Volume Title: The Measurement of Capital

Volume Author/Editor: Dan Usher, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-84300-9

Volume URL: http://www.nber.org/books/ushe80-1

Publication Date: 1980

Chapter Title: Aggregation Problems in the Measurement of Capital

Chapter Author: W. Erwin Diewert

Chapter URL: http://www.nber.org/chapters/c7171

Chapter pages in book: (p. 433 - 538)

8 Aggregation Problems in the Measurement of Capital

W. E. Diewert

8.1 Introduction and Overview

In his Introduction to this volume, Dan Usher has provided us with a rather comprehensive discussion on the purposes of capital measurement as well as the problems of defining capital in the context of a specific purpose.

With Usher's introduction in mind, the scope of the present paper can readily be defined: I will concentrate on the problems of defining and measuring capital in the context of estimating production functions and measuring total factor productivity, with particular emphasis on the associated index number problems.

However, before discussing the special problems involved in aggregating capital, I will first discuss the general problem of aggregating over goods in section 8.2 and the general problem of aggregating over sectors in section 8.3. The material presented in these sections is for the most part not new, although much of it is fairly recent and not widely known. Usher's new definition of real capital is discussed in section 8.2.6 along with some other definitions.

In section 8.4 I discuss some of the aggregation problems that are specifically associated with capital. In particular, the problem of defining capital as an instantaneous stock or a service flow is discussed along with the concomitant problems of measuring depreciation.

In section 8.5 I present some new material on the measurement of total factor productivity and technical progress. In this section, capital

W. E. Diewert is associated with the Department of Economics at the University of British Columbia, Vancouver.

The author is indebted to E. R. Berndt, C. Blackorby, P. Chinloy, M. Denny, L. J. Lau, D. Usher, and F. Wykoff for helpful comments.

does not play a more important role than any other factor of production, so that one could question its inclusion in a paper that is supposed to be restricted to capital aggregation problems. However, past discussions of technical change have emphasized the possibility that technical change may be embodied in new capital goods (see Jorgenson 1966), and thus I decided to include section 8.5.

One of the most difficult problems in the measurement of capital is the problem of new goods. This is of course not specific to capital, and so in section 8.6 I present some suggestions for solving the new goods problem in general.

In section 8.7 I briefly consider a problem that occurs when measuring capital as well as other inputs and outputs: the problem of aggregating over time; that is, How should "monthly" estimates of a capital good be constructed? or, given monthly estimates, How should we construct an annual estimate of the capital component?

In section 8.8 I conclude by making some concrete recommendations to national income accountants based on the material in the previous sections. Some mathematical proofs are contained in the Appendix.

8.2 Methods for Justifying Aggregation over Goods

8.2.1 Price Proportionality: Hicks's Aggregation Theorem

Hicks (1946, pp. 312–13) showed (in the context of twice-differentiable utility functions) that if the prices of a group of goods change in the same proportion, that group of goods behaves just as if it were a single commodity. This aggregation theorem and the homogeneous weak separability method (which will be discussed in the following section) are the two most general methods we have for justifying aggregation over goods. Alternative statements and proofs of Hicks's aggregation theorem in the consumer context can be found in Wold (1953, pp. 109–10), Gorman (1953, pp. 76–77), and Diewert (1978*a*).

Versions of Hicks's aggregation theorem also exist in the producer context, particularly in the context of measuring real value added (see Khang 1971; Bruno 1978; Diewert 1978*a*). Below, I sketch yet another version of Hicks's aggregation theorem in the producer context, a version that does not make use of any restrictive differentiability assumptions.

Suppose there are $N + M$ goods that a given firm can produce or use as an input and that the set of feasible input-output combinations of goods is a set $S \equiv \{(x,y)\} \equiv \{(x_1,x_2, \ldots ,x_N,y_1,y_2, \ldots ,y_M)\}$, where x_n represents the quantity of good n produced (used as an input if $x_n < 0$) and y_m represents the quantity of good $N + m$ produced by the firm (used as an input if $y_m < 0$). We assume that the firm can buy or sell the first N goods at the positive prices $(w_1,w_2,\ldots,w_N) \equiv w \gg 0_N$[1] and the last M goods at the positive prices $(p_1,p_2,\ldots,p_M) \equiv p \gg 0_M$. We assume that the firm behaves competitively and attempts to solve the following *microeconomic profit-maximization problem*:

$$(1) \qquad \max_{x,y}\{w \cdot x + p \cdot y : (x,y) \in S\} \equiv \Pi(w,p).$$

The solution to the above profit-maximization problem (if one exists)[2] is a function of the prices w,p that the producer is facing and is called the *(micro) profit function* Π. It can be shown (see McFadden 1978 or Diewert 1973*a*) that under suitable regularity conditions on the technology S, the profit function completely characterizes the underlying technology. This duality property will prove very useful in subsequent sections of this chapter.

The firm's *gross*[3] or *restricted*[4] or *variable*[5] profit function Π^* is defined as

$$(2) \qquad \Pi^*(w,y) \equiv \max_x\{w \cdot x : (x,y) \in S\}.$$

The usual interpretation of the maximization problem (2) is that the firm is maximizing only with respect to its variable inputs and output x, while the inputs (and or outputs) y remain fixed in the short run. It can also be shown under suitable regularity conditions on S that a knowledge of the variable profit function Π^* is sufficient to completely determine the underlying technology S.[6] Thus, we will use the variable profit function to define an aggregate technology.

Suppose the prices of the first N goods vary in strict proportion; that is,

$$(3) \qquad (w_1,w_2,\ldots,w_N) = (p_0\alpha_1, p_0\alpha_2,\ldots,p_0\alpha_N),$$

where $p_0 > 0$ is a scalar that varies over time while the proportionality constants $(\alpha_1,\alpha_2,\ldots,\alpha_N) \equiv \alpha \gg 0_N$ remain fixed over time.

We can now define a *macro technology set* S_α using the variable profit function Π^* and the vector of constants α as follows:

$$(4) \qquad S_\alpha \equiv \{(y_0,y) : y_0 \leq \Pi^*(\alpha,y), \text{ where } y \text{ is such that there exists an } x \text{ such that } (x,y) \in S\}.$$

We will see that y_0 can be interpreted as an aggregate of the components of x; that is, $y_0 = w \cdot x / p_0$. It is easy to show that the macro technology set S_α inherits many of the properties of the micro technology set S. For example, if S is a convex set[7] (which is a generalization of the Hicksian [1946, p. 81] diminishing marginal rates of transformation regularity conditions on S), then S_α is also a convex set.[8] Moreover, if S exhibits constant returns to scale,[9] then S_α also exhibits constant returns to scale.[10]

Given that the macro technology set S_α has been defined, we may now define the *macro profit maximizations problem*:

$$(5) \qquad \max_{y_0, y}\{p_0 y_0 + p \cdot y : (y_0, y) \epsilon S\} \equiv \Pi(p_0, p).$$

The following theorem shows that if the price proportionality assumption in (3) is satisfied, then the macro profit maximization problem (5) is completely consistent with the underlying "true" micro profit maximization problem (1).

(6) *Thorem*: If (x^*, y^*) is a solution to the micro profit maximization problem (1) and the price proportionality assumption (3) holds, then (y^*_0, y^*) is a solution to the macro profit maximization (5), where the aggregate y^*_0 is defined by

$$(7) \qquad y^*_0 \equiv w \cdot x^* / p_0.$$

Note that, if the vector of constants α is known, the aggregate y^*_0 can be calculated from observable price and quantity data.

The theorem above shows that, if the factors of proportionality $(\alpha_1, \alpha_2, \ldots, \alpha_N) \equiv \alpha$ remain constant over time, then the true micro technology S can be replaced by the macro technology S_α. However, in most practical situations, α will not remain constant over time, though it may be *approximately* constant, in which case the set S_α will be *approximately* constant also, and this approximate constancy may suffice for empirical work. Perhaps a concrete example would make this point clearer.

Suppose the technology of the firm can be represented by a *translog variable profit function*[11] Π^* which is defined by the following equation:

$$(8) \qquad \ln \Pi^*(w^r, y^r) \equiv \beta_0 + \sum_{i=1}^{N} \beta_i \ln w^r_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \beta_{ih} \ln w^r_i \ln w^r_h + \sum_{i=1}^{N} \sum_{j=1}^{M} \gamma_{ij} \ln w^r_i \ln y^r_j + \sum_{j=1}^{M} \delta_j \ln y^r_j + \tfrac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \delta_{jk} \ln y^r_j \ln y^r_k,$$

where $w^r \equiv (w^r_1, w^r_2, \ldots, w^r_N) \rangle\rangle 0_N$ is the vector of prices for the first N goods in period r where $r = 1, 2, \ldots, T$ ($x^r \equiv (x^r_1, x^r_2, \ldots, x^r_N)$ is the corresponding quantity vector) and $y^r \equiv (y^r_1, y^r_2, \ldots, y^r_M) \rangle\rangle 0_M$ is a vector of purchases and sales of the last M goods in period r ($p^r \equiv (p^r_1, p^r_2, \ldots, p^r_M)$ is the corresponding vector of prices). Because the logarithm of a negative number is not defined, we have temporarily changed our sign convention and made all components of y^r positive whether the corresponding goods are outputs or inputs. Thus if the $N + M$th good during period r is an output, $y^r_m > 0$ and $p^r_m > 0$; however, if the $N+m$ th good during period r is an input, set $y^r_m > 0$ equal to the absolute value of the amount of input and set $p^r_m < 0$ equal to minus the input price.

The technological parameters β_{ih} in (8) satisfy the symmetry restrictions $\beta_{ih} = \beta_{hi}$ for $1 \leq i < h \leq N$, and the δ_{jk} satisfy the restrictions $\delta_{jk} = \delta_{kj}$ for $1 \leq j < k \leq M$. In order that the translog variable profit function $\Pi^*(w,y)$ be linearly homogeneous in w, the following restrictions on the parameters in (8) must be satisfied:

$$(9) \qquad \sum_{i=1}^{N} \beta_i = 1; \ \sum_{h=1}^{N} \beta_{ih} = 0 \text{ for } i = 1,2,\ldots,N; \ \sum_{i=1}^{N} \gamma_{ij} = 0 \text{ for } j = 1,2,\ldots,M.$$

Now let the w^r prices vary approximately proportionately over time; that is,

$$(10) \qquad w^r \equiv (w^r_1, w^r_2, \ldots, w^r_N) = (p^r_0\alpha_1 e^{\epsilon^r_1}, p^r_0\alpha_2 e^{\epsilon^r_2}, \ldots, p^r_0\alpha_N e^{\epsilon^r_N}),$$

where p^r_0 represents the general level of prices of the first N goods, the α_i represent fixed factors of proportionality, and the ϵ^r_i represent perturbations in these fixed factors of proportionality.

If we deflate $\Pi^*(w^r,y^r) = w^r \cdot x^r$ by $p_0{}^r$ (which converts a nominal value added into a "real" value added), then (10) and (8) yield

$$(11) \qquad \ln w^r \cdot x^r / p^r_0 = \beta_0 + \sum_{i=1}^{N} \beta_i \ln \alpha_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \beta_{ih} \ln \alpha_i \ln \alpha_h + \sum_{j=1}^{M} \delta_j \ln y^r_j + \sum_{i=1}^{N} \sum_{j=1}^{M} \gamma_{ij} \ln \alpha_i \ln y^r_j + \tfrac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \delta_{jk} \ln y^r_j \ln y^r_k + \epsilon^r,$$

where the Hicks's aggregate approximation error ϵ^r in period r is defined as

$$(12) \qquad \epsilon^r \equiv \sum_{i=1}^{N} \left[\beta_i + \sum_{h=1}^{N} \beta_{ih} \ln \alpha_h + \sum_{j=1}^{N} \gamma_{ij} \ln y^r_j \right] \epsilon^r_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \beta_{ih} \epsilon^r_i \epsilon^r_h.$$

Note that the left-hand side of (11) is observable (if p^r_0 is known), while the right-hand side is a conventional translog production function in the quantities y^r if the error term ϵ^r is neglected. Note how shifts in the price proportionality constants $\alpha \equiv (\alpha_1, \alpha_2, \ldots, \alpha_N)$ will systematically shift this translog production function.

If the perturbations ϵ^r_i defined above are such that $E\epsilon^r_i = 0$ and $E\epsilon^r_i\epsilon^t_h = \delta_{rt}\sigma_{ih}$ for $i,h = 1,2,\ldots,N$ and $r,s = 1,2,\ldots,T$, where E denotes the expectation operator and δ_{rt} equals 0 if $r \neq t$ but $\delta_{rt} = 1$ if $r = t$, then it is trivial to show that the error ϵ^r will have a constant bias[12] that will be absorbed into the constant if regression techniques are used in order to estimate the parameters of (11). Thus, with the above stochastic assumptions,[13] the production function that corresponds to the macro technology set S_α could be unbiasedly estimated up to a scaling factor, provided that the underlying technology S could be adequately approximated by the translog variable profit function Π^* defined by (8). This last proviso will be satisfied for moderate variations in prices and quantities, since the translog variable profit function can provide a second-order approximation to an arbitrary twice-differentiable variable profit function that in turn provides a complete description of the underlying technology S under suitable regularity conditions.

Thus it appears that the assumption of approximate price proportionality provides a rather powerful justification for aggregating over commodities.

Note that the aggregation method studied in this section did not restrict the technology in any essential way; rather, the set of prices that producers faced was restricted. In the following section, a method of aggregating over commodities is outlined that depends on the technology's satisfying certain restrictive assumptions.

