

THE EXISTENCE OF AGGREGATE PRODUCTION FUNCTIONS¹

BY FRANKLIN M. FISHER

This paper discusses recent work on the existence of aggregate production functions in models in which capital goods are specific to firms and cannot be used interchangeably. It is found that this raises problems not only for capital aggregation but also for the existence of labor and output aggregates. Recent work on the question of using aggregate production functions as approximations is also discussed.

1. INTRODUCTION

AS HAVE ALL MY predecessors, I have chosen as my topic for the Irving Fisher Lecture a subject of considerable importance and interest (although my judgment in this regard is no doubt biased in favor of a subject on which I have been working). Like many, but not all of my predecessors, I have taken the invitation to deliver a lecture rather than a paper as providing an appropriate occasion for exposition and summary rather than for detailed analysis and proof.

The estimation and use of aggregate production functions has become a widespread and important practice in economic analysis. Broadly speaking, such uses fall into two main classes. On the one hand, following Solow's seminal article [22], we have had a spurt of interest in the estimation of aggregate production functions for entire economies, for manufacturing, for durables, or for more narrowly defined industry aggregates. These papers have made empirically derived inferences about the importance of technical change, embodied or disembodied; about the rate of return to investment, social or private; about the share of wages in national product; and, generally, about the technical and economic forces making for growth.

On the other hand, different in focus but not entirely divorced in development from such empirically oriented studies, is the use of aggregate production functions by a large and growing number of authors to represent the technical possibilities of an economy in which some intertemporal welfare function is to be maximized, or as a basic constituent of a one or two-sector descriptive growth model.

¹ Presented as the Irving Fisher Lecture at the meeting of the Econometric Society in Amsterdam, September, 1968. I am indebted to Robert M. Solow for comments and criticism. The work was supported by National Science Foundation Grant GS-1791.

This series of lectures stands in permanent memorial to Irving Fisher and Henry Schultz on one side of the Atlantic and to Leon Walras and Arthur Bowley on the other. In view of my topic, however, it seems fitting that I should dedicate this lecture to the memory of André Nataf.

In both sorts of problems, the use of an aggregate relationship between an output aggregate on the one hand and constructs called "labor" and "capital" on the other, is an immense convenience. The question naturally arises, however, whether it is merely that. Under what circumstances can the technical relationships of a diverse economy be appropriately subsumed in such an aggregate form? What can one say about inferences drawn from such models? Is it enough for such use that we be content with approximations?

These questions are not new ones, although the particular (and, I believe, crucial) features of the model within which I shall examine them have not been employed to answer them until quite recently. Moreover, they arise at different levels of aggregation. While I shall speak generally of an aggregate production function defined over an entire economy in which production is actually carried on by individual firms, it is clear that essentially the same problems arise at the industry level. Indeed, they also arise at the firm level with production actually carried on in individual establishments or, more fundamentally, by individual workers using individual kinds of capital. The principal difference between such cases is often merely in how closely conditions for aggregation are likely to be satisfied in practice.

Now, it is important to recognize that there is a difference between the question of the existence of an aggregate production function and such related questions as the realism of models which assume smooth substitutability between capital and labor. While the fact that different kinds of machines are different plays an important role in the discussion of both questions, the aggregation problem arises whether or not technology involves fixed coefficients, and the answers typically do not turn on this question.² The difference may, perhaps, be pointed up by remarking on two items. First, all the aggregation problems which I shall discuss would arise in principle even if all capital were physically homogeneous. While in many cases, such physical homogeneity would guarantee the existence of an aggregate, this is not invariably true. Second, while the different characters of different kinds of capital play a crucial role, capital aggregation is not the only problem which must be faced. Labor aggregation and output aggregation also turn out to require quite stringent conditions which are unrealistic at the economy-wide level.

Accordingly, I shall take all my production functions to be twice continuously differentiable. It turns out, so far as is known, to make no essential difference,³ and it greatly simplifies the exposition.

2. AGGREGATION AND EFFICIENCY: THE SIMPLEST CASE

I begin by considering the simplest possible case. There are n firms. The v th firm produces a single output $Y(v)$ by utilizing a single kind of labor $L(v)$ and a

² For a "putty-clay" technology such as analyzed by Johansen [10], the aggregation problem essentially involves the *ex post* technology. Although, naturally, an *ex post* aggregate need not correspond to an *ex ante* one, this is not in itself a problem of aggregation since the *ex ante* and *ex post* technologies generally differ at any level of aggregation.

³ See Stigum [25 and 26].

single kind of capital $K(v)$. The production function for the v th firm is

$$(2.1) \quad Y(v) = f^v(K(v), L(v)).$$

To simplify matters, all the outputs are physically indistinguishable, so that it makes sense to speak of total output from the economy Y as simply the sum of individual outputs $Y(v)$. Similarly, there is only one kind of labor, so that it makes sense to speak of total labor L (more accurately, total employment) as the sum of labor employed by the individual firms. More complicated cases will be discussed later.

Capital, on the other hand, may differ from firm to firm, although it need not do so. There is hence no immediate physical sense in which total capital can be said to exist. Further, since different kinds of capital will in general have different technical properties, each firm's production function will in general be different from that of any other firm; this is represented by the superscript v on the production functions.⁴

We are interested in the conditions under which it will be possible to write total output Y as being given by an aggregate production function:

$$(2.2) \quad Y \equiv \sum_v Y(v) = F(J, P),$$

where

$$(2.3) \quad J = J(K(1), \dots, K(n)) \quad \text{and} \quad P = P(L(1), \dots, L(n))$$

are indices of aggregate capital and aggregate labor, respectively. Indeed, it is natural, given the assumed physical homogeneity of labor, to require

$$P = L = \sum_v L(v).$$

Now, it has long been recognized that if nothing more than this is said, the conditions under which such aggregation can be performed are extremely restrictive.⁵ It was shown by Nataf [18] that, in these circumstances, the aggregate production function (2.2) will exist if and only if every firm's production function is additively separable in capital and labor, that is, if and only if every f^v can be written in the form

$$(2.4) \quad f^v(K(v), L(v)) = \phi^v(K(v)) + \psi^v(L(v)) \quad (v = 1, \dots, n).⁶$$

⁴ Clearly, the model can also be interpreted as one of embodied technical change, with the $K(v)$ representing capital goods of different vintage. While the results apply directly to such models, and while work on this problem was greatly stimulated by Solow's introduction of the embodied model [23], the importance of the aggregation problem is best brought out by the interpretation in the text in which technical change does not enter.

⁵ This matter was extensively discussed in a debate in *Econometrica* in the late 1940's. See Klein [11 and 12], May [16 and 17], Nataf [18], and Pu [20]. A summary is given in Green [7].

⁶ The sufficiency of this condition is obvious. Its necessity can be easily proved among other ways by use of Leontief's well known theorem on separable functions which states that a twice differentiable function of three variables $g(x_1, x_2, x_3)$ can be written as $G(h(x_1, x_2), x_3)$ if and only if $\partial(g_1/g_2)/\partial x_3 \equiv 0$ (where the subscripts denote differentiation). In other words, the marginal rate of substitution between x_1 and x_2 must be independent of x_3 . See Leontief [14 and 15]. While all or nearly all of the results here discussed can be proved in other ways, I have tended to rely on the Leontief conditions.

If we further impose the condition that $P = L$, then we have the even more restrictive condition that

$$(2.5) \quad \psi^v(L(v)) = cL(v) \quad (v = 1, \dots, n)$$

where c is a constant and is the same for all firms.

If this were all there were to it, the hope that an aggregate production function existed would be forlorn indeed. It is pretty clear, however, that this cannot be all.

