AGGREGATION IN PRODUCTION FUNCTIONS: WHAT APPLIED ECONOMISTS SHOULD KNOW

Jesus Felipe
Georgia Institute of Technology
Atlanta, GA 30332-0610

e-mail: jesus.felipe@inta.gatech.edu

Franklin M. Fisher
Massachusetts Institute of Technology
Cambridge, MA 02139-4307

e-mail: ffisher@mit.edu

September 19, 2001

Jesus Felipe acknowledges financial support from the Center for International Business Education and Research (CIBER) at the Georgia Institute of Technology.

AGGREGATION IN PRODUCTION FUNCTIONS: WHAT APPLIED ECONOMISTS SHOULD KNOW

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

- Dostoevsky

Abstract: This paper surveys the theoretical literature on aggregation of production functions from the point of view of the applied economist. We refer to both the Cambridge Capital controversies and to the aggregation conditions. The most important 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 an aggregate production function can be derived from micro production functions are so stringent that it is difficult to believe that actual economies satisfy them. 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, generating and thinking of GDP as GDP=F(K,L), where K and L are aggregates of capital and labor, respectively, is incorrect. And thinking of aggregate investment as a well-defined addition to capital in production is 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 endogenous growth literature 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. 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 neoclassical growth models is the aggregate production function, a relationship that is intended to describe the technological relationships at the aggregate level. The problem is that the macro production function is a fictitious entity. At the theoretical level it is built by adding micro production functions. However, there is an extensive body of literature, developed since the 1940s, which points out that aggregating micro production functions into a macro production function is extremely problematic.

The purpose of this survey is to summarize the existing literature on the conditions under which an aggregate production function exists, and to discuss the implications for applied work. This is the so-called aggregation literature and the issue at hand is referred to as the *aggregation problem*. The importance of the topic lies in the fact that it seems that all the new growth literature has overlooked this *old* problem, and aggregate production functions have made since the 1980s an untroubled reappearance in mainstream macroeconomics. The review of the literature is done from the point of view of the applied economist. Thus, 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.

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 were a series of debates during the 1950s and 1960s between Joan Robinson and her colleagues at Cambridge, England, and Paul Samuelson and Robert Solow and their colleagues at Cambridge, U.S.A. Harcourt (1969) provides an excellent summary of what the debates were about. At stake were radically different value theories, the classical and the neoclassical, and the crux of the debate was 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. Quite naturally, capital, its marginal product, and the aggregate production function became part of the debate.

On the other hand, there is a related literature on aggregation of production functions which also questioned the macro-aggregates and the notion of aggregate production function, although from an altogether different point of view. 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? The question is as follows. Suppose we have two production functions $Q^A = f^A(K_1^A, K_2^A)$ and $Q^B = f^B(K_1^B, K_2^B)$ for sectors A and B, and where $K_1 = K_1^A + K_1^B$ and $L_2 = L_2^A + L_2^B$ (K refers to capital — two types- and L to labor —assumed homogeneous) The problem is to determine whether and in what circumstances one can derive a function $K = h(K_1, K_2)$ from the production functions f^A and f^B, where the aggregator function h(•) has the property that $G(K, L) = G[h(K_1, K_2)] = \Psi(Q^A, Q^B)$, and where the function Ψ is the production possibility curve for the economy.

From the point of view of the applied practitioner, 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; (ii) to apportion total growth into the accumulation of factors of production, and technical change between two periods; and (iii) to test theories and quantify their predictions. 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?[1]

The main target of this non-technical survey is the new generation of economists, in particular those undertaking applied work. The reason is that, in the light of the conclusions derived from the Cambridge debates and from the aggregation problems (rather negative as we shall see), one can t help asking *why* macro-economists 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, 487). However, in a recently published survey on the new growth theories, Jonathan Temple 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 1998, 15). Is this a good enough reason to use an unsatisfactory device? We hope that these pages will make the recent generations of economists aware of the very serious problems that surround the aggregates of output, labor, and capital when thought as generated from an aggregate production function.

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, *clich s* , 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. Furthermore, since aggregate production functions appear to give empirically reasonable results, why shouldn t they be used? Second, and following Samuelson (1961-62), aggregate

---

[1] There are other purposes for which the measurement of 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, 3-5).

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 the choice. An evaluation of these answers is provided at the end of the paper.

The rest of the paper is structured as follows. We begin in Section 2 by clarifying what an aggregate production function is. Section 3 gives a succinct summary of Joan Robinson s complaint about the units in which aggregate capital should be measured, and which led to the Cambridge-Cambridge capital controversies. Sections 4-8 are dedicated to summarizing and discussing the aggregation results.[2] These refer to the properties that micro functions and micro variables must satisfy so that aggregation into higher levels becomes feasible. It is important to keep in mind 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. Section 9 returns to the question of why, in the light of the above results, economists continue using aggregate production functions. Section 10 concludes.

2. WHAT IS 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 analysis since the times of the classical economists, although at the time it was not expressed mathematically. However, this concept is today plagued by conceptual confusions, in particular regarding the micro and macro links. 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 macro-

[2] Excellent summaries of the literature on aggregation are Green (1964), Sato (1975), Brown (1980), Diewert (1980), and Fisher (1969a, 1993), from where most of the material in this paper draws. Certainly there are other approaches to aggregation apart from those summarized here. Reasons of space prevent us from discussing them.

model. 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, 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.[3] Thus, the macro production function is a fictitious entity, in the sense that is no macroeconomic decision-maker allocating 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 clarified that at any level of aggregation, the production function *is not* a description of the relationship between the levels of output and inputs. Rather, the production function describes the maximum level of output that can be achieved if the inputs were efficiently employed.

Today this is the accepted view, and thus there is a distinction between the plant or engineering production functions, and the economics production function (even at the micro-economic level). The former is strictly a technical relationship, expressed in physical terms, while the latter includes efficiency considerations. Thus, in the standard neoclassical microeconomic model, the economic production function embeds the technical relations, and the first-order conditions embed the economic decisions of the firm.

3. JOAN ROBINSON S COMPLAINT

What was the problem Joan Robinson pointed out in the early 1950s and which led to the Cambridge-Cambridge capital controversies (recall Bernanke s quote in the Introduction)? In the pre-Keynesian tradition going back to the physiocrats and classical economists, it was very clear that capital was conceived in two fundamentally different ways: (i) capital could be conceived

[3] 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

as a fund of resources which can 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.

