Working Paper No. 1045

The Aggregate Production Function and Solow's "Three Denials"

by

Jesus Felipe (*)
School of Economics, De La Salle University (Manila, Philippines)

and

John McCombie (**)
University of Cambridge, Cambridge, UK

March 2024

Author contact information: Felipe (jesus.felipe@dlsu.edu.ph); McCombie (jslm2@cam.ac.uk)

Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
http://www.levyinstitute.org

ISSN 1547-366X

ABSTRACT

This paper offers a retrospective view of the key pillar of Solow's neoclassical growth model, namely the aggregate production function. We review how this tool came to life and how it has survived until today, despite three criticisms that undermined its *raison d'être*. They are the Cambridge Capital Theory Controversies, the Aggregation Problem, and the Accounting Identity. These criticisms were forgotten by the profession, not because they were wrong but because of the key role played by Robert Solow in the field. Today, these criticisms are not even mentioned when students are introduced to (neoclassical) growth theory, which is presented in most economics departments and macroeconomics textbooks as the only theory worth studying.

JEL CLASSIFICATION: B22, B31, B32, B41, E13, E25

KEYWORDS: Accounting Identity, Aggregation Problem, Cambridge Capital Theory Controversies, Solow

1. INTRODUCTION

Robert Solow, one of the most notable economists of the twentieth century died December 21, 2023, at the age of 99. He was certainly a great researcher, someone interested in society's real problems, and those who studied at MIT during his tenure consider he was a great instructor.

Solow arrived at MIT in 1949 as a professor of statistics. Together with Paul Samuelson, he built one of the best regarded economics departments in the world. Both economists were instrumental in laying the foundations of the neoclassical synthesis. While this took some elements of Keynes' General Theory, it was, in some respects, a return to the pre-Keynesian world. The neoclassical synthesis had three pillars: microeconomic competitive general equilibrium theory, the Phillips curve, and Solow's (1956) growth model. The synthesis asserted the efficacy of Keynesian policy in the short run, but asserted a belief in competitive market forces and Say's Law in the long run (in particular in the case of Solow's growth model).

As an economist interested in society's real problems, Solow was concerned, *inter alia*, with the causes of unemployment (Solow 1980), and viewed the labor market as a social institution and not like the typical market for a good (Solow 1990). He took part in the two most important intellectual debates in economics of the twentieth century: the Cambridge Capital Theory Controversies (CCTCs) (mainly with Joan Robinson in the UK) and with the new classical macroeconomists (mainly with Milton Friedman in Chicago).

Solow was awarded the Nobel Prize in 1987 for his contributions to the theory of economic growth. Solow's (1956) growth model incorporated an aggregate production function (APF). This concept became a pillar of neoclassical macroeconomics and certainly of subsequent neoclassical growth models. Production functions appeared in economics in the nineteenth century. It was not, however, until Cobb and Douglas (1928) that one was estimated for the American economy using statistical methods, which required using aggregate data (indices) for output, labor, and capital. In this article, they found the elasticities of output with respect to labor and capital to be 0.75 and 0.25, respectively, close to the factor shares. Their work was initially received with great hostility, to the point that Douglas considered giving up. He moved from the original work with time series to cross-sectional data, which produced much better statistical results (McCombie 1998; Felipe and McCombie 2013, chapter 4).

Aggregate production functions continued being used during the following decades but it was not until Solow's (1956, 1957) two seminal papers that they became a standard concept in macroeconomics. This was despite some doubts about their theoretical and empirical foundations, as we shall see below. Solow's (1956) growth model is still used today to introduce students to growth theory. In Solow (1957), the APF served as the tool to empirically apportion the contributions of technical progress (total factor productivity growth or Solow residual) and those of capital and labor growth, to output growth, in the US economy between 1909 and 1949. Although total factor productivity had been calculated before, this was the first time it had been explicitly linked to production theory. The 1957 paper launched the neoclassical research program on growth, in particular growth accounting exercises and the estimation of total factor productivity growth.

The present paper offers a retrospective view of the APF, in particular of three fundamental criticisms it was subjected to, and of how, despite these criticisms, it has survived the test of time. These critiques are the Cambridge Capital Theory Controversies (CCTCs), the Aggregation Problem, and the Accounting Identity. It is argued that these three criticisms undermined the *raison d'être* of the APF, Solow's growth model, and growth accounting exercises. Without the APF, the neoclassical macroeconomics edifice would largely collapse. Hence, its putative defense at the time that these critiques were advanced became a matter of crucial importance. However, all three have long been forgotten by the profession, and not because they were fundamentally flawed but because of the key role that Solow played in the debates. Obviously, Solow was not the only one, but he was the "father" of the neoclassical growth model (for which he received the Noble Prize) and of this literature. He was aware of the three criticisms and played a most important role in dismissing them. He clearly did so in the case of the CCTCs, ignored the aggregation results, and denied the veracity of the accounting identity argument. Today, these arguments are not even mentioned when students are introduced to neoclassical growth theory, which is presented in most economics departments and macroeconomics textbooks as the only theory worth studying.[1]

It should be emphasized that the discussion here does not intend to undermine all the many important contributions that Solow made to economics. Yet, it is important to set the record straight on the above issues.

The rest of the paper is structured as follows. Section 2 discusses the CCTCs. Section 3 discusses the Aggregation Problem. Section 4 discusses the Accounting Identity critique. Section 5 concludes. The Appendix introduces the algebra of the accounting identity argument and revisits Solow (1957).

2. DENIAL 1: THE CAMBRIDGE CAPITAL THEORY CONTROVERSIES

"The student of economic theory is taught to write 0 = f (L, C) where L is a quantity of labour, C a quantity of capital and 0 a rate of output of commodities. He is instructed to assume all workers alike, and to measure L in man-hours of labour; he is told something about the index-number problem involved in choosing a unit of output; and then he is hurried on to the next question, in the hope that he will forget to ask in what units C is measured. Before ever he does ask, he has become a professor, and so sloppy habits of thought are handed on from one generation to the next" (Robinson 1953–54, 81).

"In her paper [The Production Function and the Theory of Capital] Mrs. Robinson was annoyed at many of the practices of academic economists. We have reason to be grateful for her annoyance, for she seems to have written her article the way an oyster makes pearls-out of sheer irritation" (Solow 1955–56, 101).

[1]The standard textbooks on growth theory in the 1970s such as Wan (1971), Jones (1975) and Hacche (1970), mentioned the CCTCs. Wan (1971) and Jones (1975) also mentioned he aggregation problems. By the 1990s, there was no discussion in the growth textbooks of either of these, e.g., Barro and Sala-i-Martin (1995), Jones (1998 [2002]), Aghion and Howitt (1998), Weil (2005), or Acemoglu (2009). For alternative growth theories see Sen (1970), Scott (1989), and Setterfield (2010).

The Cambridge Capital Theory Controversies (CCTCs) were a series of exchanges between two groups of economists on both sides of the Atlantic. They were theoretical. Cambridge UK economists, led initially by Robinson, questioned the very deep foundations of neoclassical macroeconomics. Cambridge USA economists primarily led by Samuelson and Solow, counterargued.

Summarizing the exchanges between the two sides is a rather complicated task. Harcourt (1969, 1972) provides an excellent summary and discussion of the literature. There are many good summaries of the CCTCs. Cohen and Harcourt (2003) and Felipe and McCombie (2013) also provide summaries. In the final analysis, the debates were about different value theories, the classical (Smith, Ricardo, Marx) *versus* the neoclassical (Böhm-Bawerk, Jevons, Clark, Wicksteed, Wicksell, Walras), and how those two theories explain prices and distribution. The debates centered on a series of issues derived from and linked to the question of whether one can use an aggregate measure of capital in a macroeconomic production function without running into apparently paradoxical phenomena.

The APF appeared in the CCTCs in Robinson's (1953–54) article. Coming from a classical tradition, capital in the production function needed to be in the form of physical, homogeneous machines. These were termed "leets" (steel spelt backwards) by Robinson. However, she argued that this could not be the case at the aggregate level. At this level, it had to be a value measure, with the consequence that this notion of capital was conflating the notions of physical capital (a list of heterogeneous machines) and financial capital (a fund of resources, "money"), which were very different in classical economics. The neoclassical economists used the first notion of capital in the study of interest and portfolio adjustment; capital in physical terms was used in the theoretical study of production (Clark 1893). Clark, however, argued that, *over long periods of time*, changes in the value of capital reflected changes in the stock of investment goods. This was imperative in order to show that, in a laissez-faire economy, each individual (factor of production) who contributed to production received the value of what he or she produced (in a normative sense, *deserved*). This was, of course, the point of the marginalist theory of production: the wage rate equals the marginal product of labor, and the return on each dollar of capital equals the marginal product of capital. Hence, each social class gets what it contributes to production (naturally, there is a clear contrast between this view of the world and that provided by the labor theory of value, in particular the Marxian version). [2]

What was the problem Robinson (1953–54) pointed out and which initiated the CCTCs? She criticized that the distinction made by the neoclassical economists between capital as a monetary fund and capital as a collection of different capital goods was completely lost after the Keynesian Revolution, in particular, in empirical analyses. In fact, according to Robinson, the two conceptions of capital were mixed. At some point in time, economists began analyzing the performance of the economy in terms of an aggregate production function with "capital" and labor as factors of production, and began discussing the remuneration of these factors in terms of their marginal productivities. As a consequence, it appeared that the division of the aggregate product between labor and capital *could be* explained in terms of marginal productivities.

