

AGGREGATION IN PRODUCTION FUNCTIONS: WHAT APPLIED ECONOMISTS SHOULD KNOW

Jesus Felipe and Franklin M. Fisher*

Asian Development Bank and Massachusetts Institute of Technology

There is no subject so old that something new cannot be said about it.

(Dostoevsky)

Glendower: I can call spirits from the vasty deep

Hotspur: Why, so can I, or so can any man;

But will they come when you do call for them?

(Shakespeare, Henry IV, Part I, Act III, Scene I)

ABSTRACT

This paper surveys the theoretical literature on aggregation of production functions. The objective is to make neoclassical economists aware of the insurmountable aggregation problems and their implications. We refer to both the Cambridge capital controversies and the aggregation conditions. The most salient results are summarized, and the problems that economists should be aware of from incorrect aggregation are discussed. The most important conclusion is that the conditions under which a well-behaved aggregate production function can be derived from micro production functions are so stringent that it is difficult to believe that actual economies satisfy them. Therefore, aggregate production functions do not have a sound theoretical foundation. For practical purposes this means that while generating GDP, for example, as the sum of the components of aggregate demand (or through the production or income sides of the economy) is correct, thinking of GDP as $GDP = F(K, L)$, where K and L are aggregates of capital and labor, respectively, and $F(\bullet)$ is a well-defined neoclassical function, is most likely incorrect. Likewise, thinking of aggregate investment as a well-defined addition to

* We are grateful to the participants at the conference 'Old and New Growth Theories: an Assessment' (Pisa, Italy, 5–7 October 2001) and to the participants at the CEPA workshop at the New School University (New York), especially to Per Gunnar Berglund, for their comments. Special gratitude goes also to Gary Mongiovi and Fabio Petri for their detailed observations, in particular regarding the Cambridge–Cambridge debates. Two anonymous referees made useful suggestions. Jesus Felipe acknowledges very helpful conversations with Anwar Shaikh and John McCombie throughout the years, and with Nazrul Islam and Joy Mazumdar, who often played devil's advocate over very pleasant discussions. Jesus Felipe also acknowledges financial support from the Center for International Business Education and Research (CIBER) and from the Georgia Institute of Technology Foundation, Georgia Institute of Technology (Atlanta, USA), where he was a faculty member between 1999 and 2002. This paper represents the views of the authors and should not be interpreted as reflecting those of the Asian Development Bank, its executive directors, or the countries that they represent. The usual disclaimer applies.

'capital' in production is also a mistake. The paper evaluates the standard reasons given by economists for continuing to use aggregate production functions in theoretical and applied work, and concludes that none of them provides a valid argument.

1. INTRODUCTION

With the surge of the new neoclassical endogenous growth literature (Barro and Sala-i-Martin (1995), Aghion and Howitt (1998)) in the 1980s there has been a renewed interest in growth and productivity, propagated by the development of new models, the availability of large data sets with which to test the new and the old growth theories (e.g. Mankiw *et al.*'s (1992) use of the Summers and Heston data set), and episodes of growth that need to be explained and which have led to important debates (e.g. the East Asian Miracle).

The pillar of these growth models, like that of the older neoclassical growth model (Solow (1956, 1957)), is the neoclassical aggregate production function, a relationship that is intended to describe the technology at the aggregate level. The problem is that the aggregate or macro production function is a fictitious entity.¹ At the theoretical level it is built by adding micro production functions. However, there is an extensive body of literature that showed that aggregating micro production functions into a macro production function is extremely problematic. This is the subject of the so-called aggregation literature and the issue at hand is referred to as the *aggregation problem*. Its importance lies in the fact that without proper aggregation we cannot interpret the properties of an aggregate production function.

An aggregate production function is a function that maps aggregate inputs into aggregate output. But what exactly does this mean? Such a concept has been implicit in macroeconomic analyses for a long time. However, it has always been plagued by conceptual confusions, in particular as to the link between the underlying micro production functions and the aggregate macro production function, the latter thought to summarize the alleged aggregate technology. Mankiw (1995), for example, defined an aggregate production function as a 'mapping from quantities of inputs into a quantity of output' (Mankiw (1995, p. 281)). This is certainly not incorrect, except for the fact that at the aggregate level the aggregation conditions matter. And Mankiw (1997, p. 103) referred to the aggregate production function as the 'economy's

¹ The standard practice, paradoxically, is to argue that the production function in the theoretical model is a micro production function. However, the empirical evidence and examples provided tend to be macro (e.g. Romer (1987)).

production technology'. This conceptualization is, on the other hand, rather questionable in the light of the aggregation problems.

To understand what an aggregate production function is one must understand what the aggregation problem involves. The issue at stake is how economic quantities are measured, in particular those quantities that represent by a single number a collection of heterogeneous objects; in other words, what is the legitimacy of aggregates such as investment, GDP, labor and capital? To take a simple example, the question is as follows. Suppose we have two production functions $Q^A = f^A(K_1^A, K_2^A, L^A)$ and $Q^B = f^B(K_1^B, K_2^B, L^B)$ for firms A and B, where $K_1 = K_1^A + K_1^B$, $K_2 = K_2^A + K_2^B$ and $L = L^A + L^B$ (K refers to capital—two types—and L to labor, assumed homogeneous). The problem is to determine whether and in what circumstances there exists a function $K = h(K_1, K_2)$ where the aggregator function $h(\bullet)$ has the property that $G(K, L) = G[h(K_1, K_2), L] = \Psi(Q^A, Q^B)$ and the function Ψ is the production possibility curve for the economy.²

It will be noted that above we have already assumed that a production function exists at the level of the firm. In one sense, this is guaranteed. As we show below, if an entity assigns the use of its various factors to different techniques of production so as to maximize output, then maximized output will depend only on the total amount of such factors, and that dependence can be written as a functional relationship (differentiability, of course, will not be guaranteed. We assume differentiability below only for convenience). That does not mean that one can aggregate over factors, and that is one part of the aggregation problem; the other one, more emphasized in this paper, is aggregation over firms—aggregation to the case where factors are not all efficiently assigned.

The paper is somewhat nihilistic. It is a non-technical survey that has two objectives. The first of these is to summarize the existing literature on the conditions under which an aggregate production function exists, i.e. the aggregation problem. There are two strands of theoretical work that deal with this issue. One falls under the umbrella of the Cambridge–Cambridge debates. The other is the aggregation literature *per se* (as described in the previous paragraph). Both are distinct issues and provide different views of the

² Two clarifications are important. First is the philosophical question of how far down or up one has to go to begin talking about the aggregation problem. From a theoretical and mathematical point of view, one can disaggregate *ad infinitum*. Second, often in the paper we refer explicitly or implicitly to macro issues as those that involve aggregation. However, most economists understand macroeconomics to involve problems concerning unemployment, inflation, economic growth etc., i.e. the fact that aggregation is involved is not the usual criterion for distinguishing between macroeconomics and microeconomics.

cause and nature of the aggregation problem, but intersect at one juncture: the problem of the measurement of capital. The second objective of the paper is to discuss the implications that these two lines of work have for applied work, and to draw lessons for applied economists.³ The importance of the topic at hand resides in the fact that it seems that all the new growth literature (as well as literature in other areas) has overlooked the *old* aggregation problem and its implications. The result has been that aggregate production functions have made an untroubled reappearance in mainstream macroeconomics since the 1980s.

The review of the literature is carried out from the point of view of the applied neoclassical economist.⁴ We have kept the technical aspects to a minimum, and have concentrated on pointing out the important results on aggregation of production functions that applied economists should be aware of and why. The main target of the survey is the new generation of neoclassical macroeconomists, since, in the light of the negative conclusions derived from the Cambridge debates and from the aggregation literature, one cannot help asking *why* they continue using aggregate production functions. Sylos Labini recently wrote: 'It is worth recalling these criticisms, since an increasing number of young and talented economists do not know them, or do not take them seriously, and continue to work out variants of the aggregate production function and include, in addition to technical progress, other phenomena, for example, human capital' (Sylos Labini (1995, p. 487)). The younger generation of economists remains ignorant of these problems, with the consequence that bad habits and bad science breed bad economics and bad policy advice. This position appears in a recently published survey on the new growth theories. Jonathan Temple, the author, concluded: '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, p. 15), italics added). Is this a good enough reason to use an unsatisfactory device? We hope that these pages will make the recent generation of

³ In this paper we discuss the aggregation literature in greater detail.

⁴ Given that we write from a neoclassical perspective, we assume that a neoclassical production function exists at the firm level (see above). However, the Cambridge, UK, economists argued that this assumption is unwarranted. The Cambridge, UK, work was also grounded on microeconomic theory. The basis was the notion of the 'choice of technique', which derived from the assumption that each firm chooses among alternative methods of production the one which yields the highest expected rate of profit. This generates the wage-profit frontier, i.e. the inverse relationship between wage and profit rates, upon which the Cambridge, UK, critique was based (Kurz (1990)).

economists aware of the very serious problems that surround the aggregates of output, labor and capital when thought of as related by an aggregate production function.

One of the first economists to offer a systematic treatment of the aggregation problem in production functions was Klein (1946a, 1946b). Klein argued that the aggregate production function should be strictly a technical relationship, akin to the micro production function, and objected to utilizing the entire micromodel with the assumption of profit-maximizing behavior by producers in deriving the production functions of the macromodel. He argued that 'there are certain equations in microeconomics that are independent of the equilibrium conditions and we should expect that the corresponding equations of macroeconomics will also be independent of the equilibrium conditions. The principal equations that have this independence property in microeconomics are the technological production functions. The aggregate production function should not depend upon profit maximization, but purely on technological factors' (Klein (1946b, p. 303)).

Klein's position, however, was rejected altogether by May (1947), who argued that even the firm's production function is not a purely technical relationship, since it results from a decision-making process.⁵ Thus, the macro production function is a fictitious entity, in the sense that there is no macroeconomic decision-maker that allocates resources optimally. The macro function is built from the micro units assumed to behave rationally. Years later, Fisher (1969a) took up the issue again, and reminded the profession that, at any level of aggregation, the production function *is not* a description of what output can be produced from given inputs. Rather, the production function describes the maximum level of output that can be achieved if the given inputs are efficiently employed.

The view is that engineering yields the technology set—the combinations of inputs and outputs that are technologically feasible. The (economic)

⁵ May argued: 'The aggregate production function is dependent on all the functions of the micromodel, including the behavior equations such as profit-maximization conditions, as well as upon all exogenous variables and parameters. This is the mathematical expression of the fact that the productive possibilities of an economy are dependent not only upon the productive possibilities of the individual firms (reflected in production functions) but on the manner in which these technological possibilities are utilized, as determined by the socio-economic framework (reflected in behavior equations and institutional parameters). Thus the fact that our aggregate production function is not purely technological corresponds to the social character of aggregate production. Moreover, if we examine the production function of a particular firm, it appears that it, too, is an aggregate relation dependent upon nontechnical as well as technical facts. It tells us what output corresponds to total inputs to the firm of the factors of production, but it does not indicate what goes on within the firm' (May (1947, p. 63)).

production function, however, describes the efficient frontier of that set. It embodies 'best-practice' use of the available input and output combinations.

From the point of view of the applied practitioner, aggregate production functions are estimated for the following purposes: (i) to obtain measures of the elasticity of substitution between the factors, and the factor-demand price elasticities—such measures are used for predicting the effects upon the distribution of the national income of changes in technology or factor supplies (Ferguson (1968)); (ii) to apportion total growth into the accumulation of factors of production and technical change between two periods (Solow (1957)); (iii) to test theories and quantify their predictions (Mankiw *et al.* (1992)); and (iv) to address policy issues (Jorgenson and Yun (1984)). Thus, from this point of view the most important question is the following: is the aggregate production function a summary of the 'aggregate' technology? That is, suppose one estimates econometrically an aggregate production function: are the estimated coefficients (i.e. input elasticities, elasticity of substitution) the technological parameters?⁶

Why do economists use aggregate production functions despite the results reviewed in this paper? It seems that since these results are rather inconvenient for an important part of neoclassical macroeconomics, the profession has chosen to ignore them and feels comfortable with the standard justifications, *clichés*, for using them. These are the following. One, based on the methodological position known as instrumentalism, is that aggregate production functions are constructed by analogy with the micro production functions and that their validity is an empirical issue (Ferguson (1971)). Furthermore, since aggregate production functions appear to give empirically reasonable results, why shouldn't they be used? Second, and following Samuelson (1961–62), aggregate production functions are seen as parables. Finally, it has been argued that, for the empirical applications where aggregate production functions are used (e.g. growth accounting and econometric estimation), there is no other choice. An evaluation of these answers is provided at the end of the paper.

The rest of the paper is structured as follows. Section 2 provides a summary of Joan Robinson's complaint about the units in which *aggregate capital* was measured in aggregate production functions, the complaint that led to the Cambridge–Cambridge capital controversies of the 1950s and 1960s. The Cambridge–Cambridge debates provided the background for the second-generation literature on aggregation, although the latter, in fact, provided a

⁶ There are other purposes for which the measurement of (aggregate) capital is a crucial issue, such as the investment function, the consumption function, budgeting and planning, and connections with the rest of the economy (Usher (1980, pp. 3–5)).

completely different view of the aggregation problem.⁷ Section 3 summarizes the initial work on aggregation, referred to as the first generation. Section 4 discusses the important Leontief, Nataf and Gorman aggregation theorems. Sections 5, 6, 7 and 8 discuss Fisher's, Sato's and Gorman's works, the second generation.⁸ These works look at the aggregation problem from different points of view and suggest alternative aggregation conditions. Section 9 returns to the question of why, in the light of the Cambridge–Cambridge debates and the aggregation results, economists continue using aggregate production functions. Section 10 provides a summary of what applied economists should know about aggregate production functions, and provides an explanation for what the parameters of putative aggregate production functions mean consistent with the non-existence of an aggregate production function.

It is important to keep in mind and stress that the aggregate production function is the result of two types of aggregation. One is aggregation over multiple inputs or outputs (i.e. different types of labors into one labor; different types of capital into one capital; different types of output into one output). The other is aggregation over firms. To motivate the question, think of the following problem. Suppose the technology of two firms is Cobb–Douglas. Can they simply be added up to generate the aggregate production function? The answer is no. What if the restriction that both production functions have constant returns to scale is added? Not yet. Are further restrictions needed? Yes.

2. THE CAMBRIDGE DEBATES AND JOAN ROBINSON'S COMPLAINT

One of the first endogenous growth papers containing empirical work was Romer (1987). In his discussion of the paper, Ben Bernanke aired the following concern: 'It would be useful, for example, to think a bit about the meaning of those artificial constructs "output," "capital," and "labor," when they are measured over such long time periods (*the Cambridge–Cambridge debate and all that*)' (Bernanke (1987, p. 203), italics added). The so-called Cambridge–Cambridge controversies are a series of debates that took place

⁷ The authors who developed the second-generation aggregation theory clearly indicate the connection between their work and the Cambridge–Cambridge capital controversies (see, for example, Sato (1975), Fisher (1993)).

⁸ Excellent summaries of the literature on aggregation are Green (1964), Sato (1975), Brown (1980) Diewert (1980) and Fisher (1969a, 1993), from which much of the material in this paper draws.

mostly during the 1950s, 1960s and 1970s between Joan Robinson and her colleagues at Cambridge, UK, and Paul Samuelson and Robert Solow and their colleagues at Cambridge, USA.

What were the Cambridge–Cambridge capital controversies about? Summarizing the exchanges between the two sides is a rather complicated task, and our principal focus here is on the implications of the aggregation literature for applied economists rather than for theoreticians. Harcourt (1969, 1972) provides an excellent summary and discussion of the literature. But if we had to do it in one sentence, we must say that, in the final analysis, the debates were about different value theories, the classical (Smith, Ricardo, Marx) and the neoclassical (Böhm-Bawerk, Jevons, Clark, Wicksteed, Wicksell, Marshall, Walras), and how those theories explain prices and distribution. As we shall see, 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.