Nothing in the above discussion prevented capital from being physically homogeneous, although this was not required. Further, nothing prevented each firm's production function from exhibiting constant returns to scale, although this was not assumed. The result just given holds therefore, even if all firms have the same technology, the same kind of capital, and constant returns. In this case, however, we should certainly expect an aggregate production function to exist. If "constant returns" means anything, it surely means that there is no difference between one big firm and two little ones—that total output does not depend on the way in which production is divided among the producing units. Hence something must be wrong.

Nevertheless, there is no mistake in the results presented. Mere identity of technologies and constant returns does *not* imply the existence of an aggregate production function as may be seen by trying to add up two identical Cobb-Douglas production functions without further restrictions. The catch is not in the way in which the problem has been analyzed but in the fact that the wrong problem has been posed.

At any level of aggregation, a production function is not a description of what output will be achieved for given levels of different inputs; that output is not unique. Rather a production function describes the maximum level of output that can be so achieved if the inputs are efficiently employed. The true content of the statement just made about constant returns, for example, is that there is no difference between one big firm and two little ones *if both are producing efficiently*. Accordingly, we must ask not for the conditions under which total output can be written in the form (2.2) no matter what, 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*.

That such considerations can make a major difference can be seen by continuing for a moment in the case in which all capital is physically homogeneous and mobile between firms. Then efficient production requires that Y be maximized given L and $K \equiv \sum_v K(v)$. Calling the value of Y when so maximized Y^{**} , it is obvious that

$$(2.6) \quad Y^{**} = F(K, L)$$

since the individual allocations of labor and capital to firms will be determined in the course of the maximization problem.⁷ This holds even if all firms have different production functions, f^v (because, for example, each firm has knowledge of a

⁷ This was pointed out by May [16 and 17].

different secret process). Moreover, it holds whether or not there are constant returns to scale.

It has been argued, however,⁸ that to allow factors to be assigned to firms so as to achieve maximum total output is to introduce institutional considerations into what ought to be a purely technological affair. In a sense this is true. If there are institutional barriers to factor mobility, then one might properly say that efficiency requires maximization within such barriers, so that the happy result of (2.6) only holds for those institutional arrangements which permit full mobility. Such arrangements are compatible with rather different social frameworks, however, since the same organization of production will be achieved in this problem by a perfectly competitive economy, a centrally planned socialist economy, and a monopolist organizing the plants under his control. Moreover, it is hard to know where to draw the line. We do not refrain from using a production function for an individual firm because it is open to that firm to behave inefficiently; yet the ability of the firm to produce efficiently also rests on institutional factors. This aspect of the problem thus seems to me to be largely one of agreeing on what we are to mean by an aggregate production function. I would prefer to speak of output produced efficiently subject only to technological constraints as given by a production function and to regard institutional barriers to efficiency as forces causing the production system in question to lie inside the efficient frontier. I recognize, however, that in dealing with actual economies, the barriers may be more important than the frontier and that it is often quite difficult to distinguish between technological and institutional constraints. (Is a slow dissemination of technical knowledge an institutional constraint in this context?)⁹

It is clear, however, that whatever one decides about the case in which all capital is homogeneous and mobile, the far more realistic case in which capital goods are different and not interchangeable provides a constraint which should be treated as technological. Accordingly, we shall again assume technology to be embodied in the capital goods and shall assume that only labor, but not capital, can be allocated to firms so as to maximize total output. This sort of distinction between fixed and movable factors (which does not always coincide precisely with conventional distinctions between labor and capital) will be maintained when we move to more complicated models.

3. CAPITAL AGGREGATION

As now constituted, the problem is the following. Given that Y is maximized with respect to the allocation of labor to firms, and denoting the resulting value of Y by Y^* , under what circumstances is it possible to write

$$(3.1) \quad Y^* = F(J, L), \quad J = J(K(1), \dots, K(n)).$$

⁸ See Klein [11 and 12], for example.

⁹ Whitaker [28] discusses capital aggregation in the context of immobile capitals but mobile labor with labor allocated to firms by arbitrary rules.

It is evident that in any case Y^* can be written as

$$(3.2) \quad Y^* = G(K(1), \dots, K(n), L),$$

since the values of the $L(v)$ will be determined in the course of the maximizing procedure, so for this simple model there is no labor aggregation problem and the entire problem is that of the existence of a capital aggregate.¹⁰

By a well known theorem of Leontief, already referred to, (3.1) and (3.2) are equivalent if and only if the marginal rate of substitution between any pair of the $K(v)$ in the production of Y^* is independent of L . The problem is to see what this implies about the original firm production functions, the f^v . If we assume strictly diminishing returns to labor, so that $f_{LL}^v < 0$ ($v = 1, \dots, n$), where the subscripts denote differentiation, then it can be shown¹¹ that a necessary and sufficient condition for capital aggregation is that every firm's production function satisfy a partial differential equation in the form

$$(3.3) \quad \frac{f_{KL}^v}{f_K^v f_{LL}^v} = g(f_L^v) \quad (v = 1, \dots, n),$$

where the function g is the same for all firms. I know of no simple way to interpret this condition directly in the general case,¹² but most of its implications are clear and interpretable enough.

In the first place, one way in which (3.3) can be satisfied is for every firm's production function to be additively separable as in (2.4). This is natural, since we already know that additive separability is a sufficient condition for capital aggregation whether or not labor is optimally allocated to firms. On the other hand, since (3.3) is a necessary as well as sufficient condition for capital aggregation, it is evident that if any one firm has an additively separable production function ($f_{KL}^v \equiv 0$), then no capital aggregate exists unless every firm has such a production function. In other words, capital aggregation is not possible if there is both a firm which uses labor and capital in the same production process and another which has a fully automated plant. (I shall henceforth assume that there is no such separability.)

This is but the first of a number of somewhat uncomfortable results with the general characteristic that if there are one or two firms with production functions having some property, then it matters very little what reasonable properties the production functions of other firms have; aggregation will not be possible. A stronger result of this kind can easily be obtained by observing that if constant returns are not assumed, there is no reason why perfectly well behaved production functions cannot fail to satisfy *any* partial differential equation in the form (3.3). If some firm has such a production function, then exact capital aggregation is impossible regardless of the nature of the production functions of other firms;

¹⁰ This problem has been studied in different ways by several authors. See, for example, Diamond [1], Fisher [2], Gorman [6], Hall [8], Nataf [19], Stigum [25], and Whitaker [27].

¹¹ See [2] for the proof of this and other statements in this section.

¹² If one assumes constant returns, (3.3) immediately becomes the condition that profits per worker (or, indeed, average product) be the same for all firms when labor is optimally allocated. See [2].

indeed, no capital aggregate in a production function sense will exist even if all firms are exactly alike and capital physically homogeneous (but immovable).¹³

It is evident on reflection, however, that this sort of problem arises in part from the lack of constant returns since without constant returns it *does* matter how production is organized into firms. Indeed, it is not hard to show that every *constant returns* production function does satisfy a partial differential equation such as (3.3), although naturally, not all constant returns production functions satisfy the same one.

Let us therefore assume constant returns for the moment and ask what conditions are then necessary and sufficient for capital aggregation. This can be done by integrating (3.3), but at least so far as sufficiency is concerned it is more revealing to take a different route.

Suppose that the production functions of the different firms differ from each other only by a *capital augmenting* technical difference, that is, that each f^v can be written

$$(3.4) \quad f^v(K(v), L(v)) = f^1(b_v K(v), L(v)) \quad (v = 1, \dots, n),$$

where the b_v are positive constants ($b_1 = 1$). In this highly restrictive case, a different capital good is equivalent in all respects to more of the same capital good. It is then natural to think of measuring capital in efficiency units and natural to suppose that with constant returns aggregation of capital will be possible since firms will differ only as to amount of efficiency capital. Indeed, this is so; it will be possible in this case to allocate labor so as to make firms differ only as to scale and, moreover, it will be efficient to do so. Constant returns will then allow firms to be added together.