8.2.2 Homogeneous Weak Separability

The second major method justifying commodity aggregation is due to Leontief (1947) and Shephard (1953, pp. 61–71; 1970, pp. 145–46; see also Solow 1955–56; Green 1964; Arrow 1974; Geary and Morishima 1973), and I will outline this method below. To cover both producer and consumer theory applications of this method of aggregation, we assume cost minimizing (instead of profit maximizing) behavior on the part of producers.

Suppose the microeconomic production (or utility) function f^* where $u = f^*(x,z)$ is output (or utility), $x \geq 0_N$ is a nonnegative N-dimensional vector of commodity inputs to be aggregated, and $z \geq 0_M$ is an M-dimensional vector of "other" commodity inputs. The producer's (or consumer's) total minimum cost function is defined as:

$$(13) \qquad C^*(u;p,w) \equiv \min_{x,z}\{p \cdot x + w \cdot z\colon f^*(x,z) \geq u\},$$

where $p \equiv (p_1,p_2,\ldots,p_N) \gg 0_N$ is a vector of positive input prices and $w \equiv (w_1,w_2,\ldots,w_M) \gg 0_M$ is a vector of "other" input prices. The Shephard (1953, 1970) duality theorem (see also Samuelson 1953–54; Uzawa 1964; McFadden 1978; Hanoch 1978; and Diewert 1971) states that, under certain regularity conditions, the total cost function C^* completely determines the production function f^*. Thus restrictions on the production function f^* translate into restrictions on the cost function C^* and vice versa. We will make use of this fact below.

To justify the aggregation of the commodities x, Shephard assumed that x was *homogeneously weakly separable*[14] from the "other" commodities z; that is, he assumed that the micro function f^* could be written as

$$f^*(x,z) = \hat{f}[f(x),z], \tag{14}$$

where $\hat{f}$ is a *macro* production (or utility) function (satisfying the same regularity conditions as f^*) and f is an *aggregator* function that is assumed to satisfy the following regularity conditions.

(15) *Conditions on* f:
i) f is defined for $x \rangle\rangle 0_N$ and $f(x) > 0$ (*positivity*);
ii) $f(\lambda x) = \lambda f(x)$ for $\lambda > 0$, $x \rangle\rangle 0_N$ (*linear homogeneity*);
iii) $f(\lambda x^1 + (1-\lambda)x^2) \geq \lambda f(x^1) + (1-\lambda)f(x^2)$ for $0 \leq \lambda \leq 1$, $x^1 \rangle\rangle 0_N$, $x^2 \rangle\rangle 0_N$ (*concavity*).

The macro function $\hat{f}$ has a cost function dual defined by

$$\hat{C}(u; p_0,w) \equiv \min_{y,z}\{p_0 y + w \cdot z: \hat{f}(y,z) \geq u\}, \tag{16}$$

where $p_0 > 0$ is the price of the aggregate, y. The aggregator function f also has a total cost function dual defined by

$$\begin{aligned} C(y; p) &\equiv \min_x\{p \cdot x: f(x) \geq y\} \\ &= y \min_{x/y}\{p \cdot x/y: f(x/y) \geq 1\} \text{ using (15.ii)} \\ &\equiv yc(p). \end{aligned} \tag{17}$$

It turns out that the unit cost function $c(p)$ satisfies the same regularity conditions as f; that is, $c(p)$ is positive, linearly homogeneous, and concave for $p \rangle\rangle 0_N$ (see Samuelson 1953–54; Diewert 1974*b*). Moreover, given a unit-cost function $c(p)$ satisfying the conditions (15), the production function dual may be defined as[15]

$$f(x) \equiv 1/max_p\{c(p): p \cdot x = 1, p \geq 0_N\}. \tag{18}$$

With the above preliminaries disposed of, we can now outline the Shephard-Solow-Arrow results. Suppose $p^* \cdot x^* + w^* \cdot z^* = C^*(u^*; p^*,w^*)$; that is, x^*,z^* is a solution to the micro cost minimization problem (13) when micro prices p^*,w^* (and utility or output u^*) prevail. If the micro function f^* is homogenously weakly separable (i.e., f^* satisfies eq. 14) and *if the functional form for the aggregator function f* *is known* (or the functional form for its unit-cost function $c(p)$ is known), then the aggregate y^* can be defined as

$$y^* \equiv f(x^*) \quad (\text{or } y^* \equiv p^* \cdot x^*/c(p^*)), \tag{19}$$

and the price of the aggregate may be defined as

$$p^*_0 \equiv p^* \cdot x^*/f(x^*) \quad (\text{or } p^*_0 \equiv c(p^*)), \tag{20}$$

and $p^*_0 y^* + w^* \cdot z^* = \hat{C}(u^*; p^*_0, w^*)$; that is, y^*, z^* is a solution to the macro cost minimization problem (16) with prices p^*_0, w^* (and utility or output u^*).[16]

There are, of course, at least two problems with this aggregation method: (*a*) the micro function f^* may not be homogeneously weakly separable in practice, and (*b*) the functional form for the aggregator function f is generally unknown. In the remainder of this section, our attention will be directed toward solving the second difficulty.

Let $p^r \gg 0_N$, $w^r \gg 0_M$ for $r = 0,1, \ldots, T$. If x^r, z^r is a solution to $\min_{x,z}\{p^r \cdot x + w^r \cdot z\colon\ \hat{f}[f(x),z] \geq u^r\}$ and if $\hat{f}$ is increasing in its first argument, it is easy to see that x^r must be a solution to the following *aggregator maximization problem.*

$$\max_x\{f(x)\colon\ p^r \cdot x \leq p^r \cdot x^r;\ x \geq 0_N\} \qquad r = 0,1, \ldots, T. \tag{21}$$

In other words, if an economic agent wishes to minimize the cost of achieving a certain utility or output level when the micro function f^* is weakly separable (i.e., $f^*(x,z) = \hat{f}[f(x),z]$) and the macro function $\hat{f}$ is increasing in its first argument, then the "intermediate input" (or "real value added" or "category subutility") $f(x)$ must be a maximum subject to an expenditure constraint.

Notice that (21) involves only the (unknown) aggregator function f and observable prices and quantities, $\{p^r, x^r\}$ for $r = 0,1, \ldots, T$. If f is differentiable, then the first-order necessary conditions for a maximum yield the following identity after the Lagrange multiplier is eliminated:

(22) *Lemma*: (Konyus and Byushgens 1926, p. 155; Hotelling 1935, pp. 71–74; Wold 1944, pp. 69–71). Suppose f *is* differentiable and $x^r > 0_N$ is a solution to $\max_x\{f(x)\colon\ p^r \cdot x \leq p^r \cdot x^r,\ x \geq 0_N\}$, where $p^r \gg 0_N$. Then

$$\frac{p^r}{p^r \cdot x^r} = \frac{\nabla f(x^r)}{x^r \cdot \nabla f(x^r)},$$

where $\nabla f(x^r)$ is the vector of first-order partial derivatives evaluated at x^r.

Corollary:[17] If f is also homogeneous of degree one (i.e., $f(\lambda x) = \lambda f(x)$ for every $\lambda > 0$), then

(23) $$\frac{p^r}{p^r \cdot x^r} = \frac{\nabla f(x^r)}{f(x^r)}.$$

The above corollary suggests the following *Method I* (due to Arrow 1974) for determining the aggregator function f given the micro data $\{p^r,x^r\}$, $r = 0,1,2,\ldots,T$: simply assume a convenient functional form for f and use the relations (23) to econometrically estimate the unknown parameters. For example, suppose that f is the *homogeneous translog function* (Christensen, Jorgenson, and Lau 1971):

(24) $$\ln f(x^r) \equiv \beta_0 + \sum_{n=1}^{N} \beta_n \ln x^r_n + \tfrac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \beta_{jk} \ln x^r_j \ln x^r_k, \quad r = 0,1,\ldots,T,$$

where $$\sum_{n=1}^{N} \beta_n = 1,\ \beta_{jk} = \beta_{kj} \text{ and } \sum_{k=1}^{N} \beta_{jk} = 0 \text{ for } j = 1,2,\ldots,N.$$

With the above parameter restrictions, f turns out to be linearly homogeneous. Application of (23) yields the following system of equations that is linear in the unknown parameters:

(25) $$\frac{p^r_n}{p^r \cdot x^r} = \frac{\beta_n + \sum_{k=1}^{N} \beta_{nk} \ln x^r_k}{x^r_n}, \quad \begin{array}{l} n = 1,2,\ldots,N \\ r = 0,1,2,\ldots,T, \end{array}$$

where $p^r \equiv (p^r_1,p^r_2,\ldots,p^r_N)$ and $x^r \equiv (x^r_1,x^r_2,\ldots,x^r_N)$. Notice that the parameter β_0 is not identified, but once the other parameters are determined, an estimate for β_0 may be obtained by solving $f(x^0) = 1$ for β_0 (*base period normalization*). For an econometric application of this method in the context of estimating a real value-added production function, see Berndt and Christensen (1973).

Instead of econometrically estimating the parameters of the aggregator function f, we may attempt to estimate the parameters of its unit cost function, $c(p)$. In this context, the following result is useful.

(26) *Lemma*: (Shephard 1953, p. 11; Samuelson 1953–54) If f satisfies (15), $p^r \cdot x^r = \min_x\{p^r \cdot x: f(x) \geq f(x^r)\} = c(p^r)f(x^r)$ for $r = 0,1,\ldots T$, and the unit cost function c is differentiable at p^r, then

$$x^r = \nabla c(p^r) f(x^r), \quad r = 0,1,\ldots,T.$$

(27) *Corollary*:[18] $x^r/p^r \cdot x^r = \nabla c(p^r)/c(p^r)$, $r = 0,1,\ldots,T$.

The above corollary suggests the following *Method II* (also due to Arrow 1974) for determining the dual c to the aggregator function f

given the micro data $\{p^r,x^r\}$, $r = 1,2,\ldots,T$: assume a functional form for $c(p)$ and use the relations (27) to estimate the unknown parameters of $c(p)$. For example, suppose that c is the *translog unit cost function* (Christensen, Jorgenson, and Lau 1971):

$$\text{(28)} \qquad \ln c(p^r) = \gamma_0 + \sum_{n=1}^{N} \gamma_n \ln p^r_n + \tfrac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \gamma_{jk} \ln p^r_j \ln p^r_k, \quad r = 0,1,\ldots,T,$$

where $\sum_{n=1}^{N} \gamma_n = 1$, $\gamma_{jk} = \gamma_{kj}$ and $\sum_{k=1}^{N} \gamma_{jk} = 0$ for $j = 1,2,\ldots,N$.

Application of Shephard's lemma (27) yields the following system of equations that is linear in the unknown parameters:

$$\text{(29)} \qquad \frac{x^r_n}{p^r \cdot x^r} = \frac{\gamma_n + \sum_{k=1}^{N} \gamma_{nk} \ln p^r_k}{p^r_n}, \quad \begin{array}{l} n = 1,2,\ldots,N, \\ r = 0,1,2,\ldots,T. \end{array}$$

Notice that the parameter γ_0 is not identified, but once the other parameters are determined an estimate for γ_0 may be obtained by solving the equation $p^0 \cdot x^0 / c(p^0) = 1$ (*base period normalization*), which makes $f(x^0) = 1$.

Note that the translog unit cost function generates an aggregator function via (18) that does not in general[19] coincide with the translog aggregator function defined by (24). Thus, in general, the two translog functional forms correspond to *different* tastes or technologies, although either functional form can approximate the same underlying (differentiable) technology to the second order.

At this point I should mention *Method III* for determining the ratio of the aggregates, $f(x^r)/f(x^0)$. This final method involves assuming a functional form for the aggregator function f that is consistent with an index number formula (which is a function of observable prices and quantities for the two periods under consideration). The method assumes that x^r is a solution to the aggregator maximization problem defined by (21),[20] and it will be studied in greater detail in section 8.2.4. When reading section 8.2.4, recall that it was the assumption of expenditure minimizing behavior (which is consistent with profit maximizing behavior), plus the assumption that the technology was homogeneously weakly separable in the x goods that led us to conclude that x^r was a solution to the aggregator maximization problem $\max_x\{f(x): p^r \cdot x \leq p^r \cdot x^r\}$, where f is the linearly homogeneous aggregator function.

In the next section, I outline another method for aggregating over goods, a method due to François Divisia (1926).

8.2.3 The Divisia Index and Various Discrete Approximations

The most frequently suggested index to be used in the measurement of total factor productivity is the Divisia (1926, p. 40) index. Let us briefly outline Solow's (1957) derivation of the index.[21]

Suppose a linearly homogeneous, concave, nondecreasing in x production function F exists where $y(t) = F(x(t); t)$, $y(t)$ is output at time t, and $x(t) \equiv (x_1(t), x_2(t), \ldots, x_N(t))$ is a vector of inputs at time t. If the production function exhibits neutral technical change (see Blackorby, Lovell, and Thursby 1976 for a formal definition), then it can be written as $F(x(t); t) = A(t)f(x(t))$, where $A(t)$ is the cumulative multiplicative shift factor for the production function at time t. If we totally differentiate the following equation

$$y(t) = A(t)f(x(t)) \tag{30}$$

with respect to time and divide by $y(t)$, we obtain

$$\frac{\dot{y}(t)}{y(t)} = \frac{\dot{A}(t)}{A(t)} + A(t) \sum_{i=1}^{N} \frac{\partial f[x(t)]}{\partial x_i} \frac{\dot{x}_i(t)}{y(t)}, \tag{31}$$

where a dot over a variable signifies a derivative with respect to time. Let $p(t) \equiv (p_1(t), \ldots, p_N(t))$ be the vector of input prices at time t relative to the price of output, which is set equal to one. Then, if inputs are being paid the value of their marginal products, $A(t)\partial f[x(t)]/\partial x_i = p_i(t)$, and if we define the ith input's share of output as $s_i(t) \equiv p_i(t)x_i(t)/y(t)$, $i = 1,2, \ldots, N$, then (31) may be rewritten as

$$\frac{\dot{A}(t)}{A(t)} = \frac{\dot{y}(t)}{y(t)} - \sum_{i=1}^{N} s_i(t) \frac{\dot{x}_i(t)}{x_i(t)}. \tag{32}$$

If $\dot{A}(t)/A(t) = 0$, there is no exogenous shift in the production function owing to technical progress, increasing returns to scale, or any other cause; that is, the growth of output is completely accounted for by the growth of inputs.