Traditionally, economists had distinguished between two notions of capital. (i) Capital could be conceived as a fund of resources that could be shifted from one use to another relatively easily. This is a fund embodying the savings accumulated in time. This is what can be called the *financial* conception of capital. (ii) Capital could be conceived as a set of productive factors, a list of heterogeneous machines, stocks etc., that are embodied in the production process and designed for specific uses, all specified in physical terms. This is what may be called the *technical* conception of capital.

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 study of production (Clark (1893)). The only one of these two concepts that could be unambiguously measured was the fund, since it was money. Real capital, on the other hand, could not be measured, since it consisted of a set of heterogeneous machines and materials. 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 produced. 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). Later on, this idea of expressing the endowment of capital goods as a single quantity, either as an *amount of value* or in the context of

a *single commodity world*, took off and became standard. This materialized in the widespread aggregate production function $Q = F(K, L)$, where Q , K and L are aggregates of output, capital and labor.⁹

2.1 Joan Robinson's complaint¹⁰

What was the problem Joan Robinson pointed out in the early 1950s and which led to the Cambridge–Cambridge capital controversies? She complained 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 Joan Robinson (1971a), they 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 (e.g. Solow (1957)). As a consequence, it appeared that the division of the aggregate product between labor and capital *could be* explained in terms of marginal productivities.¹¹

In a seminal paper, Joan Robinson (1953–54) asked the 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 (investment goods), each having specific technical characteristics, it is *impossible* to express the stock of capital goods as a homogeneous physical entity.

⁹ Of course economists at the time realized that capital consisted of heterogeneous capital goods (Cobb and Douglas (1928)). But their aggregation into a more or less homogeneous aggregate was considered an index number problem that could be solved. As indicated in section 5.1, the aggregation and index number problems are different.

¹⁰ The Cambridge capital controversies are plagued with confusions (Petri (1999)). We thank Fabio Petri and Gary Mongiovi for pointing this out. The most important one 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 aggregatability is needed for the validity of the neoclassical approach to (macro) distribution?

¹¹ Blaug indicates that ‘The notion that the functional distribution of income may be explained simply by invoking the principles of marginal productivity, as enshrined in an aggregate production function of the simple Cobb–Douglas variety, was broached virtually for the first time in Hicks’s *Theory of Wages* (1932), in particular Chapter 6 of that book’ (Blaug (1993, p. 171)).

Robinson claimed that it followed that only their values can be aggregated. Such a value aggregate, however, is not independent of the rate of profit and thus of income distribution.¹² The problem is as follows. Suppose there are n types of capital goods, denoted K_i^t , $i = 1, \dots, n$. The price of each capital good in terms of some chosen base year is P_i^0 . Then, the question is whether a measure of the total stock of capital can be defined as $K^t = \sum_{i=1}^n P_i^0 K_i^t$. In the words of Usher: 'Can time series of quantities of capital goods be combined into a single number that may be interpreted as "the" measure of *real capital* in the economy as a whole?' (Usher (1980, p. 2), italics added).¹³

This heterogeneity of capital became an important part of the controversies (at least initially and certainly for Joan Robinson). The Clarkian concept of capital, conceived as a fairly homogeneous and amorphous mass which could take different forms, Joan Robinson argued, cannot serve in a macroeconomic production function *à la* Cobb–Douglas because it is essentially a monetary value. She claimed that, although labor is not a homogeneous input, in principle it can be measured in a technical unit, man-hours of work. 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 (although see below). But what about (aggregate) capital? Joan Robinson argued that the statistics of capital used in practice had nothing to do with the previous two notions of capital. Such statistics are in dollars; and however they are deflated (to convert them into constant dollars), they continue being money, a sum of value. Therefore

¹² This problem had been mentioned by Wicksell in the nineteenth century. He claimed that the value of capital is not an appropriate measure of the aggregate capital stock as a factor of production except under extremely restrictive conditions. He was aware that the partial derivatives of any function in which capital appears in value terms cannot be used for determining the productive contributions of the factors, and hence distribution (Burmeister (1990), Pivetti (1990)). Burmeister has proved that a necessary and sufficient condition for the existence of an index of capital and a neoclassical production function defined across steady-state equilibria is that the so-called real Wicksell effect (i.e. the price-weighted sum of the changes in the physical quantities of different capital goods) be negative (Burmeister (1990)).

¹³ Usher (1980, p. 1) distinguishes between 'real capital' and the 'value of capital in current dollars', implicitly legitimizing Joan Robinson's question. A further question to be discussed refers to what this definition of aggregate real capital measures. Usher (1980, pp. 13–18) discusses four interpretations: (1) instantaneous productive capacity; (2) long-run productive capacity; (3) accumulated consumption forgone; (4) real wealth. Choice among these concepts of real capital depends on the purpose of the time series. Usher (1980, p. 18) indicates that the notion of long-run capacity is the one to use in an investment function; capital as wealth is the one to use in a consumption function; instantaneous productive capacity is the appropriate notion of capital for estimating production functions; cumulative consumption forgone is the most appropriate for a growth accounting exercise; and for the computation of capital–output ratios, the most appropriate measure is probably also instantaneous productive capacity.

she asked rhetorically: 'How can this be made to correspond to a physical factor of production?' (Robinson (1971a, p. 598)); and: 'The well-behaved production function in labor and stuff was invented, I think, to answer the question: What is a quantity of capital?' (Robinson (1975, p. 36)).¹⁴

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 and demand for factors of production, then the latter must each have a measure.¹⁵ 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 for capital as an amount of finance. Thus, capital has no natural unit, akin to those of 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 woman-hour of labor by Joan Robinson really the same as one of Queen Elizabeth II or one of Britney Spears in terms of productivity?

The argument of the Cambridge, UK 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 know, first, 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 are known *ex ante*. The consequence is that aggregate capital, the aggregate production

¹⁴ Thus, aggregate capital, viewed as a fund of money, has a homogeneous unit, but in that form it is not productive. To be productive, it must be transformed into produced means of production (i.e. a vector of capital goods). It is in this form that capital does not have a homogeneous unit.

¹⁵ The endowment of capital must be specified as a datum, independently of distribution, in order to determine the profit rate. However, since the endowment of capital measured as a value aggregate depends upon the profit rate it cannot be taken as parametric for the determination of the profit rate.

function and the marginal products of the factors can only be defined when the rate of profit is given, and it implies that they cannot be used to build a theory of the rate of profit or distribution.

Joan Robinson's critique led to a series of intertwined debates that lasted two decades and went far beyond the original question. The main debates were as follows. (i) One was around 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. 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. (ii) Another important debate involved the so-called *Wicksell effects*:¹⁶ 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) A third one was the relationship between savings and investment.¹⁷ (iv) A fourth one centered on the problems of *reswitching* and *reverse capital-deepening*. 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

¹⁶ In an economy with n different types of capital goods, the value of the capital stock is $V = \sum_{i=1}^n P_i K_i$, where P_i denotes the price of the i th capital good. The Wicksell effect is the change in the value of the capital stock from one steady state to another, i.e. dV/dr . The Wicksell effect is the sum of the price and real Wicksell effects: $dV/dr = \sum_{i=1}^n (dP_i/dr) K_i + \sum_{i=1}^n P_i (dK_i/dr)$.

¹⁷ This is related to the different solutions given by each side of the Atlantic to the knife-edge problem in the Harrod-Domar model. The Cambridge, UK, side proposed a model where capitalists and workers had different saving propensities. On the other hand, Solow (1956) proposed a model that used the aggregate production function. See Pasinetti (1974) and Solow (1988, 1994).

capital-intensive technique.¹⁸ Finally (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. 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 micromodel. Hence, this part of the Cambridge–Cambridge debate is a consequence of the aggregation problem properly considered (although in section 9 we indicate that there is an important disagreement on this issue).

If attention is restricted to the question of aggregate capital and the aggregate production function, the answers to Joan 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) (discussed in section 9).¹⁹ The other solution was to search for the technical conditions under which aggregation is possible (discussed in sections 3–8). This is the work developed by Fisher, *inter alios*. The aggregation problem became the search for the

¹⁸ The possibility of reswitching was originally discovered by Sraffa, who published his results in 1960.

¹⁹ We also mention the solution proposed by Champernowne (1953–54). This was to construct a chain index of quantity of capital in which capital goods could be measured such that the conventional production function could be constructed and the marginal productivity theory could be preserved. Joan Robinson (1953–54) had proposed to measure capital in units of labor. This, Champernowne argued, while not wrong, 'is inconvenient if we wish to regard output as a function of the quantities of labor and capital' (Champernowne (1953–54, p. 113)). The chain index compares the amounts of capital in a sequence of stationary states and it is a Divisia type of chain index. Garegnani summarized the proposal as follows: 'The device . . . with which to register, so to speak, the equilibrium value of the physical capital per worker, when the system of production in question, say *I*, first becomes profitable in the course of a monotonic change of the interest rate, and then to keep constant that value for the interval of *I* over which *I* remains profitable. It will then allow that value to change in proportion to the relative value of the capital goods of the new system at the prices of the switch point, as the economy switches to the adjacent system *II* and so on and so forth as the monotonic change of the rate of interest makes the economy switch to the appropriate systems' (Garegnani (1990, pp. 34–5)). Harcourt (1972) showed that the chain index measure of capital is not independent of distribution and prices; it cannot be constructed unless either the wage rate or the rate of profits is known. And Champernowne himself showed that reswitching prevents the unambiguous ordering of techniques in terms of capital intensity and the profit rate. Thus, the chain index solution to the capital measurement problem is unacceptable.

conditions under which macro aggregates (not only capital) exist. Joan Robinson certainly rejected both.²⁰

3. FIRST-GENERATION WORK ON AGGREGATION IN PRODUCTION FUNCTIONS

As indicated in section 1, the other strand of work critical of the notion of aggregate production function is the so-called aggregation literature. Formal work on this problem had begun a few years before Joan Robinson had ignited the capital controversies, but almost two decades after Cobb and Douglas's first estimates.

Klein (1946a) initiated the first debate on aggregation in production functions by proposing methods for simultaneously aggregating over inputs and firms regardless of their distribution in the economy. He wanted to establish a macroeconomic system consistent with, but independent of, the basic microeconomic system. He thus approached the problem assuming as given both the theory of microeconomics and that of macroeconomics, and then tried to construct aggregates (usually in the form of index numbers) which were 'consistent with the two theories' (Klein (1946a, p. 94)). The question Klein posed was whether one could obtain macroeconomic counterparts of micro production functions and the equilibrium conditions that produce supply-of-output and demand-for-input equations in analogy with the micro system.²¹ As noted above, Klein argued that the macro production function should be a purely technological relationship, and that it should not depend on profit maximization (i.e. aggregation outside equilibrium). It should depend only on technological factors.

Algebraically, Klein's problem is as follows. Suppose there are M firms in a sector, each of which produces a single product using N inputs (denoted x). Let the technology of the v th firm be representable as $y^v = f^v(x_1^v, \dots, x_N^v)$. Klein's aggregation problem over sectors can be phrased as follows: what

²⁰ Although she was not altogether unhappy with Fisher's results discussed below. See Robinson (1971b).

²¹ The standard procedure in neoclassical production theory is to begin with micro production functions and then derive equilibrium conditions that equate marginal products of inputs to their real prices. The solution to the system of equations given by the technological relationship and the equilibrium price equations yields the supply-of-output and demand-for-input equations as functions of output and input prices. And adding these equations over all firms yields the macro demand and supply equations. Note, however, as May pointed out (quoted above), that not even the micro production functions are simply technological relations but assume an optimization process by engineers and management.

conditions on the firm production functions will guarantee the existence of (i) an aggregate production function G ; (ii) input aggregator functions g_1, \dots, g_N ; and (iii) an output aggregator function F such that the equation

$$F(y^1, \dots, y^M) = G[g_1(x_1^1, \dots, x_1^M), g_2(x_2^1, \dots, x_2^M), \dots, g_N(x_n^1, \dots, x_n^M)]$$

holds for a suitable set of inputs $x_N^{M?22}$

Klein used Cobb–Douglas micro production functions. He suggested that an aggregate (or strictly, an average) production function and aggregate marginal productivity relations analogous to the micro functions could be derived by constructing weighted geometric means of the corresponding micro variables, where the weights are proportional to the elasticities for each firm. The elasticities of the macro function are the weighted average of the micro elasticities, with weights proportional to expenditure on the factor. The macro revenue is the macro price multiplied by the macro quantity, which is defined as the arithmetic average of the micro revenues (similar definitions apply to the macro wage bill and macro capital expenditure).

Klein's approach ran into a series of serious obstacles. First, Klein's problem was not the same as that of deriving the macromodel from the micromodel. In fact his macromodel does not follow from the micromodel. Both are taken as given, and it is the indices that are derived. The second problem was related to his definition of the aggregate production function as strictly a technical relationship, and the criteria for the aggregates. May's objections to Klein's attempt to define the aggregate production function as a purely technological relation have already been noted. Micro production functions do not give the output that is produced with given inputs. Rather, they give the *maximum* output that can be produced from given inputs. As Pu (1946) indicated, the macroeconomic counterpart of the equilibrium conditions holds if and only if Klein's aggregates arise from micro

²² Klein proposed two criteria that aggregates should satisfy: (i) if there exist functional relations that connect output and input for the individual firm, there should also exist functional relationships that connect aggregate output and aggregate input for the economy as a whole or an appropriate subsection; and (ii) if profits are maximized by the individual firms so that the marginal productivity equations hold under perfect competition, then the aggregative marginal productivity equations must also hold (this criterion cannot be satisfied without the first). The first criterion means that an aggregate output must be independent of the distribution of the various inputs, i.e. output will depend only on the magnitude of the factors of production, and not on the way in which they are distributed among different individual firms nor on the way in which they are distributed among the different types of factors within any individual firm. The second criterion seems to be relevant only for the construction of the aggregate production function.

variables, all of which satisfy equilibrium conditions. Otherwise, the equilibrium conditions do not hold at the macro level. Thus, Klein's aggregates cannot be independent of equilibrium conditions if they are to serve the intended purpose.²³

4. THE LEONTIEF, NATAF AND GORMAN THEOREMS

The first major result on aggregation was provided by Leontief (1947a, 1947b).²⁴ It deals with aggregation of variables into homogeneous groups. Leontief's (1947a) theorem provides the necessary and sufficient conditions for a twice-differentiable production function, whose arguments are all non-negative, to be expressible as an aggregate. The theorem states 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. For the three-variable function $g(x_1, x_2, x_3)$ Leontief's theorem says that this function can be written as $G[h(x_1, x_2), x_3]$ if and only if $\partial(g_1/g_2)/\partial x_3 \equiv 0$ where g_1 and g_2 denote the partial derivatives of g with respect to x_1 and x_2 , respectively. That is, aggregation is possible if and only if the marginal rate of substitution between x_1 and x_2 is independent of x_3 . In general, the theorem states that a necessary and sufficient condition for the weak separability of the variables is that the marginal rate of substitution between any two variables in a group shall be a function only of the variables in that group, and therefore independent of the value of any variable in any other group.

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. Think of the production function

²³ A third problem was pointed out by Walters (1963, pp. 8-9). Walters noted that Kleinian aggregation over firms has some serious consequences. The definition of the macro wage bill (i.e. the product of the macro wage rate and the macro labor) is $WL = (1/n)\sum_{i=1}^n W_i L_i$, where W_i and L_i are the wage rate and homogeneous labor employed in the i th firm and $L = \prod_{i=1}^n L_i^{\alpha_i/\sum \alpha_i}$ is the definition of the macro labor input, a geometric mean, where α_i is the labor elasticity of the i th firm. In a competitive market, all firms have the same wage rate $W^* = W_i$ for all i . Substituting the macro labor into the definition of the macro wage bill and substituting W^* for W_i yields $W = W^* \sum L_i / n \sum_{i=1}^n L_i^{\alpha_i/\sum \alpha_i}$. This implies that the macro wage will almost always differ from the common wage rate of the firms (similar issues apply to the prices of output and capital). It is therefore difficult to interpret W and to see why it should differ from W^* .