A formal proof along these lines is easy to give.¹⁴ It will suffice to consider two firms. Define $J \equiv b_1 K(1) + b_2 K(2)$ (this will turn out to be the right definition) and recall that $L = L(1) + L(2)$. The sum of the outputs of the two firms is

$$(3.5) \quad Y = f^1(b_1 K(1), L(1)) + f^1(b_2 K(2), L(2)).$$

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

$$(3.6) \quad \frac{L(1)}{b_1 K(1)} = \frac{L(2)}{b_2 K(2)} = \frac{L}{J}$$

when labor is optimally allocated. Let

$$(3.7) \quad \lambda \equiv \frac{b_1 K(1)}{J} = \frac{L(1)}{L}$$

¹³ In this extreme case, however, a competitive market would lead all firms to the same equilibrium position in the long run and capital would have the same marginal product in all uses. This makes capital essentially a long run movable factor and takes us back to the case of full maximization discussed in the preceding section.

¹⁴ This proof is due to Solow [24, pp. 104–105].

(the second equality holding when labor is optimally allocated); then

$$(3.8) \quad Y^* = f^1(\lambda J, \lambda L) + f^1((1 - \lambda)J, (1 - \lambda)L) = f^1(J, L)$$

because of constant returns.

So far, so good. The trouble is that the case of capital augmenting technical differences turns out to be the *only* case under constant returns in which a capital aggregate exists.¹⁵ Only a very limited and special kind of technical diversity can be accommodated. Just how limited that diversity is can be seen either from contemplation of the definition of capital augmentation or from the fact that if all firms differ only by a capital augmenting parameter, then when labor is optimally allocated, *average* product per worker (as well as marginal product) and profits per worker will be the same in all firms.

As already indicated, the situation is even worse in most respects when we drop the constant returns assumption. Here there are even cases (and such cases are the rule, rather than the exception) in which no capital aggregate exists whether or not firms are technically diverse. On the other hand, for a rather limited class of cases, capital aggregation is possible under somewhat wider conditions than that of merely capital augmenting technical differences. This is the class of cases in which each firm's production function can be made constant returns after a suitable (generally nonlinear) stretching of the capital axis. Thus (omitting henceforth the firm index on the factors, where the context is clear):

$$(3.9) \quad f^v(K, L) = F^v(H^v(K), L) \quad (v = 1, \dots, n)$$

where the F^v are homogeneous of degree one in their arguments and the H^v are monotonic. I have given this class of production functions the rather non-euphonious name of "capital generalized constant returns" or "CGCR" for short.¹⁶

It is easy to see that this case is really not analytically different from the constant returns case for our purposes. Since, unlike labor, capital is not allocated over firms so as to maximize output, it makes no difference whether we take the capital goods in their original units or in their transformed, stretched units as being the fundamental capital goods of the model. If we do the latter, it is evident that a necessary and sufficient condition for capital aggregation is that the production functions of the individual firms differ only in the ways in which the capital axis is stretched, in the functions H^v but not the functions F^v of (3.9). This is a generalization of capital augmenting technical differences which I have called "capital altering." Despite the fact that it is more general than capital augmentation, it allows capital aggregation only in a restricted class of cases and is itself quite restrictive. It is evident that the force of the conditions necessary for capital aggregation is not to be evaded by dropping the assumption of constant returns.¹⁷

¹⁵ This holds true except for the trivial case of all production functions additively separable. This result was proved independently by a number of authors. See the works cited in footnote 10.

¹⁶ Examples are easy to generate, the simplest being the nonconstant returns Cobb-Douglas, $AK^\alpha L^\beta$, which is, however, the only CGCR production function which is also homogeneous of some degree other than one.

¹⁷ The class of CGCR functions differing by capital altering differences is not the only nonconstant returns case permitting capital aggregation. There are some other very special cases. See [2, p. 273], for example.

One might well ask, however, whether the restrictiveness of such conditions may not be a consequence of the highly simplified nature of the model. Suppose that there are several outputs, several labor types, and several capital goods produced by or used by each firm. Unfortunately, it turns out that essentially the same results apply when there are several outputs or several labor types, although conditions such as (3.3) are replaced by rather more complicated matrix equivalents. The presence of more than one kind of capital good for each firm does make some difference, however, and this I shall now discuss.

The essentials of what is involved can be seen by examining the case of a single output, a single labor type, but two capital goods. Thus the v th firm's production function becomes $f^v(K_1, K_2, L)$, where the firm index has been omitted from the arguments. There are two questions. First, what are the conditions under which it is possible to form an aggregate of only the first capital type (for example, an equipment aggregate or a plant aggregate)? Second, when is it possible to form an aggregate of all capital together (a total capital aggregate)?

These are not the same question and the conditions for one do not imply the conditions for the other. In the formation of a K_1 aggregate only, marginal rates of substitution between K_1 for different firms must be independent of K_2 ; this is not required if K_2 is included in the aggregate. On the other hand, if K_2 is to be so included, then the marginal rate of substitution between K_1 and K_2 must be independent of L ; this is not required if K_2 is to be left out of the aggregate.

Taking first the case in which only K_1 is to be in the aggregate, the Leontief conditions turn out to be twofold. First, the Leontief conditions with respect to labor require that every firm's production function satisfy a partial differential equation in the form

$$(3.10) \quad \frac{f_{K_1 L}^v}{f_{K_1}^v f_{LL}^v} = g(f_L^v) \quad (v = 1, \dots, n),$$

where the function g is the same for all firms and does not depend on K_2 . This is just (3.3) again, as one might expect. Unfortunately, however, whereas any two-variable, constant returns production function satisfies a partial differential equation in the form of (3.3), not every three-variable function satisfies a partial differential equation in the form of (3.10), for the presence of constant returns in K_1 , K_2 , and L implies the absence of constant returns in K_1 and L alone. This already means that K_1 aggregation requires conditions more stringent than we found in the one-capital case; capital augmentation or K_1 augmentation and constant returns will not do. The mere existence of a firm with a particular constant returns production function (one not satisfying any equation in the form (3.10)) will prevent K_1 aggregation.

To make matters worse, the Leontief condition with respect to K_2 is equally stringent. Assuming (3.10) to hold, that condition becomes

$$(3.11) \quad f_{K_1 K_2}^v - \frac{f_{K_1 L}^v f_{K_2 L}^v}{f_{LL}^v} = 0 \quad (v = 1, \dots, n).$$

This condition can be interpreted as follows. Suppose a little K_2 is added to the v th firm. There are two effects. The first of these is a direct effect on the marginal product of K_1 ; it is $f_{K_1 K_2}^v$. The second is an effect through the reassignment of labor. That reassignment can be thought of in two steps. First, a certain amount of labor is withdrawn from the system as a whole; second, it is assigned to the v th firm.¹⁸ The change in labor available to the entire system has no effect on the marginal rate of substitution between K_1 in the v th firm and K_1 in any other firm, because (3.10) was assumed to hold. There is such an effect through the specific assignment of more labor to the v th firm, however; like the direct effect of the increase in K_2 , such reassignment shifts the marginal product of K_1 in the v th firm but does not affect it in any other firm. It is easy to see that the amount of reassigned labor will be $(-f_{K_2 L}^v/f_{LL}^v)$ as this is just what will be required to keep the marginal product of labor unaffected. Accordingly, the second term in (3.11), including the minus sign, is the indirect effect on the marginal product of K_1 in the v th firm. The condition (3.11) is that this must just cancel out the direct effect.

The condition (3.11) is obviously very strong. Just how strong it is may be grasped in part by considering one of its implications. Suppose we define two factors as complements if increasing one of them increases the marginal product of the other. In two-factor, constant returns production functions, the two factors must be complements, but not all pairs of factors need be so if there are more than two. Nevertheless, suppose that the v th firm has a production function with all three factors complements. It is easy to see that the left hand side of (3.11) must then be positive for that firm and no K_1 aggregate can exist.