Robinson (1953–54) asked the important question that triggered the debate: in what unit is "capital" to be measured? Robinson was referring to the use of "capital" as a factor of production in aggregate production functions. Because capital goods are a series of heterogeneous commodities (physical investment goods), each having specific technical characteristics, it is *impossible* to express the stock of capital goods as a homogeneous physical entity. Robinson claimed that it followed that only their values could be aggregated. Such a value aggregate, however, is not independent of the rate of profit and thus of income distribution.

This heterogeneity of capital became an important part of the controversies (at least initially and certainly for Robinson). The Clarkian concept of capital, conceived as a fairly homogeneous and amorphous mass which could take different forms, Robinson argued, cannot serve in a macroeconomic production function because it is essentially a monetary value. She claimed, perhaps implausibly, that although labor is not a homogeneous input, in principle it can be

[2] The Cambridge Capital Controversies were plagued with confusions. The most important is between the legitimacy of postulating an aggregate production function, and the legitimacy of the marginalist-neoclassical, or the supply-demand, approach to distribution. The important question is the second one. It is only when one accepts the marginalist approach that one may feel like using aggregate production functions whose marginal products determine distribution. So, one must ask: what kind of aggregation is needed for the validity of the neoclassical approach to (macro) distribution?

measured in a technical unit, i.e., man-hours of work (disregarding different qualities). The same goes for land (acres of land of a given quality). These are natural units, so that the marginal products of land and labor could be defined independently of the equilibrium factor prices. But what about (aggregate) capital? Robinson argued that the statistics of capital used in practice had nothing to do with the previous two notions of capital. Such statistics are values (e.g., dollars) and no matter how they are deflated (to convert them into constant dollars), they continue to be money—a sum of value. Therefore, this cannot be made to correspond to a physical factor of production.

All this matters because, it was claimed, it is impossible to get any notion of capital as a measurable quantity independent of distribution and prices. That is, if distribution is to be explained by the forces of supply of and demand for the factors of production, then the latter must each have a measure. Those measures must be homogeneous so that aggregation is possible. This was claimed not to be possible for capital as a factor of production, but only when it is an amount of finance. Thus, capital has no natural unit, akin to labor and land, which can apparently be aggregated to give a quantity of a productive service, that can then be used for the determination of their prices. We note, however, that such aggregation is illusory, and the aggregation problem is not restricted to capital. Was one labor hour by Robinson really the same as one by Queen Elizabeth II or one of Britney Spears in terms of productivity? The answer is no.

The argument of the Cambridge, England economists was that investment goods, which make up the stock of real capital are themselves produced (i.e., produced means of production). These goods are produced in a market economy where capitalists require profits. This implies that in order to provide a "quantity of capital" one must first know its price. In fact, the price of any commodity cannot be determined independently of the technical conditions of production and of the rate of profits. In other words, the price of the aggregate factor capital is affected by the distribution of income among the factors. The value of capital changes as the profit and wage rates change so that the same physical capital can represent a different value, whereas different stocks of capital goods can have the same value (Robinson 1956). So long as the capital stock is heterogeneous, its measurement requires knowledge of the relative values of individual capital goods. This can only be achieved if the price vector of the economy and the rate of profits were known *ex ante*. The consequence is that aggregate capital, the aggregate production function, and the marginal products of the factors can only be defined when the rate of profit is given, implying that they cannot be used to build a theory of the rate of profit or distribution.

Robinson's critique led to a series of intertwined debates that lasted two decades and which went far beyond the original question. The main debates were about the following five issues.

(i) The theory of distribution, in particular the neoclassical claim that the distribution of income could be derived from some technical properties of an economy, embedded in the production function, and that factor shares could be somehow linked to the marginal products of some corresponding factors.[3]

(ii) The so-called *Wicksell effects*: does a decrease (increase) in the steady-state interest rate always imply a rise (fall) in per capita steady-state consumption provided the rate of interest is greater (lower) than the rate of growth of labor?

(iii) The relationship between savings and investment.

(iv) The problems of *reswitching* and *reverse capital deepening*.[4]

[3] Can distribution be explained by wage and interest-elastic factor demand curves? Quite naturally, capital, its marginal product, and the aggregate production function became part of the debate. The point to be stressed here is that one generally thinks of the question of the determination of the wage rate or the profit rate as a question of microeconomic theory. Thus, the aggregation issue came to the fore when at some point neoclassical economists surmised that the neoclassical distribution theory might be empirically tested by using economy-wide aggregated data.

[4] The former refers to the violation of the supposedly unique inverse relation between capital intensity and the rate of profit. It was shown theoretically that the economy can move between production techniques depending on the level of the rate of profit, so that at high and low levels of profit the same technique could be utilized, thus leading to the possibility of a non-negative relationship between the rate of profit and the capital–labor ratio. The latter occurs when the value of capital moves in the same associated direction as the rate of profit. This is the case when the most profitable project is the one with a less capital-intensive technique. The reswitching problem had been discovered by Sraffa (1960).

(v) Is it, in general, possible, for the rate of profit to equal the marginal productivity of capital in equilibrium?

Note that these are all macroeconomic questions in the sense that they all involve the use of aggregates (capital, labor, output, etc.). It must be stressed that the existence of phenomena such as reswitching or reverse capital deepening only appear paradoxical if one is intent on believing that such aggregates are related by an aggregate production function satisfying the properties that one expects of micro-production functions. One's intuition as to such phenomena comes from thinking about production functions—they cannot occur in a true two-factor, one-output micro model. Hence, this part of the CCTCs is a consequence of the aggregation problem properly considered (although there were, and still are, disagreements on this issue).

If attention is restricted to the question of aggregate capital and the aggregate production function, the answers to Robinson's questions can be grouped into two main lines. The first solution was to conceive the aggregate production function in terms of a parable, the single-commodity world referred to above, following Samuelson (1961–62). The other solution was to search for the technical conditions under which aggregation is possible. This is the work developed by Fisher, *inter alios* (see section 3). The aggregation problem became the search for the conditions under which macro aggregates (not only capital) exist. Robinson certainly rejected both (though she was not altogether unhappy with Fisher's results).

What was the profession's reaction? Economists like Stiglitz (1974) dismissed the CCTCs on the grounds that reswitching was *empirically unimportant*, and thus had faith that the factor substitution mechanisms postulated by the neoclassical approach do exist. Ferguson (1969, p.xvii) explicitly made this instrumental defence with respect to the criticism about the measurement of capital as a single index in CCTCs. "Its validity is unquestionable, *but its importance is an empirical or an econometric matter* that depends upon the amount of substitution there is the system. Until the econometricians have the answer for us, placing reliance upon [aggregate] neoclassical economic theory is a matter of faith. I personally have faith" (emphasis added). This was Solow's position. However, faith is a poor argument and it does not explain why aggregate production functions generally give such good statistical results, especially in the light of the aggregation results, already known at the time (see next section). Moreover, it was argued that the criticisms of the neoclassical theory of capital raised by the phenomena of reswitching and capital reversing were only valid with reference to the neoclassical model conceived in aggregate terms. They did not, however, apply to the general equilibrium model conceived in disaggregated terms and based on the behavior of profit and utility maximizing agents.

On this, Pasinetti (2000) argued that "this proposition actually has no objective foundation; phenomena of non-convexity, reswitching of techniques and badly behaved production functions […] are not –as has been amply demonstrated – a consequence or a characteristic of any particular process of 'aggregation.' They may occur at any time and in any context, aggregated or disaggregated" (212).

Yet, Fisher (1971, 2005) argued that all this referred to the very foundations of neoclassical macroeconomics, and not about the foundations of neoclassical microeconomics. Nor did it refer to the foundations of the neoclassical theory of distribution. These issues, in his view, had nothing to do with the existence of aggregate production functions. He claimed that the Sraffians consider the existence of reswitching and reverse-capital deepening to be a decisive criticism of neoclassical theory. They believe, he claimed, that this was the deciding factor in the "Cambridge v. Cambridge" debates over capital theory. But that view fails to realize that reswitching and reverse capital deepening only appear paradoxical if one supposes that aggregates should behave the way intuition suggests they should behave—the way that factors of production and outputs behave at the micro level. But the non-existence of aggregate production functions means that such intuition simply does not apply.