²⁴ Leontief dealt with aggregation in general rather than only with production functions. For proofs of Leontief's theorem see the original two papers by Leontief; also Green (1964, pp. 10-15) or Fisher (1993, pp. xiv-xvi).

$Q = Q(k_1, \dots, k_n, L)$. This function can be written as $Q = F(K, L)$, where $K = \phi(k_1, \dots, k_n)$ is the aggregator of capital if and only if

$$\frac{\partial}{\partial L} \left(\frac{\partial Q / \partial k_i}{\partial Q / \partial k_j} \right) = 0$$

for any $i \neq j$. That is, the theorem requires that changes in labor, the non-capital input, do not affect the substitution possibilities between the capital inputs. This way, the invariance of the intra-capital substitution possibilities against changes in the labor input is equivalent to the possibility of finding an index of the quantity of capital. This condition seems to be *natural*, in the sense that if it were possible to reduce the n -dimensionality of capital to one, then it must be true that what happens in those dimensions does not depend on the position along the other axes (e.g. labor).

Note that Leontief's condition is for aggregation within a firm, or within the economy as a whole assuming that aggregation over firms is possible. As discussed later in the paper, the aggregation conditions over firms are very stringent. Is Leontief's condition stringent? 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, p. 103)).

Solow argued that there is a class of situations where Leontief's condition may be expected to hold. This is the case of three factors of production partitioned into two groups. For example, suppose $y_j = f^j(x_{0j}, x_j)$, $j = 1, 2$, where x_j is produced as $x_j = g^j(x_{1j}, x_{2j})$, so that the production of y_j can be decomposed into two stages: in the first one x_j is produced with x_{1j} and x_{2j} , and in the second stage x_j is combined with x_{0j} to make y_j . An example of this class of situations is that x_{1j} and x_{2j} are two kinds of electricity-generating equipment and x_j is electric power. In this case, the g^j -functions are capital index functions (Brown (1980, p. 389)).²⁵

The second important theorem is due to Nataf (1948). Besides the problem of aggregation of variables into homogeneous groups, there is the problem of aggregating a number of technically different microeconomic production functions. Nataf pointed out that Klein's (1946a) aggregation over sectors

²⁵ However, if there are more than two groups, Gorman (1959) showed that, not only must the weak separability condition hold, but also each quantity index must be a function homogeneous of degree 1 in its inputs. This condition is termed 'strong separability'.

was possible if and only if micro production functions were additively separable in capital and labor, e.g. log-additive Cobb–Douglas or harmonic-mean constant elasticity of substitution (CES) (thus, this is a condition on the functional form). Under these circumstances, output is then equal to a labor component plus a capital component.

The problem here is as follows. Suppose there are n firms indexed by $v = 1, \dots, n$. Each firm produces a single output $Y(v)$ using a single type of labor $L(v)$ and a single type of capital $K(v)$. Suppose that the v th firm has a two-factor production function $Y(v) = f^v\{K(v), L(v)\}$. To keep things simple, assume all outputs are physically homogeneous so that one can speak of the total output of the economy as $Y = \sum_v Y(v)$, and that there is only one kind of labor so that one can speak of total labor as $L = \sum_v L(v)$. Capital, on the other hand, may differ from firm to firm (although it may also be homogeneous). The question is: under what conditions can total output Y be written as $Y = \sum_v Y(v) = F(K, L)$ where $K = K\{K(1), \dots, K(n)\}$ and $L = L\{L(1), \dots, L(n)\}$ are indices of aggregate capital and labor, respectively? Nataf showed (necessary and sufficient condition) that the aggregates Y, L, K which satisfy the aggregate production function $Y = F(K, L)$ exist, when the variables $K(v)$ and $L(v)$ are free to take on all values, if and only if every firm's production function is additively separable in labor and capital; i.e. if every f^v can be written in the form $f^v\{K(v), L(v)\} = \phi^v\{K(v)\} + \psi^v\{L(v)\}$. Assuming this to be so, the aggregate production relation can be written $Y = L + K$, where $Y = \sum_v Y(v)$, $L = \sum_v \psi^v\{L(v)\}$ and $K = \sum_v \phi^v\{K(v)\}$. Moreover, if one insists that labor aggregation be 'natural', so that $L = \sum_v L(v)$, then all the $\psi^v\{L(v)\} = c\{L(v)\}$, where c is the same for all firms. Nataf's theorem provides an extremely restrictive condition for intersectoral or even interfirm aggregation.²⁶ It makes one rather chary about the existence of an aggregate production function unless there are some further restrictions on the problem.²⁷

Finally, Gorman (1953) developed a set of aggregation conditions over firms assuming that the optimal conditions for the distribution of given totals of inputs among firms are satisfied. These efficiency conditions require that

²⁶ For a number of applications of this result see Green (1964, ch. 5).

²⁷ Nataf's result can be proved using Leontief's theorem. By the latter, aggregation is possible if and only if the ratio of marginal products of capital in two firms are independent of all labor inputs. But in Nataf's non-optimizing set-up, the amount of labor in a given firm cannot influence the marginal product of capital in any other. Hence, Leontief's condition requires that it not influence the marginal product of capital in the given firm either. This way one obtains additive separability. The conclusion that the marginal product of labor must be constant and the same in all firms follows from the requirement that the labor aggregate is total L , so that reassigning labor among firms does not change total output.

the marginal rates of substitution between the i th and j th inputs be the same for all firms, i.e.

$$\frac{\partial F^i}{\partial x_{ki}} \bigg/ \frac{\partial F^i}{\partial x_{hi}} = \frac{\partial F^j}{\partial x_{kj}} \bigg/ \frac{\partial F^j}{\partial x_{hj}}$$

where i, j denote the firms, and k, h the inputs. Gorman showed that, if this condition holds, then a necessary and sufficient condition for the consistent aggregation of the functions $y_s = f_s(x_{1s}, \dots, x_{ms})$ to the function $y = F(x_1, \dots, x_m)$ is possible if the expansion paths for all firms at a given set of input prices are parallel straight lines through their origins.

Green (1964, pp. 49–51) provides an application of Gorman's aggregation conditions to the Cobb–Douglas case. Theorem 10 in Green (1964) says that if the expansion paths for all firms, at a given set of input prices, are parallel straight lines through their origins, then consistent aggregation of the functions $f_s(x_{1s}, \dots, x_{ms})$ to the function $y = F(x_1, \dots, x_m)$ is possible. Furthermore, there exist functions F and h_1, \dots, h_m such that $y = \sum_{s=1}^n h_s(y_s) = F(x_1, \dots, x_m)$ where the function F is homogeneous of degree 1 in its arguments. A corollary of this theorem is that, if the conditions in the theorem are satisfied and each of the functions f_s is homogeneous of degree 1, consistent aggregation is possible with $y = \sum_{s=1}^n c_s y_s$.

If the expansion paths are straight lines through the origin, the marginal rates of substitution depend only on the ratios $x_{1s}/x_{rs}, \dots, x_{ms}/x_{rs}$. And if all expansion paths are parallel, the optimal ratios will be the same for all s , and equal to the ratios of the totals $x_1/x_r, \dots, x_m/x_r$. Hence, for each r and s , x_{rs} depends only on y_s and the ratios $x_1/x_r, \dots, x_m/x_r$.

With the above background, assume a Cobb–Douglas with three inputs $y_s = A_s L_s^{\alpha_s} K_s^{\beta_s} H_s^{\gamma_s}$ where the subscript s indexes the firms, and $\alpha_s + \beta_s + \gamma_s = 1$. Then,

$$y_s = A_s L_s \left(\frac{K_s}{L_s} \right)^{\beta_s} \left(\frac{H_s}{L_s} \right)^{\gamma_s}$$

If (i) the expansion paths are parallel and (ii) the first-order conditions are satisfied, then the production functions can be written as

$$y_s = A_s L_s \left(\frac{K}{L} \right)^{\beta} \left(\frac{H}{L} \right)^{\gamma}$$

where K , H and L represent the totals. Now aggregate them:²⁸

$$\begin{aligned} y &= \sum_{s=1}^n \frac{y_s}{A_s} = \sum_{s=1}^n L_s \left(\frac{K}{L}\right)^\beta \left(\frac{H}{L}\right)^\gamma = L \left(\frac{K}{L}\right)^\beta \left(\frac{H}{L}\right)^\gamma \\ &= L^{1-\beta-\gamma} K^\beta H^\gamma = L^\alpha K^\beta H^\gamma \end{aligned}$$

5. FISHER'S AGGREGATION CONDITIONS

Fisher (1969a, 1993) observed that, taken at face value, Nataf's theorem essentially indicates that aggregate production functions almost never exist. Note, for example, that Nataf's theorem does not prevent capital from being physically homogeneous. Likewise, each firm's production function could perfectly 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. As indicated previously, identity of technologies (e.g. all of them Cobb–Douglas) and constant returns do not imply the existence of an aggregate production function. Yet intuition indicates that under these circumstances one *should* expect an aggregate production function to exist. Something is wrong here.

Fisher pointed out that one must ask not for the conditions under which total output can be written as $Y = \sum_v Y(v) = F(K, L)$ under any economic conditions, 'but rather for the conditions under which it can be so written *once production has been organized to get the maximum output achievable with the given factors*' (Fisher (1969a, p. 556)), italics in the original. This was, of course, the problem with Klein's original formulation). Thus 'the problem with Nataf's theorem is not that it gives the wrong answer but that it asks the wrong question. A production function does not give the output that can be produced from given inputs; rather, it gives the *maximum* output that can be so produced. Nataf's theorem fails to impose an efficiency condition' (Fisher (1993, p. xviii), italics in the original). Thus efficient allocation requires that Y be maximized given K and L . This is why optimization over the assignment of production to firms makes sense in constructing an aggregate production function. Competitive factor markets will do this.

²⁸ Brown (1980, pp. 397–8) shows that Gorman's conditions appear implicitly in the standard practice of using economy-wide deflators to obtain real capital measures within a sector (i.e. deflating the 'value' of capital). Unless Gorman's conditions are satisfied, the deflation process does not eliminate the price effect inherent in the value of capital. The resulting magnitude is not, therefore, a quantity or real value.

These considerations lead to an altogether different set of aggregation conditions.

Moreover, 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 Joan Robinson's problem (and the Cambridge–Cambridge debates). First, it is stressed that the aggregation problem is not only the aggregation of capital, at least the way Joan Robinson understood it (recall that for the Cambridge, UK, side the measurement of capital was problematic because the prices of capital good 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, even were there only one type of capital, labor and output aggregation problems would continue to exist. Second, from the point of view of the aggregation literature the problem is 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 once, 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 (Joan Robinson's problem). In fact, Fisher's aggregates are indeed indices, and in his view, Joan Robinson misunderstood the aggregation problem (Fisher (1993, p. xiii)). It is here that an important difference arises in the understanding of the issues at stake. For the Cambridge, UK, scholars the aggregation problem was *strictly* a problem that affected capital and the rate of profit, and it was related to the problem of income distribution. In the words of Pasinetti: '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’ (Pasinetti (2000, p. 209)).

For Fisher, on the other hand, capital does not present any special problem. Similar aggregation problems occur with labor and output.²⁹ The whole problem reduces to finding the technical conditions under which *any or all* aggregates can be generated. This view was advocated by Bliss, who concluded that: ‘there is no support whatsoever for the idea that the aggregation of capital is relatively difficult. The conditions for general capital aggregation are identical to the conditions for the aggregation of labor, or of output. We may thus conclude that the widespread belief that there is a notable, particular and distinct problem posed by capital aggregation is at best an ill-formulated idea, and at worst is based simply on ignorance’ (Bliss (1975, p. 162)).

It seems, therefore, that Joan 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 the latter sense, Joan Robinson’s remarks about labor and land being different from capital are not true. They are not physically homogeneous factors either.

The result of the above observations was a series of seminal papers on aggregation conditions along lines similar to those followed by Gorman. They have been edited and collected in Fisher (1993). To show that this way of approaching the problem makes a difference, let us consider the case in which capital is physically homogeneous, so that total capital can be written as $K = \sum_v K(v)$. Under these circumstances, efficient production requires that aggregate output Y be maximized given aggregate labor (L) and aggregate capital (K). Under these simplified circumstances, it follows that $Y^M = F(K, L)$ where Y^M is maximized output, since, as was pointed out by May (1946, 1947), individual allocations of labor and capital to firms would be determined in the course of the maximization problem (note that without optimal allocation even factor homogeneity does not help). This holds even if all firms have different production functions and whether or not there are constant returns.

²⁹ Diewert (1980) also disagrees: ‘Even if the *theory* of the aggregation of capital does not appear to be any more difficult than the aggregation of say, labor, in *practice* it is very much more difficult to construct a capital aggregate that researchers can agree is appropriate for the purpose at hand’ (Diewert (1980, p. 474), italics in the original). Diewert (1980, pp. 475–86) offers a thorough discussion of these difficulties.

5.1 Capital aggregation

In the more realistic case where only labor is homogeneous and technology is embodied in capital, Fisher proposed to treat the problem as one of labor being allocated to firms so as to maximize output, with capital being firm-specific. It is at this point that there is a link between the Cambridge–Cambridge capital controversies and the aggregation issues. In her seminal paper opening the debate, Joan Robinson (1953–54) asked how aggregate and heterogeneous capital were to be measured. Solow (1955–56) took the question from a different angle, and asked in his reply ‘under what conditions can a consistent meaning be given to the quantity of capital?’. And, ‘When if ever can the various capital inputs be summed up in a single index-figure, so that the production function can be “collapsed” to give output as a function of inputs of labor and “capital-in-general”?’ (Solow (1955–56, p. 102)). Thus Solow introduced the Leontief aggregation condition into the Cambridge controversy. However, Joan Robinson (1955–56), in her rejoinder, completely dismissed Solow’s reply by arguing that ‘it does not touch upon the problem of capital, but is concerned rather with how to treat non-homogeneous natural resources. . . . His C_1 and C_2 [the two types of capital] are two kinds of equipment, but nothing is said about the time which it takes to produce them (gestation period) or the period over which they are expected to be useful (service life). . . . None of these questions can be dealt with in terms of an index of physical equipment’ (Robinson (1955–56, p. 247)).³⁰

It was argued above that when labor and capital are homogeneous across firms and allocated optimally across firms, aggregation does not pose a special problem as regards that factor. But when capital is not homogeneous, i.e. firms use different techniques, one cannot add up heterogeneous quantities meaningfully unless there is some formula that converts heterogeneous items into homogeneous units.

Fisher’s first paper on aggregation dates back to 1965. In this context, it is important to remark that the assumption that technology is embodied in capital (i.e. capital is firm-specific) induces difficulties whether or not a capital aggregate exists for each firm. However, no such difficulties exist as to

³⁰ Usher (1980, p. 19) indicates that the aggregation problem (summarized in section 1) and the index number problem are different. The latter refers to the following. Suppose there is a function $K = h(K_1, K_2)$, where the form of h is unknown and time series of quantities of capital goods K_1 and K_2 are available. Thus, we do not have a time series of K . The prices of the capital goods, P_1 and P_2 , are proportional to the derivatives $\partial h(K_1, K_2)/K_1$ and $\partial h(K_1, K_2)/K_2$. The problem is to infer the series K from K_1 , K_2 , P_1 and P_2 .

aggregate labor *if* there is only one type of labor. The reason is that labor is assumed to be assigned to firms efficiently. Now, given that output is maximized with respect to the allocation of labor to firms, and denoting such value by Y^* , the question is: under what circumstances is it possible to write total output as $Y^* = F(J, L)$ where $J = J\{K(1), \dots, K(n)\}$, where $K(v)$, $v = 1, \dots, n$, represents the stock of capital of each firm (i.e. one kind of capital per firm)? Since the values of $L(v)$ are determined in the optimization process there is no labor aggregation problem. The entire problem in this case lies in the existence of a capital aggregate. Recalling that the weak separability condition is both necessary and sufficient for the existence of a group capital index, the previous expression for Y^* is equivalent to $Y^* = G\{K(1), \dots, K(n), L\}$ if and only if the marginal rate of substitution between any pair of $K(v)$ is independent of L .

