes a mark-up pricing policy, a là Kalecki. (The work of Hall and Hitch (1939), many years ago, established this was the most common way of price setting in industry. See also Coutts, Godley and Nordhaus (1978) and Coutts and Norman (2013).) Prices are determined by a mark-up on the unit costs of labor (the waiters and chefs) and
the ingredients of the meals together with the other capital costs (rates, etc.). Consequently, total revenue is given by:

\[ p_M M \equiv R \equiv (1 + \pi)(wL + I) \]  

(1)

where \( p_M \) is the price of the meal \((M)\), \( R \) is total revenue, and \( I \) is the value of the ingredients. The operating profit is equal to \( \Pi \equiv \pi(wL + I) \). The mark-up is determined by the state of competition from other restaurants, the overall level of affluence in the local area and it is also influenced by a target for the level of profits. Nominal wages are assumed to be determined by the state of the local labor market. The contribution of value added of the restaurant to output as reported in the national income and product accounts (NIPA) will be given by:

\[ R - I \equiv Y \equiv wL + \Pi \equiv wL + \pi(wL + I) \]  

(2)

Suppose the restaurant is flourishing and the manager considers it desirable to hire a new waiter to speed up the service, but for sake of argument, the same number of meals is served. Under this pricing policy, the increase in value added \((Y)\) in adding an extra employee, from equation (2), is definitionally equal to \( \partial Y / \partial L = (1 + \pi)w \).

So if we interpret \( \partial Y / \partial L \) as the marginal product of labor, we can see that it is less than the wage rate. This is because the hiring of the extra waiter, through the pricing policy, automatically increases profits at the same time. Consequently, \( \Pi \) is not held constant as \( L \) changes and as the neoclassical marginal productivity theory assumes. Of course, if the manager merely passes on the increased labor cost in the form of an increased price of the meal then, from equation (2) and holding \( \Pi \) constant, by definition, \( \partial Y / \partial L \equiv w \). But this is not the result of optimization using a well-behaved production function subject to a cost constraint. In fact, changes in the local labor market conditions (such as an increase in the minimum wage) that affect the wage rate of the waiter will also cause his/her putative marginal productivity to change. But the causation runs from the wage rate to the putatively marginal productivity.\(^3\)

It should be noted that this applies to a firm that is selling a marketed product to the private sector. But what about the large (public) sector of the economy where there is no independent measure of aggregate output?

Much depends upon the way it is calculated. In the early national accounts, the output was just taken to be equal to the total labor compensation. In many cases, there are measures of physical outputs (such as the number of operations in hospitals, or trials in the judicial system, which can be

---

\(^3\) Note that if prices are determined by a mark-up on unit labor costs, labor’s share is given by \( 1/(1 + \pi) \). The mark up will be determined by the state of competition in both the product and the labor market.
used), but the problem still arises as to how to price or value them. Attempts in the UK have been made to revise the output measures of government services after the Atkinson Review (2005), but insurmountable problems remain for the testing of marginal productivity.

It should be noted that the accounting identity, \( Y \equiv wL + rK \), where \( Y \) is income, holds irrespective of the degree of state of competition, whether or not there are well-defined production functions and whether or not firms optimize. If this accounting identity is partially differentiated with respect to labor, we obtain \( \frac{\partial Y}{\partial L} = w \) and \( (\frac{\partial Y}{\partial L})(L/Y) = wL/Y = a \) where \( a \) is labor’s share. The expression \( (\frac{\partial Y}{\partial L})(L/Y) = \alpha \) is the neoclassical definition of labor’s output elasticity and, under neoclassical production theory, is equal to the wage share if there are perfectly competitive markets, a well-behaved aggregate production function and factors are paid their marginal products. But from the definition of the national accounts, \( \alpha \) must be definitionally equal to the wage share, \( a \). This led Phelps Brown (1957, p.557) to comment wryly that labor’s output elasticity of the production function and the wage share “will be only two sides of the same coin”. We shall return to the important implications of this below.

On a more pragmatic note, Thurow (1975, pp. 211-230) in his “A Do-it-Yourself Guide to Marginal Productivity” raises some further problems that occur even if output can be valued independently of the inputs. The questions Thurow raises include:

- “What is the time period over which marginal products are paid?”
- “Are groups or individuals paid their marginal products?”
- “What is the theory that determines whether marginal-productivity groups are large or small, heterogeneous or homogeneous?”
- “What is the level of aggregation at which capital and labor are paid their marginal products?”
- “To what degree does the economy fit the competitive model and to what degree does it fit the monopoly model? The actual economy is a mixture, but what are the relevant proportions?”

Other questions include the problems posed by disequilibrium, uncertainty, the presence of increasing returns to scale, whether governments can in principle ever pay their employees according to their marginal productivity and to what extent do psychic-income benefits influence monetary remuneration. As Adam Smith long ago pointed out, production is characterized by the division of labour. The decisions of, say, a CEO will be influenced by the quality of the decisions of his subordinates and indeed the outcome of different views. It makes little sense to try to identify the output of an individual in these and similar circumstances.
Clearly, even ignoring the problems of the measurement of the monetary value of output independently of the value of wages, there are many other insuperable difficulties noted by Thurow in the way of providing an adequate test of the marginal productivity theory. These concerns are shared, *inter alios*, by Stiglitz (2012).

**4. The Illusion of the Aggregate Production Function**

It is somewhat paradoxical that Piketty, in spite of his reservations about the marginal productivity theory in explaining the wage rate, nevertheless at times explains the changes in the shares going to capital and labour in terms of an aggregate CES production function.