Thus, for example, the existence of any firm with a three-factor Cobb-Douglas production function in, say, plant, equipment, and labor: $AK_1^\alpha K_2^\beta L^{1-\alpha-\beta}$, prevents the construction of either a separate plant or a separate equipment aggregate for the economy as a whole (although, as we shall see, it does not prevent the construction of a full capital aggregate). Similar statements are true about most easy three-factor generalizations of two-factor, constant returns production functions. Even the existence of some pair of factors which are substitutes, moreover, does not guarantee either the satisfaction of (3.11) or the satisfaction of any condition in the form (3.10). So far as the present assumptions are concerned, the construction of a subaggregate of capital goods requires even less reasonable conditions than the construction of a single aggregate in the case earlier discussed.

What about the construction of a complete capital aggregate in the present case, however? Here the conditions, while not technically weaker, seem a trifle less stringent than those for the construction of subaggregates.

In the first place, it is natural to suspect that a necessary condition for such complete capital aggregation is that it be possible to construct such a capital aggregate for each firm taken separately. This indeed turns out to be the case and it already greatly restricts the admissible class of firm production functions. On the other hand, many firm production functions having reasonable appearance will allow such a construction; the three-factor Cobb-Douglas is one.

¹⁸ This assumes that more rather than less labor will be assigned to the v th firm as a consequence of its having an increase in K_2 . The opposite case is analogous.

Curiously, however, whereas one might expect capital aggregation over all firms to involve the condition that the individual firm aggregates all be of the same form,¹⁹ this is the only restriction which is *not* involved, at least in the constant returns (and the CGCR) case. As can be seen by applying the theorem for the two-factor, constant returns case to the present problem with each firm's individual capital aggregate counted as a single capital good, the necessary and sufficient condition for full capital aggregation in the constant returns case, given the existence of individual firm aggregates, is that all firms differ by at most a capital augmenting technical difference. Interpreted in the present context, that means that firms can differ as much as we like in the way in which their individual capital aggregate is constructed; they may not differ at all in any other way. It is as though each firm used its basic capital goods to construct an intermediate capital jelly which was then combined with labor to produce output. Capital aggregation is possible (under constant returns) if and only if the only difference among firms is in the way in which the jelly is constructed, not in how it is used thereafter.

This condition leads to a rather wider class of cases than appears from the two-factor case. For example, if one firm has a three-factor Cobb-Douglas function,

$$(3.12) \quad f^1(K_1(1), K_2(1), L(1)) = AK_1(1)^a K_2(1)^b L(1)^\gamma$$

(where the three exponents may or may not be restricted to sum to unity), aggregation will be possible over this firm and any other with production function

$$(3.13) \quad f^2(K_1(2), K_2(2), L(2)) = BK_1(2)^c K_2(2)^d L(2)^\gamma$$

or, indeed, any firm with production function in the (now completely general) form

$$(3.14) \quad f^3(K_1(3), K_2(3), L(3)) = H(K_1(3), K_2(3))L(3)^\gamma.$$

Note that the exponent of the labor term is the same in all cases; the relative importance of the $K_1(v)$ and $K_2(v)$ is allowed to vary at will. Nevertheless, the class of cases allowing aggregation is clearly still very restricted.²⁰

Thus, capital aggregation turns out to require rather stringent conditions if firms are technically diverse. This will, of course, come as little surprise to some. After all, Mrs. Robinson has argued this sort of position for years.²¹ Indeed, one might say that the wonder is that capital aggregation can be done at all, not that it can only be done in rather restricted circumstances. Nevertheless, it is true that capital aggregation can be performed in cases in which capital goods are physically quite different and production functions allowed to differ over firms.

¹⁹ A similar expectation turns out to be correct in the aggregation of labor and of output discussed below.

²⁰ One might also add that the aggregate will look a bit odd. For the three firms just described, it is

$$(3.15) \quad J = (AK_1(1)^a K_2(1)^b)^{1/(1-\gamma)} + (BK_1(2)^c K_2(2)^d)^{1/(1-\gamma)} + (H(K_1(3), K_2(3)))^{1/(1-\gamma)}.$$

The aggregate production function would be

$$(3.16) \quad Y^* = J^{1-\gamma} L^\gamma.$$

²¹ See [21], for example. Mrs. Robinson was not specially concerned with aggregation over firms.

The mere immobility and physical differentiation of capital goods does not prevent the construction of such aggregates, even though the technical diversity that can be accommodated is fairly limited. Moreover, the difficulties involved are not associated with the question of fixed coefficients versus smooth substitutability. They arise, as we have seen, in the neoclassical case.

What is perhaps more surprising than the difficulty of aggregating capital itself is the fact that the immobility and physical differentiation of capital impose nearly equally stringent conditions on the aggregation of labor types and of outputs. I now turn to this question.

4. LABOR AGGREGATION AND OUTPUT AGGREGATION²²

To consider the problem of labor aggregation, we must, of course, drop the assumption that there is only one homogeneous type of labor and reinterpret $L(v)$ as an s -component vector, $(L_1(v), \dots, L_s(v))$, where $L_j(v)$ denotes the amount of the j th type of labor employed by the v th firm. It does not matter whether there is one or more types of capital, so $K(v)$ may be interpreted either as a scalar or a vector. Clearly this provides a quite general model, save for the assumption that there is only one homogeneous output.

The case of output aggregation when there is only one labor can be handled analogously. Relabel the single homogeneous labor as $Y(v)$ and take $L(v)$ to be the vector of outputs and f^v to be a labor requirements function rather than (directly) a production function. Then the same model serves to discuss output aggregation.

Moreover, it is clear that both problems can be handled simultaneously. Let $Y(v)$ be the amount of a particular output produced by the v th firm and consider all other outputs as negative inputs. Labor aggregation in the presence of many outputs then becomes the problem of aggregating over a subset of the variable inputs. The problem of aggregating outputs in the presence of many labor types can clearly be similarly handled. Since the results in all important cases turn out to be the same, there is no need to treat the various cases in detail here.

The principal regularity condition on the firm production functions which must be imposed is that each of them have a negative definite Hessian with respect to the variable inputs—essentially that each have strictly diminishing returns to any linear combination of movable factors. This is, of course, the natural generalization of the condition $f''_{LL} < 0$ in the simple model.

On the other hand, another assumption appears required which parallels a result rather than an assumption in the capital aggregation problem. We are here primarily interested in the cross-firm aggregation problem that arises because labors or outputs are shifted over firms, given the capital stocks and production functions, to achieve efficient production. Yet there is also a labor aggregation or output aggregation problem within each firm. Thus, a labor or output aggregate might exist for each firm separately and not for all firms together. This phenomenon is the one we are analyzing. What is strange is that apparently there can exist a labor or output aggregate for all firms together without one existing for each firm

²² This section is based on Fisher [3]. See also Stigum [26].

separately. (This can happen because the effect on the marginal rate of substitution between two variable factors of changing one of the capitals is different when one considers only one firm from when one considers efficient factor reallocations over all firms.) In the case of capital aggregation, this cannot happen. One can show that a necessary consequence of aggregation over all firms is that an aggregate must exist for each firm separately; in the case of the aggregation of variable factors it is apparently possible, although I have not actually constructed an example.

Nevertheless, such a case, if it exists, is only a *curiosum*. It is hard to find much interest in an aggregate so fragile that it exists over a set of firms but not over proper subsets. If the existence of an aggregate production function depended on this sort of phenomenon, then an earthquake which swallowed one or more firms would also swallow the aggregate production function. I shall thus assume, as in the capital aggregation case, that any aggregate to be considered already exists at the firm level, even though (unlike the capital aggregation case) such an assumption may not be technically necessary for the existence of the aggregate over the entire set of firms.