Fisher then proceeded to draw the implications of this condition for the form of the original firm production function. He found that, under the assumption of strictly diminishing returns to labor (i.e. $f_{LL}^v < 0$), a necessary and sufficient condition for capital aggregation is that, if any one firm has an additively separable production function (i.e. $f_{KL}^v \equiv 0$), then *every* firm must have such a production function.³¹ This means that capital aggregation is not possible if there is both a firm which uses labor and capital in the same production process and another one which has a fully automated plant.³² More important, assuming constant returns to scale, capital-augmenting technical differences (i.e. embodiment of new technology can be written as the product of the amount of capital multiplied by a coefficient) turn out to be *the only case* in which a capital aggregate exists. This means that each firm's production function must be writeable as $F(b_v K_v, L_v)$, where the function $F(\cdot, \cdot)$ is common to all firms but the parameter b_v can differ. Under these circumstances, a unit of one type of new capital equipment is the exact duplicate of a fixed number of units of old capital equipment ('better' is equivalent to 'more'). The aggregate stock of capital can be constructed with capital measured in efficiency units.³³ Summing up: aggregate production functions exist

³¹ Here and later, such subscripts denote partial differentiation in the obvious manner.

³² Strictly speaking, Fisher found that a necessary and sufficient condition for capital aggregation is that every firm's production function satisfy a partial differential equation in the form $f_{KL}^v / f_K^v f_{LL}^v = g(f_L^v)$, where g is the same function for all firms.

³³ Fisher (1965) indicates that he could not come up with a closed-form characterization of the class of cases in which an aggregate stock of capital exists when the assumption of constant returns is dropped. Nevertheless, as he shows, there do exist classes of non-constant returns production functions which do allow construction of an aggregate capital stock. Capital aggregation is possible only under the restrictive assumption that the individual firm's production function can be made to yield constant returns after suitable 'stretching of the capital axis'.

if and only if all micro production functions are identical except for the capital efficiency coefficient. Certainly this conclusion represents a step beyond Nataf's answer to the problem. But certainly it also continues to require an extremely restrictive aggregation condition, one that actual economies do not satisfy.

To see that this condition permits aggregation, consider two firms, and define $J = b_1K(1) + b_2K(2)$ with $L = L(1) + L(2)$. The sum of the outputs of the two firms is $Y = F[b_1K(1), L(1)] + F[b_2K(2), L(2)]$. Since efficient allocation of labor requires that labor have the same marginal product in both uses, it is clear that when Y is maximized with respect to labor allocation the ratio of the second argument to the first must be the same in each of the two firms. Thus,

$$\frac{L(1)}{b_1K(1)} = \frac{L(2)}{b_2K(2)} = \frac{L}{J}$$

when labor is optimally allocated. If we let

$$\lambda = \frac{b_1K(1)}{J} = \frac{L(1)}{L}$$

(this second equality holds when labor is optimally allocated), it then follows that $Y^* = F(\lambda J, \lambda L) + F[(1 - \lambda)J, (1 - \lambda)L] = F(J, L)$ because of constant returns.³⁴

But the bite of the theorem is that the capital-augmentation condition is necessary (as well as sufficient) for capital aggregation under constant returns. Thus, an implication of Fisher's work is the importance of the aggregation level. On the one hand the aggregation problem appears both with two firms and with a thousand. On the other hand, it is fair to say that the more firms, the *more likely* it is that they will differ in ways that prevent aggregation, or that (at least) one of them will fail to satisfy the partial differential equation condition mentioned in a footnote above.

As an extension, Fisher (1965) analyzed the case where each firm produces a single output with a single type of labor but two capital goods, i.e. $Y(v) = f^v(K_1, K_2, L)$. Here Fisher distinguished between two different cases. The first case was aggregation across firms over one type of capital (e.g. plant, equipment). Fisher concluded that the construction of a sub-aggregate of capital

Likewise, if constant returns are not assumed there is no reason why perfectly well-behaved production functions cannot fail to satisfy the partial differential equation given in the preceding footnote. Capital aggregation is then impossible if any firm has one of these 'bad apple' production functions.

³⁴ This proof holds for any constant returns to scale production function. Of course this construction is only for the case in which (only) labor is optimally assigned.

goods requires even less reasonable conditions than for the construction of a single aggregate.³⁵ For example, if there are constant returns in K_1 , K_2 and L , there will not be constant returns in K_1 and L , so that the difficulties of the two-factor non-constant returns case appear. Further, if the v th firm has a production function with all three factors as complements, then no K_1 aggregate can exist. Thus, for example, if any firm has a generalized Cobb–Douglas production function (omitting the v argument) in plant, equipment and labor $Y = AK_1^\alpha K_2^\beta L^{1-\alpha-\beta}$, one cannot construct a separate plant or separate equipment aggregate for the economy as a whole (although this does not prevent the construction of a full capital aggregate).

The second case Fisher considered was that of the construction of a complete capital aggregate. In this case, a necessary condition is that it be possible to construct such a capital aggregate for each firm taken separately; and a necessary and sufficient condition (with constant returns), given the existence of individual firm aggregates, is that all firms differ by at most a capital-augmenting technical difference. That is, they can differ *only* in the way in which their individual capital aggregate is constructed.

Fisher (1982) extended the previous analysis and returned to the Cambridge–Cambridge debates by asking whether the crux of the aggregation problem derives from the fact that capital is considered to be an immobile factor. Recall that in the previous discussion Fisher had assumed a model in which each firm's technology was embodied in its capital stock, which was immobile. This is what made (aggregate) capital a heterogeneous good, and was the genesis of the aggregation problem. On the other hand, labor and output were assigned to firms in the course of the optimization process, and thus efficiently. Fisher (1982) argued that the aggregation problem only seems to be due to the fact that capital is fixed and is not allocated efficiently. This is true in the context of a two-factor production function. However, if one works in terms of many factors, all mobile over firms, and asks when it is possible to aggregate them into macro groups, it turns out that the mobility of capital has little bearing on the issue. In fact, where there are several factors, each of which is homogeneous, optimal allocation across *firms* does not guarantee aggregation across *factors*. The conditions for the existence of such aggregates are still very stringent, but this has as much to do with the necessity of aggregating over firms as with the immobility of capital. For the two-firm case, assuming each firm's production function can be written in the form $f^v[X(v), L(v)] = F^v\{\phi^v[X(v), L(v)]\}$, where $\phi^v(\cdot)$ is scalar valued, and assuming constant returns (if there are non-constant returns, no aggregate will exist in general), aggregation is permitted over some group of

³⁵ The conditions turn out to be twofold: (i) $f_{K_1 L}^v / f_{K_1}^v f_{L L}^v = g(f^v)$; (ii) $f_{K_1 K_2}^v - f_{K_1 L}^v f_{K_1}^v f_{L L}^v = 0$.

variables (outputs or factors) if and only if at least one of the following two holds:³⁶ (i) the $F^v(\cdot, \cdot)$ can be taken to be the same; (ii) the $\phi^v(\cdot)$ can be taken to be the same. A possible way of interpreting the existence of aggregates at the firm level is that each firm could be regarded as having a two-stage production process. In the first one, the factors to be aggregated, the $X_i(v)$, are combined together to produce an intermediate output, $\phi^v[X(v)]$. This intermediate output is then combined with $L(v)$ to produce the final output. Aggregation of X can be done if and only if firms are either all alike as regards the first stage of production or all alike as regards the second stage. If they are all alike as regards the first stage, then the fact that L is mobile plays no role (aggregation condition of mobile factor when the remainder are fixed). If, on the other hand, they are all alike as regards the second stage, then the fact that the X_i are mobile plays no role (aggregation condition for aggregation of fixed factors when the remainder are mobile). These conditions imply that mobility of capital permits instant aggregation over firms of any one capital type across firms. However, the fact that aggregation over firms is involved, whether or not capital is fixed, restricts aggregation to the cases described above.

When there are more than two firms, aggregation over the entire set of firms requires aggregation over every pair (the two-firm case). This implies that an aggregate over n firms exists if and only if at least one of the following two holds: (i) all the $F^v(\cdot, \cdot)$ can be taken to be the same; (ii) all the $\phi^v(\cdot)$ can be taken to be the same.

Finally, Fisher (1983) is another extension of the original problem to study the conditions under which full and partial capital aggregates, such as 'plant' or 'equipment', would exist simultaneously. Not surprisingly, the results are as restrictive as those above. Fisher showed that the simultaneous existence of a full and a partial capital aggregate (e.g. plant) implies the existence of a complementary partial capital aggregate (e.g. equipment), and that the two partial capital aggregates must be perfect substitutes.³⁷

³⁶ The assumption of constant returns implies that both $F^v(\cdot, \cdot)$ and $\phi(\cdot)$ can be taken to be homogeneous of degree 1. There are two other cases in which aggregation is trivial. The first of these occurs when, for every $v = 1, \dots, n$, $F^v(\cdot, \cdot)$ is additively separable in its two arguments à la Nataf. The second special case occurs when all of the $X(v)$ factors are assigned to a single firm and, within that firm, the marginal rate of substitution between any pair of the $X(v)$ is independent of L . In that case, an X aggregate will exist by Leontief's theorem.

³⁷ Blackorby and Schworm (1984) is an extension of Fisher (1983). By presenting an alternative formulation of the problem in which one can have both a full and a partial capital aggregate without the restrictive substitution implications derived by Fisher, they show that there need be only one partial aggregate and that, if there are two partial aggregates, they need not be perfect substitutes. The conditions nevertheless remain very restrictive.

5.2 Labor and output aggregation

Fisher (1968) extended his work on capital aggregation to the study of problems involved in labor and output aggregation, thus pointing out that the aggregation problem is not restricted to capital. Output and labor aggregation are necessary only if one wants to determine the aggregate production function. This represents another important apparent difference with respect to Cambridge, UK, in the capital debates. For the latter, as pointed out above, there is a problem with the 'symmetrical treatment' of labor and capital. This is a question that had been addressed by Wicksell over a century ago. He pointed out that while 'labor and land are measured each in terms of its own *technical* unit . . . capital, on the other hand, . . . is reckoned, in common parlance as a sum of *exchange value*' (quotation taken from Pasinetti and Scazzieri (1990, p. 139), italics in the original). The argument of Cambridge, UK, was that while labor can be expressed in physical terms (e.g. hours) to which its reward can be referred (i.e. wage per hour), for capital, which of course can be expressed in physical terms too (e.g. number of machines, or an index of physical quantity), the problem does not lie in itself. The problem lies in its reward, the rate of profit, since it is commensurate with the value of capital, not with the physical quantity of capital. But the value of capital, the argument continues, is the product of the physical quantity multiplied by its price. The latter is dependent on the rate of profit and thus on income distribution (Pasinetti (2000, p. 206)).^{38,39}

The reply from the aggregation literature is that one could equally argue that wages are also commensurate with the value of labor and not with the physical quantity. Further, wages from the demand side depend on the profitability of hiring more workers. The same is true of machines, and indeed, just as there is a supply price for labor, there is a supply price for machines.

The problem studied by Fisher is in the context of cross-firm aggregation that arises because labor or outputs are shifted over firms, given the capital

³⁸ Pasinetti (2000, p. 207) indicates that there are two possibilities. (i) Use K to represent the physical quantity of capital. In this case, the marginal product $\partial Y/\partial K$ no longer represents the rate of profit but the rental of capital, which has to be multiplied by the price of the capital good. (ii) Use K to represent the current value of capital. In this case, $\partial Y/\partial K$ is the sum of two components, one representing the variation in the physical quantity of capital, and the other representing the variation in the price of the physical quantity. These are the Wicksell effects (Burmeister (1990)).

³⁹ In personal correspondence Gary Mongiovi has pointed out that the problems involving the interdependence of prices and distribution that arise in connection with capital do not arise with labor or land. The problem with these two appears in constructing models designed for empirical testing.

stocks and production functions, to achieve efficient production. That is, now there is a vector of labor $L_1(v), \dots, L_s(v)$ and a vector of outputs $Y_1(v), \dots, Y_s(v)$ (it does not matter whether there is one or more types of capital).⁴⁰ In the simplest case of constant returns, a labor aggregate will exist if and only if a given set of relative wages induces all firms to employ different labor in the same proportion. Similarly, where there are many outputs, an output aggregate will exist if and only if a given set of relative output prices induces all firms to produce all outputs in the same proportion.

The implication of these conditions is that the existence of a labor aggregate requires the absence of specialization in employment, and 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 their scale.⁴¹

6. FISHER'S SIMULATIONS

Fisher (1969a, pp. 572–4) posed an interesting conundrum: despite the stringency of the aggregation conditions, the fact is that when one fits aggregate data on output to aggregate data on inputs the results tend to be 'good', meaning that the fit tends to be relatively high, and that in the case of the Cobb–Douglas the elasticities are close to the factor shares in output. Furthermore, the wage rate is well explained by the marginal product. Fisher sketched several possible reasons for this paradox, of which he favored the following: for unspecified reasons, firms always invest in proportion (i.e. fixed ratios) to a particular index. In such case the index would be an approximate aggregate.⁴² And likewise, if outputs were always produced and labor hired in approximately fixed proportions, then approximate output and labor aggregates would exist.

Fisher (1971a) and Fisher *et al.* (1977) are two attempts at providing an answer to the question of why, despite the stringent aggregation conditions, aggregate production functions seem to work when estimated econometri-

⁴⁰ An interesting issue in this context is that the aggregates of labor and output might exist for each firm separately, but not for all firms together. However, since this would imply some strange things about aggregation, Fisher assumed that an aggregate at the firm level exists. No similar problem arises in the case of capital, where aggregation over all firms requires the existence of an aggregate for each firm separately.

⁴¹ The 'same market basket' condition for output aggregation and the similar condition for labor aggregation are cases of the 'common aggregator' condition in Fisher (1982) (see above). Blackorby and Schworm (1988) is an extension of Fisher (1968).

⁴² This argument relates to the Houthakker–Sato aggregation conditions. See below.

cally. Likewise, the marginal product of labor appears to give a reasonably good explanation of wages. To answer the question, Fisher undertook a series of simulation analyses. The important aspect of the simulations is that the series were aggregated even though the aggregation conditions were violated. Under these circumstances, if the aggregate production function yields 'good results', one cannot take it as evidence that the aggregate production function summarizes the true technology.

In the first of these papers, Fisher (1971a) set up an economy consisting of N firms or industries ($N = 2, 4$ or 8 in the simulations), each hiring the same kind of labor and producing the same kind of output. Each firm, however, had a different kind of capital stock, and its technology is embodied in that stock. This implies that capital could not be reallocated to other firms. In the aggregation process, the conditions for successful aggregation were violated. The micro production functions were Cobb–Douglas, and labor was allocated optimally to ensure that output was maximized. This economy was simulated over 20 periods. The total labor force, the firms' technology and their capital stocks were assumed to grow at a constant rate (with a small random term to reduce multicollinearity in the subsequent regression analysis). In certain of the experiments, some of these growth rates were set equal to zero and the growth of the capital stock was allowed to vary between firms.

Fisher observed that in all his experiments (a total of 1010 runs each covering a 20-year period) the fit was around 0.99, although he pointed out that this 'reflects the fact that with everything moving in trends of one sort of another, an excellent fit is obtained regardless of misspecifications of different sorts' (Fisher (1971a, p. 312)).

The most important conclusion Fisher drew from his results 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 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, i.e. the aggregate Cobb–Douglas works well *because* labor's share is roughly constant.