This determination of the functional distribution of income between the two factors of production also involves the usual neoclassical assumptions underlying the marginal productivity theory of distribution. These are the existence of a well-behaved one sector aggregate production function, perfect competitive markets with all firms and service providers internally efficient (*i.e.* no $X$-efficiency), constant returns to scale and that factors are paid their marginal products. Piketty notes that over the last thirty years or so, capital’s share of income has risen while the ratio of capital to income has also increased. In terms of conventional neoclassical production theory, this change can simply be explained in terms of an aggregate production function where capital and labour are paid their marginal products and the elasticity of substitution is greater than unity.\(^4\)

Piketty (2014a, p. 232) starts with the simple explanation of the return to capital being its marginal product, but concedes that it also depends upon the relative bargaining power of the parties concerned. It may be higher or lower than the marginal product, especially “since this quantity is not always measurable”. Yet, at times, the main thrust of his argument rests on factor shares being determined by technology and the form of the aggregate production function. After discussing the effect of bargaining power on factor shares, this is soon ignored and Piketty is discussing the role of technology and the production function as an explanation for the changes in the functional distribution of income between capital and labour. “The relevant question is whether the elasticity between labor and capital is greater or less than one. If the elasticity [of substitution] lies between zero and one, then an increase in the capital/income ratio … leads to a decrease in the marginal product of capital large enough that the capital share decreases”. Alternatively, if the elasticity is greater than one, then an *increase* in the capital-output ratio leads to an increase in capital’s share. It is the increase, or fall in labour’s share that is observed empirically.

\[^4\] In terms of the aggregate CES production function with constant returns to scale and factors paid their marginal products, capital’s share equals \((1 - a) = \delta (K/Y)^{(\sigma - 1)/\sigma}\) where \(\delta\) is a constant and \(\sigma\) is the elasticity of substitution.
One of the difficulties with, and indeed criticisms of, Piketty’s approach and conventional economic analysis of shares, is that Piketty has a very wide definition of capital or wealth. He uses the terms interchangeably, including land, dwellings, commercial inventory, other buildings, infrastructure and financial assets (bank accounts, mutual funds, bonds, stocks, financial investments of all kinds, insurance policies, pension funds). The conventional production function has a much narrow definition of capital as including machines, buildings, etc., that contribute directly to production. As a number of authors point out, Piketty’s measure, by including capital appreciation, overstates the increase in the capital-output ratio, which may actually have fallen. Consequently, in this case, the increase in capital’s share will be caused by an elasticity of substitution of less than unity, which seems to be the case (Chrinko, 2008; Rowthorn, 2014). There is no role for changes in labour market polices, globalization, etc., to affect the functional distribution of income. It’s all down to technology. But is it?

4.1 The Cambridge Capital Theory Controversies and the Aggregation Problem

The concept of the marginal product of labour is predicated on the assumption that a well-behaved production function actually exists. There is, however, a fundamental question about whether or not this is correct. It is worth briefly discussing this which, although it was fully debated nearly fifty years ago and the problems largely confirmed, has been subject to what can be best termed collective amnesia by the majority of the economics profession. (The exceptions include the post Keynesian economist and heterodox economists. See, for example, Galbraith, 2014.)

In the 1950s and 1960s, what came to be known as Cambridge capital theory controversies took place, which, as its name suggests, centred around whether the theoretical concept of “capital” as a factor of production had any meaning outside the highly restrictive one-commodity world. The upshot was that the answer was “no”. This important debate between Cambridge, UK, and Cambridge, Massachusetts, has long been relegated to the history of economic thought, forgotten or treated as an esoteric debate in theory (Birner, 2002). In 1962, Samuelson published a paper where he purported to show that a one-commodity aggregate production function could be generalized to include more than one sector. In other words, his construct of the “surrogate production function” could be seen as what he terms a “parable” for a more complex technology than the single-sector “corn” model. The capital theory controversies, and they were entirely a matter of theory, proved that this construct was untenable.

Piketty seems to regard it as a debate over whether it is the capital-output ratio adjusts to allow balanced growth or there is some other mechanism, such as Kaldor’s (1955-6) model of the change in the savings propensities. “It was not until the 1970s that Solow’s so-called growth model effectively carried the day” (Piketty, 2014a, p.231). However, the Cambridge capital theory controversies had nothing to do with this. It is thus difficult to comprehend Piketty when he writes “In my view, the virulence – and at times sterility- of the Cambridge capital controversy was due in part to the fact that the participants on both sides lacked the
Once one moved from a one-sector model, some important standard predictions of neoclassical production theory would not hold. Importantly, an increase in the capital-labour ratio was not unambiguously associated with an increase in the rate of profit and a fall in the wage rate; the phenomenon of “capital reversing”. “Reswitching” could occur where a given technique of production could be the most profitable at two different interest rates. (See Cohen and Harcourt (2003) and Pasinetti and Scanzieri (2008) for useful summaries.) While even theoretical debates are rarely conclusive in economics, the force of the Cambridge, UK, critique was handsomely conceded by Samuelson (1966). Moreover, in the 1970s, a discussion of the capital theory controversies was even included in several undergraduate textbooks on economic growth. But the debate now has been completely ignored or forgotten. The controversy was somewhat acrimonious as Cambridge, UK, saw it more than a mere technical argument, but one that had fundamental methodological implications as to the way the capitalist system should be analyzed (Harcourt, 1976). Eventually, the problems of capital reversing and reswitching, while a logical possibility, were dismissed by mainstream economists simply as a (Ruth Cohen) curiosum and regarded by some as merely equivalent to a Giffen good (Stiglitz, 1974). They were interesting, but not important, anomalies.

But for those impatient with what Solow (1988, p. 309) saw as “a playing-out of ideological games in the language of analytical economics”, it still remains the problems posed by the more general aggregation problem, of which the Cambridge capital theory controversies have been seen as a subset (Fisher, 1969, 2005). Fisher (2005), who has done more work than most on the aggregation problem from a quintessentially neoclassical viewpoint, is firmly of the opinion that successful aggregation simply cannot be done.

Even under constant returns, the conditions for aggregation are so very stringent as to make the existence of aggregate production functions in real economies a non-event. This is true not only for the existence of an aggregate capital stock but also for the existence of such constructs as aggregate labor or even aggregate output.