These distinctions, according to Joan Robinson, were lost after the Keynesian Revolution, in particular in empirical analyses. In fact, according to Joan Robinson (1971a), they were mixed. As discussed above, at some point in time economists began analyzing the performance of the economy in terms of a 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, the division of the aggregate product between labor and capital *could be* explained in terms of marginal producitvities. But in order to explain the rate of profit in these terms, capital had to be conceived as a single magnitude. This could apparently only be if capital were measured as a value quantity, unlike labor or land, which could be measured as physical quantities, and thus had their own technical units.

In a seminal paper, Joan Robinson (1953-1954) 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, each having specific technical characteristics, it is impossible to express the stock of capital goods as a homogeneous physical entity. Robinson claimed that it followed that the only things that can be aggregated are their values (but we shall see below that this is not the case). Such a value aggregate, however, is not independent of the rate of profit and thus of income distribution. The problem is as follows. Suppose there are n types of capital goods, denoted K_i^t , i = 1 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

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, 63).

combined into a single number that may be interpreted as the measure of *real capital* in the economy as a whole? (Usher 1980, 2; italics added). [4]

The next point of contention refers to what this definition of aggregate real capital measures. Usher (1980, 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.[5]

This heterogeneity of capital became the source of the controversies. 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 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-hour of work. The same goes for land (acres of land of a given quality). These are *natural units*.[6] 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. But she argued, How can this be made to correspond to a physical factor of production? (Robinson 1971a, 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, 36). [7]

All this matters because it was claimed that 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. 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 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 (but see our remark in a footnote above regarding women-hours). This implies that

---

4 Usher (1980, 1) distinguishes between real capital and the value of capital in current dollars , implicitly legitimizing Joan Robinson s question.

5 Usher (1980, 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; and 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 instantaneous productive capacity.

6 It is difficult to see what difference this makes, however. One woman-hour of Joan Robinson s time was not the same as one of Queen Elizabeth II or of Britney Spears in terms of productivity.

7 Thus, aggregate capital, viewed as 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., real capital). It is in this form that capital does not have a homogeneous unit.

in order to provide a quantity of capital one must know, first, its price in order to yield its quantity. However, the price of the aggregate factor capital is affected by the distribution of income among the factors. The value of capital changes as the profit and wage rates change so that the same physical capital can represent a different value, whereas different stocks of capital goods can have the same value (Robinson 1956). So long as the capital stock is heterogeneous, its measurement requires knowledge of the relative values of individual capital goods. This can only be achieved if the price vector of the economy and the rate of profits were known *ex ante*. The consequence is that aggregate capital, the aggregate production function, and the marginal products of the factors can only be defined when the rate of profit is given, 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 debates that lasted two decades and which went far beyond the original question. If attention is restricted to the question of aggregate capital and the aggregate production function, the answers could be grouped into two main lines. First, the solution was to search for the technical conditions under which aggregation is possible (discussed in sections 4-8). This is, for example, the work developed by Fisher, *inter alios*. The aggregation problem became the search for the conditions under which macro aggregates (not only capital) exist. The second solution was to conceive the aggregate production function in terms of a parable, following Samuelson (1961-1962) (discussed in section 9).[8] Joan Robinson certainly rejected both.

4. FIRST-GENERATION WORK ON AGGREGATION IN PRODUCTION FUNCTIONS

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

---

[8] We also mention the solution proposed by Champernowne (1953-54). This was to construct a chain-index of quantity of capital. Joan Robinson (1953-1954) 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 1953-1954, 113). The chain index compares the amounts of capital in a sequence of stationary states. Garegnani summarizes it 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, 34-35).

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 micro- and macroeconomics, and then tried to construct aggregates (usually in the form of index numbers) which were consistent with the two theories (Klein 1946a, 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.[9] As noted in section 2, 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 ν-th firm be representable as $y^{\nu} = f^{\nu}(x_1^{\nu},...,x_N^{\nu})$. Klein s aggregation problem over sectors can be phrased as follows: what conditions on the firm production functions will guarantee the existence of: (i) an aggregate production function G; (ii) input aggregator functions g_1, , g_N ; and (iii) an output aggregator function F such that the equation $F(y^1,...,y^M) = G([g_1(x_1^1,...,x_1^M), g_2(x_2^1,...,x_2^M),...,g_N(x_n^1,...,x_N^M)]$ holds for a suitable set of inputs x_N^M ?[10]

---

[9] 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 point out (quoted in section 2), that not even the micro production functions are simply technological relations but assume an optimization process by engineers and management.

[10] 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, that is, 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 in 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.

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 elasticties 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 macro-model from the micro model. In fact his macro model does not follow from the micro model. 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 were already 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 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.[11]

5. THE LEONTIEF, NATAF, AND GORMAN THEOREMS

[11] A third problem was pointed out by Walters (1963, 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 times the macro labor) is $WL = \frac{1}{n}\sum_{i=1}^{n} W_i L_i$, 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 $W = W^* \frac{\sum L_i}{n\sum_{i=1}^{n} L_i^{\alpha_i / \sum \alpha_i}}$. 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^*.

The first major result on aggregation was provided by Leontief (1947a, 1947b).[12] 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 can be written as $G[h(x_1, x_2), x_3]$ if and only if $\frac{\partial(g_1 / g_2)}{\partial x_3} \equiv 0$ 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 must be 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 $Q = Q(k_1, \ldots k_n, L)$. This function can be written as $Q = F(K, L)$, where $K = \phi(k_1, \ldots k_n)$ 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≠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

[12] 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, 10-15); or Fisher (1993, xiv-

possibilities will depend on the amount of labor. Think for example of bulldozers and trucks, or one-ton and two-ton trucks. In these cases no quantity of capital-in-general can be defined (Solow 1955-56, 103).

Solow argued that there is a class of situations where Leontief s condition may be expected to hold. 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 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, 389).[13]

The second important theorem was due to Nataf (1948). Nataf pointed out that Klein s (1946a) aggregation over sectors 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 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 ν=1 .,n. Each firm produces a single output Y(ν) using a single type of labor L(ν), and a single type of capital K(ν). Suppose that the ν—th firm has a two-factor production function $Y(\nu) = f^{\nu}\{K(\nu), L(\nu)\}$. To keep things simple, assume all outputs are physically homogeneous so that one can speak of the total output of the economy as $Y = \sum_{\nu} Y(\nu)$, and that there is only one kind of labor so that one can speak of total labor as $L = \sum_{\nu} L(\nu)$. 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 $Y = \sum_{\nu} Y(\nu) = F(K, L)$ where $K = K\{K(1),..., K(n)\}$ and $L = L\{L(1),..., L(n)\}$ are indices of aggregate capital and labor, respectively? Nataf showed (necessary and sufficient condition) that the aggregates Y, L, K exist which satisfy the aggregate production function Y=F(K, L), when the variables K(ν) and L(ν) are free to take on all values, if

---

xvi).