We can integrate (32) (given continuous data on output, inputs, and prices) to obtain the cumulative index of total factor productivity from time $t = 0$ to time $t = T$:

$$\frac{A(T)}{A(0)} = \frac{y(T)/y(0)}{e^{\int_0^T \sum_{i=1}^{N} s_i(t)\, \dot{x}_i(t)/x_i(t)\, dt}}, \tag{33}$$

where the denominator on the right-hand side of (33) is the Divisia index of input growth between, say, time 0 and T, $X(T)/X(0)$.

Richter (1966) and Jorgenson and Griliches (1967) have generalized equations (32) and (33) by replacing the single output term

$\dot{y}(t)/y(t)$ in (32) with a share-weighted average of the growth rates of many outputs, and the term $y(T)/y(0)$ in (33) with a Divisia index of output growth.

Since the right-hand sides of (32) and (33) are in principle observable, the technical change term $A(t)$ can, in principle, be estimated.[22] But in practice data do not come in nice continuous series; rather they come at discrete intervals. Thus the continuous formulas (32) and (33) must be approximated using discrete data.

Let us now introduce some new notation that is appropriate when data come at discrete intervals. Let the vector of period r inputs be $x^r \equiv (x^r{}_1, x^r{}_2, \ldots, x^r{}_N)$ and period r prices be $p^r \equiv (p^r{}_1, p^r{}_2, \ldots, p^r{}_N)$ for $r = 0,1$.

Denote the denominator of (33) as $X(1)/X(0)$, when $T = 1$. If the input shares are approximately constant, then $\ln X(1)/X(0)$ approximately equals $\sum_{i=1}^{N} s_i \ln x^1{}_i/x^0{}_i$. For any number z close to 1, $\ln z$ can be accurately approximated by $-1 + z$, so that $X(1)/X(0)$ approximately equals $\sum_{i=1}^{N} s_i x^1{}_i/x^0{}_i$. Thus the Divisia index of input growth $X(1)/X(0)$ can be approximated by a share-weighted rate of growth of the quantity relatives $x^1{}_i/x^0{}_i$, $i = 1,2,\ldots,N$. If we choose base-period shares, the resulting index is the Laspeyres quantity index Q_L:

$$Q_L(p^0,p^1;x^0,x^1) \equiv \sum_{i=1}^{N} \frac{p^0{}_i x^0{}_i}{p^0 \cdot x^0} (x^1{}_i/x^0{}_i) = \frac{p^0 \cdot x^1}{p^0 \cdot x^0}, \tag{34}$$

where $p^0 \cdot x^r \equiv \sum_{i=1}^{N} p^0{}_i x^r{}_i$ denotes the inner product between the vectors p^0 and x^r, $r = 0,1$. On the other hand, if we choose current-period prices and base-period quantities to form shares, the resulting index is the Paasche quantity index Q_P:

$$Q_p(p^0,p^1;x^0,x^1) \equiv \sum_{i=1}^{N} \frac{p^1{}_i x^0{}_i}{p^1 \cdot x^0} (x^1{}_i/x^0{}_i) = \frac{p^1 \cdot x^1}{p^1 \cdot x^0}. \tag{35}$$

A third way of approximating the share-weighted rate of growth of inputs that apepars in (32) would be to take a geometric mean of the index Q_L and Q_P:

$$Q_2(p^0,p^1;x^0,x^1) \equiv (p^0 \cdot x^1 p^1 \cdot x^1 / p^0 \cdot x^0 p^1 \cdot x^0)^{1/2}. \tag{36}$$

The index Q_2 is Irving Fisher's (1922) ideal quantity index. The price index that corresponds to Q_2 is P_2 defined implicitly by Fisher's weak factor reversal test:

$$P_2(p^0,p^1;x^0,x^1)\, Q_2(p^0,p^1,x^0,x^1) = p^1 \cdot x^1/p^0 \cdot x^0; \tag{37}$$

that is, the product of the price index times the quantity index equals the expenditure ratio between the two periods. Fisher called P_2 and Q_2

ideal indexes because they satisfied (37) and also $P_2(p^0,p^1; x^0,x^1) \equiv Q_2(x^0,x^1; p^0,p^1)$; that is, the price and quantity indexes turn out to have the same functional form, except that the role of prices and quantities are reversed for the two indexes.

The integral expression for the Divisia index of inputs found in (33) suggests some further discrete approximations. If the input shares $s_i(t)$ remain constant between 0 and 1, then the log of the Divisia index becomes:

$$\ln e^{\int_0^1 \sum_{i=1}^{N} s_i \dot{x}_i(t)/x_i(t)\, dt} = \sum_{i=1}^{N} s_i \ln x^1_i/x^0_i.$$

Since the shares $s_i(t)$ are not generally the same for periods 0 and 1, Törnqvist (1936) suggested the following discrete approximation Q_0 to the continuous Divisia quantity index:

$$(38) \qquad \ln Q_0(p^0,p^1; x^0,x^1) \equiv \sum_{i=1}^{N} \tfrac{1}{2} \left[\frac{p^0_i x^0_i}{p^0 \cdot x^0} + \frac{p^1_i x^1_i}{p^1 \cdot x^1}\right] \cdot \ln[x^1_i/x^0_i].$$

Star and Hall (1976) develop an analytic expression for the difference between the discrete approximation Q_0 defined by (38) and the continuous Divisia index. They conclude that the approximation error will be small provided the shares do not fluctuate wildly.

The Törnqvist price index P_0 can be defined by the formula for Q_0 except that prices and quantities are interchanged; more explicitly:

$$(39) \qquad \ln P_0(p^0,p^1; x^0,x^1) \equiv \sum_{i=1}^{N} \left[\tfrac{1}{2} \frac{p^0_i x^0_i}{p^0 \cdot x^0} + \tfrac{1}{2} \frac{p^1_i x^1_i}{p^1 \cdot x^1}\right] \cdot \ln[p^1_i/p^0_i].$$

Given the price index P_0, an implicit Törnqvist quantity index $\tilde{Q}_0$ may be defined using Fisher's weak factor reversal test:

$$(40) \qquad \tilde{Q}_0(p^0,p^1; x^0,x^1) \equiv [p^1 \cdot x^1/p^0 \cdot x^0]/P_0(p^0,p^1; x^0,x^1).$$

Kloek (1967) and Theil (1968) showed that the Törnqvist indexes Q_0, P_0, and $\tilde{Q}_0$ had some good approximation properties. Kloek noted that Q_0 was not well defined if some quantities were zero, while P_0 was not well defined if some prices were zero. Thus he advocated using the price index P_0 and the quantity index $\tilde{Q}_0$, since prices are usually nonzero. I will return to this problem of zero prices and quantities in section 8.6.2.

We have now defined five reasonable-looking discrete approximations to the Divisia quantity index. The problem is that the theory of Divisia indexes outlined above does not tell us which discrete index number formula should be used in empirical applications, even though it is known that the Laspeyres and Paasche quantity indexes can differ considerably from the other indexes.[23]

It turns out that the economic theory of exact index numbers[24] enables us to discriminate more sharply among the above index number formulas. I will briefly outline this theory.

8.2.4 Exact and Superlative Index Number Formulas

Suppose the production function (or aggregator function) is $y = f(x_1, x_2, \ldots, x_N) = f(x)$, where y is output (or the aggregate), x is a vector of inputs (or goods to be aggregated), and f is a nondecreasing, linearly homogeneous and concave function. Suppose further that, given a positive vector of input prices $p \equiv (p_1, p_2, \ldots, p_N)$, the producer attempts to minimize the cost of producing a given output level. The solution to the cost minimization problem is the total cost function $C(y;p)$, which decomposes into a unit-cost function $c(p)$ times the output level owing to the linear homogeneity of the aggregator function f;[25] that is,

$$(41) \qquad C(y;p) \equiv \min_x \{p \cdot x \colon \ f(x) = y\} = c(p)y.$$

It is natural to identify $c(p)$ with the price of output; that is, as being the price of the aggregate good y.

Suppose we are given price and quantity data for two periods, p^0, p^1, x^0, x^1. Define a *price index* simply as a function P of prices and quantities, $P(p^0, p^1; x^0, x^1)$, while a *quantity index* $Q(p^0, p^1; x^0, x^1)$ is another function of prices and quantities for the two periods. We generally assume that the price and quantity indexes satisfy Fisher's weak factor reversal test; that is, P and Q satisfy

$$(42) \qquad P(p^0, p^1; x^0, x^1)\, Q(p^0, p^1; x^0, x^1) = p^1 \cdot x^1 / p^0 \cdot x^0.$$

A given functional form for a quantity index Q is defined to be *exact* for a functional form for the aggregator function f if given output levels y^0, y^1, input price vectors $p^0, p^1; x^0$ a solution to the period 0 cost minimization problem (41) and x^1 a solution to the period 1 cost minimization problem (41), then

$$(43) \qquad f(x^1)/f(x^0) = Q(p^0, p^1; x^0, x^1)$$

for all $y^0 > 0$, $y^1 > 0$, $p^0 \gg 0_N$, $p^1 \gg 0_N$.[26] Similarly, a given functional form for a price index P is defined to be *exact* for a functional form for the aggregator function f (and its derived unit cost function c) if given output levels y^0, y^1, input price vectors $p^0, p^1; x^r$ a solution to the period r cost minimization problem (41) for $r = 0,1$, then

$$(44) \qquad c(p^1)/c(p^0) = P(p^0,p^1;x^0,x^1).$$

Thus the quantity index Q equals the ratio of the "outputs" y^1/y^0, and the price index P equals the ratio of the unit costs (or the ratio of the "prices" of the "outputs") $c(p^1)/c(p^0)$, provided Q and P are exact for some f. Note that for $r = 0,1$,

$$C(y^r;p^r) \equiv \min_x\{p^r\cdot x: f(x) = y^r\} = p^r\cdot x^r = c(p^r)y^r,$$

and, using (43) and (44),

$$\begin{aligned} P(p^0,p^1;x^0,x^1)\,Q(p^0,p^1;x^0,x^1) &= [c(p^1)/c(p^0)]\,[f(x^1)/f(x^0)] \\ &= c(p^1)y^1 \,/\, c(p^0)y^0 \\ &= p^1\cdot x^1/p^0\cdot x^0, \end{aligned}$$

so that exact price and quantity indexes satisfy the weak factor reversal test (43).[27]

With the above theoretical considerations disposed of, we can now return to the problem of evaluating the five alternative discrete approximations to the Divisia quantity index. It seems that we could define two other discrete approximations to the Divisia quantity index by defining the Laspeyres and Paasche price indexes analogously to the Laspeyres and Paasche quantity indexes (defined by equations 34 and 35), except that the roles of prices and quantities are reversed and then the implicit Laspeyres and Paasche quantity indexes, Q_L and Q_P, may be defined by the weak factor reversal test (42):

$$\begin{aligned} (45) \qquad \tilde{Q}_L(p^0,p^1;x^0,x^1) &\equiv [p^1\cdot x^1/p^0\cdot x^0]/P_L(p^0,p^1;x^0,x^1) \\ &= p^1\cdot x^1/p^1\cdot x^0 \\ &\equiv Q_P(p^0,p^1;x^0,x^1), \\ \tilde{Q}_P(p^0,p^1;x^0,x^1) &\equiv [p^1\cdot x^1/p^0\cdot x^0]/P_P(p^0,p^1;x^0x^1) \\ &= p^0\cdot x^1/p^0\cdot x^0 \\ &\equiv Q_L(p^0,p^1;x^0,x^1). \end{aligned}$$

Thus the quantity index that corresponds to the Laspeyres price index is the Paasche quantity index, and the Paasche price index corresponds to the Laspeyres quantity index.

Konyus and Byushgens (1926) have shown that:[28] (a) the Laspeyres and Paasche quantity indexes are exact for a fixed coefficients (or Leontief) aggregator function of the form $f_L(x_1,x_2,\ldots,x_N \equiv \min\{x_i/a_i: i = 1,2,\ldots,N\}$, where the $a_i > 0$ are fixed coefficients, and (b) Fisher's ideal quantity index Q_2 defined by (36) is exact for a *homogeneous quadratic aggregator function* of the form $f_2(x_1,x_2,\ldots,x_N) \equiv$

$\left(\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}x_ix_j\right)^{1/2}$, where $a_{ij} = a_{ji}$ and the matrix of coefficients $[a_{ij}]$ is such that f_2 is concave and nondecreasing over the relevant range of quantities. Thus, under the assumption of cost-minimizing behavior and if the aggregator function is the homogeneous quadratic defined above, then we have

$$y^1/y^0 = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}x^1{}_ix^1{}_j\right)^{1/2} \Big/ \left(\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}x^0{}_ix^0{}_j\right)^{1/2}$$
$$\equiv Q_2(p^0,p^1; x^0,x^1).$$

Thus we can calculate the "output" ratio y^1/y^0 by calculating Q_2, *which can be evaluated without knowing what the a_{ij} coefficients are.*

On the other hand, Diewert (1976) has shown that the Törnqvist quantity index Q_0 defined by (38) is exact for the *homogeneous translog aggregator function f*, defined by

$$\ln f_0(x_1,x_2,\ldots,x_N) \equiv \beta_0 + \sum_{i=1}^{N} \beta_i \ln x_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} \ln x_i \ln x_j,$$

where the parameters β_i and $\beta_{ij} = \beta_{ji}$ are such that $f_0(x)$ is concave, nondecreasing, and linearly homogeneous over the relevant range of xs. In order that f_0 be linearly homogeneous, it is necessary and sufficient that the following restrictions be satisfied:

$$(46) \qquad \sum_{i=1}^{N} \beta_i = 1; \ \sum_{j=1}^{N} \beta_{ij} = 0 \text{ for } i = 1,2,\ldots,N;$$
$$\sum_{i=1}^{N} \beta_{ij} = 0 \text{ for } j = 1,2,\ldots,N.$$

Thus the homogeneous translog aggregator function has exactly the same number of independent parameters as the homogeneous quadratic aggregator function defined earlier, namely $N(N+1)/2$ independent parameters. Moreover, it turns out that both aggregator functions are capable of providing a second-order differential[29] approximation to an arbitrary twice continuously differentiable, linearly homogeneous function.