In a subsequent paper, Fisher *et al.* (1977) extended the simulation analysis to the case of the CES production function developed by Arrow *et al.* (1961). The simulations were similar in spirit to those in Fisher (1971a), with the corresponding complications introduced by the fact that the micro production functions were CES and have more coefficients to parameterize

(elasticity of substitution and distribution parameter). The objective was the same, i.e. to learn when the CES, despite the aggregation problems, would perform well in empirical work. The aggregate series of output, labor and capital were also generated following procedures similar to those in Fisher (1971a). And the aggregation conditions for capital were violated as in Fisher (1971a). Thus the authors stated that 'the elasticity of substitution in these production functions is an "estimate" of nothing; there is no true aggregate parameter to which it corresponds' (Fisher *et al.* (1977, p. 312)). Each firm had a different elasticity of substitution, ranging between 0.25 and 2.495. For each choice of the elasticities of substitution, the distribution parameters were chosen in two sets, half the runs having distribution parameters and substitution elasticities positively correlated, and half of them negatively correlated (ranging between 0.15 and 0.35). The objective was to generate a labor share of approximately 0.75. It must be mentioned that in this paper, besides the aggregate CES, Fisher *et al.* (1977) also estimated the Cobb–Douglas and the log-linear relationship implied by the CES with constant returns to scale, namely $\ln(Y^*/L) = H + \sigma \log w$, where σ is the elasticity of substitution. They called the latter the 'wage equation'. This is used in what they refer to as the 'hybrid estimate' of the wage equation and the production function. This was obtained imposing the elasticity of substitution estimated from the wage equation on the production function; and then they used the latter to estimate the distribution and efficiency parameters in the production function.

What conclusions did Fisher *et al.* (1977) reach? The fit in all cases was very good. They also established that the hybrid wage predictions were the best, and that the wage equation estimates of the elasticity of substitution were better than those given by the production function. Likewise, Fisher's earlier findings with Cobb–Douglas were confirmed in these simulations, i.e. the Cobb–Douglas works well when the observed factor share is fairly stable. But the authors failed to find any similar organizing principle with which to explain when the aggregate CES production function does or does not give good wage predictions. In other words, while in Fisher (1971a) the organizing principle was that the aggregate Cobb–Douglas would work when factor shares were constant, in the case of the CES they could not establish any similar 'rule'.⁴³

⁴³ Nelson and Winter (1982) also used simulation analysis to show that, in the context of their evolutionary model, they could generate a data set such that, when an aggregate Cobb–Douglas was fitted, an almost perfect fit was obtained, and with factor elasticities very close to the input shares in revenue. The model, however, is anti-neoclassical in many respects, e.g. firms are not profit-maximizers; the aggregate production function does not exist, the technology available to each firm is fixed coefficients, and firms learn about them (they do not know all possible

7. HOUTHAKKER-SATO AGGREGATION CONDITIONS

Sato (1975) provided a different set of aggregation conditions from those of Fisher. Sato's approach to the aggregation problem was based on the procedure that Houthakker had developed years before. Houthakker (1955-56) proposed an ingenious way of addressing the aggregation problem by postulating that factor proportions are distributed in a certain way among the firms over which the aggregation is to take place. He then showed, for the one-output two-variable-input case, that if individual production functions are of the fixed-coefficients type (not necessarily the same in each firm) and if the input-output ratios (the capacity density function) are distributed according to a Pareto distribution $Q = C(L/Q)^{\alpha_1-1}(K/Q)^{\alpha_2-1}$ with $\alpha_1 > 1$ and $\alpha_2 > 1$, then the aggregate production function is the Cobb-Douglas with decreasing returns to scale $Q = AL^{\alpha_1/(\alpha_1+\alpha_2+1)}K^{\alpha_2/(\alpha_1+\alpha_2+1)}$.⁴⁴ The peculiar conclusion of Houthakker's model is that if all individual firms operate according to Leontief production functions, and if efficiencies are distributed according to a Pareto distribution, then the aggregate production function will be Cobb-Douglas. In other words, while the aggregate production function has the appearance of a technology with an elasticity of substitution of unity, at the micro level there is no possibility of substitution between inputs.⁴⁵ This procedure is generally known as the efficiency-distribution approach.

Sato (1975) developed and extended the procedure introduced by Houthakker with a view to investigating how the macro behavior in production relates to the macro behaviors via the efficiency distribution, i.e. the distribution of input coefficients. He allowed for elasticities of substitution to exceed zero, and the distribution function needed no longer be Pareto.⁴⁶ This approach to the aggregation problem shows what aggregate production functions can be expected when the distribution of capital over firms with related technologies is fixed, or changes in very restricted ways. The link between the micro and the macro functions is provided by the efficiency distribution.

combinations of the input-output coefficients) by engaging in a search process. Furthermore, this search is undertaken only if the profit rate falls below a pre-established, acceptable, minimum.

⁴⁴ A simplified explanation of this model can be found in Heathfield and Wibe (1987, pp. 151-3). See also Sato (1975, pp. 10-12, 25-7).

⁴⁵ Levhari (1968) reversed Houthakker's procedure and derived the distribution of factor proportions for a CES production function.

⁴⁶ A few years before, Johansen (1972) had also used Houthakker's approach. However, Johansen had not seen the connection with Fisher's work, and there was no direct discussion of aggregation of heterogeneous capital.

Sato proceeded by splitting the aggregation problem into two sequential questions. First, suppose one has the production function $Q = Q(K_1, \dots, K_m, L)$. Then ask: can this form be compressed into a form like $Q = F(K, L)$ by aggregating the vector of K s? In this step one must find both the capital aggregate K and the macro function F . Sato called this the *existence* problem. This must be done for each distribution. This will give rise to a series of F s. The second step is to ask for the conditions that the distributions must satisfy if they are to generate the same F . This is the *invariance* problem. And as a corollary Sato asked whether two entirely different distributions can yield macro production functions $Q = F(K, L)$ identical in every respect. Sato shows that, if the efficiency distribution is stable, the resulting estimates should reflect a production function. Thus, the key of this approach lies in the stability of the distribution function. Testing this empirically is not easy. In general, the aggregate production function does not remain invariant and therefore cannot be used as a basis for the long-run growth theory. The other important characteristic of this approach is that the capital aggregate generated is the total productive capacity of the industry, which in general has no direct connection with conventionally measured capital stocks.

8. GORMAN'S AGGREGATION CONDITIONS

Finally, it is important to make a reference to Gorman's (1968) work, since he used yet another alternative method, namely the restricted profit function, to derive the aggregation conditions. Gorman also set out to find what the technologies of the individual firms should be so that aggregates of fixed factors (different classes of fixed goods) would exist. Examples of fixed factors are 'capital', 'land', 'equipment' and 'buildings'. The aggregates are referred to as the quantity of capital, land etc. These quantities are required to depend only on the amounts of the various types of equipment used in individual firms. Gorman showed that if the micro (labor optimized) variable profit functions Π^{*m} can be written as

$$\Pi^{*m}(p, w, z^m) = b(p, w)h^m(z^m) + c^m(p, w)$$

for $m = 1, 2, \dots, M$ (sectors of the economy)

then capital aggregation is possible (p is a vector of output and intermediate input prices; z^m is a vector of fixed capital input; and w is a vector of labor prices); i.e. the macro (labor optimized) variable profit function Π^* can be written as

$$\begin{aligned}
\Pi^*(p, w, z^1, \dots, z^M) &= \sum_{m=1}^M \Pi^{*m}(p, w, z^m) \\
&= b(p, w) \left[\sum_{m=1}^M h^m(z^m) \right] + \sum_{m=1}^M c^m(p, w) \\
&= \Pi^* \left[p, w, \sum_{m=1}^M h^m(z^m) \right]
\end{aligned}$$

Therefore, the separability restriction on the micro production possibility sets is sufficient to imply the existence of the aggregate.

Is the restriction on the micro variable profit functions a stringent aggregation condition? Perhaps it is not very restrictive if every z^m is a scalar, i.e. if there is only one fixed capital good for each sector. However, the restriction becomes more unrealistic from an empirical point of view as the number of fixed capital goods in each sector increases.

9. WHY DO ECONOMISTS CONTINUE USING AGGREGATE PRODUCTION FUNCTIONS?

It must be clear that the fact that neoclassical macroeconomists *use* aggregates such as investment, capital, labor, GDP (as derived from a well-behaved aggregate production function), as well as aggregate production functions, in theoretical and empirical exercises does not legitimize the existence of such constructs. Economists have learnt how to answer the inconvenient question of 'why' they use aggregate production functions despite the aggregation problems. As indicated in section I, three standard answers are the following. One answer, based on the methodological position known as instrumentalism, is that as long as aggregate production functions appear to give empirically reasonable results, why shouldn't they be used? Neoclassical macro theory deals with macroeconomic aggregates derived by analogy with the micro concepts (Ferguson (1971)). The usefulness of this approach is strictly an empirical issue. Second, and following Samuelson (1961–62), aggregate production functions are seen as useful parables. Finally, for the applications where aggregate production functions are used, there is no other choice. In the light of the aggregation results, none of these reasons seems valid.

The first argument is that despite the aggregation results and the Cambridge–Cambridge controversies, the fact is that aggregate production functions seem to work empirically, at least at times. Then, the argument goes, let us continue using them. This position is the one espoused by Ferguson (1971) in his reply to Joan Robinson (1970): 'Neoclassical theory deals with macroeconomic aggregates, usually by constructing the aggregate theory by analogy with the corresponding microeconomic concepts. Whether or

not this is useful is an empirical question to which I believe an empirical answer can be given. This is the "faith" I have, but which is not shared by Mrs Robinson. Perhaps it would be better to say that the aggregate analogies provide working hypotheses for econometricians' (Ferguson (1971, pp. 251–2)).⁴⁷ This argument, however, is based on pure instrumentalism, and thus it is indefensible on methodological grounds (Blaug (1993)).⁴⁸ Furthermore, it was dispelled by Fisher (1971a): factor shares are not constant because the underlying aggregate technology is Cobb–Douglas; rather, the aggregate Cobb–Douglas works because factor shares are constant. The fact that Fisher *et al.* (1977) could not derive a similar organizing principle for the CES does not undermine the generality of the argument: aggregate production functions do not work *because* they are a summary of the aggregate technology.

Naturally, the aggregation problem appears in all areas of economics, including consumption theory, where a well-defined micro consumption theory exists. The neoclassical aggregate production function is also built by analogy. This is Ferguson's (1971) argument. The aggregation problem is therefore viewed as being merely a nihilistic position. Again, in the light of the discussion in this paper, this argument is untenable. Employing macroeconomic production functions on the unverified premise that inference by analogy is correct appears to be inadmissible, and the concept of 'representative firm' *à la* Marshall is, in general, inapplicable. Furthermore, the difference with the case of the consumption function is that the conditions for successful aggregation in this case, while strong, do not seem to be so outlandish as those in the case of the production function. The aggregate consumption function can be shown to exist so long as either individual marginal propensities to consume are constant and about equal; or so long as the distribution of income remains relatively fixed. These seem relatively plausible. See Green (1964, ch. 5).⁴⁹

⁴⁷ This seems to be also Solow's position: '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' (Solow (1966, pp. 1259–60)).

⁴⁸ But paradoxically, after criticizing the aggregate marginal productivity theory and referring to the aggregation problems (Blaug (1993, chs 9, 10)), Blaug (1993, p. 181) defends Ferguson and argues that there is *nothing absurd* with his faith in the neoclassical parables.

⁴⁹ Interestingly, Solow indicated that 'the aggregate production function is only a *little less* legitimate concept than, say, the aggregate consumption function . . .' (Solow (1957, p. 349), italics added). Certainly we disagree. Fisher (1969a, p. 575) compares the two sets of conditions, for production and consumption functions, and concludes that the former are substantially more stringent.

The second argument sometimes given to justify the use of aggregate production functions is that the aggregate production function is to be thought of as a *parable*, following the arguments in Samuelson's (1961–62) work. Samuelson claimed that even in cases with heterogeneous capital goods, some rationalization could be provided for the validity of the neoclassical parable, which assumes that there is a single homogeneous factor referred to as capital whose marginal product equals the interest rate. Samuelson worked with a one-commodity model assuming a well-behaved, constant returns to scale production function (i.e. the surrogate production function). His surrogate production function relies on the crucial assumption that the same proportion of inputs is used in the consumption-goods and capital-goods industries; i.e. the machines required for different techniques on the surrogate production function are different with respect to engineering specifications, but, with each technique, the ratio of labor to machines required to produce its machines is the same as that required to produce homogeneous consumption goods. This means that the cost of capital is determined solely by labor embodied in the machines required for each technique and the time pattern of all techniques is the same.⁵⁰ Then, Samuelson showed that the relation between the wage rate and the profit rate would be the same as that obtained from an appropriately defined surrogate production function with surrogate capital as a single factor of production. In competitive equilibrium, the wage rate is determined by the marginal productivity of labor. The latter is a ratio of two physical quantities, independent of prices (i.e. independent of distribution). And the same for the rate of profit: it is determined by the marginal productivity of capital. It is also measured in physical quantities. Under these circumstances, since there is a well-behaved production function, there is a unique inverse relation between the intensity of the factors and the relative price, and thus, as a resource becomes more scarce, its price increases. Thus, Samuelson turned the real economy with heterogeneous goods into an imaginary economy with a homogeneous output.

However, in the light of the aggregation literature, Samuelson's parable loses its power. Furthermore, the results of the one-commodity model do not hold in heterogeneous commodity models, and Samuelson's results depend crucially on the assumption of equal proportions, as shown by Garegnani (1970) (this assumption excludes reswitching). For the surrogate function to yield the correct total product, the 'surrogate capital' would have to coincide

⁵⁰ Samuelson, apart from working with a model where there is only one consumption good, and where input coefficients are fixed at the micro level, also assumed constant returns to scale, perfect competition, that only the n th capital good is used to produce the n th capital good, and that depreciation of a capital good is independent of its age.

with the value in terms of consumption of the capital in use. The surrogate production function cannot be generally defined. In the words of Brown: 'Given that assumption, one arrives at the simplest neoclassical (Clarkian) parable, in which there is one homogeneous malleable physical capital (actually, one can measure capital in value terms in this case, but the value capital behaves like a physical quantity), no joint production, and smooth substitutability of labor and the capital aggregate. The marginal productivity relations determine the functional income distribution and all the other variables in the general equilibrium system upon which the parable is based' (Brown (1980, p. 385)).

It is important to mention that by the time the debates died during the 1970s, Samuelson (1966) had conceded important points in the debate, e.g. reswitching and reverse capital deepening (see also Samuelson (1999)). This, however, did not deter neoclassical macroeconomists from arguing that, although theoretically correct, the important point was that reswitching was *empirically unimportant* (Stiglitz (1974), Blaug (1993, p. 183)), and thus they had faith that the factor substitution mechanisms postulated by the neoclassical approach do exist.⁵¹ It was contended that the production function was empirically useful (Ferguson (1971)).

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; but that they did not apply to the general equilibrium model conceived in disaggregated terms and based on the behavior of profit- and utility-maximizing agents. This position was best summarized by Hahn (1982) in his dismissal of the Sraffian criticism of the notion of aggregate capital, who argued that, although the criticism against marginalist theory that certain properties of the one-good neoclassical model could not be generalized was correct, in no way did it affect the fully disaggregated general equilibrium micro foundations of the neoclassical approach to value and distribution.

Here lies another important disagreement. Referring to the previous comment, Pasinetti asserts that 'this proposition actually has no objective foundation; phenomena of non-convexity, reswitching of techniques and

⁵¹ This is an 'interesting' critique since the problem does not concern the empirical likelihood of capital reversing. Mongiovi indicates that 'The principle of factor substitution originated not in the observation of empirical regularities but in a process of deduction from axioms presumed by the early marginalists to be plausible. The notion of price-elastic factor and commodity demand functions has so deeply penetrated the economic intuition of our age that to doubt their existence seems to contradict the obvious. But of course these functions have never been, and can never be, directly observed' (Mongiovi (1999, p. 7)).

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’ (Pasinetti (2000, p. 212)). Sraffa (1960) had made clear the possibility of reswitching and capital reversing without relying at all on any aggregation procedure. Garegnani (1970) also presented a model in which capital reversing occurs, even though there is no aggregate production function. See also Petri’s (1999) reply to Hahn’s arguments.