One cannot escape the force of these results by arguing that aggregate production functions are only approximations. While, over some restricted range of the data, approximations may appear to fit, good approximations to the true underlying technical relations require close approximation to the stringent aggregation conditions, and this is not a sensible thing to suppose. (p.490)

The damaging implications of this for the aggregate production function are now almost never mentioned in the literature. (For a more detailed discussion of the aggregate production function, see Fisher (1992) and Felipe and Fisher (2008)). It should be noted that this is a very strong result as it is

---

historical data needed to clarify the terms of the debate. It is striking to see how little use either side made of national capital estimates done prior to World War I, they probably believed them to be compatible with the realities of the 1950s and 1960s (p.232)”. The debate had nothing to do with historical empirical estimates of the capital stock.

7 It is difficult to see how it could be anything other than ideology as it involved competing paradigms (Kuhn, 1970).
based on the premise that there exist well-defined production functions at the microeconomic level, which itself is a contentious assumption.

If these results are accepted at face value, it means that, for example, most of the discussion is about growth within the neoclassical framework, including the Solow growth model is flawed.

So why are aggregate production functions still so widely used?

One, usually unstated, reason is that aggregate production functions “work”, in that statistical estimations of them give plausible estimates of the parameters. As Solow once remarked to Fisher, “had Douglas found labor’s share to be 25 per cent and capital's 75 per cent instead of the other way around, we would not now be discussing aggregate production functions” (Fisher, 1971, p. 305).

4.2 Why Aggregate Production Functions “Work”

Economists rarely pay much attention to methodology and when they do, seem only to have read with approval Friedman’s (1953) essay on “The Methodology of Positive Economics”. The central theme of Friedman’s argument is that “the mark of ‘truly important and significant hypotheses’ is that although their assumptions will be ‘wildly inaccurate descriptive representations of reality’, they nevertheless have good predictive power ....[T]he only relevant test of the validity of a hypothesis is comparison of its predictions with experience.” (p.14).

Consequently, the implications of aggregation problem and the Cambridge capital theory controversies may be safely dismissed as ever since Cobb and Douglas’s (1928) seminal paper, aggregate production functions usually, but not always, have high predictive power in terms of good statistical fits and with the estimated output elasticities close to the factor shares. This, as we have noted, indirectly confirms that factors are paid their marginal products.

This is essentially the only theory of the determination of wages and the rate of profit discussed in both introductory and advanced textbooks. Mankiw and Taylor (2008, p.69), for example, assert that “[T]otal output is divided between the payment to capital and the payments to labor depending on their marginal productivities” (emphasis in the original). “We can now verify that if factors earn their marginal products, then the parameter α tells us how much income goes to labor and how much goes to capital” …. “The Cobb-Douglas production function is not the last word in explaining the economy’s production of goods and services or the distribution of income between capital and labor. It is, however, a good place to start” (p.71). Similar sentiments about the usefulness of the Cobb-Douglas production function are also expressed in Hoover’s (2012, pp. 326-331) macroeconomics textbook. There is not a word here about social norms, bargaining power and the role of institutions.

In the 1930s, and in subsequent years, Douglas and his colleagues undertook a large number of cross-industry estimations of the Cobb-Douglas production function. These found that the output elasticities were virtually identical to the factor shares. Douglas (1976) reports some of these
estimates for the 1950s and 1960s for Australia using cross-industry data where estimated output elasticities are virtually identical to the factor shares (the average value of labour’s output elasticity is 0.58 and its share is 0.54) and that the R²s are near unity (all are over 0.980).

Douglas (1976, pp. 913-914, emphasis added), summarizes the implication of these as follows: “the approximate coincidence of the estimated coefficients with the actual shares received also strengthens the competitive theory of distribution and disproves the Marxian. Many of the original objections have been answered. Some remain”.

The fundamental problem is that the production function is an engineering relationship and should be expressed in physical terms. This is how students are introduced to the theory in introductory economics textbooks. In Mankiw (2010), for example, the production function is that of a bakery and the output is the number of loaves. Typically, use is then made of Euler’s theorem for a linear homogeneous function using the price in dollars of each physical unit of output to demonstrate how the value of output, under the usual neoclassical assumptions, is equal to the value of the total compensation of the two inputs, namely, capital and labour. Ferguson (1971, p. 250), in a debate with Joan Robinson, wrote that he “assumed a production function relating physical output to the physical inputs of heterogeneous labor, heterogeneous machines and heterogeneous raw materials” (emphasis added). Indeed, this is also true of Cobb and Douglas (1928) who, in the opening sentence of their classic article, wrote of the “volume of physical production”. But then a legerdemain occurs in the introductory, and even the more advanced, textbooks. The discussion next progresses seamlessly to the whole economy and aggregate factor shares, where the output is a constant-price value measure and the “price” is a price deflator. But the (erroneous) implication is that the results of the physical one-sector production function still follow through unaffected.

The problem is that in practice the aggregate production function has to be estimated using constant-price value data for both output (confusingly, sometimes called the volume of output) and the capital stock. The identity \( Y = wL + rK \) must hold for any state of competition, whether or not there are constant returns to scale and importantly, even if the aggregate production function does not exist. If the identity is differentiated and then integrated at any point of time, then the result is a Cobb-Douglas relationship given by:

\[
Y = wL + rK = B w^a L^{-a} r^{1-a} K^{(1-a)} = AL^a K^{(1-a)}
\]

where \( B \) is the constant of integration and \( a \) and \( (1-a) \) are the factor shares.\(^8\) Equation (3) has no behavioural content at all. However, when cross-sectional observations are used in statistical

\(^8\) Note that this is different from the identity derived from neoclassical production theory where the value of output is \( pQ \) where \( p \) is the price in, say, £s per unit output. It is theoretically possible to recover the physical
estimation, \( a, (1 - a) \), \( w \) and \( r \) may differ. However, it can be seen that if these differences are not great, as empirically they are not, then one will get a near prefect statistical fit. No wonder Douglas and his colleagues found the estimates of the supposed output elasticities almost identical to the factor shares and obtained \( R^2 \)'s of over 0.98 – this is simply because they were in fact estimating the factor shares. Equation (3) shows that if one were to estimate a putative Cobb-Douglas production function, the “output elasticities” will be close to the factor shares, which would be misleadingly interpreted as confirming that factors of production are paid their marginal products.