Petri (1999, 21) indicated that the greatest difficulty in the Cambridge debates was, in fact, a lack of proper communication. While the Cambridge, UK critiques were aimed at the traditional neoclassical versions, based on the notion of capital as a single factor, the Cambridge, US side replied that their theory, in its rigorous versions—that is, the neo-Walrasian formulations—did not need to aggregate anything. Though, as Petri (1999) indicates: "[a] majority of economists did not realize that the main reason for the shift [from the traditional marginalist models to the neo-Walrasian models] was precisely the difficulties of the traditional conception of capital as a single factor, and that therefore they no longer had the right to assume that things work out *as if* capital could be treated in the traditional way" (51, italics in original). This has an important implication: the only versions of neoclassical macroeconomic theory compatible with the analysis of long-period positions are the ones dependent on capital as a single-value factor. The decreasing demand curve for labor and the decreasing aggregate investment schedule can be justified only on the basis of this notion of capital. The justifications given by neoclassical authors, therefore, seem odd in claiming that standard one-good models are simplifications whose micro-foundations are the neo-Walrasian models; neoclassical one-good models are not simplifications of neo-Walrasian disaggregated analyses.

Samuelson (1961–62) tried to save the notion of an aggregate production function in his paper on the surrogate production function. However, Garegnani (1970) pointed out that Samuelson's assumptions to derive the aggregate production function were untenable and that it was difficult to discern with which he agreed.

Often, controversies don't come to an unambiguous and conclusive end. This was the case with the CCTCs. This is notwithstanding "A Summing Up" by Samuelson (1966), when he conceded the defeat of the US side. Yet, this was not the end of the debate as the next generation of neoclassical economists continued making efforts at defending the notion of the aggregate production function. Of course, without this defense, much of macro, growth, and even labor, neoclassical economics (under the Keynesian hat that Solow practiced), would have to be seriously questioned.[5]

During the 1960s and 1970s, the aggregate production function became a key part of the (neoclassical) economics toolkit. Many interpreted this as a *de facto* defeat of Cambridge UK, especially as the implications of the CCCTs disappeared from view. Our view is that this is incorrect. As Joan Robinson died in 1983 and the profession moved on (not only in the USA, but

[5] Recently, Schefold (2020) has argued that reswitching and reverse capital deepening are empirically rare. He shows this by using input-output data. Petri (2022) raises doubts about the use of input-output data to make a point about a discussion that was about physical quantities.

also in the UK), the Cambridge UK arguments criticizing the aggregate production function ended up being ignored.

3. DENIAL 2: THE AGGREGATION PROBLEM

"I have never thought of the macroeconomic production function as a rigorous justifiable concept. In my mind, it is either an illuminating parable, or else a mere device for handling data, to be used so long as it gives good empirical results, and to be abandoned as soon as it doesn't, or as soon as something better comes along" (Robert Solow 1966, 1259–60).

The aggregate production function was discussed during the CCTCs but this was not the core of the debate, and the Cambridge, UK arguments were not the most important criticism against it. This came through a parallel literature, the aggregation problem.

During the 1940s, a group of economists began asking whether and how micro-productions could be aggregated into an aggregate production function. Among these economists were Klein (1946a, b), Leontief (1947a, b), May (1946, 1947), Nataf (1948), and Pu (1946). This was a very different approach to that of CCTCs but was obviously dealing with the same problem. These authors provided the necessary aggregation conditions using different methods and under different conditions. The problem for the aggregate production function was that this literature showed these conditions were unbelievably restrictive.

Solow's reply to Robinson (1953–54) about the units of capital came in the form of the aggregation conditions (Solow 1955–56). To do so, he referred to the Leontief (1947a, b) aggregation conditions (aggregation of variables into homogeneous groups within a firm, or within the economy as a whole, assuming that aggregation over firms is possible). This condition is that aggregation is possible if and only if the marginal rates of substitution among variables in the aggregate are independent of the variables left out of it. In the context of aggregation in production theory (in the simplest case of capital aggregation), the theorem means that aggregation over capital is possible if and only if the marginal rate of substitution between every pair of capital items is independent of labor. The theorem requires that changes in labor, the non-capital input, do not affect the substitution possibilities between the capital inputs. It will hold for cases such as brick and wooden buildings, or aluminum and steel fixtures. But most likely this condition is not satisfied in the real world, since in most cases, the technical substitution possibilities will depend on the amount of labor. Think, for example, of bulldozers and trucks, or one-ton and two-ton trucks. In these cases, no quantity of capital-in-general can be defined (Solow 1955–56, 103). Solow argued that there is a class of situations where Leontief's condition may be expected to hold when the production function can be decomposed into two stages. One example is that, in the first stage there are two kinds of electricity-generating equipment that, combined, generate electric power. In the second stage this is combined with another input to make the output.

Besides the problem of aggregating variables into homogeneous groups, there is the problem of aggregating a number of technically different microeconomic production functions. This is the question Nataf (1948) addressed. Nataf showed that the output, labor, and capital aggregates, which satisfy an aggregate production function, exist if and only if every firm's production function is additively separable in labor and capital. Obviously, Nataf's theorem provides an extremely restrictive condition for intersectoral or even interfirm aggregation. It makes one rather nervous about the existence of an aggregate production function unless there are further restrictions on the problem.

How did Robinson reply to Solow on the problem at hand? "As far as I can understand Professor Solow's note, it does not touch upon the problem of capital, but is concerned rather with how to treat non-homogeneous natural resources, for instance, varieties of Marshall's meteoric stones" And "[i]n my opinion Professor Solow is barking up the wrong tree in seeking either the problem or the solution amongst index number;" and furthermore: "[n]one of these questions can be dealt with in terms of an index of physical equipment, and each of them is important for the analysis of distribution and of capital accumulation" (Robinson 1955–56, 247). The disagreement could not have been bigger.

That Solow was aware of the aggregation problems and the fact that he chose to discount their importance is also demonstrated the following statement in his famous growth accounting exercise for the US (Solow 1957, 312), he wrote: "…it takes something more than the usual 'willing suspension of disbelief' to talk seriously of the aggregate production function. But the aggregate production function is only a little less legitimate a concept than, say, the aggregate consumption function…" However, Felipe and Fisher (2003) argued that the conditions to aggregate micro-production functions into an aggregate production function are much more stringent than those for consumption.

Though not the only one, probably the most important work on aggregation in production functions was developed between the 1960s and the 1980s by Franklin Fisher (see also Sato 1975). He argued in his first paper on the subject (Fisher 1965) that if the aggregation problem was as conceived in the 1940s, it was the end of the story as it was virtually impossible to generate an aggregate production function. For example, Nataf's (1948) theorem essentially indicates that aggregate production functions almost never exist. It should be noted Nataf's theorem does not prevent capital from being physically homogeneous. Likewise, each firm's production function could exactly exhibit constant returns to scale, thus implying that output does not depend on how production is divided among different firms, or even have identical technologies with the same kind of capital.

Therefore, Fisher proposed to look at the problem from a different angle, namely by imposing optimization conditions on firms' behavior into the aggregation problem. The production function describes the maximum level of output that can be achieved if the given inputs are efficiently employed. This approach leads to a way of looking at the aggregation problem that is significantly different from the discussions of the 1940s, and in particular from Robinson's problem (and the CCTCs).

First, it was stressed that the aggregation problem is not only the aggregation of capital, at least the way Robinson understood it (for the UK side the measurement of capital was problematic because the prices of capital goods change when distribution changes). It was pointed out that there exist equally important labor and output aggregation problems. Furthermore, there would be aggregation problems even if each type of capital were physically homogeneous and the same in all firms. Indeed, were there only one type of capital, labor, and output, aggregation problems would continue to exist.

Secondly, from the point of view of the aggregation literature the problem was whether an economy-wide (or a sector or, indeed, a firm) production function can be constructed that exhibits the properties needed to establish downward-sloping factor demand functions. Therefore, perhaps a useful and clarifying way to think about the Cambridge debates and the aggregation problem is to consider whether the measurement of capital problem relates to the interdependence of prices and distribution (i.e., the problem that underpins the reswitching and capital reversing debate), or whether it emerges out of the need to justify the use of the neoclassical aggregate production function in building theoretical models, and in empirical testing. Both problems can be present at the same time, of course, but they are not the same issues.