Petri (1999, p. 21) indicates 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, USA, side replied that their theory, in its rigorous versions, i.e. the neo-Walrasian formulations, did not need to aggregate anything. But as Petri 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’ (Petri (1999, p. 51), italics in the original). This has an important implication. This is that the only versions of neoclassical macroeconomic theory compatible with the analysis of long-period positions are the ones depending 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. What therefore seems odd is the justifications given by authors 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.

A variation of the parable argument is that the aggregate production function should be understood as an *approximation*. It is evident that Fisher’s (exact) aggregation conditions are so stringent that one can hardly believe that actual economies will satisfy them. Fisher (1969b), therefore, asked: what about the possibility of a *satisfactory approximation*? The motivation behind the question is very simple. In practice, what one cares about is whether aggregate production functions provide an adequate approximation to reality over the values of the variables that occur in practice. Thus suppose the values of capital and labor in the economy lie in a bounded set. And suppose further that the requirement is that an aggregate production function exists within some specified distance of the true production function for all points in the bounded set. Does this new restriction help the conditions for aggregation? One possible way to answer this question is by requiring that the exact

conditions hold only *approximately* (e.g. for approximate capital aggregation it suffices that all technical differences among firms be approximately capital augmenting). Is this a useful solution? Fisher showed it is not. The reason is that in reality there will be differences that are not approximately capital augmenting. Therefore, the interesting question is whether there are cases where the exact aggregation conditions are not approximately satisfied but in which an aggregate production function gives a satisfactory approximation for all points in the bounded set. Fisher (1969b) proved that *the only way* for approximate aggregation to hold without approximate satisfaction of the Leontief conditions is for the derivatives of the functions involved to wiggle violently up and down, an unnatural property not exhibited by the aggregate production functions used in practice.

The third argument given for the use of aggregate production functions is that there is no other option if one is to answer the questions for which the aggregate production function is used, e.g. to discuss productivity differences across nations. This argument acquires shape with the remark that it is hoped that the results be more or less qualitatively correct, and that they provide some guide to orders of magnitude (Solow (1988, p. 314)). This reasoning is a by-product of the instrumentalist position and it also clashes with the results of the aggregation work. Of course, if one insists on a research program whose goal is, for example, to split overall growth into the alleged contribution of technical progress and factor accumulation (i.e. growth accounting) at the country level, surely one needs an aggregate production function in order to allegedly relate aggregate output to aggregate inputs (and thus to speak of a country's multi-factor or total factor productivity). But if one realizes that the whole meaning of aggregates such as investment, GDP, labor and capital is questionable, as Fisher (1987) pointed out, the legitimacy of the research program collapses. And even at the conceptual level, the objective behind a growth accounting exercise for purposes of estimating total factor productivity growth is by no means universally shared (e.g. Kaldor (1957), Pasinetti (1959), Nelson (1973, 1981), Nelson and Winter (1982), Scott (1989), Fisher (1993)).⁵²

Finally, as indicated above, the notion of production function is fundamental as the basis for the aggregate neoclassical theory of distribution. In this model, the distribution of the product between the social classes is explained purely in technical terms (i.e. optimization, marginal productivities and capital-labor ratios), and thus the notion of aggregate production

⁵² Fisher (1993) indicates that as far back as 1970 he had already called 'into question the use of aggregate production functions in macroeconomic applications such as Solow's famous 1957 paper' (Fisher (1993, p. xiii)).

function is fundamental (Ferguson (1968)). To think of the distribution of output in terms of, for example, class conflict (or any other political, historical, sociological or psychological forces) is unthinkable, almost anathema, for many economists, on the grounds that it is not scientific (i.e. akin to sociology or political science, not *like* economics). On this see Blaug (1993, ch. 9) and Pasinetti (2000).⁵³

10. CONCLUSIONS: WHAT SHOULD APPLIED ECONOMISTS KNOW ABOUT AGGREGATE PRODUCTION FUNCTIONS?

This paper has aimed at providing a survey of the dense literature on aggregation in production with a view to drawing lessons for the applied economist. It is difficult to find an optimistic note on which to close. As far back as 1963, in his seminal survey on production and cost, Walters had concluded: 'After surveying the problems of aggregation one may easily doubt whether there is much point in employing such a concept as an aggregate production function. The variety of competitive and technological conditions one finds in modern economies suggest that one cannot approximate the basic requirements of sensible aggregation except, perhaps, over firms in the same industry or for narrow sections of the economy' (Walters (1963, p. 11)). More recently Burmeister, also after surveying the literature, concluded: 'I am not very optimistic [. . .] I have one revolutionary suggestion: Perhaps for the purpose of answering many macroeconomic questions—particularly about inflation and unemployment—we should disregard the concept of a production function at the macroeconomic level. The economist who succeeds in finding a suitable replacement will be a prime candidate for a future Nobel prize' (Burmeister (1980, pp. 427–8)).⁵⁴

Here is a summary of the main conclusions and lessons.

⁵³ Blaug claims that 'it would be a great advantage if the phrase "marginal productivity theory of distribution" were banished from the literature' (Blaug (1993, p. 171)). Regarding the possibility of testing it he argues that the marginal productivity theory is a 'highly abstract theory: it is formulated in terms so general as to make it virtually useless for answering specific questions about say, the structure of wages in the labor market' (Blaug (1993, pp. 174–5)). And: 'the famous or infamous marginal productivity of wages has never been spelled out in sufficient detail to be of much use in accounting for the observed pattern of relative wages. No wonder, therefore, that it has rarely been tested, and even where efforts have been made to put it to the test, the results have been inconclusive' (Blaug (1993, p. 176)). Thurow (1975, pp. 211–30) poses these problems nicely in the form of a series of questions.

⁵⁴ Recent works on unemployment where the aggregate production function plays a key role are Rowthorn (1999) and Blanchard and Wolfers (2000). Applied international trade theory also requires the traditional conception of capital.

(i) We have discussed some of the issues raised during the Cambridge–Cambridge capital controversies as well as the problems derived from the aggregation conditions. Although the starting point of both literatures is radically different, the conclusions, for purposes of applied economists, seem to *converge*: the notion of aggregate production function is rather problematic. The problem of aggregation of production functions is more serious than in other areas (e.g. consumption). The work on aggregation points out that aggregates such as investment, capital, labor and output do not have a sound theoretical foundation in the sense that it is not possible to aggregate variables and still preserve the central neoclassical claims. The conditions for successful aggregation are so stringent that one can hardly believe that actual economies satisfy them. If no optimization condition is imposed on the problem, Nataf's theorem indicates that aggregation over sectors is possible if and only if micro production functions are additively separable in capital and labor. Even imposing efficiency conditions as in Fisher's work, the aggregation conditions remain extremely restrictive. The existence of a labor aggregate requires that all firms employ different labor in the same proportion. This requires the absence of specialization in employment. Similarly, where there are many outputs, an output aggregate will exist if and only if all firms produce all outputs in the same proportion. This requires the absence of specialization in production, i.e. all firms must produce the same market basket of outputs differing only in their scale. In the Houthakker–Sato approach, the possibility of aggregation depends on the stability of distribution of the input–output coefficients. And in Gorman's framework it depends on the separability of the micro variable profit functions. All of them are extremely restrictive conditions. If the aggregates cannot be generated, the aggregate production function will not exist.

(ii) Economists act, however, as if aggregate output and capital were, in fact, generated from a well-behaved aggregate production function. This is simply and plainly wrong. In other words, investment, for example, means something in the national accounts and in the $GDP = C + I + G$ identity. However, the *relationship* $GDP = F(K, L)$ between aggregate output (GDP) and aggregate inputs (K, L) used in theoretical and applied macroeconomic work does not have, in general, a meaningful interpretation. This implies that the statement that there must be some connection between aggregate output and aggregate inputs, and that this is what the aggregate production function shows, has no theoretical basis. This should provide a clear answer to the question raised by Bernanke and quoted in section 2. However, as Fisher indicates, 'this has not discouraged macroeconomists from continuing to work in such terms' (Fisher (1987, p. 55)). This attitude

is prevalent in all areas of macroeconomics, but even more acute in growth theory.⁵⁵

(iii) ~~The~~ Cambridge–Cambridge controversies and the aggregation literature pose serious problems for whole areas of neoclassical macroeconomics, such as international trade, growth, investment and labor theories. The validity of the analyses in these fields depends crucially on the existence of the aggregates questioned in this survey. For example, thinking of aggregate investment as a well-defined addition to ‘capital’ in production is a mistake. This conceptualization, however, appears in many places, e.g. in Chirinko’s recent survey on investment, where he claims: ‘The demand for capital is derived from elementary economic principles, and is determined by the equality between the expected marginal benefits and costs from an additional unit of capital. This equality can be transformed so that the desired or optimal capital stock (K_t^*) depends on price variables, quantity variables, and autonomous shocks’ (Chirinko (1993, p. 1877)). And below: ‘A fundamental issue in investment research is the translation of the demand for the stock of capital into a demand for the flow of investment’ (Chirinko (1993, p. 1905)). Without reference to a well-behaved demand function for ‘capital’ one cannot derive the negative elasticity of investment with respect to the interest rate. Indeed, one cannot think of ‘investment’ as an addition to ‘capital’ in the sense of an increase in a well-defined factor of production. Hence, the rationale that the interest rate acts as the price bringing investment into line with full-employment savings vanishes. And Fisher, in his reply to Joan Robinson (1971b), indicated that ‘If aggregate capital does not exist, then of course one cannot believe in the marginal productivity of *aggregate* capital’ (Fisher (1971b, p. 405), italics in the original).⁵⁶ Further, Fisher’s (1971a) simulations

⁵⁵ In one of the very few cases where authors recently have bothered to mention the possible problems for applied work derived from the aggregation question, Basu and Fernald (1997) nevertheless argue: ‘The theorems of Fisher (1993) would seem to assure the existence of an aggregate production function. Fisher’s theorems do not apply to our setup, however, since factors are not necessarily allocated efficiently to maximize output’ (Basu and Fernald (1997, p. 266)). If this is true, then for sure Nataf’s theorem applies, and aggregation becomes a far more stringent problem. However, we believe this is a remarkable misunderstanding of Fisher’s conclusions which, if anything, ensure the non-existence of the aggregate production function.

⁵⁶ Unlike in the case of GDP above, the interpretation of a stock of capital constructed through the perpetual inventory method (i.e. adding past investments and subtracting the depreciation) can be a can of worms. First, even the subtraction of the depreciation is not so simple unless firms use Hotelling’s depreciation (Fisher and McGowan (1983)), which they do not. Otherwise, capital stock measurements are highly questionable figures. Second, it is much less clear the purpose of estimating the stock of capital for its own sake than the purpose of estimating total GDP (e.g. calculating the growth rate of the economy). A recent attempt at constructing capital stock figures for the USA is by Jorgenson (2001).

questioned the finding that the marginal product of labor explains the wage rate. As long as factor shares are sufficiently constant, the aggregate Cobb–Douglas marginal product (i.e. output per man) will give good wage predictions even though the underlying technical relationships are not consistent with the existence of any aggregate production function.

(iv) Economists use aggregate production functions for purposes without intrinsic content, e.g. to measure the aggregate elasticity of substitution, a concept which does not exist since there is no true aggregate parameter to which it corresponds. Likewise, the reasons given for continuing to use aggregate production functions are fallacious and thus unacceptable, e.g. that they work empirically; or that in order to perform growth accounting one needs to assume their existence. Mermaids do not exist simply because one insists on studying them!⁵⁷

(v) The aggregation problem is present whether there are only two firms or a thousand. However, it is fair to say that the more firms, the higher the likelihood that these firms will differ in ways that prevent aggregation.

(vi) Intuitions based on micro variables and micro production functions will often be false when applied to aggregates (e.g. growth models with micro foundations applied to study countries). In this sense, what, for example, is the meaning of multi-factor productivity, and Solow's residuals, in country-level growth accounting exercises or regressions pooling many countries? (Islam (1995, 1999, 2001), Prescott (1998)).⁵⁸ As indicated in section 1, without proper aggregation we cannot interpret the properties of an aggregate production function, which rules the behavior of (aggregate) total factor productivity.

(vii) The revival of growth theory during the last two decades no doubt has produced important discussions, and seemingly interesting empirical results.

⁵⁷ One discussant at a conference where a previous version of the paper was presented pointed out that mermaids do exist. The dictionary indicates that a mermaid is a *legendary* marine creature having the head and upper body of a woman and the tail of a fish (*italics added*).

⁵⁸ Regarding the estimation of total factor productivity growth, Nadiri commented in his survey: 'The conclusion to be drawn from this brief discussion 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' (Nadiri (1970, pp. 1145–46), *italics added*).

However, authors do not realize that they are using a tool whose lack of legitimacy was demonstrated decades ago. The consequence is that these empirical results are unjustifiable and even misleading. For example, an important aspect emphasized by the new models is the idea of increasing returns at the aggregate level. However, the aggregate production functions derived theoretically have constant returns to scale (recall that the aggregation conditions strongly depend on the existence of constant returns to scale at the micro level, and that with non-constant returns aggregates do not exist in general).⁵⁹

(viii) At the empirical level, and contrary to widespread belief (recall Fisher's justification of his simulations), production functions, when estimated econometrically, tend to yield, in general, poor results, a point made recently by Sylos Labini (1995) discussing estimations with the Cobb–Douglas function. This has been corroborated by McCombie (1998) and Felipe and Adams (2002), who subjected the original Cobb–Douglas (1928) data set to a series of stability tests. The results indicate that the famous regression is very fragile. Furthermore, adding a linear time trend to it to account for technical progress (something not done by Cobb and Douglas in their original work) yields very poor and questionable results (e.g. negative elasticity of capital). With today's econometric tools, nobody would conclude that this data set indicates that the elasticity of labor was 0.75 and that of capital 0.25 in the USA during the period analyzed, much less go as far as Douglas's (1976) extreme of claiming that the 'approximate coincidence of the estimated coefficients with the actual shares received also strengthens the competitive theory of distribution and disproves the Marxian' (Douglas (1976, p. 914)). Likewise, Temple (1998) applied robustness tests to the Mankiw *et al.* (1992) regression, and showed that the results were rather weak. As a corollary, if it is difficult to justify the existence of the aggregate production function as a summary of the technical relationships, one wonders how one can test theories that depend on the existence of such

⁵⁹ We also have to make a reference to the important fact that the new endogenous growth models have introduced as a factor of production a very problematic concept, namely that of the 'physical quantity of human capital' (H). What are the logical foundations or conditions under which this can be represented? As Petri (2001, p. 13) indicates, usually it is represented as influencing the efficiency units of labor (L) through a multiplicative effect HL . This procedure, however, is very vague and lacks rigor. The notion of human capital has to do with acquiring know-how, something different from increasing the amount of an input. Adding human capital is akin to adding more or better software to a computer. But then different quantities of human capital mean that one is dealing with different kinds of labor, and aggregatability and measurement of increases of the stock of human capital become very dubious notions.

a construct, or estimate the degree of returns to scale and the elasticity of substitution.⁶⁰

(ix) If the notion of aggregate production function is so problematic, it follows that we lose the sense of what it is supposed to measure in applied work (as noted in section 6, Fisher *et al.* (1977) indicated that the aggregate elasticity of substitution is an estimate of nothing). Then, is there any alternative interpretation of the estimates of aggregate production functions that does not presuppose the existence of an aggregate technology? (Black-orby and Schworm (1984, p. 647)). This question has been answered in the positive by Felipe (2000, 2001a, 2001b), Felipe and Adams (2002), Felipe and Holz (2001) and Felipe and McCombie (2001, 2002a, 2002b, 2002c, 2003), following the seminal work of Shaikh (1974, 1980) and Simon (1979). The answer, however, is most discouraging. They show that the *ex post* income accounting identity that relates the value of output (VA) to the sum of the wage bill (wL , where w is the average wage rate and L is employment) plus *total* profits (rK , where r is the average *ex post* profit rate and K is the stock of capital), i.e. $VA \equiv WL + rK$, can be easily rewritten through a simple algebraic transformation as $VA = A(t) F(K, L)$.⁶¹ The precise functional form (Cobb–Douglas, CES, translog etc.) corresponding to the data set in question will depend on the paths of the factor shares and of the weighted average of the wage and profit rates (where the weights are the factor shares).