What about estimates of aggregate production functions using time-series data?

We can express the argument as follows where the “direction of causation” runs from the identity to the putative production function:

\[
Y \equiv wL + rK \Rightarrow \hat{Y} \equiv a_iw_i + a_i\hat{L}_i + (1 - a_i)r_i + (1 - a_i)\hat{K}_i \Rightarrow \\
Y = F(K, L, t) \Rightarrow \text{Cobb-Douglas; CES; translog production functions.} \quad (4)
\]

The accounting identity may be expressed in growth rates which as above may be approximated by a mathematical functional form, such as those given by a Box-Cox mathematical transformation of which the Cobb-Douglas (if factor shares are constant) and CES functions are particular cases.

The identity will also give a good fit to time-series data provided the weighted logarithm of the wage rate and profit rate can be accurately proxied by a time trend. This will often have to be a non-linear function as the wage rate and the profit rate have a strong cyclical component. The use of a linear time-trend can give such poor statistical results that it often gives the impression that a behavioural equation is being estimated. It should be noted that this critique does not just apply to the Cobb-Douglas production function. If the identity has changing factor shares due to, say, the relative change in the bargaining power of firms and workers due to globalization, a better transformation of the accounting identity may be given by a CES relationship as in equation (4) (Felipe and McCombie, 2001; Simon, 1979a).

This argument was first articulated in the case of the Cobb-Douglas production function by Phelps Brown (1957), as we mentioned above, and his argument was formalized by Simon and Levy (1963). Simon (1979b)\(^9\) thought the criticism sufficient to mention in his Nobel Prize acceptance speech and the criticism was applied to Solow’s (1957) paper by Shaikh (1974, 1980). Samuelson (1979) also noted it with respect to Paul Douglas’s estimations. The criticism was extended and applied to other

---

\(^9\) As Simon (1979b, p.497) put it “Fitted Cobb-Douglas functions are homogeneous, generally of degree close to unity and with a labor exponent of about the right magnitude. These findings, however, cannot be taken as strong evidence for the classical theory, for the identical results can readily be produced by mistakenly fitting a Cobb-Douglas function to data that were in fact generated by a linear accounting identity (value of goods equals labor cost plus capital cost), (see E. H. Phelps-Brown). The same comment applies to the SMAC [CES] production function”.  

16
studies by Felipe and McCombie and their series of papers brought together in Felipe and McCombie (2013). The argument is deceptively simple and has been subject to a number of criticisms, or rather misunderstandings (notably that it only applies to the Cobb-Douglas production function) and criticisms that beg the question by assuming *a priori* that an aggregate production function theoretically exists. These are considered in Felipe and McCombie (2013, chapter 12) and Felipe and McCombie (2014) and none of them are found compelling.

Felipe and McCombie (2009) examine empirical estimates of labour demand functions, which are derived from the aggregate production functions using the assumption that factors are paid their marginal products. They confirm that from knowledge of the values of accounting identity alone, one can determine the values of the estimated parameters before the regression is run. Fisher (1971) undertook a simulation exercise where the aggregation problems meant that the well-defined micro-production Cobb-Douglas production functions could not be aggregated to give an aggregate production function. Yet, the estimations of the aggregate wage equation derived from the simulated data suggested that a well-behaved aggregate Cobb-Douglas production function with output elasticities equal to factor shares did exist. However, Fisher concluded that the causation ran from the stability of the factor shares to the Cobb-Douglas production function and not *vice versa*. In a later paper, Fisher et al. (1977) repeated the same exercise for the CES production function. They found that the simulations gave some well-defined estimates of the “aggregate elasticity of substitution”, which, as Fisher et al. (1977) point out, does not exist. Moreover, the estimates of the some of the “aggregate elasticities of substitution” actually lay outside the range of those of the individual shares. While Fisher did not find an organizing principle behind this result, as he did with the Cobb-Douglas production function, it is clear that it is the change in the factor shares that determines the elasticity of substitution and not *vice versa*. Felipe and McCombie (2006) also undertook a simulation analysis where the underlying micro production functions were Cobb-Douglas but the output elasticity of capital was 0.75 (instead of the usual 0.25) and of labor was 0.25 (instead of 0.75). Prices were determined by a mark-up of 1.333 and used to aggregate the micro-production functions. When the aggregate production function was estimated the “output elasticities” of capital and labor were now 0.25 and 0.75, each equal to the relevant factor share.

What are the implications? The use of the aggregate production function to determine the output elasticities and hence indirectly test and often supposedly confirm the marginal productivity theory of distribution by comparing them to the factor shares is without foundation. The fact that capital’s share has increased while the capital-output ratio has fallen (if in fact it has) is not caused by an aggregate elasticity of substitution of less than unity. It is the change in factor shares and the accounting identity that gives, or causes, the misleading impression of an aggregate elasticity of substitution (which “does not exist” as Fisher et al. (1977) remind us) of a particular value.

We have already noted that Piketty is well aware of the limitations of the aggregate production function and the role of the paradigm in determining what are the legitimate questions.
“All economic concepts, irrespective of how ‘scientific’ they pretend to be, are intellectual constructions that are socially and historically determined, and which are often used to promote certain views, values or interests.” [ … ] “In particular, the notion of the aggregate capital stock $K$ and of an aggregate production function $Y = F(K, L)$ are highly abstract concepts. From time to time I refer to them. But I certainly do not believe that such gross oversimplified concepts can provide an adequate description of the production structure and the state of property and social relations for any society” Piketty (2015, p.70). It is a pity that he did not dispense with them altogether.