It was also shown in Diewert (1976) that the implicit Törnqvist quantity index $\tilde{Q}_0$ defined by equation (40) above is exact for the aggregator function $\tilde{f}_0$ that has as its dual the translog unit cost function $c_0(p)$ defined by

$$\ln c_0(p_1,p_2,\ldots,p_N) = \beta^*{}_0 + \sum_{i=1}^{N} \beta^*{}_i \ln p_i$$
$$+ \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta^*{}_{ij} \ln p_i \ln p_j,$$

where $\sum_{i=1}^{N} \beta^*_i = 1, \beta^*_{ij} = \beta^*_{ji}$, and

$$\sum_{i=1}^{N} \beta^*_{ij} = 0 \text{ for } j = 1,2, \ldots ,N.$$

(The restrictions $\sum_{j=1}^{N} \beta^*_{ij} = 0$ for $i = 1,2, \ldots ,N$ follow from the symmetry restrictions $\beta^*_{ij} = \beta^*_{ji}$.) The translog unit cost function can provide a second-order differential approximation to an arbitrary twice continuously differentiable unit cost function, which in turn is capable of completely describing the corresponding linearly homogeneous aggregator function.

The fixed coefficients aggregator function has a linear unit cost function (equal to $\sum_{i=1}^{N} a_i p_i$) which can provide only a first-order approximation to an arbitrary twice-differentiable unit-cost function. Thus the Törnqvist price index P_0 should be preferred to the Laspeyres and Paasche price indexes P_L and P_P, respectively.

Thus the economic theory of exact index numbers has enabled us to discriminate somewhat between the five discrete approximations to the Divisia quantity index that we have considered: the indexes Q_2, Q_0, and $\tilde{Q}_0$ are to be preferred to Q_L and Q_P, since the former are exact for functional forms for the underlying aggregator function (or its dual unit cost function) that are more flexible than the very restrictive fixed coefficients aggregator function.

That the indexes Q_2, Q_0, and $\tilde{Q}_0$ are approximately equivalent can be demonstrated in another way (which does not depend on the assumption that the producer attempted to minimize the cost of producing the aggregate during the two periods). Diewert (1978*b*) showed that when prices in the two periods are equal (i.e., $p^0 = p^1 = p$) and quantities are also equal (i.e., $x^0 = x^1 = x$), then

$$(47) \qquad \begin{aligned} &Q_2(p,p;x,x) = Q_0(p,p;x,x) = \tilde{Q}_0(p,p;x,x) \\ &\nabla Q_2(p,p;x,x) = \nabla Q_0(p,p;x,x) = \\ &\nabla \tilde{Q}_0(p,p;x,x) \text{ and} \\ &\nabla^2 Q_2(p,p;x,x) = \nabla^2 Q_0(p,p;x,x) = \\ &\nabla^2 \tilde{Q}_0(p,p;x,x); \end{aligned}$$

that is, the three "better" quantity indexes differentially approximate each other to the second order at any point where the two price vectors are equal and the two quantity vectors are equal.[30] Thus for small changes in prices and quantities between the two periods, the three indexes should give the same answer to the second order.[31] The equalities in (47) can be derived simply by evaluating and differentiating the appropriate index number formula—no assumptions about minimizing behavior are required.

Diewert (1976) defined a price index (quantity index) to be *superlative* if it is exact for a unit cost function c (aggregator function f) capable of providing a second-order differential approximation to an arbitrary twice-differentiable linearly homogeneous function. Since a linearly homogeneous translog function can provide a second-order approximation to an arbitrary twice-differentiable linearly homogeneous function (see Lau 1974), it can be seen that P_0 defined by (39) is a superlative price index and Q_0 defined by (38) is a superlative quantity index. In general, "superlative" indexes are exact for "flexible" functional forms for the underlying aggregator function.

It is easy to show that the three "better" quantity indexes are superlative indexes. Are there any other superlative indexes? The answer is yes, as the following examples show.

For $r \neq 0$, define the quadratic mean of order r price index P_r as

$$(48) \qquad P_r(p^0,p^1;x^0,x^1) \equiv \left[\frac{\sum_{k=1}^{N} (p^1{}_k x^1{}_k / p^1 \cdot x^1)(p^0{}_k / p^1{}_k)^{r/2}}{\sum_{i=1}^{N} (p^0{}_i x^0{}_i / p^0 \cdot x^0)(p^1{}_i / p^0{}_i)^{r/2}} \right]^{1/r}.$$

It can be shown (Diewert 1976) that P_r is exact for the quadratic mean of order r unit cost function,

$$c_r(p) \equiv \left(\sum_{i=1}^{N} \sum_{j=1}^{N} b_{ij} p_i^{r/2} p_j^{r/2} \right)^{1/r}.$$

Since c_r can approximate an arbitrary unit cost function to the second order, P_r is a superlative price index.

For $r \neq 0$, define the quadratic mean of order r quantity index Q_r as

$$(49) \qquad Q_r(p^0,p^1;x^0,x^1) \equiv \left[\frac{\sum_{i=1}^{N} (p^0{}_i x^0{}_i / p^0 \cdot x^0)(x^1{}_i / x^0{}_i)^{r/2}}{\sum_{j=1}^{N} (p^1{}_j x^1{}_j / p^1 \cdot x^1)(x^0{}_j / x^1{}_j)^{r/2}} \right]^{1/r}.$$

It can similarly be shown that Q_r is exact for the quadratic mean of order r aggregator function,[32]

$$f_r(x) \equiv \left(\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i^{r/2} x_j^{r/2} \right)^{1/r}, \text{ and}$$

that Q_r is a superlative quantity index.

The reason for our notation P_2 and Q_2 for the Fisher ideal price and quantity indexes should now be evident: they are special cases of (48) and (49) when $r = 2$.

We can now also explain our reason for choosing the notation P_0 and Q_0 for the Törnqvist price and quantity indexes: it can be shown that $f_r(x)$ tends to $f_0(x)$, the homogeneous translog functional form, as r tends to zero under certain conditions, which we explain below. Thus Q_0 is in some sense a limiting case of Q_r.

It is no loss of generality to choose units of measurement for "output" y so that $\Sigma_i\Sigma_j a_{ij} = 1$. Let us further redefine the a_{ij} as $a_{ii} = \beta_i + 2\beta_{ii}r^{-1}$, and $a_{ij} \equiv 2\beta_{ij}r^{-1}$ for $i \neq j$, where $\Sigma_i\beta_i = 1$, $\beta_{ij} = \beta_{ji}$ *and* $\Sigma_i\beta_{ij} = 0$ for $j = 1,2,\ldots,N$. Then the equation that defines the quadratic mean of order r becomes

$$y = \left[\sum_{i=1}^{N} \beta_i x^r{}_i + 2r^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} x^{r/2}{}_i x^{r/2}{}_j \right]^{1/r}.$$

Now raise each side of this equation to the power r, subtract 1 from each side of the resulting equation, divide both sides by r, and upon making use of the restrictions on the β_i and β_{ij}, we may write the result as

$$\frac{y^r - 1}{r} = \Sigma_i\beta_i \frac{(x^r{}_i - 1)}{r} + \tfrac{1}{2}\Sigma_i\Sigma_j\beta_{ij} \frac{(x^{r/2}{}_i - 1)}{r/2} \frac{(x^{r/2}{}_j - 1)}{r/2}.$$

If we take limits of both sides of this equation as r tends to zero, we obtain (since, using L'Hospital's rule, $\lim \lambda \to 0 (x^\lambda - 1)/\lambda = 1n\, x$)

$$1n\, y = \Sigma_i\beta_i 1n\, x_i + \tfrac{1}{2}\Sigma_i\Sigma_j\beta_{ij} 1n\, x_i 1n\, x_j,$$

which is the homogeneous translog functional form since, the β_i and β_{ij} satisfy the restrictions (46). The above proof that f_0 is a limiting case of f_r owes much to suggestions made by L. J. Lau.

The following theorems indicate that it does not matter which superlative index is used in empirical work with time series data: they will all give virtually the same answer.

(50) *Theorem* (Diewert 1978*b*): For any $r \neq 0$, $P_r(p^0,p^1; x^0,x^1) = P_0(p^0,p^1; x^0,x^1)$ and the first- and second-order partial derivatives of the two functions coincide provided $p^0 = p^1 \gg 0_N$ (all price components are positive) and $x^0 = x^1 > 0_N$ (at least one quantity component is positive).

(51) *Theorem* (Diewert 1978*b*): For any $r \neq 0$, the quantity index Q_r differentially approximates Q_0 to the second order at any point where the prices and quantities for the two periods are equal; that is, $Q_r(p^0,p^1; x^0,x^1) = Q_0(p^0,p^1; x^0,x^1)$, and the first- and second-order partial

derivatives of the two functions coincide provided $p^0 = p^1 > 0_N$ and $x^0 = x^1 \gg 0_N$.

Thus, if changes in prices and quantities are small, all the superlatives indexes P_r and Q_r will give virtually the same answer, even if economic agents are not engaging in optimizing behavior.[33] Some empirical evidence on the degree of closeness of the various indexes to each other is available in Fisher (1922), Ruggles (1967), and Diewert (1978*b*).

Theorems (50) and (51) suggest that using the *chain principle* (i.e., the base is changed to the previous period $t - 1$ rather than maintaining a fixed base when calculating the change in the aggregate going from period $t - 1$ to period t) in calculating aggregates will minimize the differences between the various index number formulas, since the changes in prices and quantities will generally be small between adjacent periods.[34] Furthermore, as we saw in the previous section, the Paasche, Laspeyres, and any superlative index number can be regarded as discrete approximations to the continuous-line integral Divisia index, which has some useful optimality properties from the viewpoint of economic theory (see Malmquist 1953; Wold 1953; Richter 1966; and Hulten 1973 on these optimality properties). These discrete approximations will be closer to the Divisia index if the chain principle is used.

8.2.5 Two-Stage Aggregation

To reduce the number of commodities, macroeconomic models generally employ index numbers of prices and quantities. However, very often an index number that is used in an economic model has been constructed in two or more stages, and thus the question arises: Does the two-stage procedure give the same answer as the single-stage procedure? It is true that the usually employed Paasche and Laspeyres indexes have this property of consistency in aggregation, but these index numbers are consistent only with very restrictive functional forms for the underlying aggregator function, as we have seen in section 8.2.4.

Diewert (1978*b*) shows that superlative indexes have an approximate consistency-in-aggregation property. This result was obtained by utilizing some results due to the Finnish economist Vartia (1974, 1976), who proposed a discrete approximation to the continuous Divisia price or quantity index that has the following two remarkable properties: (*a*) the price index and the corresponding quantity index (which is defined by the same formula except that prices and quantities are interchanged) satisfy Fisher's (1922) factor-reversal test (i.e., the product of the price and quantity indexes equal the expenditure ratio for the two periods under consideration) and (*b*) the price or quantity index has the property of *consistency in aggregation*.

Vartia defines an index number formula to be consistent in aggregation if the value of the index calculated in two stages necessarily coincides with the value of the index as calculated in an ordinary way, that is, in a single stage.

As we saw in section 8.2.4, the economic theory of index numbers is concerned with rationalizing functional forms for index numbers with functional forms for the underlying aggregator function. Diewert (1978*b*) shows that the Vartia I price and quantity indexes are consistent only with a Cobb-Douglas aggregator function. This is perhaps not surprising, since thus far the only way the two-stage method of calculating index numbers has been justified from the viewpoint of the economic theory of index numbers has been to assume that the underlying aggregator function is weakly separable in the same partition that corresponds to the two stages.[35] Thus, to justify the two-stage method of constructing index numbers for any partition of variables, one so far has had to assume that the aggregator function is weakly separable in any partition of its variables; but then the results of Leontief (1947) and Gorman (1968*b*) imply that the aggregator function is strongly separable in the coordinatewise partition of its variables. If we also assume that aggregator function is linearly homogeneous, then, using Bergson's (1936) results, it can be seen that the aggregator function must be a mean of order r (Hardy, Littlewood, and Polya 1934); that is, a CES function. However, it turns out that the Vartia price and quantity indexes are exact only for a mean of order 0 (or Cobb-Douglas) aggregator function.

In spite of the rather negative result that the Vartia I price and quantity indexes are exact only for a Cobb-Douglas aggregator function, Diewert (1978*b*) shows that for small changes in prices and quantities these indexes have some rather good approximation properties. I outline these results due to Vartia and Diewert below.