A by-product of these differences is the implicit acknowledgement that the aggregation process does not lead to physical quantities (Robinson's problem). In fact, Fisher's aggregates are indeed indices, and in his view, Robinson misunderstood the aggregation problem (Fisher 1992, xiii). It is here that an important difference arises in the understanding of the issues at stake. For the UK scholars, the aggregation problem *strictly* affected capital and the rate of profit, and was related to the problem of income distribution. In the words of Pasinetti (2000):

> The problem that arises in the case of capital has not so much to do with the difficulty of finding practical means to carry out aggregation with a fair degree of approximation; it is more fundamentally the conceptual difficulty of having to treat an aggregate quantity expressed in value terms (capital) in the same way as other aggregate quantities (land and labor) which are instead expressed in physical terms. The two types of aggregate quantities do not belong to the same logical class, and can thus neither be placed on the same level nor be inserted symmetrically in the same function […] It becomes a fundamental and indeed abyssal conceptual diversity concerning the "factors" labor and land on the one hand, and the factor "capital" on the other." (209)

For Fisher, on the other hand, capital does not present any special problem. Similar aggregation problems occur with labor and output. The whole problem can be reduced to finding the technical conditions under which *any or all* aggregates can be generated.

It seems, therefore, that Robinson and her followers understood the aggregation problem in terms of what could be termed "natural" aggregation in some physical sense. This is not the same as aggregation of productive factors, as conceptualized by Fisher. And certainly, if one understands the aggregation problem in this latter sense, Robinson's remarks about labor and land being different from capital are not true. They are not physically homogeneous factors either.

The outcome of Fisher's efforts in a series of important papers was that the conditions to generate an aggregate production function with neoclassical properties remained unbelievably unrealistic. Fisher assumed that well-behaved production functions exist at the level of the firm. There are then types of aggregation problem involved in the existence of an aggregate production function for the economy as a whole, or even for some sector thereof. The first of these consists of aggregation over different factors to form aggregate labor or aggregate capital and of aggregation over different products to form aggregate output. This set of problems exists even at the level of the firm.

The second problem consists of aggregation over firms. This is trivial if all factors and all outputs are continually reassigned to firms to maximize efficiency, but this does not happen in practice (except, perhaps, when comparing situations of long-run competitive equilibrium). In particular, different types of capital goods are associated with different techniques, and such capital goods are relatively fixed in place rather than constantly reallocated over firms.

Fisher's aggregation conditions for the three arguments in the production function indicate that: (i) aggregate production functions exist if and only if all micro-production functions are identical except for the capital efficiency coefficient. Certainly, this conclusion is a step beyond Nataf's answer to the problem. Yet, it also continues requiring an extremely restrictive aggregation condition, one that actual economies do not satisfy; (ii) the existence of a labor aggregate requires the absence of specialization in employment; and (iii) the existence of an output aggregate requires the absence of specialization in production—indeed all firms must produce the same market-basket of outputs differing only in scale.

Moreover, the conditions required for aggregation yield results such as:

- Except under constant returns, aggregate production functions are unlikely to exist at all.
- 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.
- Since aggregate production functions could not exist, neither could by implication the *aggregate* marginal productivity theory. But "this hardly stops one from considering the marginal productivity of well-defined individual capital goods" (Fisher, 1971a, 405). Nevertheless, it is difficult to think of any (engineering) micro production function where a homogeneous capital good's contribution to a homogeneous output could be, or has been, readily identified.

Wilson (2009) provided a test of the Leontief-Solow-Fisher conditions for the existence of a single aggregate capital stock, for any unit of production (e.g., firm, industry, economy), formed of separate quantities of heterogeneous capital. Wilson focuses on the relationship between heterogeneous capital services and productivity at the firm level. The paper first derives an empirical, firm-level production function specification incorporating heterogeneous capital services based on standard neoclassical production theory. Wilson then directly tested the Leontief-Solow-Fisher conditions for a single aggregate capital stock. The ability to express a firm's total capital services with a single measure, even if that measure weights heterogeneous capital goods by their relative marginal products, requires that individual capital services each be weakly separable with labor (Solow 1955–56) and that their services be expressed in common units (Fisher 1965). These two conditions together require that different capital services be perfectly substitutable. These conditions have long been viewed by many economists as unrealistic. Solow himself, referring to the first of the two, commented that it “will not often be even approximately satisfied in the real world” (1955–56, 103). Wilson’s (2009) results strongly rejected both conditions. First, he found strong evidence of complementarities and substitutabilities between capital types. In fact, there seems to be a particular pattern in these complementarities and substitutabilities: using any reasonable division of types of capital into high- and low- tech categories, the data indicate that high-tech capital goods tend to be complementary with low-tech capital goods, and substitutable with other high-tech capital. He also found complementarities and substitutabilities between a number of capital types and labor. For instance, software was found to be especially labor saving, while general purpose machinery and trucks were especially labor augmenting.

Also, Fisher (1992, xiii) recalled that in the 1970 Econometric Society Meetings, where he was a participant in a session that featured Solow and Robinson, he (Fisher) made it clear that he called “into question the use of aggregate production functions in macroeconomic applications such as Solow’s famous 1957 paper.”

Fisher and Solow collaborated on a paper where they conducted simulations for the CES production function-aggregation of micro CES production functions (Fisher, Solow, and Kearl 1977). This followed a similar paper by Fisher (1971) for the Cobb-Douglas case. The most important conclusion Fisher (1971) drew from his simulations was the observation that, as long as the labor share happened to be roughly constant, the aggregate production function would yield good results, even though the underlying technical relationships are not consistent with the existence of any aggregate production function. And this conclusion remained true even in cases where the underlying variables showed a great deal of relative movement. This suggests that the (standard) view that constancy of the labor share is due to the presence of an aggregate Cobb-Douglas production function is wrong. The argument runs the other way around, that is, the aggregate Cobb-Douglas works well *because* labor’s share is roughly constant.

In this paper, Fisher also argued that, since aggregate production functions could not exist, neither could by implication the *aggregate* marginal productivity theory. However, he stressed that "this hardly stops one from considering the marginal productivity of well-defined individual capital goods" (Fisher, 1971, p.405). We have doubts about this statement, as it is difficult to think of any (engineering) micro production function where a homogeneous capital good's contribution to a homogeneous output could be, or has been, readily identified.

It is worth reading Fisher (1971) and Fisher, Solow, and Kearl (1977), as they show clearly the lack of legitimacy behind the notion of aggregate production function. Yet, unlike Fisher (1971), who was very clear in his assessment of the aggregation problem, Fisher, Solow, and Kearl (1977, 319) concluded their paper very ambiguously, and surprisingly, by arguing that "For many problems, aggregate production functions are simply too useful to pass up, especially since they can work, as our experiments show. Our parting advice is to handle them the way the old garbage man tells the young garbage man to handle garbage wrapped in plastic bags of unknown provenance: 'gingerly, Hector, gingerly.'"

At least a segment of the profession was aware of the aggregation results. Commenting on this problem in his survey on total factor productivity, Nadiri (1970) argued as follows: "The conclusion […] is that aggregation is a serious problem affecting the magnitude, the stability, and the dynamic changes of total factor productivity. We need to be cautious in interpreting the results that depend on the existence and specification of the aggregate production function…That the use of the aggregate production function gives reasonably good estimates of factor productivity is due mainly to the narrow range of movement of aggregate data, rather than the solid foundation of the function. In fact, the aggregate production function does not have a conceptual reality of its own" (1145–46).

More recently, Felipe, Mehta, and McCombie (2024) also used simulations to show the different estimates obtained when using physical and monetary values. While in the first case, estimates correspond to the true elasticities, in the second case, the estimates tend to the factor shares as a result of the underlying accounting identity. This is what most of the profession has been estimating since Cobb and Douglas (1928) as the data used are values and not physical quantities.

The CCTCs and the aggregation problems are summarized in Felipe and Fisher's (2003) survey. When Felipe and Fisher started corresponding in the mid-1990s, Fisher confessed that he had thought his work on aggregation was lost and had ended up being inconsequential. He was surprised to find somebody working on the issue. Like the CCTCs, the aggregation problem is not mentioned at all and the profession goes seamlessly from the micro-production function, where output, labor, and capital are physical entities, to the aggregate production function, whose arguments are also treated as if they were physical quantities in theoretical expositions, but are not (they cannot be in empirical applications), and without any discussion of aggregation issues (e.g., the production function of tomatoes, the production function of banking services, and the production function of an oil refinery).[6] Some authors even make the point in their expositions that output, labor, and capital are homogenous, hence there is no aggregation problem. Of course!

4. DENIAL 3: THE ACCOUNTING IDENTITY

"As Solow once remarked to me we would not now be concerned with the question [the existence of the aggregate production function] had Paul Douglas found labor's share of American output to be twenty-five per cent and capital's share seventy-five percent instead of the other way around" (Fisher 1969, 572).

"Mr. Shaikh's article is based on misconception pure and simple" (Solow 1974, 121).