Given these conclusions then the logical step is to examine how the pay of, say, the top one percent is actually determined in practice, looking at the institutional framework within which these salaries are determined. As we shall show this involves using a completely different framework and discarding the neoclassical paradigm.

5. The Determination of the Pay of the CEOs

The increase in overall inequality in incomes has generally been explained in terms of labour market forces; the increasing wage premium for college graduates, the effect of technical change on the increased demand for skills (technological change is skill-biased and so favors the educated workers), the effect of globalization (the liberalization of product markets and the increased mobility of capital between countries benefit capital at the expense of labour) and the weakening of labour and product market policies and institutions (OECD, 2011; Autor, 2014). But these explanations, such as those based on the supply and demand for skills, are not adequate to explain the rapid rise of the extreme top tail of the earnings distribution. The evidence seems to point to the fact that the increase in the share of the top tail has been the result of rent extraction and the pay setting institutions (Bivens and Mishel, 2013).

Compelling evidence that these high salaries are largely rents is that the increase in the top one percent in the US has been mirrored in UK, Australia and Canada, but not to such an extent in the other advanced countries, such as continental Europe, Korea and Japan. The experiences of Japan, Germany and Sweden, where the share of the top one percent since the 1930s either depicts an L-shaped curve or is flat, are very different from those of US, and UK, where the pattern of inequality follows a U-shaped curve. Alvaredo et al. (2013) suggest that different institutional arrangements and policies may be the reason why similar countries exhibit “such diverging patterns” in inequality. They maintain that “purely technological stories based solely upon the supply and demand of skills can hardly explain such diverging patterns” (Alvaredo et al., 2013, p.5).

Arguments in support of the contention that CEOs are paid their marginal products in competitive markets are equally unconvincing. Kaplan (2012) asks how is that other groups such as private corporate lawyers, hedge fund investors, and private equity investors have achieved equal significant increases. He further argues that CEO compensation has risen slower than the average incomes of the top households, an argument quoted with approval by Mankiw (2013). But as Bivens
and Mishel (2013) and Mishel and Davis (2014) have shown, if one uses the earners and not households as the comparator, CEO compensation has actually risen faster. But even if Kaplan is correct, how does this necessarily demonstrate that top incomes are determined in a competitive market for talent? The rapid growth of the income could be largely the result of comparability with CEOs remuneration and influenced by the fact that social norms about the pay of the top earners have changed since the mid-1970s.

Furthermore, in the US and the UK the rapid increase in the size and profits of the financial sector have driven up top salaries in this sector. In 2008, in the US, the finance sector earned a quarter of GDP and 40 percent of profits. Philippon and Reshef (2012) have estimated that the most significant factor in determining wages in this sector just prior to the subprime crisis was deregulation. This led for a short time to an increase in this sector’s profits before the subprime crisis through a rapid increase in leverage and risk taking, the latter caused by the development of new financial instruments such as CDOs. Philippon and Reshef (2012) find that the excess wage in finance– the difference between the amount employees actually earned in this industry, compared with the amount they are predicted to make (given their relative education level, the skill premium and the relative risk of unemployment) reached 40 percent, which can largely be attributed to rents.

But clearly, to understand why CEOs’ income has risen so dramatically, it is necessary to examine how their salaries are actually determined in practice. There is now great deal of evidence as to how top executives’ pay is set in reality, which is a far cry from the simplistic marginal productivity theory. As Bebchuk and Fried (2004 and 2005) have shown, CEOs’ salaries are determined by supposedly independent remuneration committees and directors on behalf of the shareholders. These committees, which, in fact, can hardly be described as independent (Bebchuk and Fried, 2004), are responsible for not only setting the base salary but also bonus schemes, such as stock options and restricted stock, to incentivize the CEO to act in the best interests of the shareholders (Conyon, 2006). There are basically two competing explanations as to whether this is successful. One view is that “optimal contracts” have been introduced for CEOs, and other highly paid executives, and have largely solved the principal-agent problem.

The other view is articulated by Bebchuk and Fried (2003, 2004, 2005 ) who dismiss the optimal contracts literature, referring to it disparagingly as the “official story”. (See also the informative review of Bebchuk and Fried (2004) by Weisbach (2007)). Their central hypothesis is that the determination of executive pay is the result of a process of remuneration committee capture, whereby the CEOs, in effect, succeed in setting their own compensation. Bebchuk and Fried (2004) call this process “the managerial power approach”, which is presented as a more convincing alternative to the optimal contracting theory.

10 However, as Haldane et al., (2010) the conventional way that output of the finance sector is calculated in the NIPA is likely to have provided an overestimate in the run up to the subprime crisis.
According to the optimal contracting approach, CEOs earn what is termed their “reservation utility”, which is the remuneration that prevents them from quitting and going somewhere else. According to the managerial power approach, the CEO compensation is set as high as possible, subject to an “outrage factor”, that has, for some reason, changed over time. According to the principal-agent approach, the use of options and restricted shares are seen as incentives given to solve an agency problem. CEOs’ compensation is linked to the financial performance of their firms as reflected in their share valuation. According to the managerial power approach, whatever their rationale, options and restricted stock only transfer rents to executives and do not act as an incentive to get value-maximizing strategies adopted.

Much of the impetus for the rapid increase in the use of stock options as a substantial part of CEOs’ remuneration came from the work of two influential business economists, Jensen and Murphy. Under the standard belief that the best judge of the performance of corporations are financial markets, Jensen and Murphy (1990a and 1990b) encouraged the remuneration committees of companies to award CEOs high compensation (they thought that, at the time, CEOs were underpaid), using stock options in order to attract and retain the best and most talented individuals and to use monetary incentives to align the conflicting interests. This “pay for performance” was seen as the best solution to the principal-agent problem. It aligns shareholders’ and CEOs’ interests because, so the argument goes, CEOs are rewarded only if they pursue the principals’ interests, which will be reflected in the firms’ share price.