Define the Vartia (1974, 1976)[36] price index $P_V(p^0,p^1;x^0,x^1)$ as

$$(52)\qquad \ln P_V(p^0,p^1;x^0,x^1) \equiv \sum_{i=1}^{N} L(p^1_i x^1_i, p^0_i x^0_i)/L(p^1\cdot x^1, p^0\cdot x^0)\ \ln(p^1_i/p^0_i),$$

where the logarithmic mean function L introduced by Vartia (1974) and Sato (1976*a*) is defined by $L(a,b) \equiv (a-b)/(\ln a - \ln b)$ for $a \neq b$ and $L(a,a) \equiv a$.

The Vartia quantity index $Q_V(p^0,p^1;x^0,x^1)$ is defined by

$$(53)\qquad \ln Q_V(p^0,p^1;x^0,x^1) \equiv \sum_{i=1}^{N} L(p^1_i x^1_i, p^0_i x^0_i)/L(p^1\cdot x^1, p^0\cdot x^0)\ \ln(x^1_i/x^0_i)$$
$$= \ln P_V(x^0,x^1;p^0,p^1);$$

that is, the price and quantity indexes have the same functional form except that the role of prices and quantities are interchanged. Vartia shows that P_V and Q_V satisfy the factor-reversal test and have the property of consistency in aggregation.

Since the price index P_0 defined in section 8.2.3 resembles somewhat the Vartia price index P_V defined by (52), the following theorems may not be too surprising.

(54) *Theorem* (Diewert 1978*b*): The Vartia price index differentially approximates the superlative price index P_0 to the second order at any point where the prices and quantities for the two periods are equal; that is, $P_V(p^0,p^1;x^0,x^1) = P_0(p^0,p^1;x^0,x^1)$, and the first- and second-order partial derivatives of the two functions coincide provided that $p^0 = p^1 \gg 0_N$ and $x^0 = x^1 \gg 0_N$.

(55) *Theorem* (Diewert 1978*b*): The Vartia quantity index differentially approximates the superlative quantity index Q_0 to the second order at any point where the prices and quantities for the two periods are equal.

Thus $P_V(p^0,p^1;x^0,x^1)$ will be close to $P_0(p^0,p^1;x^0,x^1)$ provided p^0 is close to p^1 and x^0 is close to x^1. If we call an index that can approximate a superlative index differentially to the second order at any point where $p^0 = p^1$ and $x^0 = x^1$ a *pseudosuperlative* index, it can be seen that the Vartia price and quantity indexes are pseudosuperlative.

Recall theorems (50) and (51). Theorems (54) and (50) imply that the Vartia price index P_V approximates all the superlative indexes P_0 and P_r, while theorems (55) and (51) imply that the Vartia quantity index Q_V approximates all the superlative indexes Q_0 and Q_r to the second order.

For many years it was thought that the indexes P_0 and Q_0 had the property of consistency in aggregation. However, although P_0 and Q_0 are not consistent in aggregation, the results above show why they are approximately consistent in aggregation: each P_0 subindex can be approximated to the second order by a Vartia index of the same size, while the "macro" P_0 index can be approximated to the second order by a "macro" Vartia index. Thus the macro index of the subindexes is approximated to the second order by a Vartia macro index of Vartia subindexes which is identically equal to a Vartia index of the original micro components, which in turn approximates to the second order a P_0 index in the micro components. Therefore, for time-series data where indexes are constructed by chaining observations in successive periods, we would expect P_0 and Q_0 to be approximately consistent in aggregation.

The same conclusion holds for the quadratic mean of order r price indexes P_r and quantity indexes Q_r: they will be approximately consistent in aggregation, since each P_r approximates P_V and each Q_r approximates Q_V.

Some empirical evidence is available that tends to support the theoretical results above. Parkan (1975) compared the price indexes P_0, P_2, and $\tilde{P}_0$ (defined implicitly by the weak factor reversal test, using Q_0 as the quantity index) and the quantity indexes Q_0, Q_2, and $\tilde{Q}_0$ using some Canadian postwar consumption data on thirteen goods. He also calculated the nonparametric price and quantity indexes defined by Diewert (1973*b*, p. 424). Parkan then computed all four price indexes and all four quantity indexes in two stages, calculating four subaggregates in each case, then aggregating these subaggregates using the same index number formula. It was found that the resulting total of eight price indexes generally coincided to three significant figures, and the eight quantity indexes similarly closely approximated each other. Similar empirical results are reported in Diewert (1978*b*). The theoretical results cited above provide an explanation for this rather convenient empirical phenomenon.

To summarize, the arguments above show that constructing aggregate price and quantity indexes by aggregating in two (or more) stages will give approximately the same answer that a one-stage index would, provided that either a superlative index or the Vartia index is used.[37]

8.2.6 The Measurement of Real Input, Real Output, and Real Value Added

In this section we will study the various definitions of real output, input, and value added that economists have proposed, including the definition of real capital that Usher proposed in the introduction to this volume. We shall also indicate how various index number formulas can be used to closely approximate the various notions of real input and output.

First it is necessary to recall the definition of the firm's *variable profit function* from section 8.2.1:

$$(56) \qquad \Pi(x,p) \equiv \max_y\{p \cdot y: (x,y) \in S\},$$

where S is the firm's technologically feasible set, $(x,y) \equiv (x_1,x_2, \ldots ,x_N, y_1,y_2, \ldots ,y_M)$ is a vector that indicates the firm's production or input demand for each of the $N + M$ goods, and $p \equiv (p_1,p_2, \ldots ,p_M) \gg 0_M$ is a vector of positive prices. In this section we will assume that the x goods are all inputs and that $x \gg 0_N$. On the other hand, negative components of y will continue to indicate that the corresponding good is used as an input. With these sign conventions, it can be shown (see Diewert 1973*a*; Gorman 1968*a*; or Lau 1976) that if S is a nonempty, closed, convex set with certain boundedness and free disposal properties, then $\Pi(x,p)$ will be a nonnegative, nondecreasing, and concave function in x for any fixed p; that is, $\Pi(x,p)$ regarded as a function of x will have the usual regularity properties that a neoclassical production function possesses.

Thus a *real input index* X can sensibly be defined as

$$X(x^0,x^1;p^*) \equiv \Pi(x^1,p^*)/\Pi(x^0,p^*), \tag{57}$$

where $x^0 \equiv (x^0{}_1,\ldots,x^0{}_N)$ is period 0 input, $x^1 \equiv (x^1{}_1,\ldots,x^1{}_N)$ is period 1 input, and $p^* \gg 0_M$ is a reference price vector. Sato (1976*b*, p. 438) calls X defined by (57) a *true index of real value added*, and he notes that the definition does not require any assumption of optimizing behavior on the part of the producer with respect to inputs (although profit-maximizing behavior with respect to outputs and intermediate inputs in the y goods is of course required). Sato also notes that a separability assumption on the technology is required in order to make X defined by (57) independent of p^*; that is, we require that $\Pi(p,x) = r(p)f(x)$ for some functions r and f, which implies that the x inputs are separable from y.

We now study the problem of approximating (57) by observable data; that is, by means of an index number formula. However, it is first necessary to present some general material taken from Diewert (1976) that will be used repeatedly in this section.

Let z be an N-dimensional vector and define the quadratic function $f(z)$ as

$$\begin{aligned} f(z) &\equiv a_0 + a^Tz + \tfrac{1}{2}z^TAz \\ &= a_0 + \sum_{i=1}^{N} a_iz_i + \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}z_iz_j, \end{aligned} \tag{58}$$

where the a_i,a_{ij} are constants and $a_{ij} = a_{ji}$ for all i,j.

The following lemma is a global version of the Theil (1967, pp. 222–23) and Kloek (1967) local result.

(59) *Quadratic approximation lemma*: if the quadratic function f is defined by (58), then,

$$f(z^1) - f(z^0) = \tfrac{1}{2}[\nabla f(z^1) + \nabla f(z^0)]\cdot(z^1 - z^0), \tag{60}$$

where $\nabla f(z^r)$ is the vector of first-order partial derivatives of f evaluated at z^r.

This result should be contrasted with the usual Taylor series expansion for a quadratic function,

$$\begin{aligned} f(z^1) - f(z^0) = [\nabla f(z^0)]\cdot(z^1 - z^0) + \tfrac{1}{2}(z^1 \\ - z^0)\cdot\nabla^2 f(z^0)(z^1 - z^0), \end{aligned}$$

where $\nabla^2 f(z^0)$ is the matrix of second-order partial derivatives of f evaluated at an initial point z^0. In the expansion (60), a knowledge of $\nabla^2 f(z^0)$ is not required, but a knowledge of $\nabla f(z^1)$ is required. Actually, (60) holds as an equality for all z^1,z^0 belonging to an open set if and only if f is a quadratic function, provided f is once continuously differentiable (cf. Lau 1979).

Suppose we are given a homogeneous translog aggregator function (Christensen, Jorgenson, and Lau 1971) defined by

$$\ln f(x) \equiv \beta_0 + \sum_{n=1}^{N} \alpha_n \ln x_n + \tfrac{1}{2} \sum_{j=1}^{N} \sum_{k=1}^{N} \beta_{jk} \ln x_j \ln x_k,$$

where $\sum_{n=1}^{N} \beta_n = 1,\ \beta_{jk} = \beta_{kj}$ and $\sum_{k=1}^{N} \beta_{jk} = 0$ for $j = 1,2,\ldots,N$.

Recall that Jorgenson and Lau have shown that the homogeneous translog function can provide a second-order approximation to an arbitrary twice continuously differentiable linearly homogeneous function. Let us use the parameters that occur in the translog functional form to define the following function f^*:

$$f^*(z) \equiv \alpha_0 + \sum_{j=1}^{N} \alpha_i z_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_{ij} z_i z_j. \tag{61}$$

Since the function f^* is quadratic, we can apply the quadratic approximation lemma (59), and we obtain

$$f^*(z^1) - f^*(z^0) = \tfrac{1}{2}[\nabla f^*(z^1) + \nabla f^*(z^0)] \cdot (z^1 - z^0). \tag{62}$$

Now we relate f^* to the translog function f. We have

$$\begin{aligned} \partial f^*(z^r)/\partial z_j &= \partial \ln f(x^r)/\partial \ln x_j \\ &= [\partial f(x^r)/\partial x_j]\,[x^r{}_j/f(x^r)]. \\ f^*(z^r) &= \ln f(x^r), \\ z^r{}_j &= \ln x^r{}_j, \text{ for } r = 0,1 \text{ and } j = 1,2,\ldots,N. \end{aligned} \tag{63}$$

If we substitute relations (63) into (62) we obtain

$$\begin{aligned} \ln f(x^1) - \ln f(x^0) = \tfrac{1}{2} &\left[\hat{x}^1 \frac{\nabla f(x^1)}{f(x^1)} + \hat{x}^0 \frac{\nabla f(x^0)}{f(x^0)} \right] \\ &\cdot [\ln x^1 - \ln x^0], \end{aligned} \tag{64}$$

where $\ln x^1 \equiv [\ln x^1{}_1, \ln x^1{}_2, \ldots, \ln x^1{}_N]$, $\ln x^0 \equiv [\ln x^0{}_1, \ln x^0{}_2, \ldots, \ln x^0{}_N]$, $\hat{x}^1 \equiv$ the vector x^1 diagonalized into a matrix, and $\hat{x}^0 \equiv$ the vector x^0 diagonalized into a matrix.

Now suppose the firm's variable profit function can be adequately approximated by a translog variable profit function (see section 8.2.1 for a definition), which we will denote as $\Pi^*(x,p)$.

If p^* is fixed, then $\ln \Pi^*(x,p^*)$ is quadratic in $\ln x_1, \ln x_2, \ldots, \ln x_N$, and we can apply the identity (64) to obtain the following equality (define $f(x) \equiv \Pi^*(x,p^*)$):

$$
\begin{aligned}
(65)\qquad & \ln \Pi^*(x^1,p^*) - \ln \Pi^*(x^0,p^*) \\
& = \tfrac{1}{2}\left[\hat{x}^1 \frac{\nabla_x \Pi^*(x^1,p^*)}{\Pi^*(x^1,p^*)} + \hat{x}^0 \frac{\nabla_x \Pi^*(x^0,p^*)}{\Pi^*(x^0,p^*)}\right] \\
& \cdot [\ln x^1 - \ln x^0].
\end{aligned}
$$

To proceed further, we need to make two additional assumptions: (*a*) the technology S is a cone (so that constant returns to scale prevail), and hence $\Pi^*(x,p)$ is linearly homogeneous in x, and (*b*) the producer attempts to minimize input costs, or alternatively to maximize nominal value added (or variable profits) subject to an expenditure constraint on inputs. Thus we assume that $x^r \gg 0_N$ is a solution to the maximization problem

$$(66)\qquad \max_x\{\Pi^*(x,p^r):\ w^r \cdot x \leq w^r \cdot x^r,\ x \geq 0_N\},$$

where $w^r \gg 0_N$ for $r = 0,1$[38] where $w^r \equiv (w^r_1, \ldots, w^r_N)$ and w^r_n is the nth input price in period r. The first-order conditions for the two maximization problems, after elimination of the Lagrange multipliers, yield the relations $w^r/w^r \cdot x^r = \nabla_x \Pi^*(x^r,p^r)/x^r \cdot \nabla_x \Pi^*(x^r,p^r)$ for $r = 0,1$. Since Π^* is linearly homogeneous in $x, x^r \cdot \nabla_x \Pi^*(x^r,p^r)$ can be replaced by $\Pi^*(x^r,p^r)$, and the resulting relations are

$$(67)\qquad w^r/w^r \cdot x^r = \nabla_x \Pi^*(x^r,p^r)/\Pi^*(x^r,p^r), \quad r = 0,1.$$

Now return to equation (65). Assume that the components of the (constant) output price vector $p^* = (p^*_1, \ldots, p^*_N)$ are defined by

$$(68)\qquad p^*_n \equiv (p^0_n\, p^1_n)^{1/2}, \quad n = 1,2, \ldots, N.$$

Substitution of (68) into (65) and differentiation of the translog variable profit function evaluated at the points (x^1,p^*) and (x^0,p^*) yields the following equation:

$$
\begin{aligned}
(69)\qquad & \ln \Pi^*(x^1,p^*) - \ln \Pi^*(x^0,p^*) \\
& = \tfrac{1}{2}\left[\hat{x}^1 \frac{\nabla_x \Pi^*(x^1,p^1)}{\Pi^*(x^1,p^1)} + \hat{x}^0 \frac{\nabla_x \Pi^*(x^0,p^0)}{\Pi^*(x^0,p^0)}\right] \\
& \cdot [\ln x^1 - \ln x^0] \\
& = \tfrac{1}{2}\left[\hat{x}^1 \frac{w^1}{w^1 \cdot x^1} + \hat{x}^0 \frac{w^0}{w^0 \cdot x^0}\right] \cdot [\ln x^1 - \ln x^0]
\end{aligned}
$$

using (67), or

$$(70) \qquad X(x^0,x^1;p^*) = \Pi^*(x^1,p^*)/\Pi^*(x^0,p^*) = Q_0(w^0,w^1,x^0,x^1);$$

that is, if the technology can be represented by a constant returns to scale, translog variable profit function and the reference output prices p^*_n are chosen to be the geometric mean of the output prices prevailing during the two periods; then the real input index $X(x^0,x^1;p^*)$ is equal to the Törnqvist quantity index of the inputs x^0 and x^1. Note that this result does *not* require the technology to be separable; that is, we do not require that $\Pi^*(p,x) = r(p)f(x)$. However, the above result did require us to pick a very specific reference vector p^*.