This assumption, of course, imposes an already strong condition on the production functions of the individual firms. Taking the labor aggregation case with a single homogeneous output for simplicity and continuing to omit the firm argument in Y , K , and L , the assumption means that every firm's production function can be written in the form:

$$(4.1) \quad Y = f^v(K, L) = F^v(K, \phi^v(L))$$

where ϕ^v is a scalar-valued function (recall that L is now a vector).

Given such restrictions, however, it turns out that the conditions for labor or output aggregation are weaker than those for capital aggregation, in the sense that the existence of a capital aggregate for the entire set of firms implies the existence of a labor or output aggregate. This is not too surprising. Recall the constructive proof (given in the preceding section) that, under constant returns, capital augmentation (which we know to be necessary for capital aggregation) implies the existence of a capital aggregate. There, essentially, we could take every f^v as of the same form, absorbing the differences in the efficiency parameter multiplying capital. It turned out that the aggregate production function then also had that same form. If that form is assumed to permit labor (or output) aggregation before capital aggregation, then it certainly continues to do so afterwards. Nevertheless, the content of the theorem is not restricted to constant returns nor, indeed, to the existence of a full capital aggregate, as opposed to a subaggregate. Thus, it is a bit stronger than this argument indicates.²³

Unfortunately, the fact that labor or output aggregation turns out to require technically weaker conditions than does capital aggregation (given the restriction assumed) does not make it terribly likely that labor or output aggregates exist. Even aside from the strong requirement that the appropriate aggregate already

²³ Details may be found in Fisher [3]. The precise form of the results does appear to depend on whether there is more than one output when labor is being aggregated, or the reverse, but the sense of it does not.

exist at the firm level, the conditions under which a labor or output aggregate exists are quite unrealistic whether or not a capital aggregate exists. For constant returns and some related technologies,²⁴ those conditions can be described in several ways.

The simplest way to state such conditions (but perhaps not the most revealing one) is to say that they amount to the requirement that the individual aggregating functions, ϕ^p , can be taken to be the same. The functions F^p can differ unrestrictedly. It is as though (for the labor case) we interpreted (4.1) as saying that each firm uses its different labors without capital to make a composite variable factor and then combines the composite with capital to produce output. A labor aggregate exists over all firms taken together if and only if the production of the composite variable factor from individual labors is the same for all firms. The way in which the composite is used with capital to produce output is not restricted.

It is interesting to note that this is just the reverse of the analogous condition for capital aggregation. In the case of aggregation over several capitals, we found that the requirement was that the use of composite capital together with labor to produce output should be the same for all firms. The construction of composite capital from individual capitals was not restricted.

Since the functions, F^p , are unrestricted (save to preserve constant returns), stating the conditions for labor or output aggregation in this way makes it appear that a wide variety of technical differences among firms can be accommodated. It is evident, for example, that if one firm has a Cobb-Douglas production function, a labor aggregate can exist even if other firms have production functions in very different forms.

Unfortunately, however, even though the technical differences which can be accommodated are indeed wider than in the capital aggregation case, the realism of the cases that can be accommodated is not much greater. This can be seen by considering an alternate (but equivalent) way of stating the necessary and sufficient conditions for labor or output aggregation.

Under constant returns, it is easy to show that the individual aggregating functions, ϕ^p , can be taken to be homogeneous of degree one. This means, however, that the ratio in which a particular firm hires any two labor types or produces any two outputs depends only on the relative wages of all labors or the relative prices of all outputs. If (and only if) the ϕ^p are all the same, the same set of relative wages for labor will lead to the employment of all labor types in the same relative proportions for all firms; similarly, the same set of relative prices for outputs will lead to the production of all outputs in the same relative proportions for all firms.²⁵ Thus the existence of a labor aggregate requires the absence of specialization in employment; faced with the same set of wages, all firms must hire the same mix of labor types, differing only as to scale of total employment. Similarly, the existence

²⁴ Essentially, those technologies for which the individual aggregating functions, ϕ^p , can all be taken to be homogeneous of degree one. There seems no point in going into great detail here. Such technologies include CGCR technologies as a special case.

²⁵ Note, incidentally, that this form of the condition implies the existence of individual firm aggregates, ϕ^p .

of an output aggregate requires the absence of specialization in production; faced with the same set of relative prices, all firms must produce the same market basket of outputs, differing only as to scale of production. The mix of labor types employed and the market basket of outputs produced are not constant, since they depend on relative wages and relative prices, respectively; but they are the same over all firms. Obviously, this is not a realistic condition for firms in different, rather narrowly defined industries.

What about the construction of subaggregates? Is that any easier? This question is of some importance for three reasons. First, there are natural subaggregates which one might want to discuss. The employment of skilled labor or the production of nondurables comes to mind. Even if all labors or all outputs cannot be aggregated, it would be useful if some could be.

Second, we have been pretending that the conventional distinction between capital and labor exactly corresponds to the crucial distinction in this model between fixed and movable factors. If some kinds of capital are, in fact, movable, one might want to aggregate them or aggregate the true labors without aggregating all movable factors together. I shall take up this matter in the next section.

Third (a more narrowly technical matter), if we consider the problem of labor aggregation in the presence of many outputs or the problem of output aggregation in the presence of many labors, and agree to call alternative outputs negative factors, then we are in fact considering the question of forming an aggregate from some proper subset of the variable factors. So, in a way, the existence of subaggregates is what we have been discussing all through this section.

This fact points to what the conditions for the existence of a subaggregate must be (under constant returns). There are two. First, the subaggregate in question must exist at the firm level. Second, the subaggregating functions must be the same for all firms. Equivalently, faced with the same set of relative wages for the labors to be included in the subaggregate, all firms must hire *those* labors in the same relative proportions. Faced with the same set of relative prices for the outputs to be included in the subaggregate, all firms must produce *those* outputs in the same proportions.

These are the same conditions, suitably applied, that were involved in the construction of a full aggregate. They can be more realistic in the case of subaggregates, however, for a firm producing zero of some specified list of outputs can be regarded as producing any given market basket of them at zero level. Thus, some specialization in production (or, similarly, in employment) is allowed when only a subaggregate is to be constructed. Any firm which produces any output in the subaggregate, however, must produce all of them in the same proportions as any other firm, so specialization is not permitted within the production of the outputs (hiring of the labors) in the subaggregate.

One further point before proceeding. The fact that the conditions for subaggregation turn out to be the same as those for full aggregation applied to a subset of labors or outputs makes one suspect that the existence of a full aggregate might well imply the existence of a subaggregate. This is so in the same sense that the existence of a capital aggregate implies the existence of a labor or an output

aggregate. That is, under constant returns, *provided a particular subaggregate exists for every firm individually*, the existence of a full aggregate (or indeed of any larger subaggregate) for the set of firms as a whole implies the existence of the subaggregate in question for the set of firms as a whole. This is fairly easy to see from the preceding discussion; it is in sharp contrast to the case of capital aggregation already discussed where no such relationship holds.²⁶

5. HOW FIXED IS FIXED CAPITAL?²⁷

So far, I have spoken as though the distinction between the elements of K and the elements of L exactly corresponded to the conventional distinction between capital and labor. Clearly, this need not be the case. The crucial distinction in the model is that between fixed and movable factors, the elements of K being firm-specific, so to speak, while the elements of L are shuffled over firms to achieve efficiency. Some kinds of human capital, however (economics professors, for example), may be capable of only one type of work in a specific place, while (rather more importantly, perhaps) some types of physical capital may be mobile over firms. Aside from such items as typewriters and other office equipment used by nearly all firms, production equipment may in some cases be thought of as mobile in the long run as efficient reallocation thereof takes place through depreciation and reinvestment. This means that such immobile labor should be counted as an element of K and such mobile capital as an element of L in our model. Since conventional definitions still make it interesting to speak of a capital aggregate or a labor aggregate rather than a fixed factor or movable factor aggregate, we must therefore ask what becomes of our results if fixed and movable factors are to be aggregated together.