This “optimal contracting”, which is closely aligned to the “maximizing shareholder value” approach, has been widely adopted in the US. The success of the management of the firms was to be judged largely, or solely, in terms to the share price of the firm. Typically, top executives have been given options to buy shares not at the then prevailing price, but at some time in the future, when the share price is likely to be higher, supposedly due to the CEOs’ efforts. It is notable that in 2004, on the basis of evidence of the actual effect of the stock options, Jensen et al., (2004) had a complete volte face and completely changed their minds. (See also Stout, 2014). As we shall see, by then, it was too late.

Consequently, we have an answer to the question posed above, namely what was the cause of the dramatic rise in CEOs pay over the last thirty years or so? If one were to search for an, or indeed the most, important proximate factor in the growth of CEO pay relative to the mean wage, one need look no further than the widespread use of stock options. The use of stock options was introduced in addition to the CEOs’ salary as there was no corresponding reduction in the latter when the stock options were introduced.

Starting from the 1980s, there is a high correlation between CEO’s remuneration and stock prices. Figure 1 shows the consequences of the move towards a much greater part of the remuneration of CEOs being tied up with stock options and, hence, being closely correlated in the value of the stock prices.
In the following table, we report the results of regressing the logarithm of CEO annual compensation on the logarithm of S&P Index series over the period 1965 to 2014. The regression results reveal the strong and statistically significant impact of the growth of the S&P index on that of the top executives’ pay, with over 80 percent of the variation of the latter explained.\footnote{The regression results pass all the usual diagnostic texts. Tests for structural breaks (Clemente-Montañés-Reyes and Zivot-Andrews unit root tests) reveal that a breakpoint in the (ln) CEO’s annual compensation series occurred in 1993. Both the augmented Dickey-Fuller test (t-statistic equal to -3.462, 5% critical value being -2.955) and the Johansen tests for cointegration (t-statistic equal to 16.1067 for the null of no cointegration, 5% critical value being 15.41; t-statistic equal to 0.4939 for the null hypothesis of at most one cointegrating equation, 5% critical value being 3.76) reject the null hypothesis of no cointegration.}

The regression analysis starts by assessing the estimated impact of the lagged level of the S&P Index on the CEO’s annual compensation both without and with a time trend (columns I and II). It is found that the time trend is statistically significant and that the S&P index has affected positively the level of the CEO’s pay, and is statistically significant. The same occurs even when we control for the structural break. Empirical tests reveal that there has been a structural break in 1993: before and after that, the autonomous growth of CEO compensation is positive and significant, and equal to 5 and 2 percent per annum, respectively (columns III and IV). Finally, we investigated whether or not there had been any change in the slope coefficient of the S&P index. It is found that the slope has changed and has actually increased after 1993, but by a small amount (column V).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Indices of US CEO remuneration and the S&P 500 index}
\textit{(Data Source: Economic Policy Institute)}
\end{figure}
Table 1. CEO’s Annual Compensation and S&P 500 Index (1965-2014). OLS Regressions

<table>
<thead>
<tr>
<th>Dependent variable: (ln) CEO compensation (in millions of 2014 dollars)</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ln) S&amp;PIndex t-1</td>
<td>1.756***</td>
<td>.804***</td>
<td>.5498***</td>
<td>.6501***</td>
<td>.5424***</td>
</tr>
<tr>
<td></td>
<td>(17.70)</td>
<td>(8.22)</td>
<td>(4.27)</td>
<td>(4.52)</td>
<td>(3.77)</td>
</tr>
<tr>
<td>(ln)S&amp;PIndex tbreak _t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.0921***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.76)</td>
</tr>
<tr>
<td>time trend</td>
<td>.0492***</td>
<td>.0419***</td>
<td>.0501***</td>
<td>.0501***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15.52)</td>
<td>(8.25)</td>
<td>(9.33)</td>
<td>(8.88)</td>
<td></td>
</tr>
<tr>
<td>structural break dummy</td>
<td>.5275**</td>
<td>59.58**</td>
<td>74.20***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td>(2.66)</td>
<td>(3.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time trend* structural break dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.0296**</td>
<td>-.0371***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.64)</td>
<td>(-3.56)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-10.22***</td>
<td>-102.03***</td>
<td>-102.80***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-15.08)</td>
<td>(-17.46)</td>
<td>(-9.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-8.34)</td>
<td>(-8.74)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.8184</td>
<td>0.9496</td>
<td>0.9574</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9643</td>
<td>0.9692</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test p value</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Superscripts */**/*** denote 10, 5, and 1 percent significance levels.
Figures in parenthesis are the t-statistics. The regressions are controlled for heteroskedasticity.
*Structural break* is a dummy that is zero before 1993 and 1 otherwise; *time trend* * structural break dummy* is an interaction variable that creates a dummy counter of 0 before the break and time period number after the break.

Bebchuk and Grinstein (2005) run regressions attempting to explain the rapid rise of CEO compensation, over the period 1993-2003, but solely in terms of standard industry variables. They conclude that “the growth in pay levels has gone far beyond what can be explained by the changes in market cap and industry mix” (p.302).

Why did performance related pay prove ineffective and merely lead to rapid increases in CEOs’ remuneration? The answer is that in the US, in reality, the structure of a corporation is such that CEOs have enormous influence over the board of directors, who are supposed to be independent and to supervise the CEOs’ conduct and remuneration. Directors often receive large direct and indirect benefits, which are largely at the CEOs’ discretion. Moreover, there are often interlocking pay committees with CEOs being on each other’s remuneration committees, even if at several times removed. Consequently, the CEOs’ remunerations are effectively mutually determined. There are spillover effects into the public sector where large pay increases of the top managers are justified by reference to comparable private-sector pay, often judged merely by the size of the organization rather than any reference to its profitability (Bebchuk and Fried, 2003).