Note that we can associate an *implicit input price deflator* $W(w^0,w^1,x^0,x^1,p^*)$ with the real input index X:

$$(71) \qquad W(w^0,w^1,x^0,x^1,p^*) \equiv w^1\cdot x^1/w^0\cdot x^0\ X(x^0,x^1;p^*).$$

Under the assumptions that justified (70), we can see that the implicit input price deflator $W(w^0,w^1,x^0,x^1;p^*) = \tilde{P}_0(w^0,w^1,x^0,x^1)$, the implicit Törnqvist price index for the inputs (defined as $w^1\cdot x^1/w^0\cdot x^0\ Q_0(w^0,w^1,x^0,x^1)$).

It is also possible to define an input price deflator directly. To do this, we need to define the *joint cost function*[39] C as

$$(72) \qquad C(y,w) \equiv \min_x\{w\cdot x\colon (x,y)\epsilon S\}.$$

Now define the *input price deflator* W as

$$(73) \qquad W(w^0,w^1\,;y^*) \equiv C(y^*,w^1)/C(y^*,w^0),$$

where the input price vectors w^0 and w^1 have been defined above and y^* is a reference output vector that is held constant during the two periods. As was the case with the real input index, the input price deflator $W(w^0,w^1;y^*)$ is independent of the reference vector y^* (p^* in the case of the input index) if and only if the technology is separable (i.e., if $C(y,w) = g(y)c(\mathrm{w})$).

To obtain a specific functional form for W, we may proceed in a manner entirely analogous to our earlier treatment for X. First assume that the firm's technology can be adequately approximated by a translog joint cost function[40] that exhibits constant returns to scale. Then, assuming optimizing behavior on the part of the producer, we can repeat equations (65) to (70) with the obvious changes in notation, and we obtain the following equality:

$$(74) \qquad W(w^0,w^1;y^*) = C^*(y^*,w^1)/C^*(y^*,w^0) = P_0(w^0,w^1,x^0,x^1),$$

where C^* is a translog joint cost function, $y^* \equiv (y^*_1, y^*_2, \ldots, y^*_M)$ and $y^*_m \equiv (y^0_m y^1_m)^{1/2}$ for $m = 1,2,\ldots,M$, and $P_0(w^0,w^1,x^0,x^1)$ is the Törnqvist price index for the inputs.

An *implicit real input index* X can be defined as

$$(75) \qquad X(w^0,w^1,x^0,x^1;y^*) \equiv w^1 \cdot x^1 / w^0 \cdot x^0 \; W(w^0,w^1;y^*).$$

Obviously, if the input price deflator W is defined by (74), then the corresponding implicit real input index X is numerically equal to $\tilde{Q}_0(w^0,w^1,x^0x^1)$, the implicit Törnqvist quantity index of inputs.[41]

We now turn our attention to the output side. Define the producer's *real output index* Y as

$$(76) \qquad Y(y^0,y^1;w^*) \equiv C(y^1,w^*)/C(y^0,w^*),$$

where y^0 and y^1 are the output (and intermediate input) vectors are periods 0 and 1, C is the producer's joint cost function defined earlier by (72), and $w^* \gg 0_N$ is a reference input price vector. As usual, $Y(y^0,y^1,w^*)$ is independent of w^* if and only if the technology is separable.

Again, we can assume that the firm's technology is approximated by a transiog joint cost function C^* that exhibits constant returns to scale. Assuming optimizing behavior, we can repeat equations (65) to (70), with the obvious changes in notation, and obtain the following equality:

$$(77) \qquad Y(y^0,y^1;w^*) = C^*(y^1,w^*)/C^*(y^0,w^*) = Q_0(p^0,p^1,y^0,y^1),$$

where C^* is the firm's translog joint cost function, $y^0 \gg 0_N$ and $y^1 \gg 0_M$ are the output vectors produced by the firm during the two periods, p^0 and p^1 are the corresponding output price vectors,[42] and the reference input price vector $w^* \equiv (w^*_1,\ldots,w^*_N)$ is defined by $w^*_n \equiv (w^0_n w^1_n)^{1/2}$, where $w^0 \equiv (w^0_1,\ldots,w^0_N) \gg 0_N$ and $w^1 \equiv (w^1_1,\ldots, w^1_N) \gg 0_N$ are the input price vectors for the two periods. $Q_0(p^0,p^1,y^0,y^1)$ is the Törnqvist quantity index for the outputs.

Note that we can associate an *implicit output price deflator* $P(p^0,p^1, y^0,y^1;w^*)$ with the real output index Y:

$$(78) \qquad P(p^0,p^1,y^0,y^1;w^*) \equiv p^1 \cdot y^1 / p^0 \cdot y^0 \; Y(y^0,y^1;w^*).$$

If the real output index Y is defined by (77), then the corresponding implicit output price deflator defined by (78) is numerically equal to $\tilde{P}_0(p^0,p^1,y^0,y^1)$, the implicit Törnqvist price index of outputs.

However, an output price deflator can be defined directly. Following Fisher and Shell (1972), define the firm's *output price deflator* P as

$$(79) \qquad P(p^0,p^1;x^*) \equiv \Pi(x^*,p^1)/\Pi(x^*,p^0),$$

where p^0 and p^1 are the output (and intermediate input) price vectors facing the producer during periods 0 and 1, respectively, Π is the producer's variable profit function defined earlier by (56), and $x^* \rangle\rangle 0_N$ is a reference input vector. Archibald (1977, p. 61) calls P the fixed input quantity output price index. He also shows that it satisfies certain tests, and he develops some bounds for it (along with two other alternative price indexes), utilizing the techniques developed by Pollak (1971) in his discussion of the cost-of-living index.

As usual, assume that the firm's technology can be approximated by a translog variable profit function Π^* exhibiting constant returns to scale, repeat the analysis inherent in equations (65) to (70) with the obvious changes in notation, and obtain the following equality:

$$(80) \qquad P(p^0,p^1;x^*) = \Pi^*(x^*,p^1)/\Pi^*(x^*,p^0) = P_0(p^0,p^1,y^0,y^1),$$

where Π^* is the firm's translog variable profit function, $p^0 \rangle\rangle 0_M$ and $p^1 \rangle\rangle 0_M$ are the price vectors for outputs (and intermediate inputs) the firm faces during periods 0 and 1, y^0 and y^1 are the corresponding quantity vectors,[43] and the reference input quantity vector $x^* \equiv (x^*_1, \ldots, x^*_N)$ is defined by $x^*_n \equiv (x^0_n x^1_n)^{1/2}$, where $x^0 \equiv (x^0_1, \ldots, x^0_N) \rangle\rangle 0_N$ and $x^1 \equiv (x^1_1, \ldots, x^1_N) \rangle\rangle 0_N$ are the input vectors for the two periods. $P_0(p^0,p^1,y^0,y^1)$ is the Törnqvist price index in output prices.

Note that we can associate an *implicit real output index Y* with the output price deflator P:

$$(81) \qquad \tilde{Y}(p^0,p^1,y^0,y^1;x^*) \equiv p^1 \cdot y^1/p^0 \cdot y^0\, P(p^0,p^1;x^*).$$

If the output price deflator P is defined by (80), then the corresponding implicit real output index Y defined by (81) is numerically equal to $\tilde{Q}_0(p^0,p^1,y^0,y^1)$, the implicit Törnqvist quantity index of outputs.

The astute reader will by now have noticed that the definitions given above for real output (input) and output (input) price deflators are entirely analogous to the Konyus (1939) and Allen (1949) definitions for the real cost of living and real income:[44] instead of holding a scalar constant (utility), a vector (of inputs or outputs) is held constant. The astute reader will also know that an alternative approach to the Konyus-Allen approach to defining quantity indexes has ben provided by Malmquist (1953). Malmquist's approach has been extensively used by Pollak (1971) and Blackorby and Russell (1978) in the context of consumer theory, and by Bergson (1961), Moorsteen (1961), and Fisher and Shell (1972) in the context of producer theory. I outline this approach below.

Define the producer's *input distance function* D as

(82) $$D[y,x] \equiv \max_\lambda\{\lambda:\ (y,x/\lambda)\epsilon S,\ \lambda \geq 0\},$$

where S is the firm's technological set, y is a given vector of outputs, and $x \rangle\rangle 0_N$ is a given vector of inputs. The interpretation of $D[y,x]$ is that it is the proportion by which the input vector x can be deflated with the resulting deflated input vector just big enough to produce the output vector y. If S is a nonempty, closed, convex set with certain free disposal properties, then it can be shown that $D[y,x]$ is a positive, increasing, linearly homogeneous, and concave function in x for $x \rangle\rangle 0_N$ and, moreover, the distance function can be used to characterize the technology just as the variable profit function or joint cost function was used.[45] We can use the distance function to define the Malmquist (1953), Moorsteen (1961), Fisher and Shell (1972, p. 51), and Usher *real input index* as

(83) $$X_M(x^0,x^1,y^*) \equiv D[y^*,x^1]/D[y^*,x^0],$$

where y^* is a reference output vector and x^0,x^1 are the vectors of inputs utilized by the firm during the two periods under consideration. The interpretation of X_M is straightforward: pick a reference output vector y^*, deflate x^r by $\lambda^r > 0$ so that $(y^*,x^r/\lambda^r)$ is just on the boundary of the production possibility set S for $r = 0,1$, and then measure the volume of inputs in period 1 relative to period 0 by the ratio λ^1/λ^0. The resulting Malmquist real input index $X_M(x^0,x^1,y^*)$ will in general depend on the reference output vector y^*; it will be independent of y^* if and only if the technology is such that $D[y,x] = h(y)f(x)$, which is a separability property that turns out to be equivalent to the earlier separability of outputs from inputs property discussed earlier in this section.[46]

The Malmquist real input index X_M defined by (83) has at least one major advantage over the (Konyus) real input index X defined earlier by (57): the Malmquist index is defined solely by the technology and does not require any assumption that the producer competitively maximizes profits.

However, to evaluate X_M using observable data, it will be necessary to assume cost-minimizing behavior plus a particular functional form for the firm's distance function.

(84) *Theorem*: Assume that the firm's technology can be represented by a translog distance function D^*, where D^* is defined by

(85) $$\ln D^*[y,x] \equiv \alpha_0 + \sum_{i=1}^{N} \alpha_i \ln x_i + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \alpha_{ih} \ln x_i \ln x_h$$

$$+ \sum_{i=1}^{N} \sum_{j=1}^{M} \beta_{ij} \ln x_i \ln y_j + \sum_{j=1}^{M} \gamma_j \ln y_j$$

$$+ \tfrac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \gamma_{jk} \ln y_j \ln y_k,$$

where $\sum_{i=1}^{N} \alpha_i = 1$, $\alpha_{ih} = \alpha_{hi}$, $\sum_{h=1}^{N} \alpha_{ih} = 0$ for $i = 1,2,\ldots,N$,

$\sum_{i=1}^{N} \alpha_{ij} = 0$ for

$j = 1,2,\ldots,M$ and $\gamma_{jk} = \gamma_{kj}$.[47] Suppose that the quantity vector x^0 is a solution to the cost minimization problem, $\min_x\{w^0 \cdot x\colon (y^0,x) \epsilon S\}$, while x^1 is a solution to the period 1 cost minimization problem, $\min_x\{w^1 \cdot x\colon (y^1,x) \epsilon S\}$, where y^0,y^1 are the output vectors produced during periods 0 and 1, $w^0 \gg 0_N$ and $w^1 \gg 0_N$ are the input price vectors facing the producer during periods 0 and 1 and S is the firm's technology set. Then

$$(86) \qquad X_M(x^0,x^1,y^*) = D^*[y^*,x^1]/D^*[y^*,x^0] = Q_0(w^0,w^1,x^0,x^1),$$

where $y^* \equiv (y^*_1,\ldots,y^*_M)$, $y^*_m \equiv (y^0_m y^1_m)^{1/2}$ for $m = 1,2,\ldots,M$, and $Q_0(w^0,w^1,x^0,x^1)$ is the Törnqvist quantity index for the inputs.