13 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 one in its inputs. This condition is termed strong separability .

$\psi^{\nu}\{L(\nu)\}$
and only if every firm s production function be additively separable in labor and capital, that is, every f^{ν} can be written in the form $f^{\nu}\{K(\nu), L(\nu)\} = \phi^{\nu}\{K(\nu)\} + \psi^{\nu}\{L(\nu)\}$. Assuming this to be so, the aggregate production relation can be written Y=L+K, where, $Y = \sum_{\nu} Y(\nu)$, $L = \sum_{\nu} \psi^{\nu}\{L(\nu)\}$, and $K = \sum_{\nu} \phi^{\nu}\{K(\nu)\}$. Moreover, if one insists that labor aggregation be "natural", so that $L = \sum_{\nu} L(\nu)$, then all the $\psi^{\nu}\{L(\nu)\}$ = $c\{L(\nu)\}$, where c is the same for all firms. Nataf s theorem provides an extremely restrictive condition for intersectoral aggregation.[14] Without imposing any further restriction on the problem, it makes one rather chary about the existence of an aggregate production function.[15]

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 the marginal rates of substitution between the i-th and j-th inputs be the same for all firms, that is, $\frac{\partial F^i}{\partial x_{ki}} / \frac{\partial F^i}{\partial x_{hi}} = \frac{\partial F^j}{\partial x_{kj}} / \frac{\partial F^j}{\partial x_{hj}}$, where i, j denote the firms, and k, h the inputs. Gorman showed that if this conditions holds, then a necessary and sufficient condition for the consistent aggregation of the functions $y_{\upsilon} = f_{\upsilon}(x_{1\upsilon},..., x_{m\upsilon})$ to the function $Y = F(x_1,...x_m)$ 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, 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} . x_{ms}) to the function y = F(x_1 . x_m) is possible. Furthermore, there exist functions F and h_1 h $_m$ such that $y = \sum_{n=1}^{n} h_s(y_s) = F(x_1....x_m)$ where

14 For a number of applications of this result see Green (1964, chapter 5).

15 Nataf s result can be proven using Leontief s theorem. By the latter, aggregation is possi ble if and only if the ratio of marginal products of capital in two firms independent of all labor inputs. But in Nataf s non-optimizing setup, 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 function F is homogeneous of degree one 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 one, 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 $x_{1s}/x_{rs}.....x_{ms}/x_{rs}$. 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 $x_1/x_r.....x_m/x_r$. Hence, for each r and s, x_{rs} depends only on y_s and the ratios $x_1/x_r.....x_m/x_r$.

With the above background, assume a Cobb-Douglas with three inputs $y_i = A_i L_i^{\alpha_i} K_i^{\beta_i} H_i^{\gamma_i}$ where the subscript i indexes the firms, and $\alpha+\beta+\gamma=1$ for each i . If *(i)* the expansion paths are parallel and *(ii)* the first-order conditions are satisfied, then the production functions can be written as $y_i = A_i \left(\frac{K}{L}\right)^{\beta} \left(\frac{H}{L}\right)^{\gamma}$. Now aggregate them: $y = \sum_{i=1}^{n} \frac{y_i}{A_i} = \sum_{i=1}^{n} L_i \left(\frac{K}{L}\right)^{\beta} \left(\frac{H}{L}\right)^{\gamma} = L\left(\frac{K}{L}\right)^{\beta} \left(\frac{H}{L}\right)^{\gamma} = L^{1-\beta-\gamma} K^{\beta} H^{\gamma} = L^{\alpha} K^{\beta} H^{\gamma}$.[16]

6. 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.

[16] Brown (1980, 397-398) 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.

Fisher pointed out that one must ask not for the conditions under which total output can be written as $Y = \sum_{\nu} Y(\nu) = F(K, L)$ 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, 556; italics original. This was, of course, the problem with Klein s original formulation). The reason is that 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, xviii; italics 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. 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 different from the discussions in the 1940s, and in particular different from Joan Robinson s problem: the aggregation problem is not just the aggregation of capital, at least the way Joan Robinson thought of it. There would be aggregation problems even if all capitals *were* physically homogeneous. Further, there exist equally important labor and output aggregation problems.

A by-product of these differences is the implicit acknowledgement that the aggregation process does not lead to physical quantities (Joan Robinson s problem). But this is not *the* issue. Fisher s aggregates are indeed indices, and in his view, Joan Robinson misunderstood the aggregation problem (Fisher 1993, xiii). It is here that an important difference arises in the understanding of the aggregation problem. For Joan Robinson and her followers 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, 209). For Fisher, on the other hand, capital does not present any special problem. Similar aggregation problems occur with labor and output. The whole problem reduces to finding the technical conditions under which *all* aggregates can be generated.

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 this latter sense, Joan Robinson s remarks about labor and land being different from capital are not true. They are not physically homogeneous either.[17]

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 s consider the case in which capital is physically homogeneous, so that total capital can be written as $K = \sum_{\nu} K(\nu)$. 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. This holds even if all firms have different production functions and whether or not there are constant returns.

6.1. 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 connection between the Cambridge-Cambridge capital controversies and the aggregation issues. In her seminal paper opening the debate, Joan Robinson (1953-54) asked how was aggregate and

[17] To insist upon the differences between both treatments of the aggregation issue, the problem with the symmetrical treatment of all factors of production is that, as Pasinetti points out, for conceptual reasons, the two factors are not symmetrical to each other (Pasinetti 200, 206). Labor can be expressed in physical terms (e.g., hours) to which its reward can be referred (wage per hour). The problem with capital, which of course can be expressed in physical terms too (e.g., number of machines, or an index of physical quantity), 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 its *physical* quantity of capital. But the value of capital 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. However, one can argue that wages are also commensurate with the value of labor and not its physical quantity. Further, wages, from the demand side, depend on the profitability of hiring more

heterogeneous capital 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? (Solow 1955-56, 102). And, When if ever can he 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? 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, 247).[18]

It was argued above that when labor and capital are homogeneous across firms, aggregation does not pose a special problem. 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 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 $J = J\{K(1),..., K(n)\}$? Where $K(\nu)$, ν=1 n, represents the stock of capital of each firm (i.e., one kind of capital per firm). Since the values of $L(\nu)$ 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

---

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.