Accordingly, in this section I briefly examine the case in which some, but not all, capital goods are movable (the case in which some labors are fixed can be similarly handled). Does such mobility make aggregation easier? The answer turns out to depend on whether we are considering capital aggregates including both fixed and mobile factors or whether we consider aggregates of purely fixed or purely mobile factors.

In the case of partial capital aggregates including only the mobile capital goods, the answer is clear. If one compares the conditions for subaggregation discussed above for capitals and for labors, it is obvious that the conditions for such aggregation are rather less stringent when movable factors are involved than when fixed factors are. (Whether they are any more likely to be satisfied in practice is another matter.) The primary reason for this is the fact that the construction of a subaggregate of fixed factors when there are fixed factors left out, involves conditions such as (3.11) which are very special and cannot, for example, be satisfied if all

²⁶ The existence of a full aggregate also implies the existence of subaggregates for all firms which exist for each one in various nonconstant returns cases. Curiously, as opposed to the capital case where subaggregation was hindered if all factors were complements, subaggregation in the labor or output cases turns out to be aided by such pervasive complementarity. See Fisher [3] for details.

²⁷ This section is based on Fisher [4].

factors are complements. The construction of a subaggregate of movable factors involves no such strong conditions. Hence it seems easier to aggregate a given subset of capital goods if that subset is movable than if it is fixed.

Moreover, a similar remark applies (and for the same reason) if the subaggregate to be constructed includes all the fixed factors. Leaving mobile factors out of the aggregate just does not involve the same strong condition as does omitting fixed factors. Hence it is easier to form an aggregate from a given group of capitals if the *remaining* ones are mobile than if the remaining ones are fixed.

One would naturally expect some such result, for it is easy to see that mobility cannot hamper aggregation. The conditions for aggregation when capitals are fixed are conditions such that, when they are satisfied, aggregation becomes possible no matter what values the various capital arguments have. Hence those same conditions must imply the possibility of aggregation if some of the capital arguments just happen to have those values which occur when they are shifted over firms to achieve efficiency. Thus, if aggregation is possible with capitals fixed, it remains so when some (or all) capitals become mobile. It is therefore not too surprising that mobility helps aggregation in the cases indicated.

What is perhaps surprising is that in the rather more general (and important) case where mobile and fixed capitals are to be included in the *same* aggregate, the mobility of some capitals does not change the aggregation conditions. Yet I have shown this to be true, at least for the case of constant returns when all fixed capitals and some mobile ones are to be aggregated. Despite the mobility of some of the capital goods, capital augmentation (or the generalization thereof discussed at the end of Section 3) remains necessary and sufficient for the construction of a full capital aggregate. In this, perhaps the most important case, mobility does not hurt, but it does not help either.²⁸

6. APPROXIMATIONS AND THE REAL WORLD: DO EXACT RESULTS MATTER?

So far, I have been concerned with the conditions that are sufficient and, especially, necessary for an aggregate production function to give an exact representation of a diverse set of firms with technology embodied in fixed capital goods. I have discussed those conditions assuming that the values of the variables are restricted only by the requirements of economic sense and of efficient allocation of movable factors. Obviously, the requisite conditions turn out to be terribly strong.

Yet such exact results may not be of much practical force. What we really care about is whether aggregate production functions provide an adequate approximation to reality over the values of the variables that occur in practice. This is especially so for the empirical studies of production functions, technical change, growth, and related subjects; it is less so for theoretical studies of optimal capital accumulation and growth where the variables are less likely to be restricted in

²⁸ There are some differences in the extension of the results to nonconstant returns (it matters whether a production function is CGCR because a stretching of the fixed capital axis brings it back to constant returns or whether stretching of a mobile capital axis is involved), but they do not seem worth going into in detail. See Fisher [4].

range. Yet even there, one is perhaps interested less in exact results than in good approximations.

Now, there are two ways in which the values of the variables might be restricted. The first of these is simply a restriction as to range. We know that the values of the various capitals and labors observed in practice are restricted, by the finiteness of the present economy, to lie in a bounded set. It might be the case that such restriction, together with the relaxing of our requirements from exact results to good approximations, leads to a substantial relaxation of our conditions for aggregation. This is the possibility that I shall discuss in the present section.

On the other hand, whether or not the range of the variables is restricted, there may be some other conditions that, so to speak, reduce the dimensionality of the problem, just as the requirement that labor be optimally allocated reduced the problem from one in which the independent variables included each firm's employment of labor to one in which the only exogenous labor variable was the total employed by all firms. I shall return to this possibility below.

For the moment, then, I ask the following question:²⁹ Suppose the elements of K and L are restricted to lie in a bounded rectangular region, S . Suppose further that we no longer require that an aggregate production function exist which exactly equals the true disaggregated production function but rather require that an aggregate production function exist which comes within some specified distance, ϵ , of the true production function for all points in S .³⁰ To what extent can the rather stringent conditions already discussed be relaxed?

There is one obvious way in which such relaxation is possible. It is easy to see that if we are only interested in approximate results, we need only require that our exact conditions hold approximately. Thus, in the simplest case, it will clearly suffice for approximate capital aggregation that all technical differences among firms be approximately capital augmenting. (Naturally, how close to capital augmentation the situation must be depends on how close an approximation in the results is required.) As is usually (but not always) the case, small errors have small consequences.

Unfortunately, this is not a terribly helpful result. The reason for being unhappy with capital aggregation, for example, is not merely that one thinks technical differences are not likely all to be exactly capital augmenting but that one thinks there are some differences that are not anything like capital augmenting. It is not much comfort to know that small deviations from capital augmentation can be tolerated when one believes that such deviations may well be large.³¹

The interesting question, therefore, is whether there are any other cases—cases in which our exact conditions are *not* approximately satisfied but in which an aggregate production function gives a satisfactory approximation for all points

²⁹ The following discussion is based on Fisher [5], which deals with a more general problem than the existence of aggregate production functions.

³⁰ Similar results hold if the standard of approximation is not absolute but relative to the size of output.

³¹ Honesty requires me to state that I have no clear idea what technical differences actually look like. Capital augmentation seems unduly restrictive, however. If it held, all firms would produce the same market basket of outputs and hire the same relative collection of labors.

in S . Naturally, the answer depends on what one means by a "satisfactory approximation" as well as on how badly our conditions are violated. Nevertheless, the general answer appears to be in the negative.

Without going into great detail,³² it turns out that the only way in which such approximations could result would be if we were willing to accept production functions which were very irregular in a well defined sense. More particularly, assuming twice differentiable production functions (and it seems vain to hope that differentiability makes any substantial difference here), such approximations could only result if either the true production function or its aggregate approximation had first or second derivatives exhibiting very large rates of change both up and down on every closed rectangular subregion of S . Put not quite precisely, those derivatives would have to fail to satisfy a Lipschitz condition in the limit as the required approximation became closer and closer to the truth. Moreover, the violation of the Lipschitz condition would have to take place both above and below. In less technical language, the derivatives would have to wiggle violently up and down all the time.³³ There is nothing wrong with functions having such properties, but we do not ordinarily expect production functions to exhibit this kind of behavior. Certainly, it is not exhibited by the aggregate production functions used in practice.

Naturally, there is a bit more to it than this. In particular, how "irregular" things have to be depends on how closely the aggregate production function is required to approximate the true state of affairs as well as on how badly our exact aggregation conditions are violated and the functional forms and range of variables involved. Nevertheless, there can be no presumption that our exact conditions don't matter simply because one is interested in reasonable approximations over a limited domain. Indeed, I think the presumption must be the other way. Such an escape from the stringency of the conditions will be available, if at all, only in rather special cases.