[6] Neoclassical growth theory was almost dormant between 1956 and the mid-1980s. It certainly saw some theoretical progress but this was largely based on Solow's original paper. The great revolution came with Romer's work on endogenous growth, and the introduction of increasing returns to scale and imperfect competition. Yet, this did not affect at all the notion of aggregate production function. Somewhat ironically, Ben Bernanke, discussing one of Romer's early papers, made the following point: "It would be useful, for example, to think about the meaning of those artificial constructs "output", "capital", and "labor", when they are measured over such long periods (the Cambridge-Cambridge debates and all that" (Bernanke 1987, 203).

The final blow to the concept of aggregate production function came from another angle, this time the empirical one. Very early on (in the 1940s), some economists suspected that there was something additionally wrong with the production function. The clear exposition of the problem was provided by Phelps Brown (1957). He showed that the *accounting identity* that relates the value of output to the sum of the payments to labor and capital can be rewritten in a form that looks like a production function. The “coefficients” of this rewritten accounting identity are the factor shares. Hence, this is what the regression estimates, with a very high goodness of fit. There is no mystery about these results. We show the algebra of the argument in the appendix because, although we believe it is the most powerful argument to question the legitimacy of the aggregate production function (certainly for empirical work), it is not known to many economists, certainly those educated in the neoclassical tradition.

The reason behind the accounting identity critique was implicit in the CCTCs and the aggregation problem but it was not clear until it came in the form of this (the identity) critique, though its original proponents did not say so explicitly. The problem lies in the nature of the data and it does not rely on aggregation problems; it would exist if value data is used at any level of aggregation.

What is the fundamental issue with the nature of the data? It has been generally accepted since Cobb and Douglas (1928) that a production function is a technological relationship between *physical* inputs and *physical* output. Cobb and Douglas (1928, 139) explicitly referred to the *volume of physical production*. Yet, the reality is that, due to the lack of physical data, researchers use value (or monetary) data for output, the capital stock and intermediate inputs (employment is the only variable expressed as a physical quantity). These series are related through an accounting identity. The appendix shows that the identity can be rewritten in a form that looks like a production function. Yet, it is just an identity. This invalidates all testing and standard interpretations of aggregate production functions. The critique applies if value data is used at any level of aggregation. Adding this to the aggregation problems (i.e., that aggregate production functions “do not exist”) poses an insuperable problem.

Phelps-Brown's (1957) paper was followed by a short note by Simon and Levy (1963) showing how the Cobb-Douglas production function was derived from the accounting identity. Solow was aware of this work because in 1971, he corresponded with Simon (Carter 2011). In these exchanges, Solow vigorously defended the view that the identity and the production function were different. Simon (1979a) disagreed and thought the argument was correct and so important that he referred to it in his Nobel lecture.[7]

Robinson (1970) was of course aware of Phelps-Brown's argument and its relevance. She made the following remark: "It [the aggregate production function] must have needed an even tougher hide to survive Phelps-Brown's article on 'The Meaning of the Fitted Cobb-Douglas Function' than to ward off Cambridge Criticism of the marginal productivity theory of distribution" (Robinson 1970, 317). Robinson (1971, 601–2) was also aware of Fisher's (1971) simulations on the question of what lies behind the apparent success of the Cobb-Douglas function explaining factor shares. She pointed out that this question had been taken up and answered by Phelps-Brown (1957). That the estimated coefficients coincide with the factor shares is ensured by the underlying accounting identity. She concluded: "It seems then, that the controversy is over" (Robinson 1971, 602).

Shaikh (1974) published an important short paper also showing the relationship between the identity and the production function, with particular reference to Solow (1957). Shaikh's argument was similar to one that Hogan (1958) had made. He argued that Solow's procedure to estimate production functions, adjusted for technical progress, was tautological. Solow (1958) admitted this, although differentiating between a good and bad tautology. The importance of this critique initially passed unnoticed.

Shaikh showed that Solow's (1957) exercise simply mirrored the accounting identity. No production function was needed, although Solow assumed that here was one underlying his analysis. Shaikh illustrated his point by created an input-output data set that traced the word HUMBUG, and with the same profit share as in Solow (1957). Shaikh showed that the HUMBUG data could be extremely well represented statistically by a Cobb-Douglas production

[7] See also Simon (1979b).

function having constant returns to scale, neutral technical progress, and marginal products equal to the factor rewards, all embedded in the identity. Solow (1974), a master of the rhetoric, dismissed Shaikh's arguments because, he [Solow] showed that Shaikh's HUMBUG data did not correspond to real data -when he fitted a Cobb-Douglas regression to these data, results were very poor. Shaikh was not allowed by *The Review of Economics and Statistics* to reply at the time. He later published a complete reply as a book chapter (1980).[8] We show and discuss the algebra of Solow (1957) and Shaikh's (1974) arguments in the appendix.

In the ironies of history, Samuelson (1979) also came to rediscover the accounting identity argument. This was a paper written to honor none other than Paul Douglas. Yet, surprisingly, the argument used was that the Cobb-Douglas function that had been introduced 50 years earlier was simply the accounting identity in disguise, the same argument used by previous scholars. Samuelson (1979, 934) himself labeled his findings as "critical." This paper was also ignored by the profession, including Solow. Using the original Cobb and Douglas (1928) data set, Felipe and Adams (2005) showed that Samuelson was correct. In other words, the Cobb-Douglas regression could not give any other result (except for some interesting issues discussed by Felipe and Adams): an elasticity of labor of 0.75 (equal to the labor share) and an elasticity of capital of 0.25 (equal to the capital share).

Solow (1987) came back to the argument a second time claiming that the critique also applies to a micro production function estimated using physical data, which would be wrong because certainly we should be able to learn something from the relationship with such data. Felipe and McCombie (2013, 181-6) showed that Solow was wrong because the critique does not equally work with physical data. The reason is that while the accounting identity for value added is unique and determines the 'production function' regression results (that the estimates are the factor shares), no such (unique) identity exists for physical quantities. In this latter case, one can write an infinite number of artificial identities but the regression with physical quantities will pick up the true elasticities. Felipe, Mehta, and McCombie (2024) also show the algebraic derivation, and provide simulations comparing estimation results with the variables in physical units and in monetary units.

[8] Personal conversation with Anwar Shaikh.

The argument is devastating for the empirical work because it is easy to show that both growth accounting exercises and econometric estimations of production functions (and forms derived from production functions) reduce to manipulations of the identity. Econometric estimation, *if done properly* (see the appendix), must indicate that the elasticities equal the factor shares. This, however, cannot be taken as evidence that factor markers are perfectly competitive, or of constant returns to scale. The estimates are not elasticities in the proper sense of the term and, consequently, production function estimations are not tests of the neoclassical theory,[9] and much less disprove the Marxian theory as Douglas (1976) claimed.

It is worth noting that while the CCTCs became a well-known debate (however neglected) and the aggregation problem (uncontested) was also known, most of the profession has ignored the accounting identity critique. It was never openly debated, much less shown to be wrong. It is also puzzling that the authors who wrote about it did not cite the previous works.[10] If the aggregation issues debated in the Cambridge Controversies and in the Aggregation Problem literature could be ignored for being "theoretical," the accounting identity critique was a *coup de grâce* impossible to snub.

We have spent over 30 years developing the accounting identity argument and shown in multiple papers that many well-established empirical results in the areas of growth (the convergence literature, tests of increasing returns to scale and imperfect competition), productivity (growth accounting and total factor productivity), and employment (estimation of the neoclassical labor-demand function), which the profession has taken for decades as solid and robust, are flawed for being driven by the accounting identity (Felipe and McCombie 2013).

---

[9] In the words of Solow (1974, 121): "…when someone claims that production functions work, he means (a) that they give a good fit to input-output data without the intervention of the factor shares and (b) that the function so fitted has partial derivatives that closely mimic the observed factor shares."

[10] Shaikh (1974) mentioned Phelps-Brown (1957) and Simon and Levy (1963).

5. CONCLUSIONS: NO REPENTANCE

[On the Cambridge Controversy] "…that whole episode now seems to me to have been a waste of time, a playing-out of ideological games in the language of analytical economics" (Solow 1988, 309).

"The main result of that 1957 exercise was startling" (Solow 1988, 313).

Repentance typically requires an admission of guilt for committing a wrong or for omitting to do the right thing. Solow knew the three criticisms, exchanged correspondence with Herbert Simon, and was a colleague of both Frank Fisher's and Paul Samuelson's. Yet he chose to ignore, dismiss, or deny, the critiques. Of course, it is possible that he genuinely had faith in the concept of aggregate production function, and believed in the need to construct models that depended on it. We are skeptical of this possibility, and find it difficult to understand the profession's draw to arguments such as "it is parable", or it is "useful", to justify the use of aggregate production functions and the refusal to acknowledge other growth models.[11]

At the time of receiving the Nobel Prize he still claimed: "…these total-factor-productivity growth calculations require not only that market prices can serve as a rough-and-ready approximation of marginal products, but that aggregation does not hopelessly distort these relationships" (Solow 1988, 314). Two decades later, in one of his final papers, and in reference to the neoclassical growth model, Fisher (2007) was very clear in his title "Is growth theory a real subject?". The reader should not have problems guessing what his answer was.