[18] Usher (1980, 19) indicates that the aggregation problem (summarized in the Introduction) 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

a group capital index, the previous expression for Y^* is equivalent to $Y^* = G\{K(1),...,K(n), L\}$ if and only of the marginal rate of substitution between any pair of $K(\nu)$ 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 for under the assumption of strictly diminishing returns to labor (i.e., $f^{\nu}_{LL} < 0$), a necessary and sufficient condition for capital aggregation is that if any one firm has an additively separable production function (i.e., $f^{\nu}_{KL} \equiv 0$), then *every* firm must have such a production function.[19] 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.[20] 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 times a coefficient) turns out to be *the only case* in which a capital aggregate exists. This means that each firm s production function must be writteable as $F(b_{\nu}K_{\nu}, L_{\nu})$, where the function $F(\cdot,\cdot)$ is common to all firms, but the parameter b_{ν} 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.[21] Summing up: aggregate production functions exist if and only if all micro production functions are identical except for the capital efficiency coefficient. Certainly this conclusion represents a step ahead with respect to Nataf s answer to the problem. But certainly it continues to require an extremely restrictive aggregation condition, one that actual economies do not satisfy.

---

proportional to the derivatives $\frac{\partial h(K_1, K_2)}{K_1}$ and $\frac{\partial h(K_1, K_2)}{K_2}$. The problem is to infer the series K from K_1, K_2, P_1, and P_2.

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

20 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 $\frac{f^{\nu}_{KL}}{f^{\nu}_{K}\, f^{\nu}_{LL}} = g(f^{\nu}_{L})$, where g is the same function for all firms.

21 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 nonconstant returns production functions which do allow construction of an aggregate capital stock. On the other hand, 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.

To see this more clearly, 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.[22]

A corollary of Fisher s work is the importance of the aggregation level. On the one hand the aggregation problem appears equally with two firms and with one thousand. One 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., $Y(\nu) = f^{\nu}(K_1, K_2, L)$. Here Fisher distinguished between two different cases. First, aggregation across firms over one type of capital (e.g., plant, equipment). Fisher concluded that the construction of a sub-aggregate of capital goods requires even less reasonable conditions than for the construction of a single aggregate.[23] 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 2-factor non-constant returns case appear. Further, if the ν—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 ν 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).

---

22 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.

23 The conditions turn out to be twofold: (i) $\frac{f^{\nu}_{K_1L}}{f^{\nu}_{K_1}\, f^{\nu}_{LL}} = g(f^{\nu}_L)$; (ii) $f^{\nu}_{K_1K_2} - \frac{f^{\nu}_{K_1L}}{f^{\nu}_{K_1}\, f^{\nu}_{LL}} = 0$.

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 s (1982) paper 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, as Joan Robinson had argued, is 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. The aggregation problem seems to appear due to the fact that capital 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. 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, and assuming constant returns (if there are nonconstant returns, no aggregate will exist in general), aggregation is permitted over some group of variables if and only if: (i) such an aggregate exists at the level of each firm separately; and (ii) either the firm level aggregates are the same in both firms or the two firms production functions differ by at most an aggregate-augmenting technical change. 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. Summing up, when all factors are mobile, what is important is that aggregation across firms plays a role in the aggregation problem as much as does the fixity of capital. Aggregate production functions hardly exist whether or not capital is mobile.

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.[24]

6.2. Labor and Output Aggregation

Fisher (1968) extended his work on capital aggregation to the study of problems involved in labor and output aggregation, thus acknowledging that the aggregation problem is not restricted to capital. The problem studied here is in the context of cross-firm aggregation that arises because labors or outputs are shifted over firms, given the capital stocks and production functions, to achieve efficient production. That is, now there is a vector of labors $L_1(\nu),...,L_S(\nu)$ and a vector of outputs $Y_1(\nu),...,Y_S(\nu)$ (it does not matter whether there is one or more types of capital).[25]

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 labors 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 corollary 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, i.e., all firms must produce the same market-basket of outputs differing only in their scale.[26]

6.3. Are Approximations Valid?

It is evident that Fisher s 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 capitals and labors in the economy lie in a bounded set. And suppose further that the

[24] 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.

[25] 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.

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 argued 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.

Fisher (1969a, 572-574) posed an interesting conundrum, namely, that 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.[27] And likewise, if outputs were always produced and labor hired in approximately fixed proportions, then an approximate output and labor aggregates would exist.

7. FISHER S SIMULATIONS

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 econometrically. Likewise, the marginal product of labor appears to give a reasonable 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

[26] 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).

[27] This argument relates to the Houthakker-Sato aggregation conditions. See section 8 below.

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, 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, that is, 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, that is, 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, 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 approximately of 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 are 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. [28]

8. 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

[28] Nelson and Winter (1982) also used simulation analysis to show that, in the context of their evolutionary model, they can generate a data set such that, when an aggregate Cobb-Douglas is fitted, an almost perfect fit is 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 combinations of the input-output coefficients) by engaging

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 $Q = C\left(\frac{L}{Q}\right)^{\alpha_1 - 1}\left(\frac{K}{Q}\right)^{\alpha_2 - 1}$ with $\alpha_1 > 1$ and $\alpha_2 > 1$, then the aggregate production function is the Cobb-Douglas with decreasing returns to scale $Q = AL^{\alpha_1/(\alpha_1+\alpha_2+1)}K^{\alpha_2/(\alpha_1+\alpha_2+1)}$.[29] 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.[30] 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.[31] 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.

Sato s approach consists in splitting the aggregation problem into two sequential questions. First, suppose one has the production function $Q = Q(K_1, ..., K_n, L)$. 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 which the distributions must satisfy if they are to

---

in a search process. Furthermore, this search is undertaken only if the profit rate falls below a pre-established, acceptable, minimum.

29 A simplified explanation of this model can be found in Heathfield and Wibe (1987, 151-153). See also Sato (1975, 10-12 and 25-27).