The implication of this argument is that the precise form $VA = A(t) F(K, L)$ corresponding to the particular data set $VA \equiv wL + rK$ has to yield a perfect fit if estimated econometrically (because all that is being estimated is an identity); the putative elasticities have to coincide with the factor shares;

⁶⁰ For example, Solow claimed: ‘When someone claims that aggregate production functions work, he means (i) that they give a good fit to input–output data without the intervention of factor shares and (ii) that the function so fitted has partial derivatives that closely mimic observed factor shares’ (Solow (1974, p. 121)). This means, implicitly, that the aggregate production function can be tested.

⁶¹ $A(t)$ is a function of time. Note that r is not the ‘user cost of capital’, but the *ex post* profit rate that makes the accounting identity hold always. The income accounting identity $VA = wL + rK$ does not follow from Euler’s theorem (which at the aggregate level would require the aggregate production function to exist), and thus there is no reason why the wage and profit rates have to coincide with the respective aggregate marginal productivities (which we have argued do not exist). The identity simply shows how value added is divided between wages and total profits. The argument was supposedly refuted by Solow (1974). Shaikh (1980) replied. Solow (1987) again tried to attack it. For a reply see Felipe (2001b). Interestingly, Samuelson (1979) rediscovered the argument and used it in order to cast doubt on Douglas’s results (Felipe and Adams (2002)).

and the marginal products have to coincide with the factor prices. However, given that all this follows from an algebraic transformation of an accounting identity, it says nothing about the nature of production, returns to scale and distribution. This is true for any data set. In other words, the standard hypotheses about production embedded in the production function (i.e. returns to scale, elasticity of substitution, marginal theory of factor pricing) are untestable because they cannot be refuted (see the Felipe and McCombie (2003) discussion of Kim and Lau's (1994) supposed tests and refutation of these hypotheses).

A corollary of this argument is that *the problem* with the empirical work undertaken during the last 70 years is that the production functions estimated were not the ones that corresponded to the particular data sets used. For example, often, applied economists have fitted the Cobb–Douglas form including a time trend. This, in most cases, has yielded very poor results, as noted in (vii) above. The ‘solution’, in general, is to fit the Cobb–Douglas but including a trigonometric function instead of the linear time trend.⁶² This will provide, in most cases, an almost perfect approximation to the accounting identity with the results mentioned above. For example, Felipe and McCombie (2001), by referring to the accounting identity, find the ‘organizing principle’ with which to explain when the aggregate CES production function will work, a problem Fisher *et al.* (1977) could not solve.⁶³

⁶² The Cobb–Douglas form will work in most cases because factor shares tend to be constant (Fisher (1971a), Shaikh (1980)). Felipe and Holz (2001) show, using Monte Carlo simulations, that the Cobb–Douglas with a linear trend is very robust to variations in the factor shares. What makes it yield very poor results very often is the fact that the linear time trend provides a very poor approximation to the weighted average of the wage and profit rates. This is one of the approaches suggested by Burmeister (1980, p. 427) in order to understand why aggregate production functions sometimes work.

⁶³ This organizing principle is that the labor share follow the path $a_t = L_t^\rho / V_t$, where $V_t = [\delta L_t^\rho + (1 - \delta)K_t^\rho]$, δ is the distribution parameter and $\sigma = 1/(1 + \rho)$ is the elasticity of substitution. In other words, the CES production function will ‘work’ if the labor share follows such a path. On the other hand, this is a natural result, since that path is precisely the share predicted by the CES. This suggests, consistent with the results of Fisher (1971a), that two-factor aggregate production functions tend to ‘work’ if their predictions about labor’s share happen to be approximately correct. On the other hand, the result is at odds with the failure of Fisher *et al.* (1977) to find that the aggregate CES ‘works’ if the wage is a log-linear function of output per person-hour, since that property is known to be equivalent to the CES with labor’s share following the indicated path. In any event, it should come as no surprise that historically there is very little empirical evidence about the CES form. The organizing principle seems to be so complicated (we know that in most cases factor shares are relatively constant) that one should not expect actual economies to satisfy it. The evidence about the CES comes from the side equations (Fisher *et al.* (1977)) which, as Felipe and McCombie (2001) show, can also be interpreted in terms of the accounting identity.

Felipe (2001a) reassessed the debate around the quantification of the effect of government expenditures on infrastructure on private sector output and productivity, and Felipe (2001b) reinterprets the evidence that the recent empirical endogenous growth literature has provided (e.g. Romer (1987)), namely, whether there the assumption of increasing returns, made by many new growth models, has empirical support. In both cases the functional forms estimated, derived from an aggregate production function, can be directly derived as algebraic transformations of the income accounting identity. The puzzles discussed in the literature (e.g. size of the estimated parameters) disappear: the regressions estimated have to be homogeneous of degree 1 in K and L . Examples are provided where a Cobb–Douglas with a linear time trend yields very poor results (e.g. negative elasticity of capital). However, when the equations are re-estimated substituting a function of sines and cosines for the linear trend, results improve dramatically and approach those described above.

Felipe and McCombie (2002a) have re-evaluated the recent literature, following Hall (1988), on the estimation of alleged markups (the ratio of price to marginal cost) and the discussion of whether markets are competitive. Felipe and McCombie show that the regression proposed by Hall to estimate the parameter that he defines as the markup is a transformation of the income accounting identity, but misspecified due to omitted variable bias. It is shown that if the regression were correctly specified, the alleged markup should be unity simply because of the identity, not because markets are competitive.

Felipe (2000) and Felipe and McCombie (2002b, 2003) analyzed the debate about the sources of growth in East Asia and China, i.e. low total factor productivity growth in most countries in the region, in particular in Singapore. The functional forms used can also be derived as algebraic transformations of the income accounting identity and thus the analysis undertaken loses the implications that the authors tried to establish.

In other analyses economists do not estimate directly the aggregate production function. Rather, the latter is part of a model (i.e. the production function together with a series of assumptions) such that the form estimated empirically does not look like a production function. This is the case of the recent work by Mankiw *et al.* (1992), where the authors attempted to test Solow's (1956) model by estimating the steady-state income per capita equation. Felipe and McCombie (2002c) show that, because the authors' starting point was an aggregate production function, the equation estimated can also be derived as a transformation of the income accounting identity. Furthermore, empirically Mankiw *et al.* (1992) pooled time series and cross-section data for over 100 countries, thus assuming not only that different countries

use roughly the same production function at a given point in time, but that such a worldwide aggregate production function does in fact exist.

From this perspective, it also follows that Solow's residual (total factor productivity growth), the *Holy Grail* of the neoclassical growth model, is nothing but a weighted average of the growth rates of the wage (\hat{w}_t) and profit rates (\hat{r}_t), i.e. $a_t \hat{w}_t + (1 - a_t) \hat{r}_t$, where the weights are the factor shares (a_t and $1 - a_t$).⁶⁴ Although this is known (Jorgenson and Griliches (1967)), and it is referred to as the dual measure of productivity, what seems not to be appreciated is that $a_t \hat{w}_t + (1 - a_t) \hat{r}_t$ is a tautology because it follows from an identity and has no relationship to the notion of aggregate production function, which *does not* exist. Therefore, $a_t \hat{w}_t + (1 - a_t) \hat{r}_t$ cannot unambiguously be interpreted as a measure of productivity growth. If anything, it can be interpreted as a measure of distributional changes. If neoclassical macroeconomists realized this simple fact (i.e. that what is behind what they calculate as the difference between the growth of output and the weighted average of the growth of inputs is the weighted average of the wage and profit rates; this is the growth of output 'not explained' by the growth of inputs, in a tautological sense), perhaps they *would not*: (a) try to 'explain' Solow's residual the way they do (Islam (1999, 2001)); (b) proxy it with a linear trend, as is done in hundreds of papers, since they would know that what they need is a complicated function of sines and cosines; or (c) urge the profession to develop a theory of total factor productivity growth (Prescott (1998)).⁶⁵ Prescott has recently claimed that: 'The neoclassical production function is the cornerstone of the theory [neoclassical growth theory] and is used in virtually all applied aggregate analyses. . . . Another beauty of this construct is that it deals with well-defined aggregate inputs and outputs. A final beauty is that it is based on a lot of theory. . . . Thus, this theory is a theory of the income side of national income and production accounts (NIPA), as well as a theory of production' (Prescott (1998, p. 532)). We disagree with this view.

Certainly one can (stubbornly) argue that macroeconomists use the relationship $GDP = F(K, L)$ without reference to the aggregation literature because such a relationship is simply a general representation of how (aggregate) inputs and output are related, and therefore it need not be derived from micro principles. The issue, however, is whether doing this means anything at all, given what we have argued. We do not think so. The aggregate production function is either a representation of the technology derived from

⁶⁴ This follows by expressing the identity $VA = wL + rK$ in growth rates.

⁶⁵ On the relationship between this residual and business cycles see Felipe and McCombie (2002a).

micro relationships, or else is taken to be an approximation for empirical purposes. We have shown that neither one is correct. From a theoretical point of view, without proper aggregation we cannot interpret the properties of the aggregate production function. Of course one can simply *assume* that the alleged aggregate production function has certain properties (e.g. positive marginal products), but that will not make it so. More to the point, aggregate production functions do not generally exist. So just what is *it* that is to have the properties being assumed? And empirically, we have seen that the identity argument governs the empirical results. The arguments and examples in the papers above show that the empirical foundations of production, distribution and productivity of neoclassical macro analyses are rather weak, and that estimating aggregate production functions is a pointless exercise because if estimated correctly it cannot be statistically refuted.⁶⁶

(x) Finally, is there anything that can be done? Perhaps a complete answer cannot be provided, but this should not prevent one from concluding that economists ought to be much more careful in using the tool. The problem with the aggregate production function, i.e. that economists continue using it, does not lie in itself. Rather, the issue at stake is a whole way of thinking, central to which is the (macro) neoclassical approach to distribution. This is what must be questioned. It may be argued that one does not need an aggregate production function to study growth unless one insists that the only possible conception of growth is the neoclassical model. Although this is true, the problems remain since most models, if not all, use or refer to, one way or another, aggregate capital or aggregate capital-labor ratios (e.g. Kaldor (1957), Pasinetti (1974)). If, as Temple (1999, cited in section 1) indicates, many economists regard the aggregate production function as a useful device to explain output levels and growth rates, it is mostly because this is the only way it is explained in graduate and undergraduate courses (Barro and Sala-i-Martin (1995), Jones (1998)). Economists follow Popper and Lakatos and tend to hang on to an established research program despite anomalies if no better alternative program is available. While this attitude may be justified in some instances (e.g. that the existence of positively inclined demand curves does not invalidate most arguments in economics: Blaug (1993, p. 182)), this cannot be the case here. The aggregation problem and its consequences, and the impossibility of testing empirically the aggregate production function

⁶⁶ Lavoie (2000) has shown how some widely used unemployment specifications derived by Layard *et al.* (1991), e.g. the relationship between increases in real wages and unemployment, can be easily derived by manipulating the income accounting identity, thus depriving them of their alleged behavioral interpretation.

as discussed in (ix), are substantially more serious than a mere anomaly. Macroeconomists should pause before continuing to do applied work with no sound foundation and dedicate some time to studying other approaches to value, distribution, employment, growth, technical progress etc., in order to understand which questions can legitimately be posed to the empirical aggregate data.

REFERENCES

- Aghion P., Howitt P. (1998): *Endogenous Growth Theory*, MIT Press, Cambridge, MA.
- Arrow K., Chenery H. B., Minhas B. S., Solow R. M. (1961): 'Capital-labor substitution and economic efficiency', *Review of Economics and Statistics*, 43, pp. 225-50.
- Barro R. J., Sala-i-Martin X. (1995): *Economic Growth*. McGraw Hill, New York.
- Basu S., Fernald J. G. (1997): 'Returns to scale in U.S. production: estimates and implications', *Journal of Political Economy*, 105 (2), pp. 249-83.
- Bernanke B. S. (1987): 'Comment', *NBER Macroeconomics Annual 1987*, MIT Press, Cambridge, MA, pp. 202-5.
- Blackorby C., Schworm W. (1982): 'Aggregate investment and consistent intertemporal technologies', *Review of Economic Studies*, 49, pp. 595-614.
- Blackorby C., Schworm W. (1984): 'The structure of economies with aggregate measures of capital: a complete characterization', *Review of Economic Studies*, 51, pp. 633-50.
- Blackorby C., Schworm W. (1988): 'The existence of input and output aggregates in aggregate production functions', *Econometrica*, 56 (3), pp. 613-43.
- Blanchard O., Wolfers J. (2000): 'The role of shocks and institutions in the rise of European unemployment: the aggregate evidence', *Economic Journal*, 110 (March), pp. C1-C33.
- Blaug M. (1993): *The Methodology of Economics. Or How Economists Explain*, 2nd edn, Cambridge University Press, Cambridge.
- Bliss C. J. (1975): *Capital Theory and the Distribution of Income*, North-Holland, Amsterdam.
- Brown M. (1980): 'The measurement of capital aggregates: a post-reswitching problem', In Usher D. (ed.): *The Measurement of Capital*, University of Chicago Press, Chicago, pp. 377-420.
- Brown M. (1987): 'Cobb-Douglas functions', in Eatwell J., Millgate M., Newman P. (eds): *The New Palgrave. A Dictionary of Economics*, Macmillan, London, pp. 460-1.
- Burmeister E. (1980): 'Comment', in Usher D. (ed.): *The Measurement of Capital*, University of Chicago Press, Chicago, pp. 420-31.
- Burmeister E. (1990): 'Wicksell effects', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 257-61.
- Champernowne D. G. (1953-54): 'The production function and the theory of capital: a comment', *Review of Economic Studies*, 21 (2), pp. 112-35.
- Chirinko R. (1993): 'Business fixed investment spending: modeling strategies, empirical results, and policy implications', *Journal of Economic Literature*, 31, pp. 1875-911.
- Clark J. B. (1893): *The Distribution of Wealth*, Macmillan, New York.
- Cobb C. W., Douglas P. H. (1928): 'A theory of production', *American Economic Review*, 18, Suppl. (March); *Papers and Proceedings of the Fortieth Annual Meeting of the American Economic Association*, December 1927, pp. 139-65.
- Cohen A. J. (1993): 'What was abandoned following the Cambridge capital controversies? Samuelson, substance, scarcity, and value', in de Marchi N. (ed.): *Non-natural Social Science: Reflecting on the Enterprise of More Heat than Light*, Annual Supplement, *History of Political Economy*, 25, Duke University Press, Durham, NC, pp. 202-29.