To finalize, we stress that it is important to set the record straight about how the aggregate production function came to life and how it has survived until today, despite its lack of theoretical foundations and that, empirically, it appears to work simply because it mimics an accounting identity.

[11] Much to our surprise, this is true not only of 'neoclassical' of economists but also of many economists who declare themselves 'non-neoclassical'.

The last chapter in our book (Felipe and McCombie 2013, chapter 12) is dedicated to why the criticisms of the aggregate production function have been ignored. There are many possible reasons for this complete neglect of a serious challenge to economic orthodoxy. The simplest explanation is that the arguments were ignored because its implications, if accepted, were too destructive of conventional economics for neoclassical economists to contemplate. This is a not uncommon initial response in all sciences – the key difference with economic 'science' being that this can also be the final response.

"The conceptual basis for believing in the existence of a simple and stable relationship between a measure of aggregate inputs and a measure of aggregate output is uncertain at best. Yet an aggregate production function is a very convenient tool for theoretically exploring some of the determinants of economic growth, and it has served as a framework for some interesting empirical studies" (Nelson 1964, 575). We think that positions such as this would be unacceptable in other fields. Decades later, an economist of the next generation espoused an identical view: "Arguably the aggregate production function is the least satisfactory element of macroeconomics, yet many economists seem to regard this clumsy device as essential to an understanding of national income levels and growth rates" (Temple 1999, 15). That the economics profession accepts this type of arguments is problematic. As we argue in our book (Felipe and McCombie 2013), they are "Not Even Wrong."

REFERENCES

Acemoglu, D. 2009. *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press.

Aghion, P., and Howitt, P. 1998. *Endogenous Growth Theory*, Cambridge MA: MIT Press.

Barro, R. J., and Sala-i-Martin, X. 1995. *Economic Growth*, New York: McGraw Hill.

Basu, S., and Fernald, J.G. 1997. "Returns to scale in US production: estimates and implications." *Journal of Political Economy* 105(2): 249-83.

Bernanke, B. 1987. "Comment." *NBER Macroeconomics Annual 1987*. MIT Press. Cambridge, MA: 202-5.

Carter, S. 2011. "On the Cobb-Douglas and all that …": the Solow-Simon correspondence over the aggregate neoclassical production function." *Journal of Post Keynesian Economics* 34(2): 255–74.

Clark, J. B. 1893. *The Distribution of Wealth*. New York: Macmillan.

Cobb, C. and P. Douglas. 1928. "A Theory of Production." *The American Economic Review* 18(1): 139–65.

Cohen, A. J. and G. C. Harcourt. 2003. "Retrospectives: Whatever happened to the Cambridge capital theory controversies?" *Journal of Economic Perspectives* 17(1): 199–214.

Douglas, P. 1976. "The Cobb-Douglas Production Function once again: Its History, Its Testing, and Some New Empirical Values." *Journal of Political Economy* 84(5): 903–16.

Felipe, J. and F. G. Adams. 2005. "A THEORY OF PRODUCTION. The Estimation of the Cobb-Douglas Function: A Retrospective View." *Eastern Economic Journal* 31(3): 427–46.

Felipe, J. and F. M. Fisher. 2003. "Aggregation in Production Functions: What Applied Economists Should Know." *Metroeconomica* 54(2–3): 208–62.

Felipe, J., A. Mehta, and J. S. L. McCombie. 2024. "The Estimation of Production Functions with Monetary Values." Working Paper No. 1036 (January). Levy Economics Institute of Bard College, New York.

Felipe, J. and J. S. L. McCombie. 2001. "The CES Production Function, the Accounting Identity, and Occam's Razor." *Applied Economics* 33(10): 1221–32.

———. 2003. "Some Methodological Problems with the Neoclassical Analysis of the East Asian Miracle." *Cambridge Journal of Economics* 54(5): 695–721.

———. 2013. *The aggregate production function and the measurement of technical progress: “Not even wrong.”* Cheltenham, UK and Northampton, MA: Edward Elgar.

———. 2020 “The illusions of calculating total factor productivity and testing growth models: from Cobb-Douglas to Solow and Romer?” *Journal of Post Keynesian Economics* 43(20): 470–513.

Ferguson, C.E. 1969. *The Neoclassical Theory of Production and Distribution*, Cambridge: Cambridge University Press (2nd edition, 1971).

Fisher, F. M. 1965. “Embodied Technical Change and the Existence of an Aggregate Capital Stock.” *Review of Economic Studies* 32: 263–88.

———. 1969. “The Existence of Aggregate Production Functions.” *Econometrica* 37(4): 553–77.

———. 1971. “Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment.” *Review of Economics and Statistics*, 52(4): 305–25.

________. 1971a “Reply” to Robinson, J. The Existence of Aggregate Production Functions: Comment, *Econometrica* 39(2): 405

———. 1992. *Aggregation: Aggregate Production Functions and Related Topics,* edited by J. Monz. New York: Harvester-Wheatsheaf and Cambridge, MA: MIT Press.

———. 2005. “Aggregate Production Functions -A Pervasive, but Unpersuasive, Fairytale.” *Eastern Economic Journal*, Symposium on the Aggregate Production Function, 31(3): 489–91.

———. 2007. “Is growth theory a real subject?” In *Economic growth: New directions in theory and policy*, edited by P. Arestis, M. Baddeley, and J. S. L. McCombie, 8–13. Edward Elgar.

Fisher, F. M., R. M. Solow, and J. M. Kearl. 1977. “Aggregate Production Functions: Some CES Experiments.” *Review of Economic Studies* 44(2): 305–20.

Garegnani, P. (1970): “Heterogeneous Capital, the Production Function and the Theory of Distribution”, *Review of Economic Studies* 37: 407-36.

Hacche, G. 1979. *The Theory of Economic Growth: An Introduction,* London: Macmillan.

Hall, R. E. 1988. “The Relation between Price and Marginal Cost in U.S. Industry.” *Journal of Political Economy* 96(5): 921–47.

———. 1990. "Invariance Properties of Solow's Productivity Residual." In *Growth/Productivity/Employment: Essays to Celebrate Bob Solow's Birthday,* edited by Peter Diamond, chapter 5, 71–112. Cambridge, MA: MIT Press.

Hall, R. E., and C. I. Jones. 1999. "Why do some countries produce so much more output than others?" *Quarterly Journal of Economics* 114 (1): 83–116.

Harcourt, G. C. 1969. "Some Cambridge Controversies in the Theory of Capital" *Journal of Economic Literature* 7(2): 369–405.

———. 1972. *Some Cambridge Controversies in the Theory of Capital*. Cambridge University Press.

Hogan, W.P. 1958. "Technical Progress and Production Functions." *Review of Economics and Statistics* 40(4): 407–11.

Jones, C. I. 1997. "Convergence Revisited." *Journal of Economic Growth* 2: 131–53.

———. 1998 [2002]. *Introduction to Economic Growth* (2^{nd} edition). New York: W.W. Norton

Jones, H.G. 1975. *An Introduction to Modern Theories of Economic Growth*, Middlesex: Nelson.

Kaldor, N. 1957. "A model of economic growth." *The Economic Journal* 67(268): 591–624.

Klein, L. 1946a "Macroeconomics and the Theory of Rational Behavior." *Econometrica* 14(2): 93–108.

———. 1946b. "Remarks on the Theory of Aggregation." *Econometrica* 14(4): 303–12.

Leontief, W. W. 1947a. "Introduction to a Theory of the Internal Structure of Functional Relationships." *Econometrica*. 15(4): 361–73.

———. 1947b. "A Note on the Interrelationship of Subsets of Independent Variables of a Continuous Function with Continuous First Derivatives." *Bulletin of the American Mathematical Society* 53: 343–50.

Mankiw, N. G., D. Romer, and D. N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *The Quarterly Journal of Economics* 107(2): 407–37. doi: 10.2307/2118477.

May, K. 1946. "The Aggregation Problem for a One-Industry Model." *Econometrica* 14(4): 285-98.

———. 1947. "Technological Change and Aggregation." *Econometrica*, 15(1): 51-63.

McCombie, J. S. L. 1998. " 'Are There Laws of Production?' An Assessment of the Early Criticisms of the Cobb-Douglas Production Function." *Review of Political Economy* 10(2): 141–73.

Nadiri, M. I. 1970. "Some approaches to the theory and measurement of total factor productivity: A survey*." Journal of Economic Literature* 8(4): 1137–77.

Nataf, A. 1948. "Sur la Possibilité de Construction de Certains Macromodèles." *Econometrica*, 16(3): 232–44.