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

generate the same F. This is the *invariance* problem. And as a corollary Sato asked whether two entirely 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. 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.[32]

9. WHY DO ECONOMISTS CONTINUE USING AGGREGATE PRODUCTION FUNCTIONS?

It must be clear that the fact that macroeconomists *use* aggregates such as investment, capital, labor, GDP, 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 the Introduction, three standard answers are the following. One, 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?[33] Neoclassical macro theory deals with macroeconomic aggregates derived by analogy with the micro concepts. The usefulness of this approach is strictly an empirical issue. Second, and following Samuelson (1961-62), aggregate production functions are seen as useful parables.

---

[31] 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.

[32] Finally, it is important to make a reference to Gorman s (1968) work, since he used an alternative method, namely, the restricted profits function. Gorman also set out to find what the technologies of the individual firms be like so that aggregates of fixed factors (different classes of fixed goods), 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. Algebraically, define $y = (y^1 .....y^R)$ as the fixed goods vector of the economy. The problem is to find a vector of aggregates $Y(y) = [Y^1(y^1),..., Y^R(y^R)]$, whose r-th component measures the quantity of the r-th factor , for the R fixed factors, such that the production possibility set S(y) can be written as $\tilde{S}[Y(y)]$. Gorman showed that S(y)= $\tilde{S}[Y(y)]$ if and only if the gross profit function for the economy as a whole can be written as the sum of the firms gross profit functions, that is, $\sum_t g_t(p, y_t) = G[p, Y(y)]$.

[33] This seems to be 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, 1259-1260).

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 s continue using them. This position is the one espoused by Ferguson (1971). This argument, however, is based on pure instrumentalism, and thus it is indefensible on methodological grounds (Blaug 1993). Furthermore, it was implicitly dispelled by Fisher (1971): 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.[34]

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 built by analogy. This is Ferguson s (1971) argument. The aggregation problem is therefore viewed as a being 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 la Marshall is, in general, inapplicable. Furthermore, the difference from the case of the consumption function is that the conditions for successful aggregation in the consumption case, while strong, do not seem so outlandish as do those in the production case. 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, chapter 5).

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 Samuelson s (1961-1962) 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 homogenous factor referred to as capital, and 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; that is, 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. 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). For the surrogate function to yield the correct product, the surrogate capital would have to coincide with the value in terms of consumption of the capital in use. The surrogate production function cannot be generally defined.

It is important to mention that by the time the debate died during the 1970s, Samuelson (1966) had conceded important points in the debate (e.g., reswitching and reverse capital deepening).[35] This, however, did not deter neoclassical macroeconomists from arguing that, although theoretically correct, the important point was that reswitching was empirically

---

[34] The irony of this argument is that, in practice, it is not true. The reality is that estimations of aggregate production functions tend to yield poor results. For example, a standard finding for the Cobb-Douglas with a linear time trend is a negative elasticity of capital. See Sylos Labini (1995) and Felipe and Adams (2001).

[35] *Reswitching* refers to the violation of the unique inverse relation between capital intensity and the rate of profit. It was shown theoretically that 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. *Reverse capital deepening* occurs when the value of capital moves in the same direction as the rate of profit. This is the case when the most profitable project is the one associated with a less capital-intensive technique. These points were conceded by Samuelson (1966).

unimportant (Stiglitz 1974). 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; and that they do not apply to the general equilibrium model conceived in disaggregated terms and based on the behavior of profit and utility maximizing agents. Here there is another apparent disagreement. Referring to the previous comment, Pasinetti asserts that This proposition actually has no objective foundation; phenomena of non-convexity, re-switchings of techniques and badly behaved production functions [ ] are not —as has been amply demonstrated — a consequence or a characteristic of any particular process of aggregation . They may occur at any time and in any context, aggregated or disaggregated (Pasinetti 2000, 212). On the other hand, Fisher (1971b), in a reply to Joan Robinson, argued that the marginal productivity theory of distribution is a microeconomic one and that is perfectly valid to consider the marginal productivity of well-defined individual capital goods.

The final argument given for the use of aggregate production functions is that there is no other option is one is to answer the questions the aggregate production function for which is used. This is a variant of the instrumentalist position and also clashes with the results of the aggregation work. Of course, if one insists on a research program whose goal is to split overall growth into the alleged contribution of technical progress and factor accumulation (i.e., growth accounting), surely one needs an aggregate production function that relates aggregate output to aggregate inputs, and thus speak of multi-factor or total factor productivity. Likewise, 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 on technical terms (i.e., optimization, marginal productivities, and capital-labor ratios), and thus the notion of aggregate production function is fundamental (Ferguson 1968). To think of the distribution of output in terms of, for example, class conflict, is unthinkable for many economists. See the recent work of Pasinetti (2000).[36]

But if one realizes that 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).[37] The same occurs with the recent work by Mankiw et al. (1992), where the authors pooled time-series and cross-section data for over 100 countries. There is no single reason to believe that what they estimated was an aggregate production function.

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

This paper has provided 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 (even before the works of Fisher and Sato), 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, 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, 427-428).[38]

This is a summary of the main conclusions and lessons:

(i) We have discussed in passing 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 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. 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

---

[36] Although it is important to stress that the marginal productivity theory of distribution is purely microeconomic. If the aggregate production function does not exist, certainly the marginal productivities of *aggregate* labor and capital do not exist. See the exchange between Robinson (1971b) and Fisher (1971b).

[37] 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, xiii).

the problem, Nataf 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 like in Fisher, the aggregation conditions remain extremely restrictive. The existence of a labor aggregate requires that all firms to employ different labors 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 to 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. And in the Houthakker-Sato approach, the possibility of aggregation depends on the stability of distribution of the input-output coefficients.

(ii) Economists act, however, as if these constructs were 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 Y=C+I+G identity. Likewise, there is nothing wrong (from the point of view discussed in this paper) with generating a stock of capital through the perpetual inventory method. However, the *relationship* Y=F(K,L) between aggregate output (Y) and aggregate inputs (K, L) used in theoretical and applied macroeconomic work does not have, in general, a meaningful interpretation. And thinking of aggregate investment as a well-defined addition to capital in production is a mistake. 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 the Introduction.[39] However, as Fisher indicates this has not discouraged macroeconomists from continuing to work in such terms (Fisher 1987, 55). This attitude is prevalent in all areas of macroeconomics, but even more accused in growth theory.