7. BUT IT DOES MOVE, ALL THE SAME

Despite all this, there is, after all, considerable evidence that aggregate production functions may be appropriate approximations. I do not put much weight on the fact that estimates of aggregate production functions from output and factor data tend to yield high R^2 's; the aggregates are highly correlated with the variables in the true specification, and, anyway, high correlations are fairly easy

³² See Fisher [5].

³³ The problem is analogous to (and indeed derived from) the problem of whether a differentiable function of one variable can be close to zero everywhere on a bounded interval and yet have its derivative not generally close to zero, that is, not be reasonably flat. The answer is clearly yes, but not if the function must also look "regular." The function $f(x) = \lambda \sin(x/\lambda)$, for $\lambda > 0$, is everywhere between $+\lambda$ and $-\lambda$. Its derivative is $f'(x) = \cos(x/\lambda)$ which is not everywhere close to zero when λ is small. The derivative fluctuates up and down, however, with the fluctuations becoming more and more frequent as we force the function closer to zero by taking smaller and smaller values of λ . As we approach the limit, the second derivative fails to exist, and, more important, the first derivative wiggles violently up and down in any small closed interval.

to generate. On the other hand, there is apparently nothing about the aggregation or estimation procedure involved in such studies to guarantee that the production functions so estimated will give approximately the correct picture of factor shares. Yet, clearly, this is the case. As Solow once remarked to me, we would not now be concerned with this question had Paul Douglas found labor's share of American output to be twenty-five per cent and capital's share seventy-five instead of the other way round.

How do we account for this? It is hard to be sure, but I have fairly strong suspicions.

It is clear, of course, that if something is at work behind the scenes which reduces the dimensionality of the problem in ways not accounted for in our analysis, then aggregate production functions may work well even though we conclude that they ought to work badly. The simplest example of this is the case in which, for reasons unspecified by us, firms always invested in proportion to a particular index, J . In that case, J would clearly be a suitable capital aggregate regardless of our analysis, so long as that behavior continued. If firms invested approximately in fixed ratios, then J would be an approximate aggregate. Similarly, if outputs were always produced or labors always hired in approximately fixed proportions, then an approximate output or labor aggregate would exist.

To take a different example, suppose that all capital goods were produced and rented (but not sold) under conditions of monopoly with prices administered. If all such prices were fixed over time, then firms using those goods as inputs would have to adjust to make the marginal rates of substitution between pairs of capital goods equal to the fixed price ratios. Even if each firm used a different capital good, but produced the same output as other firms, such equality would come about through the equating of the marginal revenue products with factor prices. In such a case, a capital aggregate might not exist in general because the marginal rates of substitution among pairs of capitals would not generally be independent of labor. Nevertheless, a capital aggregate would appear to exist because the marginal rates of substitution in question would be fixed.³⁴ Naturally, the fixing of those rates could not be plausibly regarded as a purely technical matter, and the aggregate would only be appropriate so long as the capital goods industry failed to change relative prices.

Similarly, if the relative prices of outputs or of labors were administered and more or less stable, an output or a labor aggregate would appear to exist and would, indeed, be perfectly appropriate so long as such price behavior persisted.

Finally, there is the possibility that there are systematic forces at work which we simply haven't thought of. Suppose, for example, in the one-output, one-labor case, that all technical differences were indeed capital augmenting. As we know, a capital aggregate would then exist and an aggregate production function would give good results. An observer who had not considered the possibility that labor is efficiently allocated (by the market or by planners) would be rather puzzled, however. He could observe the good results, but, unless he thought every firm's

³⁴ This is, of course, a special case of Hicks-Leontief aggregation. See Hicks [9, p. 33] and Leontief [13].

production function were additively separable, he would also believe that such good results could not generally be expected. Similarly, in the present instance, it is possible that some systematic phenomenon is at work which, like the efficient allocation of labor, restricts the dimensionality of the problem and widens the class of cases in which aggregate production functions give appropriate answers.

These three possibilities are not mutually exclusive and there may be some element of truth in all of them. For the present, anyway, I tend to favor the first over the latter two.

In a way, the third hypothesis—that there are forces at work which we haven't taken into account—is not a hypothesis at all. Unless one is willing to be specific about what such forces are, it is impossible to refute the contention that they are there. I have nevertheless listed this possibility for two reasons. First, the example of the efficient allocation of labor and its effects on aggregation makes one realize that there very well may be plausible systematic effects left out of consideration and points toward the nontrivial way in which such effects might operate. Second, there is, after all, something real to be explained. To the extent that the other two possibilities (approximately constant relative proportions and approximately constant relative prices) do not succeed in accounting for the observed facts, then we must admit our ignorance in this regard. I point out, however, that if this turns out to be the case, then reliance on aggregate production functions is reliance on the continuation of badly understood phenomena. An investigation of this problem ought then to have very high priority in future research.

Turning to the hypothesis that it is administered and sticky prices that make aggregate production functions look good, I believe this should have some weight, but it is hard to say how much. The difficulty is knowing how stable relative prices have to be in order to give aggregate production functions the appearance of working. It is also not trivial to construct an index of stability for relative prices which has reasonably desirable properties (such as symmetry, freedom from scale, independence of original distribution of relative prices, and so forth). I have constructed one such index for the stability of relative output prices and of relative capital goods prices in the United States, disaggregating to fifteen outputs and six capital goods. For what it is worth, that index shows a high degree of stability relative to the rather foolish null hypothesis of independence over time, but whether that degree is high or low for the question at hand, I simply do not know, and I have thus decided not to report the details. Certainly, there are many administered prices in modern economies but whether that alone accounts for the performance of aggregate production functions seems rather doubtful.³⁵ Still, sticky relative prices may be part of the explanation.

The hypothesis that seems the most promising is, as indicated, the possibility that the ratios of items in a particular aggregate do not change very much so that the aggregate is approximately correct over the observed data. In particular, if the economy does not depart too far from balanced growth, this will be the case, but it should be noted that balanced growth is only a leading very special case of

³⁵ One could hope to go some way toward finding out by simulation studies, but I have not thus far done so because the simpler hypothesis about to be discussed seems more promising in this regard.

this kind of thing. Neither exponential growth nor constant ratios among labor, output, and capital (or particular labors, particular outputs, and particular capitals) are required. All that it takes for a capital aggregate are constant ratios of capital goods; for a labor aggregate, constant ratios of labor types are involved; and for an output aggregate, we require constant ratios among outputs. This does not seem a wild departure from the facts and, indeed, real economies may not depart too far from balanced growth, but again, how big a departure will leave aggregate production functions with good results cannot be settled easily.

I am currently in the process of conducting simulation experiments with the simple one output, one labor model and the results, while very preliminary, are quite suggestive. Briefly, they indicate that an aggregate production function does quite well for small, plausible, unsystematic movements in the ratios of the capitals. A relatively large, systematic trend in such ratios of more than a few per cent a year causes things to break down rather badly. As one would expect, such sensitivity to systematic change is worse the farther the departure from exclusively capital augmenting technical differences. A plausible extent of such departure can be tolerated, however, so long as capital ratios do not systematically move by amounts more than a very casual glance suggests are typical of real economies in the short run.

I hope to be more precise about these results in a later paper, but, if they hold up, they certainly suggest that the good performance of aggregate production functions may be due to the fact that there is not a great deal of movement in the relative variables.³⁶ This makes conclusions drawn from such aggregate functions valid so long as things keep on moving pretty much together; it makes such conclusions suspect if extended to movements between balanced growth paths; and it suggests that conclusions for large aggregates drawn from internationally derived estimates may be particularly shaky.

8. WHY DOES IT MATTER?

This leads directly to the final question which I wish to consider, that of the consequences of all this. Ought one to be particularly concerned over these results? Further, why should one be more concerned with aggregation difficulties in production functions than in other areas of economic analysis—consumption functions, for example.

The answers to these questions seem to me to be related and they bear, as they must, on the uses to which production functions are put.