Nelson, R. 1964. "Aggregate Production Functions and Medium Range Growth Projections." *The American Economic Review* 54(5): 575–606.

Pasinetti, L. 1959. "On Concepts and Measures of Changes in Productivity." *The Review of Economics and Statistics* 41: 270–86.

———. 2000. "Critique of the Neoclassical Theory of Growth and Distribution." *Moneta e Credito*, Banca Nazionale del Lavoro Quarterly Review, No. 210 (June): 187–232.

Petri, F. 1999. "Professor Hahn on the 'neo-Ricardian' criticism of neoclassical economics," in *Value, Distribution, and Capital. Essays in Honour of Pierangelo Garegnani,* edited by G. Mongiovi and F. Petri, 19–68. Routledge.

________. 2022. "What Remains of the Cambridge Critique? On Professor Schefold's Theses." *Contributions to Political Economy* 41(1): 77–109.

Phelps-Brown, E. H. 1957. "The Meaning of the Fitted Cobb-Douglas Function." *The Quarterly Journal of Economics* 71(4): 546–60.

Pu, S. S. 1946. "A Note on Macroeconomics." *Econometrica,* 14(4): 299–302.

Robinson, J. 1953-54. "The Production Function and the Theory of Capital." *Review of Economic Studies*, 21(2): 81-106.

———. 1955–56. "The Production Function and the Theory of Capital—A Reply." *Review of Economic Studies*, 23(2): 247.

———. 1956. *The Accumulation of Capital*. London: Macmillan.

———. 1970. "Capital Theory Up to Date: A Reply." *Canadian Journal of Economics* 3(2): 309–17.

———. 1971. "The Measure of Capital: The End of the Controversy." *The Economic Journal* 81(323): 597–602

Samuelson, P. 1961–62. “Parable and Realism in Capital Theory: The Surrogate Production Function.” *Review of Economic Studies*,29: 193–206.

———. 1966. “A Summing Up.” *Quarterly Journal of Economics* 80(4): 568–83.

———. 1979. “Paul Douglas’s Measurement of Production Functions and Marginal Productivities.” *Journal of Political Economy* 87(5): 923–39.

Sato, K. 1975. *Production Functions and Aggregation*. North-Holland.

Scott, M. 1989. *A new view of economic growth*. Clarendon Press.

Sen, A. (Ed.). 1970. *Growth economics*. Penguin Modern Economic Readings.

Setterfield, M. (Ed.). 2010. *Handbook of alternative theories of economic growth*. Edward Elgar.

Schefold, B. 2020. “What remains of the Cambridge critique of capital theory, if reswitching and reverse capital deepening are empirically rare and theoretically unlikely?” *European Journal of Economics and Economic Policies: Intervention* 17(2) 2020: 220–40.

Shaikh, A. 1974. “Laws of Production and Laws of Algebra: The Humbug Production Function.” *The Review of Economics and Statistics* 56(1): 115–20

———. 1980. “Laws of Production and Laws of Algebra: Humbug II.” In *Growth, Profits and Property*, edited by E. Nell, 80–95. Cambridge: Cambridge University Press.

Simon, H. 1979a. “Rational Decision Making in Business Organizations.” *The American Economic Review* 69(4): 493–513.

———. 1979a. “On Parsimonious Explanations of Production Relations.” *The Scandinavian Journal of Economics* 81(4): 459–71.

Simon, H. A. and F. K. Levy. 1963. “A Note on the Cobb-Douglas Function.” *Review of Economic Studies* 30(2): 93–94.

Solow, R. 1955–56. “The Production Function and the Theory of Capital.” *Review of Economic Studies*, 23(2): 101–8.

———. 1956. “A contribution to the theory of economic growth.” *The Quarterly Journal of Economics* 70(1): 65–94.

———. 1957. “Technical change and the aggregate production function.” *The Review of Economics and Statistics* 39(3): 312–20.

———. 1958. “Technical Progress and Production Functions: Reply.” *Review of Economics and Statistics* 40(4): 411–13.

———. 1966. “Review of Capital and Growth.” *American Economic Review* 56(5): 1257–60.

———. 1974 “Law of production and laws of algebra: the humbug production function: a comment.” *The Review of Economics and Statistics* 56(1): 121.

———. 1980. “On Theories of Unemployment.” *American Economic Review*, 70(1): 1–11.

———. 1988. “Growth Theory and After.” *American Economic Review* 78(3): 307–17.

———. 1990. *The Labor Market as a Social Institution*. Cambridge, Mass.: Basil Blackwell.

Sraffa, P. 1960. *Production of Commodities by Means of Commodities: A Prelude to a Critique of Economic Theory*. Cambridge: Cambridge University Press.

Stiglitz, J. E. 1974. “The Cambridge-Cambridge Controversy in the Theory of Capital: A View from New Heaven: A Review Article.” *Journal of Political Economy* 82(4): 893–903.

Temple, J. 1999. “The New Growth Evidence.” *Journal of Economic Literature* 37: 112–56.

Wan, H. Y. 1971. *Economic Growth,* New York*:* Harcourt Brace Jovanovich

Weil, D.N. 2005. *Economic Growth*. Englewood Cliffs, NJ: Prentice-Hall.

Wilson, D. J. 2009. “IT and Beyond: The Contribution of Heterogeneous Capital to Productivity.” *Journal of Business and Economic Statistics* 27(1): 52–70.

APPENDIX

The Accounting Identity and the Aggregate Production Function

For a full exposition of the argument, see Felipe and McCombie (2013). Felipe, Mehta, and McCombie (2024) provide simulations. The derivation below is for time series data. There is a similar derivation for cross-sectional data.

The value of constant-price output (V) is definitionally the sum of the payments to labor (the wage bill, denoted W) and capital (all profits, denoted P), that is:

$$V_t \equiv W_t + P_t \quad \text{(A1)}$$

This identity holds at any level of aggregation (firm, sector, economywide) and it is unrelated to a production function. (A1) in growth rates (denoted ^) is: $\hat{V}_t \equiv a_t \hat{W}_t + (1 - a_t)\hat{P}_t$, where $a_t = \frac{W_t}{V_t}$ is the labor share and $1 - a_t = \frac{P_t}{V_t}$ is the capital share.

Given that we can write the wage bill as the product of an average wage rate times employment ($W = wL$), and total profits as the product of an average profit rate times the "value of capital," ($P = rJ$), (A1) becomes $V_t \equiv w_t L_t + r_t J_t$, and the growth rate of value added is:

$$\hat{V}_t \equiv a_t \hat{w}_t + (1 - a_{t)}\hat{r}_t + a_t \hat{L} + (1 - a_t)\hat{J}_t \quad \text{(A2)}$$

Equation (A2) continues being an identity, and has nothing to do with the existence or not of an aggregate production function and market conditions such as perfectly competitive markets, or the degree of returns to scale. It likewise does not depend on whether market conditions are perfectly competitive markets, or on the degree of returns to scale. Equation (A2) can be also be written the way it appears in the neoclassical literature with a user cost of capital (υ) instead of the average profit rate (r). In this case, we would have to add to the identity a variable that would denote extra (monopoly) profits (Ω), i.e., $V_t \equiv w_t L_t + \upsilon_t J_t + \Omega_t$. This does not change the nature of the argument.

Equation (A2) can be rewritten in levels as some specific form of $V_t \equiv A_t F(L_t, J_t)$, which looks like a production function but is just the identity in a different form. The easiest case to show is when factor shares happen to be constant (i.e., $a_t = a; 1 - a_t = 1 - a$). This constancy is not the result of the fact that the underlying technology of the economy is Cobb-Douglas, as Fisher (1971) remarked. It could be the result of firms setting prices as a constant mark-up on unit labor costs.

If the shares of labor and capital happen to be constant, equation (A2) becomes, after integration:

$$V_t \equiv B w_t^a r_t^{1-a} L_t^{1-a} J_t^{1-a} \qquad \text{(A3)}$$

where B is the constant of integration. This expression is the original identity rewritten in a different form. Different paths of the factor shares over time would lead to different forms that would look like the standard production functions. This is discussed in greater depth in Felipe and McCombie (2013, chapter 5). See also Felipe and McCombie (2001) for a discussion of the CES, and Felipe and McCombie (2003) for the case of the translog function.

It is obvious that nobody would estimate equation (A3). What researchers estimate instead is:

$$V_t = C L_t^{\gamma} J_t^{\delta} e^u \qquad \text{(A4)}$$

an approximation to (A3) and, hence, to the identity (A1). The whole problem lies in properly proxying $B w_t^a r_t^{1-a}$, akin to the game "Where's Waldo?" The problem is also a case of omitted-variable bias though in this the econometrician knows the omitted variable. See Felipe, Mehta, and McCombie (2024).