(iii) 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 using them are fallacious and thus unacceptable, e.g., that they work empirically; or that in order

[38] Recent works on unemployment where the aggregate production function plays a key role are Rowthorn (1999) and Blanchard and Wolfers (2000).

[39] In one of the very few cases where authors recently have bothered to mention the possible problems derived from aggregation for applied work, Basu and Fernald (1997), nevertheless, argued: [ ] 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, 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

to perform growth accounting one needs to assume their existence. Mermaids do not exist simply because one insists on studying them!

(iv) The aggregation problem appears equally with two firms and with one 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.

(v) Intuitions based on micro variables and micro production functions will often be false when applied to models with macro production functions. For example, what is the meaning of multi-factor productivity?

(vi) The revival of growth theory during the last two decades no doubt has produced important discussions, and seemingly interesting empirical results. However, authors do not realize that they are using a tool whose invalidity 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).[40]

(vii) At the empirical level, and contrary to widespread belief, 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 (2001), who subjected the original Cobb-Douglas (1928) data set to a series of stability tests. The results indicate that the regression is very fragile. Furthermore, adding a linear time trend to the regression to account for technical progress yields very poor and questionable results. With today s econometric tools, nobody would conclude that this data set indicates that the elasticity of labor is 0.75 and that of capital 0.25. 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 construct, or estimate the degree of returns to scale and the elasticity of substitution.[41]

---

production function. Certainly it is true that if all factors were optimally allocated the aggregation problem would disappear. But this result is not due to Fisher.

[40] 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 . What are the logical foundations, or conditions under which this can be represented?

[41] For example, Solow claimed: When someone claims that aggregate production functions work, he means (a) that they give a god fit to input-output data without the intervention of factor shares and (b) that

(viii) If the notion of aggregate production function is so problematic, it follows that we lose the sense of what it is supposed to measure. Then, is there any alternative interpretation of the estimates of aggregate production functions that does not presuppose the existence of an aggregate technology? (Blackorby and Schworm 1984, 647). This question has been answered in the positive by Felipe (2000, 2001a, 2001b), Felipe and Holz (2001), and Felipe and McCombie (2001a, 2001b, 2002a, 2002b). The answer, however, is most discouraging. They show that the ex-post income accounting identity that relates output to inputs, i.e., VA=wL+rK, can be easily rewritten through a simple algebraic transformation as VA=A(t) F(K,L).[42] The precise functional form (Cobb-Douglas, CES, translog, etc.) corresponding to the data set in question will depend on the path 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 particular form VA=A(t) F(K,L) corresponding to the data set at hand has to yield a perfect fit (because all that is being estimated is an identity); the putative elasticities have to coincide with the factor shares; and the marginal products have to coincide with the factor prices This is true for any data set. 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. The corollary of this argument is that the problem with the empirical work undertaken during the last seventy years is that the production functions estimated were not the correct ones corresponding to the particular data sets used. For example, most often applied economists have fitted the Cobb-Douglas form including a time trend. This, in most cases, has yielded very poor results. The solution is to fit the Cobb-Douglas but including a complex time trend (i.e., a trigonometric function) instead of the linear time trend. This will provide, in most cases, an almost perfect approximation to the accounting identity with the results mentioned above (see the examples in the papers cited above). The arguments and examples in these papers 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.[43]

---

the function so fitted has partial derivatives that closely mimic observed factor shares (Solow 1974, 121). This means, implicitly, that the aggregate production function can be tested.

[42] VA is value added; w is the wage rate; L is employment; r is the ex-post profit rate; K is the value of the stock of capital; and A(t) is a function of time. Note that r is not the user cost of capital, but the profit rate that makes the accounting identity hold always. The income accounting identity VA=wL+rK does not follow from Euler s theorem, and thus there is no reason why the wage and profit rates have to coincide with the respective marginal productivities. Furthermore, recall that, in general, aggregate production functions do not exist.

[43] 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 from their alleged behavioral interpretation.

(ix) 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. It may nevertheless 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. If, as Temple (1999, cited in the Introduction) 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.[44] See, for example, the work of the Cambridge (UK) economists such as Kalecki, Kaldor, Keynes and Sraffa (Pasinetti 1974, 1994), ignored by mainstream neoclassical macroeconomics during the last few decades, or that of Scott (1989).[45]

[44] It is very difficult to publish in prestigious journals work that is not formulated using the standard language and tools, in the case of growth, an aggregate production function (Lavoie 1992, 14-19).

[45] Certainly no model is free from criticism. We mention these as alternatives that do not use aggregate production functions.

REFERENCES

Arrow, K., H.B. Chenery, B.S. Minhas, and R.M. Solow. 1961. "Capital-Labor Substitution and Economic Efficiency." Review of Economics and Statistics, Vol. XLIII, pp. 225-250.

Basu, Susanto, and J.G. Fernald. 1997. Returns to Scale in U.S. Production: Estimates and Implications. Journal of Political Economy, Vol.105, No.2, pp. 249-283.

Bernanke. B. 1987. Comment. NBER Macroeconomics Annual 1987, pp. 202-205, The MIT Press, Cambridge, MA.

Blackorby, Charles, and William Schworm. 1982. Aggregate Investment and Consistent Intertemporal Technologies. Review of Economic Studies , XLIX: 595-614.

________. 1984. The Structure of Economies with Aggregate Measures of Capital: A Complete Characterization. Review of Economic Studies, LI: 633-50.

________. 1988. The Existence of Input and Output Aggregates in Aggregate Production Functions. Econometrica, Vol.56, No.3 (May): 613-43.

Blanchard, Olivier, and Justin Wolfers. 2000. The Role of Shocks and Institutions in the Rise of European Unemployment: The Aggregate Evidence. The Economic Journal, 110 (March): C1-C33.

Blaug, Mark. 1993. *The Methodology of Economics. Or How Economists Explain*. Cambridge University Press. Second edition (reprinted).

Bliss, C.J. 1975. *Capital Theory and the Distribution of Income*. Amsterdam: North-Holland.

Brown, Murray. 1980. The Measurement of Capital Aggregates: A Post- reswitching Problem. In Dan Usher (ed.), *The Measurement of Capital*, pp. 377-420. Chicago and London: The University of Chicago Press.

________. 1987. Cobb-Douglas Functions. In Eatwell, Millgate and Newman (eds.), *The New Palgrave. A Dictionary of Economics*, pp.460-1. Basingstoke: Macmillan.