Thus the Törnqvist quantity index $Q_0(w^0,w^1,x^0,x^1)$ can be interpreted as either a Malmquist real input index (86) *or* a Konyus real input index (70). However, we note that the Malmquist interpretation requires fewer assumptions: constant returns to scale are not required, nor are producers required to competitively optimize with respect to inputs. Thus the Törnqvist quantity index can be given a strong economic justification.

Obviously, once we have defined the Malmquist real input index X_M, we can define an *implicit (Malmquist) input price deflator* W_M as

$$(87) \qquad W_M(w^0,w^1,x^0,x^1,y^*) \equiv w^1 \cdot x^1 / w^0 \cdot x^0 \, X_M(x^0,x^1,y^*).$$

If X_M is defined by (86), then (87) becomes $\tilde{P}_0(w^0,w^1,x^0,x^1)$, the implicit Törnqvist price index for inputs.

With minor modifications, the entire Malmquist procedure from equation (82) to equation (87) can be repeated, except that outputs replace inputs; that is, define the producer's output distance function d as $d[y,x] \equiv \min_\lambda\{\lambda\colon (y/\lambda,x) \epsilon S\}$, define the Malmquist real output index as $Y_M(y^0,y^1,x^*) \equiv d[y^1,x^*]/d[y^0,x^*]$, assume that the producer's technology can be adequately represented by a translog (output) distance function d^* and that the producer is revenue maximizing with respect to outputs and intermediate inputs only, and finally show that

(88) $$Y_M(y^0,y^1,x^*) = d^*[y^1,x^*]/d^*[y^0,x^*] = Q_0(p^0,p^1,y^0,y^1),$$

where $Q_0(p^0,p^1,y^0,y^1)$ is the Törnqvist quantity index for the outputs and the nth component of the reference input vector, $x^*_n \equiv (x^0_n x^1_n)^{1/2}$, $n = 1,2,\ldots,N$.[48] Thus again the Törnqvist index can be given a strong economic justification.

8.3 Methods for Justifying Aggregation over Sectors

8.3.1 Aggregation without Optimizing Behavior

Suppose there are M firms in a sector, each of which produces a single product using N inputs. Let the firm technologies be representable by means of firm production functions f^m, where $y^m = f^m(x^m_1,\ldots,x^m_N)$ denotes the amount of output producible by firm m using input quantities $x^m_1,\ldots,x^m_N$ for m $= 1,2,\ldots,M$. Klein's (1946*a*) aggregation over sectors problem[49] can be phrased as follows: What conditions on the firm production function's f^m will guarantee the existence of: (*a*) an aggregate production function G, (*b*) input aggregator functions $g_1,\ldots,g_N$, and (*c*) an output aggregator function F such that the following equation holds for a suitable set of inputs x^m_n?

(89) $$F(y^1,\ldots,y^M) = G[g_1(x^1_1,\ldots,x^M_1),g_2(x^1_2,\ldots,x^M_2),\ldots,g_N(x^1_N,\ldots,x^M_N)].$$

Klein (1946*b*) explicitly asked that (89) hold without the assumption of profit-maximizing behavior by producers, since in the real world monopolistic practices may be prevalent and thus it would be preferable to be able to derive an aggregate production function without the assumption of competitive behavior.

Unfortunately, Nataf (1948) demonstrated that the conditions on the micro production functions f^m, $m = 1,2,\ldots,M$, that are necessary to derive (89) are very stringent: the f^m must be strongly separable; that is, f^m must have the structure $f^m(x^m_1,\ldots,x^m_N) = h^m[k^m_1(x^m_1) + \ldots + k^m_N(x^m_N)]$, where the h^m and k^m_n are monotonically increasing functions.[50]

The restrictive separability assumptions on the micro production functions[51] required for Klein-Nataf sectoral aggregation seem to limit the usefulness of the method from an empirical point of view. A more promising method is outlined in the following section.

8.3.2 Aggregation with Optimizing Behavior

Bliss (1975, p. 146) notes that if all producers are competitive profit maximizers and face the same prices, then the group of producers can be treated as if they were a single producer subject to the sum[52] of the individual production sets. Thus, if the time period is chosen to be long enough so that all inputs and outputs are variable, there is no problem of aggregation over sectors (provided producers are behaving competitively).

This extremely simple aggregation criterion does not seem to have been stressed very much in the literature, but it is certainly explicit in the contributions of Hotelling (1935), May (1946), Pu (1946) and Cornwall (1973, p. 512), though not stated as elegantly as the above criterion noted by Bliss.

8.3.3 Aggregation with Optimizing Behavior for Some Goods: Vintage Production Functions

Assume we have M sectors in an economy (or M firms in an industry or M plants or processes in a firm) and the production possibilities set for the mth sector is denoted by S^m, $m = 1,2, \ldots ,M$. Define the sectoral variable profit functions Π^m as

$$\Pi^m(p,x^m,z^m) \equiv max_y\{p \cdot y\colon (y,x^m,z^m)\epsilon S^m\}, \quad m = 1,2, \ldots ,M, \tag{90}$$

where $p \rangle\rangle 0_K$ is a vector of output (and or intermediate input) prices each sector faces for the first K goods, $x^m \geq 0_N$ is a vector of "labor" inputs utilized by the mth sector, and z^m is a vector of fixed "capital" inputs that could be specific to the mth sector for $m = 1,2, \ldots ,M$. Following Solow (1964), we could interpret the Π^m as being dual to the "vintage" production functions f^m (of a single firm) that utilize the "vintage" capital inputs z^m in addition to labor inputs x^m. Assume further that the firm has a fixed vector of labor inputs $x \rangle\rangle 0_N$ to allocate across the M processes. The firm will then wish to solve the following vintage or micro labor maximization problem (which defines the aggregate variable profit function Π):

$$\max_{x^1,x^2, \ldots ,x^M} \left\{ \sum_{m=1}^{M} \Pi^m(p,x^m,z^m)\colon \sum_{m=1}^{M} x^m \leq x, x^m \geq 0_N \right\} \tag{91}$$

$$\equiv \Pi(p,x,z^1,z^2, \ldots ,z^M). \tag{92}$$

A generalized[53] Solow (1964), Fisher (1965), and Stigum (1967, 1968) vintage capital aggregation problem is: Under what conditions on the "vintage" technologies S^m (or Π^m) do there exist functions $\hat{\Pi}$ and g such that

(93) $$\Pi(p,x,z^1,z^2,\ldots,z^M) = \Pi[p,x,g(z^1,z^2,\ldots,z^M)],$$

where $g(z^1,z^2,\ldots,z^M)$ can be interpreted as a capital aggregate?

This problem is difficult to solve because the aggregate variable profit function $\Pi(p,x,z^1,\ldots,z^M)$ is not related to the micro variable profit functions $\Pi^m(p,x^m z^m)$ in any very obvious way. However, it is possible to derive a problem that is equivalent to (93) and then apply some results from Gorman (1968*a*) to the equivalent problem. Below, I indicate how this equivalent problem can be derived and solved.

If the firm is given a vector of positive labor prices, $w \gg 0_N$, the firm can optimize with respect to the labor inputs. Thus define the following vintage or micro (labor optimized) variable profit functions Π^{*m}:

(94) $$\Pi^{*m}(p,w,z^m) \equiv \max_x\{\Pi^m(p,x,z^m) - w\cdot x: x \geq 0_N\}, \quad m = 1,2,\ldots,M,$$

where $p \gg 0_K$ is the vector of output prices each sector faces for the first K goods and z^m is the vector of fixed capital inputs specific to the mth sector. The (labor optimized) variable profit functions Π^{*m} can be used to provide a complete characterization of the sectoral technologies S^m (under appropriate regularity conditions; see Gorman 1968*a* or Diewert 1973*a*) just as the original variable profit functions Π^m can be used.

The firm or macro (labor optimized) variable profit function Π^* dual to Π is defined as

(95) $$\Pi^*(p,w,z^1,\ldots,z^M) \equiv \max_x\{\Pi(p,x,z^1,\ldots,z^M) - w\cdot x: x \geq 0_N\}.$$

We may now state the problem that is equivalent to the original generalized Solow vintage capital aggregation problem (93): Under what conditions on the "vintage" technologies S^m (or equivalently Π^m or Π^{*m}) do there exist functions Π^* and g^* such that

(96) $$\Pi^*(p,w,z^1,\ldots,z^M) = \Pi^*[p,w,g^*(z^1,\ldots,z^M)],$$

where $g^*(z^1,\ldots,z^M)$ can be interpreted as a capital aggregate?

This problem is reasonably easy to solve because the aggregate (labor optimized) variable profit function Π^* is related in a simple manner to the micro (labor optimized) variable profit functions Π^{*m}.

(97) *Theorem*: If the micro variable profit functions $\Pi^m(p,x^m,z^m)$ are concave[54] and increasing in x^m for $m = 1,2,\ldots,M$, then

(98) $$\Pi^*(p,w^*,z^1,\ldots,z^M) = \sum_{m=1}^{M} \Pi^{*m}(p,w^*,z^m),$$

where $w^* \rangle\rangle 0_N$ is a vector of Lagrange multipliers or shadow prices for the maximization problem (91).

The decomposition (98) allows us to immediately prove the following theorem.

(99) *Theorem* (Gorman 1968*a*): If the micro (labor optimized) variable profit functions Π^{*m} can be written as

(100) $$\Pi^{*m}(p,w,z^m) = b(p,w)h^m(z^m) + c^m(p,w),$$
$$m = 1,2,\ldots,M,$$

then capital aggregation is possible; that is,

(101) $$\Pi^*(p,w,z^1,\ldots,z^M) = \sum_{m=1}^{M} \Pi^{*m}(p,w,z^m)$$
$$= b(p,w)\left[\sum_{m=1}^{M} h^m(z^m)\right]$$
$$+ \sum_{m=1}^{M} c^m(p,w)$$
$$\equiv \Pi^*\left[p,w, \sum_{m=1}^{M} h^m(z^m)\right].$$

Thus the separability restrictions (100) on the micro production possibility sets are sufficient to imply the existence of a capital aggregate; Gorman (1968*a*) shows that these conditions are also necessary under suitable regularity conditions on the technology (which do not involve differentiability restrictions).

How restrictive in practice are the restrictions (100)? They are not very restrictive at all if every z^m is a scalar (i.e., there is only one fixed capital good for each sector), for in this case functions of the form $b(p,w)h^m(z^m) + c^m(p,w)$ can provide a second-order approximation to an arbitrary twice-differentiable variable profit function $\Pi^{*m}(p,w,z^m)$. However, the restrictions (100) become progressively more unrealistic from an empirical point of view as the number of fixed capital goods in each sector increases.

8.3.4 Aggregation with Optimizing Behavior for Some Goods: Putty-Putty Production Functions

In the analysis of the previous section, we did not assume that there was any particular relationship between the sectoral production possibility sets S^m. In this section we will assume that these sets are related in the following manner: before the mth producer chooses a vector of fixed inputs z^m, the set of technological possibilities open to him ex ante is S, a set of feasible input and output combinations open to all producers $m = 1,2,\ldots,M$. However, once the mth producer chooses his vector of fixed inputs z^m, his production possibilities set is $S(z^m)$, a subset of S. If we know the ex ante production possibilities set plus the economy's *distribution of fixed inputs*, then we can readily calculate the economy's ex post production possibilities set for the variable inputs and outputs. Often (see Johansen 1959), the ex ante production function is taken to be Cobb-Douglas (putty), while the ex post production possibility sets $S(z^m)$ are taken to be of the fixed coefficients variety (clay), but this putty-clay model can readily be generalized to a putty-putty model (see the contributions by Fuss and McFadden 1978).

Houthakker (1955–56), Johansen (1959, 1972), Cornwall (1973), Sato (1975), and Fuss and McFadden (1978) have all made substantial contributions to the theory of aggregation sketched above.

To make the above discussion more concrete, I will outline in some detail Johansen's (1972) contribution. Johansen's basic theoretical model is presented in chapter 2, where the various types of production function are defined and related to each other. The *ex ante micro production function* $\emptyset$ gives the maximum output y, given amounts of two variable inputs x_1,x_2 and capital z invested in the sector; that is, $y = \emptyset(x_1,x_2,z)$. The *ex post micro production function* for a particular firm or sector is defined as $y = \emptyset(x_1,x_2,z)$; that is, the capital input is fixed at z. Johansen restricts the functional form for $\emptyset$ to be such that the ex post micro production function is of the fixed coefficients variety; that is, $0 \leq y \leq \bar{y}$, $x_1 = \xi_1 y$, $x_2 = \xi_2 y$, where ξ_1 and ξ_2 depend on z. The short-run macro production function, $F(X_1,X_2)$, is defined as

$$(102) \qquad F(X_1,X_2) \equiv \max_{x^1{}_1,x^1{}_2,\ldots,x^M{}_1,x^M{}_2} \left\{ \sum_{m=1}^{M} \emptyset(x^m{}_1,x^m{}_2,z^m) : \sum_{m=1}^{M} x^m{}_1 \leq X_1;\ \sum_{m=1}^{M} x^m{}_2 \leq X_2 \right\},$$

where the short-run ex post production function of the mth firm is $\emptyset(x^m{}_1,x^m{}_2,z^m)$ and X_j is the available factor supply for the jth variable input, $j = 1,2$. Thus this maximization problem takes the short-run micro production functions as given and maximizes industry output subject to restrictions on the availability of variable inputs. With Johansen's restrictive a priori assumptions on the functional form for the micro production function $\emptyset$, this maximization problem simplifies to:

$$(103) \qquad \max_{y^1,\ldots,y^M} \left\{ \sum_{m=1}^{M} y^m : \sum_{m=1}^{M} \xi^m{}_1 y^m \leq X_1;\ \sum_{m=1}^{M} \xi^m{}_2 y^m \leq X_2;\ 0 \leq y^m \leq \bar{y}^m \right\}.$$

Finally, Johansen defines the long-run macro production function. This function may be constructed from the ex ante micro function $\emptyset$ as follows: given (variable) input-capital ratios u_1 and u_2, choose the optimal scale plant; that is, choose z so as to maximize output per unit of capital, $\emptyset(u_1z,u_2z,z)/z$, along a ray in input space. Denote the solution to this maximization problem as $z = g(u_1,u_2)$. Now, given aggregate amounts of inputs X_1,X_2 and Z, the long-run macro production function ψ is defined as $\psi(X_1,X_2,Z) \equiv Z\emptyset[u_1g(u_1,u_2),u_2g(u_1,u_2), g(u_1,u_2)]/g(u_1,u_2)$, where $u_1 \equiv X_1/Z, u_2 \equiv X_2/Z$. Note that ψ is homogeneous of degree one even though the micro function need not be (if $\emptyset$ is linearly homogeneous, then $\psi = \emptyset$).