- Diewert E. W. (1978): 'Hicks' aggregation theorem and the existence of a real value-added function', in Fuss M., McFadden D. (eds): *Production Economics: A Dual Approach to Theory and Applications*, North-Holland, Amsterdam, pp. 17–51.
- Diewert E. W. (1980): 'Aggregation problems in the measurement of capital', in Usher D. (ed.): *The Measurement of Capital*, University of Chicago Press, Chicago, pp. 433–528.
- Douglas P. (1976): 'The Cobb–Douglas production function once again: its history, its testing, and some empirical values', *Journal of Political Economy*, 84, pp. 903–15.
- Eatwell J. (1990): 'Walras's theory of capital', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 247–56.
- Felipe J. (2000): 'On the myth and mystery of Singapore's "zero TFP"', *Asian Economic Journal*, 14 (2), pp. 187–209.
- Felipe J. (2001a): 'Aggregate production functions and the measurement of infrastructure productivity: a reassessment', *Eastern Economic Journal*, 27 (3), pp. 323–44.
- Felipe J. (2001b): 'Endogenous growth, increasing returns, and externalities: an alternative interpretation of the evidence', *Metroeconomica*, 52 (4), pp. 391–427.
- Felipe J., Adams F. G. (2002): '“A Theory of Production”'. The estimation of the Cobb–Douglas function: a retrospective view', Manuscript, Georgia Institute of Technology and Northeastern University.
- Felipe J., Holz C. (2001): 'Why do aggregate production functions work? Fisher's simulations, Shaikh's identity, and some new results', *International Review of Applied Economics*, 15 (3), pp. 261–85.
- Felipe J., McCombie J. S. L. (2001): 'The CES production function, the accounting identity, and Occam's razor', *Applied Economics*, 33 (10), pp. 1221–32.
- Felipe J., McCombie J. S. L. (2002a): 'A problem with some recent estimations and interpretations of the markup in manufacturing industry', *International Review of Applied Economics*, 16 (2), pp. 187–215.
- Felipe J., McCombie J. S. L. (2002b): 'Productivity in China before and after 1978 revisited', *Zagreb International Review of Economics and Business*, 5 (1), pp. 17–43.
- Felipe J., McCombie J. S. L. (2002c): 'Why are some countries richer than others? A reassessment of Mankiw–Romer–Weil's test of the neoclassical growth model', Asian Development Bank, Economics and Research Department, working Paper no. 19 (August).
- Felipe J., McCombie J. S. L. (2003): 'Methodological problems with the neoclassical analysis of the East Asian miracle', *Cambridge Journal of Economics*, forthcoming.
- Ferguson C. E. (1968): 'Neoclassical theory of technical progress and relative factor shares', *Southern Economic Journal*, 34 (4), pp. 490–504.
- Ferguson C. E. (1971): 'Capital theory up to date: a comment on Mrs. Robinson's article', *Canadian Journal of Economics*, 4 (2), pp. 250–4.
- Fisher F. M. (1965): 'Embodied technical change and the existence of an aggregate capital stock', *Review of Economic Studies*, 32, pp. 263–88.
- Fisher F. M. (1968): 'Embodied technology and the existence of labor and output aggregates', *Review of Economic Studies*, 35, pp. 391–412.
- Fisher F. M. (1969a): 'The existence of aggregate production functions', *Econometrica*, 37 (4), pp. 553–77.
- Fisher F. M. (1969b): 'Approximate aggregation and the Leontief conditions', *Econometrica*, 37 (4), pp. 457–69.
- Fisher F. M. (1971a): 'Aggregate production functions and the explanation of wages: a simulation experiment', *Review of Economics and Statistics*, 53 (4), pp. 305–25.
- Fisher F. M. (1971b): 'Reply', *Econometrica*, 39 (2), p. 405.
- Fisher F. M. (1982): 'Aggregate production functions revisited: the mobility of capital and the rigidity of thought', *Review of Economic Studies*, 49 (4), pp. 615–26.
- Fisher F. M. (1983): 'On the simultaneous existence of full and partial capital aggregates', *Review of Economic Studies*, 50 (1), pp. 197–208.
- Fisher F. M. (1987): 'Aggregation problem', in Eatwell J., Millgate M., Newman P. (eds): *The New Palgrave. A Dictionary of Economics*, Macmillan, London, pp. 53–5.

- Fisher F. M. (1993): *Aggregation. Aggregate Production Functions and Related Topics*. MIT Press, Cambridge, MA.
- Fisher F. M., McGowan J. J. (1983): 'On the misuse of accounting rates of return to infer monopoly profits', *American Economic Review*, 73 (1), pp. 82–97.
- Fisher F. M., Solow R., Kearn J. M. (1977): 'Aggregate production functions: some CES experiments', *Review of Economic Studies*, 44 (2), pp. 305–20.
- Garegnani P. (1970): 'Heterogeneous capital, the production function and the theory of distribution', *Review of Economic Studies*, 37, pp. 407–36.
- Garegnani P. (1990): 'Quantity of capital', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 1–78.
- Gorman W. M. (1953): 'Community preference fields', *Econometrica*, 21, pp. 63–80.
- Gorman W. M. (1954): 'Klein aggregates and conventional index numbers', *Econometrica*, 22, pp. 113–15.
- Gorman W. M. (1959): 'Separable utility and aggregation', *Econometrica*, 27, pp. 469–81.
- Gorman W. M. (1968): 'Measuring the quantities of fixed factors', in Wolfe J. N. (ed.): *Value, Capital, and Growth. Papers in Honour of Sir John Hicks*, Edinburgh University Press, Edinburgh, pp. 141–72.
- Green H. A. J. (1964): *Aggregation in Economic Analysis. An Introductory Survey*, Princeton University Press, Princeton, NJ.
- Hagemann H. (1990): 'Capital goods', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 123–7.
- Hahn F. H. (1982): 'The neo-Ricardians', *Cambridge Journal of Economics*, 6, pp. 352–74.
- Hall R. (1988): 'The relation between price and marginal cost in U. S. industry', *Journal of Political Economy*, 96, pp. 921–47.
- Harcourt G. C. (1969): 'Some Cambridge controversies in the theory of capital', *Journal of Economic Literature*, 7 (2), pp. 369–405.
- Harcourt G. C. (1972): *Some Cambridge Controversies in the Theory of Capital*, Cambridge University Press, Cambridge.
- Heathfield D. F., Wibe S. (1987): *An Introduction to Cost and Production Functions*, Humanities Press International, Atlantic Highlands, NJ.
- Hennings K. H. (1990): 'Capital as a factor of production', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 108–22.
- Houthakker H. S. (1955–56): 'The Pareto distribution and the Cobb–Douglas production function in activity analysis', *Review of Economic Studies*, 23, pp. 27–31.
- Hulten C. R. (2000): 'Total factor productivity: a short biography', *NBER Working Paper 7471*, (January).
- Islam N. (1995): 'Growth empirics: a panel data approach', *Quarterly Journal of Economics*, 110 (4), pp. 1127–70.
- Islam N. (1999): 'International comparisons of total factor productivity: a review', *Review of Income and Wealth*, 45 (4), pp. 493–518.
- Islam N. (2001): 'Different approaches to international comparison of total factor productivity', in Hulten C., Dean E., Harper M. (eds): *New Developments in Productivity Analysis*, University of Chicago Press for the National Bureau of Economic Research, Chicago, IL, pp. 465–502.
- Johansen L. (1972): *Production Functions*, North-Holland, Amsterdam.
- Jones C. I. (1998): *Introduction to Economic Growth*, Norton, New York.
- Jorgenson D. W. (2001): 'Information technology and the U. S. economy', *American Economic Review*, 91 (1), pp. 1–32.
- Jorgenson D. W., Griliches Z. (1967): 'The explanation of productivity change', *Review of Economic Studies*, 34 (3), pp. 249–83.
- Jorgenson D. W., Yun K. Y. (1984): 'Tax policy and capital allocation', Harvard Institute of Economic Research Discussion Paper 1107.
- Kaldor N. (1957): 'A model of economic growth', *Economic Journal*, 67 (268), pp. 591–624.

- Kim J.-I., Lau L. (1994): 'The sources of economic growth of the East Asian newly industrialized countries,' *Journal of the Japanese and International Economies*, 8, pp. 235–71.
- Klein L. R. (1946a): 'Macroeconomics and the theory of rational behavior,' *Econometrica*, 14 (2), pp. 93–108.
- Klein L. R. (1946b): 'Remarks on the theory of aggregation,' *Econometrica*, 14 (4), pp. 303–12.
- Kurz H. (1990): 'Factor price frontier,' in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 155–60.
- Lavoie M. (2000): 'Le chômage d'équilibre: réalité ou artefact statistique?', *Revue Economique*, 51 (6), pp. 1477–84.
- Layard R., Nickell S. J., Jackman R. (1991): *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford University Press, Oxford.
- Leontief W. W. (1947a): 'Introduction to a theory of the internal structure of functional relationships,' *Econometrica*, 15 (4), pp. 361–73.
- Leontief W. W. (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, pp. 343–50.
- Levhari D. (1968): 'A note on Houthakker's aggregate production function in a multifirm industry,' *Econometrica*, 36 (1), pp. 151–4.
- Mankiw G. (1995): 'The growth of nations,' *Brookings Papers on Economic Activity*, 1 (September), pp. 275–325.
- Mankiw G. (1997): 'Comment,' in Bernanke B. S., Rotemberg J. (eds): *NBER Macroeconomics Annual 1997*, MIT Press, Cambridge, MA, pp. 103–7.
- Mankiw G., Romer D., Weil D. (1992): 'A contribution to the empirics of economic growth,' *Quarterly Journal of Economics*, 107 (2), pp. 407–37.
- May K. (1946): 'The aggregation problem for a one-industry model,' *Econometrica*, 14 (4), pp. 285–98.
- May K. (1947): 'Technological change and aggregation,' *Econometrica*, 15 (1), pp. 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, pp. 141–73.
- Mongiovi G. (1999): 'Introduction,' in Mongiovi G., Petri F. (eds): *Value, Distribution, and Capital. Essays in Honour of Pierangelo Garegnani*, Routledge, London, pp. 1–16.
- Nadiri M. I. (1970): 'Some approaches to the theory and measurement of factor productivity: a survey,' *Journal of Economic Literature*, 8 (December), pp. 1137–77.
- Nataf A. (1948): 'Sur la possibilité de construction de certains macromodèles,' *Econometrica*, 16 (3), pp. 232–44.
- Nelson R. (1973): 'Recent exercises in growth accounting: new understanding or dead end?', *American Economic Review*, 73 (2), pp. 462–8.
- Nelson R. (1981): 'Research on productivity growth and productivity differences: dead ends and new departures,' *Journal of Economic Literature*, 19 (September), pp. 1029–64.
- Nelson R., Winter S. G. (1982): *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA.
- Pasinetti L. (1959): 'On concepts and measures of changes in productivity,' *Review of Economics and Statistics*, 41, pp. 270–86.
- Pasinetti L. (1974): *Growth and Income Distribution. Essays in Economic Theory*, Cambridge University Press, Cambridge.
- Pasinetti L. (2000): 'Critique of the neoclassical theory of growth and distribution,' *Moneta e Credito, Banca Nazionale del Lavoro Quarterly Review*, 210 (June), pp. 187–232.
- Pasinetti L., Scazzieri R. (1990): 'Capital theory: paradoxes,' in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 136–47.
- Petri F. (1999): 'Professor Hahn on the "neo-Ricardian" criticism of neoclassical economics,' in Mongiovi G., Petri F. (eds): *Value, Distribution, and Capital. Essays in Honour of Pierangelo Garegnani*, Routledge, London, pp. 19–68.

- Petri F. (2001): 'Theory of output growth and of per capita output growth: with or without say's law?', Manuscript, Dipartimento di Economia Politica, Università di Siena.
- Pivetti M. (1990): 'Wicksell's theory of capital', in Eatwell J., Milgate M., Newman P. (eds): *The New Palgrave. Capital Theory*, Macmillan, London, pp. 262–9.
- Prescott E. C. (1998): 'Needed: a theory of total factor productivity', *International Economic Review*, 39 (3), pp. 525–52.
- Pu S. S. (1946): 'A note on macroeconomics', *Econometrica*, 14 (4), pp. 299–302.
- Robinson J. (1953–54): 'The production function and the theory of capital', *Review of Economic Studies*, 21 (2), pp. 81–106.
- Robinson J. (1955–56): 'The production function and the theory of capital—a reply', *Review of Economic Studies*, 23 (2), p. 247.
- Robinson J. (1956): *The Accumulation of Capital*, Macmillan, London.
- Robinson J. (1970): 'Capital theory up to date', *Canadian Journal of Economics*, 3 (2), pp. 309–17.
- Robinson J. (1971a): 'The measure of capital: the end of the controversy', *Economic Journal*, 81 (323), pp. 597–602.
- Robinson J. (1971b): 'The existence of aggregate production functions: comment', *Econometrica*, 39 (2), p. 405.
- Robinson J. (1975): 'The unimportance of reswitching', *Quarterly Journal of Economics*, 89 (1), pp. 32–9.
- Romer P. (1987): 'Crazy explanations for the productivity slowdown', *NBER Macroeconomics Annual 1987*, MIT Press, Cambridge, MA, pp. 163–202.
- Rowthorn R. (1999): 'Unemployment, wage bargaining and capital–labor substitution', *Cambridge Journal of Economics*, 23, pp. 413–25.
- Samuelson P. (1961–62): 'Parable and realism in capital theory: the surrogate production function', *Review of Economic Studies*, 29 (June), pp. 193–206.
- Samuelson P. (1966): 'A summing up', *Quarterly Journal of Economics*, 80 (4), pp. 568–83.
- Samuelson P. (1979): 'Paul Douglas's measurement of production functions and marginal productivities', *Journal of Political Economy*, 87, pp. 923–39.
- Samuelson P. (1999): 'The special thing that I learned from Sraffa', in Mongiovi G., Petri F. (eds): *Value, Distribution, and Capital. Essays in Honour of Pierangelo Garegnani*, Routledge, London, pp. 230–7.
- Sato K. (1975): *Production Functions and Aggregation*, North-Holland, Amsterdam.
- Scott M. F. (1989): *A New view of Economic Growth*, Clarendon Press, Oxford.
- Shaikh A. (1974): 'Laws of production and laws of algebra: the humbug production function', *Review of Economics and Statistics*, 56 (61), pp. 115–20.
- Shaikh A. (1980): 'Laws of production and laws of algebra: humbug II', in Nell E. J. (ed.): *Growth, Profits and Property, Essays in the Revival of Political Economy*, Cambridge University Press, Cambridge, pp. 80–95.
- Simon H. (1979): 'On parsimonious explanations of production relations', *Scandinavian Journal of Economics*, 81, pp. 459–74.
- Solow R. (1955–56): 'The production function and the theory of capital', *Review of Economic Studies*, 23 (2), pp. 101–8.
- Solow R. (1956): 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, 70, pp. 65–94.
- Solow R. (1957): 'Technical change and the aggregate production function', *Review of Economics and Statistics*, 39, pp. 312–20.
- Solow R. (1966): 'Review of *Capital and Growth*', *American Economic Review*, 56 (5), pp. 1257–60.
- Solow R. (1974): 'Laws of production and laws of algebra: the humbug production function: comment', *Review of Economics and Statistics*, 56, p. 121.
- Solow R. (1987): 'Second thoughts on growth theory', in Steinherr A., Weiserbs D. (eds): *Employment and Growth: Issues for the 1980s*, Martinus Nijhoff, Dordrecht.
- Solow R. (1988): 'Growth theory and after', *American Economic Review*, 78 (3), pp. 307–17.

- Solow R. (1994): 'Perspectives on growth theory', *Journal of Economic Perspectives*, 8 (1), pp. 45–54.
- Sraffa P. (1960): *Production of Commodities by Means of Commodities: A Prelude to a Critique of Economic Theory*, Cambridge University Press, Cambridge.
- 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), pp. 893–903.
- Sylos Labini P. (1995): 'Why the interpretation of the Cobb–Douglas production function must be radically changed', *Structural Change and Economic Dynamics*, 6, pp. 485–504.
- Temple J. (1998): 'Robustness tests of the augmented Solow model', *Journal of Applied Econometrics*, 13, pp. 361–75.
- Temple J. (1999): 'The new growth evidence', *Journal of Economic Literature*, 37 (March), pp. 112–56.
- Thurow L. C. (1975): *Generating Inequality*, Macmillan, London.
- Usher D. (1980): 'Introduction', in Usher D. (ed.): *The Measurement of Capital*, University of Chicago Press, Chicago, IL, pp. 1–21.
- Walras L. (1926): *Elements of Pure Economics* (definitive edition, Jaffé translation), Allen and Unwin, London [1954].
- Walters A. A. (1963): 'Production and cost functions', *Econometrica*, 31 (1–2), pp. 1–66.

Jesus Felipe
Asian Development Bank
P.O. Box 789
0980 Manila
Philippines
E-mail: jfelipe@adb.org

Franklin M. Fisher
Massachusetts Institute of Technology
Cambridge
MA 02139-4307
USA
E-mail: ffisher@mit.edu