Burmeister, Edwin. 1980. Comment. In Dan Usher (ed.), *The Measurement of Capital*, pp. 420-31. Chicago and London: The University of Chicago Press.

________. 1990. Wicksell Effects. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.257-261. Macmillan.

Champernowne, D.G. 1953-1954. The Production Function and the Theory of Capital: A Comment. Review of Economic Studies, 21,2: 112-135.

Cobb, C.W., and P.H. Douglas. 1928. A Theory of Production. American Economic Review , Vol.18, Supplement (March), Papers and *Proceedings of the Fortieth Annual Meeting of the American Economic Association*, December 1927: 139-165.

Diewert, Erwin W. 1978. Hicks Agrgegation Theorem and the Existence of a Real Value-Added Function. In Melvyn Fuss and Daniel McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, pp.17-51. North Holland.

________. 1980. Aggregation Problems in the Measurement of Capital. In Dan Usher (ed.), *The Measurement of Capital*, pp. 433-528. Chicago and London: The University of Chicago Press.

Eatwell, John. 1990. Walras s Theory of Capital. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.247-256. Macmillan.

Felipe, Jesus. 2000. On the Myth and Mystery of Singapore s Zero TFP . Asian Economic Journal, Vol.14, No.2 (June): 187-209.

Felipe, Jesus. 2001a. Aggregate Production Functions and the Measurement of Infrastructure Productivity: A Reassessment. Eastern Economic Journal, Vol.27, No.3 (Summer). Forthcoming.

Felipe, Jesus. 2001b. Endogenous Growth, Increasing Returns, and Externalities: An Alternative Interpretation of the Evidence. Metroeconomica, Vol.52, No.4 (November). Forthcoming.

Felipe, Jesus, and F. Gerard Adams. 2001. A THEORY OF PRODUCTION . The Estimation of the Cobb-Douglas Function: A Retrospective View. Georgia Institute of Technology and Northeastern University. Manuscript.

Felipe, Jesus, and Carsten Holz. 2001. Why Do Aggregate Production Functions Work? Fisher s Simulations, Shaikh s Identity, and Some New Results. 200. International Review of Applied Economics, Vol.15, No.3 (July): 261-285.

Felipe, Jesus, and JSL McCombie. 2001a. The CES Production Function, the Accounting Identity, and Occam's Razor. Applied Economics, Vol.33, No.10 (August): 1221-1232.

Felipe, Jesus, and JSL McCombie. 2001b. How Sound are the Micro-Foundations of the Aggregate Production Function? Georgia Institute of Technology and Cambridge University. Manuscript.

Felipe, Jesus, and JSL McCombie. 2002a. A Problem with Some Recent Estimations and Interpretations of the Markup in Manufacturing Industry. International Review of Applied Economics. Forthcoming.

Felipe, Jesus, and JSL McCombie. 2002b. Methodological Problems with the Neoclassical Analysis of the East Asian Miracle. Cambridge Journal of Economics. Forthcoming.

Ferguson, Charles E. 1968. Neoclassical Theory of Technical Progress and Relative Factor Shares. Southern Economic Journal, 34, 4: 490-504.

________. 1971. Capital Theory up to Date: A Comment on Mrs. Robinson s Article. Canadian Journal of Economics, Vol.IV, No.2 (May): 250-254.

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

Fisher, F.M. 1968. Embodied Technology and the Existence of Labor and Output Aggregates. Review of Economic Studies, Vol.35: 391-412.

________. 1969a. The Existence of Aggregate Production Functions. Econometrica, Vol.37, No.4 (October): 553-77.

________. 1969b. Approximate Aggregation and the Leontief Conditions. Econometrica, Vol.37, No.4 (October): 457-69.

________. 1971a. "Aggregate Production Functions and the Explanation of Wages: A Simulation Experiment." The Review of Economics and Statistics, Vol.LIII, No.4 (November): 305-25.

________. 1971b. Reply. Econometrica, Vol.39, No.2 (March): 405.

________. 1982. Aggregate Production Functions Revisited: The Mobility of Capital and the Rigidity of Thought. Review of Economic Studies, 49, 4 (October): 615-626.

________. 1983. On the Simultaneous Existence of Full and Partial Capital Aggregates. Review of Economic Studies, 50, 1 (January): 197-208.

________. 1987. Aggregation Problem. In Eatwell, Millgate and Newman (eds.), *The New Palgrave. A Dictionary of Economics*, pp.53-5. Basingstoke: Macmillan.

________. 1993. *Aggregation. Aggregate Production Functions and Related Topics*. Cambridge, MA: The MIT Press.

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

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

________. 1990. Quantity of Capital. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.1-78. Macmillan.

Gorman, W.M. 1953. Community Preference Fields. Econometrica, 21: 63-80.

________. 1954. Klein Aggregates and Conventional Index Numbers. Econometrica, 22: 113-115.

________. 1959. Separable Utility and Aggregation. Econometrica, 27: 469-481.

_________. 1968. Measuring the Quantities of Fixed Factors. In J.N. Wolfe (ed.), *Value, Capital, and Growth. Papers in Honour of Sir John Hicks*, pp.141-172. Edinburgh University Press.

Green, H.A. John. 1964. *Aggregation in Economic Analysis. An Introductory Survey*. Princeton: Princeton University Press.

Hagemann, Harald. 1990. Capital Goods. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.123-127. Macmillan.

Harcourt, G.C. 1969. Some Cambridge Controversies in the Theory of Capital. Journal of Economic Literature, 7, 2: 369-405.

Heathfield, David F., and S ren Wibe. 1987. *An Introduction to Cost and Production Functions*. Humanities Press International, Inc. NJ: Atlantic Highlnads.

Hennings, K.H. 1990. Capital as a Factor of Production. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.108-122. Macmillan.

Hicks, J. 1946. *Value and Capital*. London: Oxford University Press.

Houthakker, H.S. 1955-1956. The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis. Review of Economic Studies, 23: 27-31.

Hulten, Charles R. 2000. Total Factor Productivity: A Short Biography. NBER Working Paper No.7471 (January).

Johansen, L. 1972. *Production Functions*. Amsterdam: North Holland.

Kaldor, N. 1957. A Model of Economic Growth. *The Economic Journal*, 67, 268. December: 591-624.

Klein, Lawrence R. 1946a. Macroeconomics and the Theory of Rational Behavior. Econometrica, Vol.14, No.2 (April): 93-108.