This derivation leads to the following results: (i) γ will tend to a, and δ to $(1 - a)$; (ii) a very high fit, potentially 1; and (iii) a "correctly" estimated production function has to indicate always that the estimated elasticities are equal to the factor shares, which of course add up to 1. If results deviate significantly from these, it will simply mean that the approximation of $B w_t^a r_t^{1-a}$ is poor.

It is a matter of finding "Waldo," the correct approximation to return to the identity. A typical specification of the production function (A4) with time series data has traditionally included a time trend under the argument that it proxies technical progress and its coefficient is an estimate of the rate of technical progress. Seldom does this work because the time trend is not a god proxy for $Bw_t^a r_t^{1-a}$, the weighted average of the wage and profit rates.

Versions of this derivation serve to explain Jones's (1998 [2002]) work on total factor productivity (Felipe and McCombie 2013, chapter 6), Mankiw, Romer, and Weil's (1992) revival of Solow's model and the conditional convergence regression (Felipe and McCombie 2013, chapter 7), or Hall's (1988, 1990) estimations of the mark-up and the degree of market power (Felipe and McCombie 2013, chapter 10). The expressions these authors estimated were derived from production functions. We show that they all can be derived as algebraic transformations of (A1). See also Felipe and McCombie (2020). The argument also serves to explain estimates of the neoclassical labor demand function (e.g., the negative relationship between employment and wages). In the literature, it is derived from the production function. We show that the same expression can be derived by transforming the identity (Felipe and McCombie 2013, chapter 11).

Finally, equation (A2) also provides an explanation of the "Solow" residual (SR), or total factor productivity (TFP) growth. Equation (A2) can be rewritten as: $SR_t \equiv \hat{V}_t - a_t\hat{L} - (1 - a_t)\hat{J}_t \equiv a_t\hat{w} + a_t\hat{r}_t$. This is tautologically true, but unless one shows that there is an aggregate production function, it is impossible to argue that $SR_t \equiv \hat{V}_t - a_t\hat{L} - (1 - a_t)\hat{J}_t$ measures the rate of technical progress. At best, and taken at face value, $SR_t \equiv a_t\hat{w} + a_t\hat{r}_t$ can be interpreted as a measure of distributional changes.

"Technical Change and the Aggregate Production Function" (Solow 1957) and Shaikh's (1974, 1980) HUMBUG production function

This discussion allows us to analyze, in some detail, Solow (1957). Solow's approach consisted of estimating the production function in intensive form, i.e., as $(\frac{V_t}{L_t}) \equiv A_tF(L_t/J_t)$. In order to estimate this function, Solow argued that he needed to "deflate" the function in order to correct for its upward shift over time, supposedly due to the rate of technical progress. In other words, the general form of the function to be estimated was $(\frac{V_t}{L_t})/A_t \equiv F(L_t/J_t)$.

Consequently, Solow first needed to construct an index of A. The fact that Solow (1957) was not aware he was only manipulating the accounting identity may be inferred from the fact that he estimated five separate regressions of this general functional form. These were all variants of the specification $q_t = c + bk_t$, where $q_t = (\frac{y_t}{A_t})$, $y_t = (\frac{V_t}{L_t})$, and $k_t = (\frac{J_t}{L_t})$. In all cases, the correlation coefficient was 0.99 (not surprising as we show below). Solow, however, was surprised about the goodness of fit. As can be seen, Solow used A, the level of TFP and constructed as an index $A_t = V_t/F(\frac{J_t}{L_t})$, to "deflate" the left-hand side of his regressions. In effect, he first calculated the annual growth of TFP as $\hat{A}_t = \hat{y}_t - a_t\hat{k}_t$.[12] He assigned a value of one to A in the first year and constructed the rest of the annual series of the index by using the subsequent growth rates of A.

To further understand Solow's procedure, recall that (A3) can be written as $(\frac{V}{L}) \equiv y \equiv Mk^{1-a}$ where $M \equiv Bw^a r^{1-a}$. This implies $q \equiv (\frac{y}{A}) \equiv k^{1-a}$. Therefore, using the constructed index for A and then regressing lnq on lnk must yield a near perfect fit (and likewise, regressing q on k will also give a close fit) as the relationship is merely estimating a reformulation of the accounting identity. The procedure is self-evidently tautological.

We, consequently, find it surprising that Solow's method to construction the TFP series as $A = V/F(J, L)$ has uncritically survived the test of time. Surprisingly, Solow (1958) did admit that his method was based on a tautology, but referred to it as a "good" tautology.[13] We note that the

12 Solow actually used a slightly different approach to calculate growth rates, but it makes no significant difference to the argument.

13 Solow (1958) is a reply to Hogan (1958), who pointed out the nature of the tautological procedure. Hogan noted that it is obvious that the coefficient b in the regression $lnq = c + blnk$ will be the capital share. Solow claimed that, if the capital share had been exactly constant, then indeed the procedure would have been a bad tautology. However, insofar as the capital share showed some variation, he argued that it was a good tautology. We think this was not a particularly convincing argument.

same method to obtain a series for A was used four decades later by Jones (1997) and Hall and Jones (1999), with the same problematical consequences (see Felipe and McCombie 2020).

As indicated in the main text, Shaikh (1974) (like Hogan (1958)), argued that Solow's (1957) procedure to estimate production functions, adjusted for technical progress, was tautological. Shaikh showed that Solow's exercise simply mirrored the accounting identity. Shaikh illustrated his point by created an input-output data set that traced the word HUMBUG and showed that these data could be extremely well represented statistically by a Cobb-Douglas production function with constant returns to scale, neutral technical progress, and marginal products equal to the factor rewards. The issue was that all these results were embedded in the identity. Inexplicably, Solow (1974) dismissed the argument because when he fitted a Cobb-Douglas with a liner time trend to the HUMBUG data, results were very poor: "If this were the typical outcome with real data, we would not now be having this discussion. The humbug seems to be on the other foot."[14]

Shaikh (1980) explained clearly why Solow (1974) obtained very poor results when he fitted the Cobb-Douglas to the HUMBUG data set: the linear time trend was a very poor proxy for the weighted average of the wage and profit rates. Shaikh, instead of the linear time trend, added a trigonometric function that captured the weighted average of the wage and profit rates -nothing in neoclassical economics says that technical progress needs to be proxied by a linear time trend. Back to the identity simply as a result of the laws of algebra.[15]

Solow's (1957) other key result was that the residual measure of TFP growth through $SR_t \equiv \hat{V}_t - a_t\hat{L} - (1 - a_t)\hat{J}_t$ accounted for almost 90 percent of the overall non-farm sector growth of the US during 1909-49, while the remaining 10 percent was the result of factor accumulation

[14] The estimated regression was: $\ln\left(\frac{V}{L}\right) = -0.1409 + 0.00352 * t - 0.33071 * (\frac{J}{L})$, with $R^2 = 0.0052$ and all coefficients statistically insignificant.

[15] This is the complex function that Shaikh added to the Cobb-Douglas to proxy the weighted average of the growth rates of the wage and profit rate ($SR_t \equiv a_t\hat{w} + a_t\hat{r}_t$): $\hat{A} = a_0 + a_1 * t + \sum_{i=1}^{3} b_i \left[\cos\left(\frac{c_i \pi t}{2}\right) + d_i \sin\left(\frac{e_i \pi t}{2}\right)\right]$, where $a_0 = 0.8565$, $a_1 = -3.966 \times 10^{-3}$, $b_1 = -0.0325$, $b_2 = 0.0435$, $b_3 = 0.0206$, $c_1 = 0.4$, $c_2 = 0.6$, $c_3 = 0.6$, $d_1 = 0.035$, $d_2 = -0.032$, $d_3 = -0.0295$, $e_1 = 0.5$, $e_2 = 0.8$, $e_3 = 0.4$.

growth. As noted above, Solow (1988) still labeled this result as "startling". Kaldor (1957) and Pasinetti (1959) explained the conceptual problems involved in the decomposition of output growth into a part attributable to factor accumulation and another one attributable to the residual (productivity).

However, as in the case of the estimation of the production function, a consideration of the identity (A2) shows that Solow's result becomes self-evident. Given the stylized fact that the rate of return is roughly constant (as it was in the US during the period under consideration), then $SR_t \simeq a_t \widehat{w}_t$. As factor shares were also roughly constant, this result implies that $\widehat{w}_t \simeq \widehat{V}_t - \widehat{L}_t$. Then it follows that $SR_t \simeq a(\widehat{V}_t - \widehat{L}_t)$. Consequently, as a took a value of about 0.70 to 0.75 in the national accounts, it follows that SR_t (TFP growth) *must* account for about three quarters of the growth of output per worker. This is merely due to the algebra of the accounting identity. Moreover, given that employment growth is small compared with that of output, TFP growth will also explain a similar proportion of output growth. In light of this, it may be questioned whether this result is particularly surprising. In fact, it is surprising that anyone should find it startling.