Bebchuk and Fried (2004) analyze in detail the performance-related-pay schemes, with a view to determining whether these resemble more the optimal contracting approach (according to the principal-agent theory) or the so-called managerial power approach. They found that the structure of
the compensation schemes provides compelling evidence for the managerial power approach. Performance pay in the private sector is often linked to the overall increase in the value of the company’s shares, not how the company performs relative to the stock market overall. Ideally, CEOs’ compensation should reflect only the degree to which the company performance that has been affected by their actions. If the value of all shares increases, as it happens during a stock market boom, then additional compensation should go only to the CEOs of those companies whose stock prices rose more rapidly than the average. But this never occurs in practice. In reality, CEOs receive stock options with a fixed price and can achieve considerable payments for these, even if their stock increases less than the market (Bebchuk and Fried, 2004). Moreover, many of the arrangements for CEOs’ pay are far from transparent, which is the opposite of what one would expect if the principal-agent problem was to be minimized.

The remarkably small number of financial linkages that connect most of the world’s international firms has been dramatically demonstrated by Vitali et al. (2011). They used complex network analysis to trace the cross holdings between 43,060 transnational corporations and found that 147 of these companies had control of 40 percent of the value of transnational corporations and 737 had control of 80 percent. It can be seen that this close interrelationship not only poses severe economic stability problems, but also how a very small network of top managers could come to set their own salaries based on a circularity notion of comparability. (Bivens and Mishel (2013, pp. 63-71) and Alvaredo et al. (2013, pp. 9-11) present a more detailed discussion of linkages between CEOs remuneration.)

In other words, according to the evidence, the rapid increase in CEOs’ remuneration has been driven more by rent extraction than the result of a well-functioning competitive market for senior executives. Moreover, while changes in income distribution need not be a zero sum game, there is overwhelming evidence that the rise in the share of the top one percent has been at the expense of the remaining 99 percent. The relationship between work effort and pay in the neoclassical schema (work is seen merely as a disutility) is over simplistic. Many CEOs and top earners gain a great deal of utility through the power and prestige of their positions, and it is doubtful whether their work effort would decrease if their earnings were taxed more or their salaries were less.

We have discussed one plausible account as to what accounts for the rapid increase in CEO pay and the share of the top one percent. But this has been associated with and, indeed, has caused the rise of financialisation that has had a significant adverse effect on the growth of the economy. We next turn to an examination of this.
6. Financialisation and Inequality

The first thing that is necessary to do is to define exactly what is meant by “financialisation”, especially as it is usually taken to be more than just the increase in the size of the finance sector, rapid though that might have been. There have been definitions but the most quoted is probably that of Epstein (2005, p.3) where he states that “financialization means the increasing role of financial motives, financial markets, financial actors and financial institutions in the operation of the domestic and international economies”. It encompasses more than just the relative increase in the financial sector, although the latter is important.

The beginnings of financialisation, defining it broadly to include the increasing role of financial motives, may be traced back to the mid-1970s. This was the start of rise of neo-liberalism, the rolling back of the state, and the privatization of many nationalized industries. Of particular importance was the deregulation of the financial industry, which was started by Thatcher and Regan in the early 1980s.

What may be termed the financialisation of the non-finance sector in the US has also been dramatic. It started with the merger boom of the 1960s and 1970s and a change in corporate objectives. The various “subunits” of these mergers came to be seen as “tradable assets that should be evaluated”, which paved the way for the financialisation of non-finance firms. Furthermore, there was the rise of neoliberalism with Thatcher and Regan. This led to financial deregulation throughout the 1990s and a consequence was that many traditional manufacturing firms branched into financial services and these came to be responsible for the major share of their corporate profits (Lin and Tomaskovic-Devey, 2013). Notable examples include General Motors and Ford. Both initially had a financial arms devoted to providing credit for customers who wished to purchase their cars. But, in the 1980s, they branched out into more general financial services.

Crotty (2003) terms this the “neoliberal paradox”, which is a contradiction that specifically arises where competition facing firms in the product market makes it difficult for these non-financial companies to achieve high rates of return, so they shift to the provision of financial services. But then the non-finance companies have to tackle increasing pressures to increase their pay-out ratios or find their stock prices decline and face the threat of a takeover. Crotty (2003) writes of a shift from the Chandlerian (1990) view of the (US) nonfinancial corporations as a combination of illiquid assets assembled to pursue long-term growth to viewing it as a “portfolio” of liquid subunits, where the objective is not to innovate but to maintain a high stock price.

Also in the US one of the most important results driving financialisation was the change in corporate objectives which was linked to the change in CEOs’ remuneration. In a sense, there was a reverse causation. The rapid increase in the use of stock options has increased the degree of financialisation of the economy that has had a deleterious impact on the growth of the US economy. This has been demonstrated by, inter alios, Lazonick (2014) who compares two periods, one from the
end of the Second World War until the late 1970s and the other period from then until the present. He argues that during the former period, the leading financial paradigm was the so-called “retain and reinvest (earnings) approach”. Under this model, corporate profits were used to invest directly in the company. This led to what he sees as shared prosperity. The growth of the firm benefited both the employees and the shareholders. However, since the late 1970s, the prevailing approach has been the so-called “downsize and distribute” regime, which has resulted in the extraction of value from the firm in preference to the creation of value. This has been achieved largely by share buybacks by the firm, which as Lazonick (2014) observes, are mostly open market repurchases of shares. The evidence suggests, often come at the expense of productive investment. But what has made open-market repurchases increase so much over the past three decades or so is the rise of stock-based pay, which has created the incentives for top managers to use this to drive up the price of the shares, especially just before they cash in their stock options.

Lazonick (2014) estimates that this buying up of shares took as much as 54% of corporate profits for 449 companies listed on the S&P 500 index and 37% of the remaining profits were paid out in dividends, which left little retained profits for investment. The reason is straightforward: of the 500 highest-paid executives in the US (who earned $30 million each), 42 percent of their remuneration came from stock options and 41 percent from stock awards.