________. 1946b. Remarks on the Theory of Aggregation Econometrica, Vol.14, No.4 (October): 303-12.

Lavoie, Marc. 1992. *Foundations of Post-Keynesian Economic Analysis*. Edward Elgar.

________. 2000. Le Ch mage D quilibre: R alit ou Artefact Statistique? Revue Economique, 51, 6 (November): 1477-1484.

Layard, R., S.J. Nickell, and R.Jackman. 1991. *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.

Leontief, W.W. 1947a. Introduction to a Theory of the Internal Structure of Functional Relationships. Econometrica, 15, No.4 (October): 361-73.

________. 1947b. A Note on the Interrelationship of Subsets of Independent Variables of a Continuous Function with Continuous First Derivatives. Bulletin of the American Mathematical Society, 53: 343-50.

Levhari, David. 1968. A Note on Houthakker s Aggregate Production Function in a Multifirm Industry. Econometrica, Vol.36, No.1 (January): 151-4.

Mankiw, G., D. Romer, and D.N. Weil. 1992: A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics, 107, 2 (May): 407-437.

May, Kenneth. 1946. The Aggregation Problem for a One-Industry Model. Econometrica, Vol.14, No.4 (October): 285-98.

________. 1947. Technological Change and Aggregation. Econometrica, Vol.15, No.1 (January): 51-63.

Mankiw, G., D. Romer, D., and D.N. Weil. 1992: A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics, 107, 2 (May): 407-37.

McCombie, JSL. 1998. Are There Laws of Production? : An assessment of the early Criticisms of the Cobb-Douglas production Function. Review of Political Economy, Vol.10: 141-173.

Nataf, Andr . 1948. Sur la Possibilit de Construction de Certains Macromod les. Econometrica, Vol.16, No.3 (July): 232-44.

Nelson, Richard. 1973. Recent Exercises in Growth Accounting: New Understanding or Dead End? American Economic Review, 73, 2 (June): 462-468.

________. 1981. Research on Productivity Growth and Productivity Differences: Dead Ends and New Departures. Journal of Economic Literature . Vol. XIX (September): 1029-1064.

Nelson, Richard, and S.G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press.

Pasinetti, Luigi L. 1959. On Concepts and Measures of Changes in Productivity. The Review of Economics and Statistics, Vol.41: 270-286.

________. 1974. *Growth and Income Distribution. Essays in Economic Theory*. Cambridge University Press.

________. 1994. The Structure of Long-Term Development: Concluding Comments. In L.L. Pasinetti and R.M. Solow (eds.), *Economic Growth and the Structure of Long-Term Development*, pp. 353-362. St. Martin s Press.

________. 2000. Critique of the Neoclassical Theory of Growth and Distribution. Moneta e Credito. Banca Nazionale del Lavoro Quarterly Review,No.210, June: 187-232.

Pasinetti, L., and R. Scazzieri. 1990. Capital Theory: Paradoxes. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.136-147. Macmillan.

Pivetti, Massimo. 1990. Wicksell s Theory of Capital. In John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave. Capital Theory*, pp.262-269. Macmillan.

Pu, Shou Shan. 1946. A Note on Macroeconomics. Econometrica, Vol.14, No.4 (October): 299-302.

Robinson, Joan. 1953-54. The Production Function and the Theory of Capital. Review of Economic Studies, Vol.21, No.2: 81-106.

________. 1955-56. The Production Function and the Theory of Capital A Reply. Review of Economic Studies, 23 (2): 247.

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

________. 1970. Capital Theory Up to Date. Canadian Journal of Economics, Vol.III, No.2 (May): 309-17.

________. 1971a. The Measure of Capital: The End of the Controversy. The Economic Journal, Vol.81, No.323 (September): 597-602.

________. 1971b. The Existence of Aggregate Production Functions: Comment. Econometrica, Vol.39, No.2 (March): 405.

________. 1975. The Unimportance of Reswitching. Quarterly Journal of Economics, 89, 1 (February): 32-39.

Romer, P. 1987. Crazy Explanations for the Productivity Slowdown. NBER Macroeconomics Annual 1987, pp. 163-202. The MIT Press, Cambridge, MA.

Rowthorn, Robert. 1999. Unemployment, Wage Bargaining and Capital-Labor Substitution. Cambridge Journal of Economics, 23: 413-425.

Samuelson, Paul. 1961-1962. Parable and Realism in Capital Theory: The Surrogate Production Function. *Review of Economic Studies* , June 1962: 193-206.

________. 1966. A summing Up. Quarterly Journal of Economics, Vol.80, No.4: 568-583.

________. 1979. Paul Douglas s Measurement of Production Functions and Marginal Productivities. Journal of Political Economy, Vol.87: 923-39.

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

Scott, Maurice F. 1989. *A New View of Economic Growth*. Oxford: Clarendon Press.

Solow, Robert. 1955-56. The Production Function and the Theory of Capital. Review of Economic Studies, 23 (2): 101-8.

________. 1956. A Contribution to the Theory of Economic Growth. Quarterly Journal of Economics, Vol.70: 65-94.

________. 1957. Technical Change and the Aggregate Production Function. Review of Economics and Statistics, Vol.39: 312-20.

________. 1966. Review of *Capital and Growth* . American Economic Review, Vol.56, No.5 (December): 1257-60.

________. 1974. Law on Production and Laws of Algebra: The Humbug Production Function: A Comment. The Review of Economics and Statistics, Vol. LVI, No. 61 (February): 121.

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, Vol.82.

Sylos Labini, Paolo. 1995. Why the Interpretation of the Cobb-Douglas Production Function Must be Radically Changed. Structural Change and Economic Dynamics, 6: 485-504.

Temple, Jonathan. 1998. Robustness Tests of the Augmented Solow Model. Journal of Applied Econometrics, 13: 361-75.

________. 1999. The New Growth Evidence. Journal of Economic Literature , Vol.XXXVII (March): 112-156.

Theil, H. 1954. *Linear Aggregation of Economic Relations*. Amsterdam.

Usher, Dan. 1980. Introduction. In Dan Usher (ed.), *The Measurement of Capital*, pp. 1-21. Chicago and London: The University of Chicago Press.

Walters, A.A.. 1963. Production and Cost Functions. Econometrica, Vol.31, Number 1-2 (January-April): 1-66.