# 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**

luglio 23, 2001

# AGGREGATION IN PRODUCTION FUNCTIONS: WHAT APPLIED ECONOMISTS SHOULD KNOW

**Abstract:** This paper surveys the theoretical literature on aggregation of production functions (e.g., Klein, Leontief, Nataf, Gorman, Fisher, Sato, etc.) from the point of view of the applied economist. The most important conclusions of this literature are summarized, and the problems that derive from incorrect aggregation that economists should be aware of are discussed. The most important result 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. The paper then evaluates the standard reasons given for the use of 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 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 growth models is the aggregate production function, a relationship that intends to describe the technological relationships at the aggregate level. The subject of this survey is whether this is true or not, and the implications for applied work. This is the so-called aggregation problem. The question is how, starting from some microeconomic relationship, say $u = f(x, y)$, where u, x, and y refer to the micro units, a corresponding aggregate relation $U = F(X, Y)$ can be established, obtained as macro=$\Sigma$(micro). The macro representation will be possible only if the production function f is of a specific type, or if the variables U, X, Y are constructed in a specific fashion. For example, 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). Bernanke's comments acknowledge that there is something odd with the standard macro aggregates, despite the fact that economists (in particular the new generation) use them as if no controversy about it had ever existed.

On the other hand, and to compound the problem, 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 the aggregation problem, that is, how economic quantities are measured, and in particular those quantities which 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? 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. 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?

The standard justifications for the use of aggregate production functions 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? Second, and following Samuelson (1961-62), aggregate production functions are seen as parables. Third, for the applications where aggregate production functions are used, there is no the choice. Thus, 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 the applied economist's viewpoint the most important question in this context 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., elasticities, elasticity of substitution) the technological parameters? An evaluation of these answers will be provided at the end of the paper.

The purpose of this survey is to review the theory of aggregation in the context of production functions, and it is done from the point of view of the applied economist. What important results on aggregation should applied economists be aware of and why? The picture that emerges is that it is not all that clear that the aggregate production function does indeed provide answers to the above questions.

The rest of the paper is structured as follows. Section 2 poses a non-trivial question: what is an aggregate production function? This question appeared in the early writings on aggregation

discussed in the next sections. Is the production function strictly a technical relationship like the micro production function? Section 3 gives a succinct historical account of how the notion of capital has evolved in time. This section shows that the term 'capital' has, in most instances, been an idea with unclear and not well-defined meaning(s). None of the issues discussed in these two sections deserves a fair number of pages (if any at all) in today's standard graduate and undergraduate macroeconomics textbooks, or in those specialized in growth, one of the areas where the aggregation problem would seem to be of utmost importance.

Sections 4-8 are dedicated to the following question: how can one proceed, if at all, to construct useful production functions for a sector as a whole, or for the entire economy? This question is entirely a technical issue, in the sense that it asks for the mathematical properties that the micro functions and micro variables must satisfy so that aggregation into higher levels becomes feasible, i.e., under what circumstances can the technical relationships of a diverse economy be appropriately subsumed in such an aggregate form? It is important to keep in mind that the aggregate production function is the result of two types of aggregation. One at the level of the multiple inputs (i.e., different types of labors and different types of capital into one labor and one capital); and aggregation at the level of the functions of all firms. To motivate the question, simply think of the following problem. Suppose the technology of two firms is Cobb-Douglas. Can they 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 4 summarizes the first-generation work on aggregation in production, represented by the debate between Klein (1946a, 1946b), Pu (1946) and May (1946, 1947). Section 5 summarizes the important theorems by Leontief, Nataf, and Gorman. Then the paper moves to the second-generation work on aggregation. The main difference between the two generations is that the work of the second one is set clearly in the context of the Cambridge capital controversies, and was motivated by the question of how to aggregate capital. Section 6 discusses the work of Fisher, and section 7 his seminal simulation work. Section 8 discusses the Houthakker-Sato work. The conclusion of this literature is rather discomforting: aggregate production functions do not have a sound theoretical foundation. Section 9 returns to the question of why, in the light of the above results, economists continue using aggregate production functions. Section 10 concludes.