Companies, it is true, sometimes use tender offers to buy back shares from shareholders when the company deems the share price to be undervalued. However, as noted before, the vast majority of bought back shares are purchased on the open market and since 1982, they can do this without any danger of the SEC charging it with stock price manipulation.

We have seen how the use of stock options in the compensation of CEOs rose dramatically from the 1980s onwards. This was paradoxically partly the result to the wave of hostile takeovers in the 1980s, which was seen as a consequence of the executives’ failure to maximize returns to shareholders, and, as we have seen, the use of stock options was supposed to give an incentive to CEOs to “maximize shareholder value”. Thus, profits that should have gone to physical investment have merely gone into stock manipulation. The extent of the buyback is that its value has exceeded new share issues and that “in aggregate, the stock market is not functioning as a source of funds for corporate investment” (Lazonick, 2014, p. 9).

In particular, the increase in financialisation over the last few decades has also reduced the labour’s share of income and increased CEOs earnings, thereby exacerbating income inequality. The quantitative effects of this in explaining income inequality are comparable to those of de-unionization, technological change and globalization. Since the early 1970s, financial income (interest, dividends, and capital gains) has become a significant stream of revenue for US corporations including manufacturing and non-finance firms (Lin and Tomaskovic-Devey, 2013, p.1287). But it has also shaped the objectives of the firms. In fact, the growth of short-term profits and a focus on the stock price of the firm become (as discussed above) more important than long-term investment and
increasing market share and there is evidence that financial income reduces physical investment. (Orhangazi, 2008; Crotty, 2003).

Using data for the US from 1970 to 2008, Lin and Tomaskovic-Devey (2013) test three hypotheses: that the increased dependence on revenues from financial services causes (i) the decline in labour’s share, (ii) increases the executives’ share of compensation and (iii) increases earnings dispersion amongst employees. Turning to the first hypothesis, they find that the effect of financialization has a statistically negative effect on labour’s share, with a 1 per cent increase in the variable leading to between about 1 and 3.7 percent decline in labour’s share. Financialization accounts for 58% of the total decline in labour’s share between 1970-2008.

The second set of regression results has the executives’ share of compensation as the dependent variable. Financialization has a positive effect in the period 1971-1997, although paradoxically the effect is quantitatively smaller in the period 1999-2008. Union density has a negative effect, suggesting union power had some role in restricting the growth in the share of executive compensation. Lin and Tomaskovic-Devey (2013) argue that the increase in financialisation in the US economy has actually modified the social framework within which the distribution of income between economic actors gets shaped. Specifically, they suppose that “firms’ increasing reliance on financial, rather than production, income decoupled the generation of surplus from production and sales, strengthening owners’ and elite workers’ negotiating power against other workers. The result was an incremental exclusion of the general workforce from revenue-generating and compensation setting processes” (Lin and Tomaskovic-Devey, 2013, p.1285). Thus, this is a broad factor explaining the rapid rise in the compensation of the CEOs. Lin and Tomaskovic-Devey (2013) argue that the definition of marginal productivity is tautological and that productivity is an “organizational outcome” embedded in a particular “social configuration” (p. 1315). Moreover, it is clear that the changes in executive compensation is the result of “identifiable ideological, political and institution developments since the late 1970s” (p.1315). The marginal productivity theory is simply used as an ex-post justification for the huge bonuses and salaries of executives. We find it difficult to disagree.

The financialisation thesis maintains that many of these forces are interrelated and should be interpreted as part of a new economic configuration that has been explicitly promoted by financial sector interests (Palley, 2007, p. 11). The impact of financialisation has been to increase the importance of the financial sector relative to the non-financial sector, to transfer income from the non-financial sector to the financial sector, and to contribute to increased inequality and wage stagnation. It has been associated with slow growth, increased financial fragility, a rise in financial systemic risk; again, with a rise in corporate and household debt and the change in corporate objectives and is associated with the rise in the share of the top one percent.
7. Conclusions

The last three or four decades have seen an explosion in the pay of not only the CEOs but also of managers in the non-private sector. What was once considered an unacceptable salary for the top earners compared to the average remuneration has now become commonplace. The whole question of the remuneration of top executives and managers is one that involves a consideration of how these payments are determined and social norms about what is acceptable. These social norms are not those of the society as a whole, but rather those involved in the determination of these salaries.

Clearly, an important question is how are these social norms (or moral outrage) determined and how and why do they change over time.\textsuperscript{12}

What is clear, however, is that any defence for the rapid increase in the earnings of the top one percent based on the notion of marginal productivity by neoclassical economists and the concept of “just deserts” is untenable. We have highlighted the theoretical and insurmountable problems concerning the marginal productivity theory of factor pricing and the related concept of the aggregate production function. But what is also telling is that for the neoclassical approach grounded in the need for microfoundations, and using extensively the individual representative agent, it is impossible to test whether the remuneration of a specific individual represents his or her contribution to society. We have considered the way that CEOs are actually remunerated. It is clear that the rapid increase in their pay, and that of the top one percent, represents a change in societal values, a concept that fits uncomfortably within neoclassical economics. The widening increase in income inequality in the UK and the US reflects a significant change in the structure of the economy with the development of financialisation and the increasing influence of the finance sector, both economically and politically. An interesting question for further research is whether or not the less dramatic rise in the share of the top ten percent in continental Europe, Korea and Japan reflects significant differences in these factors and the way the high earners are rewarded.

References


\textsuperscript{12} The Securities and Exchange Commission adopted in 2015 a rule that requires a public company to disclose the ratio of the compensation of its chief executive officer to the median compensation of its employees. The new regulation, required by the Dodd-Frank Act, helps inform shareholders when voting on the purely advisory “say on pay.” Whether it will have any major effect on the pay of CEOs remains to be seen.


Chandler, A. (1990), Scale and Scope: The Dynamics of Industrial Capitalism, Cambridge MA: Harvard University Press


Clark, J.B. (1899), The Distribution of Wealth, New York: Macmillan