The first such use may be regarded as primarily descriptive. We estimate production functions to give some idea of what is happening to productivity or to use as one piece of a forecasting model, and so forth. In this kind of use, a production function plays a role not very different from that played by consumption functions and aggregation difficulties ought perhaps not to seem more alarming in the former than in the latter case. In both cases we know that such difficulties

³⁶ This seems even more plausible if we recall that recession years in which there may be such movements are frequently excluded from the estimation process or are otherwise adjusted for capacity utilization.

are there in principle; in both we recognize that some degree of approximation is involved: in both we realize (if we are careful) that such approximations will get worse the more things move around.

If there is a difference here, it is perhaps in the confidence one can have as to the usefulness of the approximations. In the case of a linear consumption function (which I take merely as the simplest example), we know that an aggregate will be all right so long as either individual marginal propensities to consume are about equal or so long as the distribution of income remains relatively fixed. The former possibility does not seem outlandish, and the second has been roughly true for some time. In the production function case, however, the parallel to the equality of marginal propensities is (in the simplest example) the condition that all technical differences be approximately capital augmenting and this simply does not seem very plausible when aggregating over different industries. Similarly, rough constancy of the ratios of different capital goods, of different labor types, and of different outputs (if that is what accounts for the success of aggregate production functions), seems somehow an even less well understood phenomenon than that of rough constancy of the income distribution and therefore a less reliable reed on which to lean. Still, I am unwilling to press this last point and this descriptive use of aggregate production functions may well not be much worse than similar uses for other aggregates.

Production functions are not merely used in a direct descriptive way, however. For one thing, they are used together with marginal productivity theory to generate equations to help explain factor prices and employment. This indirect use has no parallel in the case of consumption functions and it means that production functions bear a relatively heavy burden in the analysis of growth. At the very least, this means that we ought to be extra careful about the analysis of a functional relationship on which we are going to put such weight.

Still, so far as empirical work and forecasting is concerned, aggregate production functions often give good results not only in the relation of outputs to inputs but also in their wage and factor share implications. Again, so long as the relevant ratios don't move too much, this is not accidental.

The real problem seems to me to come in analyses in which those ratios can be expected to move—in long run forecasting or in theoretical, descriptive, or normative growth models in which the characteristics and existence of an aggregate production function are of crucial importance. Such studies cannot afford to assume that balanced growth or constancy of ratios will continue, for they are concerned with the very forces that change such behavior. For such studies, the fact that an aggregate production function cannot be expected to exist even as an approximation is clearly very serious.

Does this vitiate the usefulness of such models? Not necessarily. Theorems derived for one and two sector growth models may very well have more general application. If they do, however, it will be because insights and methods developed in the analysis of such models carry over to less aggregative studies. It will not be because those models themselves directly summarize the technical relationships of highly diverse economies.

In short, it seems to me important to worry about aggregation in production functions because production functions are themselves important. They, and their implications, play central roles not only in empirical work but in theoretical analysis. Just because it is possible to use aggregate production functions for grand statements about long run growth and technical change, it is important to be careful about the foundation for such statements. At present, that foundation seems solid only insofar as relatively small changes are concerned. The analyses which I have here summarized have convinced me that there is at least need for great caution in this area. It may be recalled that Solow's seminal article [22, p. 312] called for "more than the usual 'willing suspension of disbelief' to talk seriously of the aggregate production function." That suspension has clearly led to very fruitful results. I am, however, finding it increasingly difficult to maintain. The conditions for the existence of aggregate production functions, at least when widely diverse industries are included, seem very, very strong.

Massachusetts Institute of Technology

REFERENCES

- October x
October x
- [1] DIAMOND, P. A.: "Technical Change and the Measurement of Capital and Output," *Review of Economic Studies*, 32 (1965), 289-298.
- [2] FISHER, F. M.: "Embodied Technical Change and the Existence of an Aggregate Capital Stock," *Review of Economic Studies*, 32 (1965), 263-288.
- [3] ———: "Embodied Technology and the Existence of Labour and Output Aggregates," *Review of Economic Studies*, 35 (1968), 391-412.
- [4] ———: "Embodied Technology and the Aggregation of Fixed and Movable Capital Goods," *Review of Economic Studies*, 35 (1968), 417-428.
- [5] ———: "Approximate Aggregation and the Leontief Conditions," *Econometrica*, 37 (1969), 457-469.
- [6] GORMAN, W. M.: "Capital Aggregation in Vintage Models," in J. N. Wolfe (ed.), *Value, Capital, and Growth: Papers in Honour of Sir John Hicks*. Edinburgh: University of Edinburgh Press, 1968.
- [7] GREEN, H. A. J.: *Aggregation in Economic Analysis: an Introductory Survey*. Princeton: Princeton University Press, 1964.
- [8] HALL, R. E.: "Technical Change and Capital from the Point of View of the Dual," *Review of Economic Studies*, 35 (1968), 35-46.
- [9] HICKS, J. R.: *Value and Capital*. Oxford: Clarendon Press, 1946 (2nd edition).
- [10] JOHANSEN, L.: "Substitution versus Fixed Production Coefficients in the Theory of Economic Growth: A Synthesis," *Econometrica*, 27 (1959), 157-176.
- [11] KLEIN, L. R.: "Macroeconomics and the Theory of Rational Behaviour," *Econometrica*, 14 (1946), 93-108.
- [12] ———: "Remarks on the Theory of Aggregation," *Econometrica*, 14 (1946), 303-312.
- Jan x
x
- [13] LBONTIEF, W. W.: "Composite Commodities and the Problem of Index Numbers," *Econometrica*, 4 (1936), 39-59.
- [14] ———: "A Note on the Interrelation of Subsets of Independent Variables of a Continuous Function with Continuous First Derivatives," *Bulletin of the American Mathematical Society*, 53 (1947), 343-350.
- October x
- [15] ———: "Introduction to a Theory of the Internal Structure of Functional Relationships," *Econometrica*, 15 (1947), 361-373.
- [16] MAY, K. O.: "The Aggregation Problem for a One-Industry Model," *Econometrica*, 14 (1946), 285-288.
- [17] ———: "Technological Change and Aggregation," *Econometrica*, 15 (1947), 51-63.
- April x
- [18] NATAF, A.: "Sur la Possibilité de Construction de Certains Macromodèles," *Econometrica*, 16 (1948), 232-244.

- [19] ———: "Possibilités d'Agrégation de Fonctions de Production à Variables Capital et Main-d'Oeuvre," *Review of Economic Studies*, 34 (1967), 219–226.
- [20] PU, SHOU-SHAN: "A Note on Macroeconomics," *Econometrica*, 14 (1946), 299–302.
- [21] ROBINSON, J.: "The Production Function and the Theory of Capital," *Review of Economic Studies*, 21 (1953), 81–106.
- [22] SOLOW, R. M.: "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39 (1957), 312–320.
- [23] ———: "Investment and Technical Progress," Ch. 7 in K. J. Arrow, S. Karlin, and P. Suppes (eds.), *Mathematical Methods in the Social Sciences, 1959*. Stanford: Stanford University Press, 1960.
- [24] ———: "Capital, Labor, and Income in Manufacturing," in National Bureau of Economic Research Conference on Research in Income and Wealth, *The Behavior of Income Shares* (Studies in Income and Wealth, 26). Princeton: Princeton University Press, 1964, 101–128.
- [25] STIGUM, B. P.: "On Certain Problems of Aggregation," *International Economic Review*, 8 (1967), 349–367.
- [26] ———: "On a Property of Concave Functions," *Review of Economic Studies*, 35 (1968), 413–416.
- [27] WHITAKER, J. K.: "Vintage Capital Models and Econometric Production Functions," *Review of Economic Studies*, 33 (1966), 1–18.
- [28] ———: "Capital Aggregation and Optimality Conditions," *Review of Economic Studies*, 35 (1968), 429–442.