All the production functions above can be constructed if (a) we know the functional form for the ex ante micro function $\emptyset$ and (b) we know the distribution of capital stocks across firms, $\{z^1,z^2, \ldots ,z^M\}$. However, because of Johansen's assumption of no substitution between variable inputs for a fixed capital stock, he is able to combine (a) and (b) by specifying a capacity distribution $\{\xi^m{}_1,\xi^m{}_2,y^m\}$ $m = 1,2, \ldots ,M$ for the industry, where $\xi^m{}_j$ is the jth input-output coefficient for the mth firm and y^m is the capacity for the mth firm. From a discrete capacity distribution for the input-output coefficients, it is a short step to a continuous capacity distribution, and in Johansen (1972), various functional forms for continuous capacity distributions are assumed (including the case of the Pareto distribution pioneered by Houthakker, 1955–56) and the resulting macro production functions are calculated. The important conclusion that there can be considerable substitution at the macro level (even though there is none at the micro level) is emphasized by Johansen.

However, the putty-clay restrictions Johansen places on the functional form for $\emptyset$ are unduly restrictive. Instead of assuming a distribution of input-output coefficients, an empirically richer and computationally simpler model would result if we assumed a "flexible" functional form for $\emptyset$ (or its dual) and a multivariate distribution of fixed inputs.[55] For example, let us combine output(s) and variable inputs in the vector y (where, as usual, inputs are indexed with a negative sign), let z be a vector of fixed inputs, and write the ex ante micro production function in implicit form as $\emptyset(y,z) = 0$. Let p^0 be a vector of variable output and input prices and define a firm's variable profit function as $\pi(p^0;z^0) \equiv \max_y\{p^0 \cdot y\colon \phi(y;z^0) = 0\}$. Now, if π is differentiable with respect to p, it turns out that the firm's (variable) profit maximizing supply and demand functions, $y(p^0;z^0)$, can be obtained by differentiating π with respect to p;[56] that is, $y(p^0;z^0) = \nabla_p\pi(p^0;z^0)$. Now suppose there are M firms in the industry and let the multivariate distribution of fixed inputs in the industry be represented by the multivariate density function $f(z)$; that is, the number of firms having a combination of fixed factors falling between z_1 and z_2 is approximated by the number $M\int_{z_1}^{z_2} f(z)dz$. The short-run industry variable supply and demand functions $Y(p^0)$ can be obtained by integrating the firm supply and demand functions over the distribution of fixed inputs; that is, $Y(p^0) \equiv M\int_z \nabla_p \pi(p^0;z)\, f(z)dz$, and the industry short-run profit function may be defined as $\Pi(p^0) \equiv M\int_z \pi(p^0; z) f(z)dz$, which has as its dual the industry short-run transformation function $F(Y) = 0$. The Houthakker-Johansen putty-clay assumption could be tested in this framework by assuming an appropriate functional form for $\pi(p;z)$.

Unfortunately, this approach to the problem of aggregation over sectors requires (a) information on aggregate variable outputs and inputs, and their prices, and (b) detailed information on the distribution of fixed inputs by firm, information that is not generally available.

This completes our discussion of the general problems of aggregation over goods and over sectors. We now turn our attention to some specific problems associated with the aggregation of capital that we have not yet discussed.

8.4 The Aggregation of Capital

8.4.1 Capital: A Stock or a Flow of Services

Jorgenson and Griliches (1967, p. 257) note that: "an almost universal conceptual error in the measurement of capital input is to confuse the aggregation of capital stock with the aggregation of capital service." They go on to note that the aggregation procedure appropriate for measuring real investment is *not* appropriate for measuring real capital:

> In converting estimates of capital stock into estimates of capital services we have disregarded an important conceptual error in the aggregation of capital services. While investment goods output must be aggregated by means of investment goods or asset prices, capital services must be aggregated by means of service prices.
>
> The prices of capital services are related to the prices of the corresponding investment goods; in fact, the asset price is simply the discounted value of all future capital services. Asset prices for different investment goods are not proportional to service prices because of differences in rates of replacement and rates of capital gain or loss among capital goods. [Jorgenson and Griliches 1967, p. 267]

Thus, in the Jorgenson-Griliches framework, the user cost of capital in period t, p_t must be distinguished from the purchase cost Q_t of the capital good. The easiest way of deriving the rental price p_t from the purchase price Q_t is to pretend that firms lease all their capital goods at rental price p_t from the "leasing" firm. Competition presumably forces the "leasing" firm to earn the going rate of return r, after corporate income tax, on its leasing activities; thus we have the following equality: {purchase cost of one unit of the capital good plus corporate and property tax expenses minus rental received during the period} $(1 + r) =$ depreciated value of the capital good next period; in symbols we have:

$$(104) \qquad \{Q_t + u_t[p_t - v_t\delta Q_t - x_tQ_t] + x_tQ_t - p_t\}(1 + r) = (1 - \delta)Q_{t+1},$$

where u_t is corporate income tax rate, v_t is the proportion of depreciation allowable for tax purposes, δ is the one-period combined depreciation and obsolescence rate for the capital good, x_t is the property tax rate, and Q_{t+1} is next period's expected purchase price for one unit of the capital good. Now equation (104) may be solved for p_t:[57]

$$(105) \qquad p_t = \{rQ_t + \delta Q_{t+1} - (Q_{t+1} - Q_t) - (1 + r)u_tv_t\delta Q_t + (1 + r)(1 - u_t)x_tQ_t\}/(1 + r)\ (1 - u_t).$$

If "leasing" firms do not exist, then the rental formula (105) can also be derived by setting up the firm's profit maximization problem. For example, consider the following specific profit maximization problem:

$$(106) \qquad \max_{\substack{w.r.t.\\ Y,W,B,K}} \quad p_yY - w_WW - w_BB - \left\{Q_t - \frac{(1-\delta)}{1+r}Q_{t+1} + x_tQ_t\right\}K - u_t\{p_yY - w_WW - w_BB - v_t\delta Q_tK - x_tQ_tK\},$$

subject to $Y = f(W,B,K)$, where p_y is the price of one unit of output Y, w_W is the white-collar wage rate, w_B is the blue-collar wage rate, W and B are the inputs of white- and blue-collar labor, and f is the firm's production function. The maximand (106) can be rewritten (after some algebraic manipulations) as:

$$(107) \qquad (1 - u_t)\ \{p_yY - w_WW - w_BB - p_tK\},$$

where p_t is defined by (105). Thus, whether "leasing" firms exist or not, p_t defined by (105) is the appropriate user cost of a unit of (corporate) capital, and it appears to be the price which should be used to weight that particular component of the capital stock K, just as p_y is the appropriate price to weight Y, and so on.

Once rental prices for the various components of the capital stock have been determined, the aggregation techniques discussed in sections

8.2 and 8.3 can be used to form estimates of the aggregate stock of capital services (or components of this aggregate). This is essentially the procedure followed by Jorgenson and Griliches (1967).

Under the assumptions above, rental prices of capital goods can be treated *symmetrically* (in the theory of production and productivity analysis) to output and variable input prices. Thus it seems that the aggregation of capital is no more difficult than the aggregation of anything else, such as labor, intermediate goods, or output. This is the position taken by Bliss (1975, p. 144).

However, even if in *theory* the aggregation of capital does not appear to be any more difficult than the aggregation of say, labor, in *practice* it is very much more difficult to construct a capital aggregate that researchers can agree is appropriate for the purpose at hand.[58]

In the following section we will consider some of the practical difficulties (and points of controversy) involved in the construction of capital aggregates using rental price formulas similar to (105) above. In the present section, I will attempt to relax somewhat the simplifying assumptions that allowed us to construct the rental price formula (105).

Thus far my treatment of the capital aggregation problem has made two fundamental simplifying assumptions: (*a*) a depreciated durable good is measured in units of the undepreciated good (i.e., *vintages* are not distinguished), and (*b*) a durable good is assumed to *evaporate* or depreciate at a rate that is independent of "normal" use (and independent of the vintage of the capital good); specifically I have assumed a *constant evaporation rate model.*

Jorgenson (1965, p. 51) has argued that the assumptions above are not as restrictive as they might first appear from the viewpoint of empirical applications; but, nonetheless, since they are restrictive (cf. Feldstein and Rothschild 1974), I shall indicate how they can be relaxed in a model where capital appears as both an input and an output.

The very general model of producer behavior that I propose to utilize was developed by J. R. Hicks in *Value and Capital* (1946, chap. 15; see also Malinvaud 1953; Bliss 1975).

Hicks (1946, pp. 193–94) assumes that producers make production plans at the beginning of period 1 that will extend to period n. The plan consists of a list of inputs and outputs for each period, where period 1 inputs include the firm's existing stocks of durable equipment, distinguished by physical characteristics and vintages. Hicks thinks of period n as the period when the firm winds up its affairs and sells all its remaining durable equipment, so that the list of period n outputs will include the firm's depreciated capital stock that will be left over at the end of the period (or at the beginning of the following period). Thus, if we assume that $n = 1$, the Hicksian intertemporal production model reduces to the following profit maximization model:[59]

(108) $$\max_{z,y,x}\{(1+r)^{-1}p_z \cdot z + p_y \cdot y - p_x \cdot x : (z,y, -x) \in S\} \equiv \Pi([1+r]^{-1}p_z,p_y,p_x),$$

where $x =$ a nonnegative N_1-dimensional vector of current period inputs, including the firm's beginning of the *current* period stocks of durable equipment.

$y =$ a nonnegative N_2-dimensional vector of *current* period outputs.

$z =$ a nonnegative N_3-dimensional vector whose components represent how much durable equipment the firm will have available to it at the beginning of the *following* period,

$p_x =$ an N_1-dimensional vector of nonnegative *current* period (purchase) prices for inputs.

$p_y =$ an N_2-dimensional vector of nonnegative *current* period prices for outputs,

$p_z =$ an N_3-dimensional vector of nonnegative *expected following* period prices for the firm's (depreciated) durable equipment.

$r =$ the one-period interest rate at which the firm can borrow or lend, and

$S =$ the firm's production possibility set, which is assumed to be convex, nonempty, and closed.

We note that Hicks (1946, p. 230) assumes that S is *smoothly convex* (i.e., that the boundary of the convex set Y can be described by a twice-differentiable surface), while von Neumann (1945–46, p. 2) and Morishima (1969, pp. 29–94) assume that S is a *polyhedral convex* set (i.e., S can be described as the set of all convex combinations of a finite number of activities).

If producer durables evaporate at a constant rate that is independent of the firm's utilization of other inputs, then the shape of the production possibility set S will be restricted, and it is easy to see that the profit maximization model (108) reduces to a profit maximization problem similar to (106). However, in general, a firm can prolong the life (and hence the value) of its durable equipment by spending more on inputs of maintenance labor and on inputs of replacement parts. Thus we should distinguish at least two broad types of labor input: production labor and maintenance labor. Increased inputs of the first type of labor will generally lead to smaller outputs of capital equipment available at the beginning of the following period, and vice versa for maintenance labor.

The model of short-run profit maximization defined by (108) is capable of being interpreted in several ways, depending on how narrowly capital goods are classified.

If capital goods are not distinguished according to vintage, then the dimensionality of the x and z vectors will be relatively small, and (108) reduces to a *variable evaporation rate model.*

If capital goods are distinguished not only according to their vintage but also by their physical condition (e.g., trucks may be classified according to how many miles they have been driven, structures may be distinguished by whether or not they have been painted recently, and so on), then the dimensionality of the x and z vectors will be enormous, and (108) may be termed a *"vintage" variable evaporation rate model.*

Some special cases of the very general Hicksian intertemporal model of production have appeared in the literature: (*a*) Taubman and Wilkinson (1970) assume that the physical amount of depreciation per unit of capital per unit of time depends on an index of capital utilization; (*b*) Schworm (1977) assumes that depreciation depends on an index of capital utilization (miles driven in the case of his empirical example using truck data) *and* units of maintenance; while (*c*) Epstein (1977) actually implements a highly aggregated model based on equation (108) using aggregate United States manufacturing data, but his empirical results are not very favorable to the Hicksian intertemporal model (perhaps owing to aggregation problems). On the other hand, Schworm is able to derive a formula for the rental price of capital that is similar to (105) above, except that utilization and maintenance variables also appear in the formula; but the other output and input variables pertaining to the firm do *not* appear in his formula, which makes the construction of rental prices for components of the capital stock much easier than in the more general Hicksian intertemporal model of production.

How can we in fact construct a capital aggregate based on the Hicksian short-run profit maximization model (108), and how will the resulting aggregate differ from a capital services aggregate constructed by means of an index number formula using rental prices similar to (105) above as weights? In the context of the Hicksian model, it is clear that we can construct several capital aggregates that must be carefully distinguished: (*a*) a current-period capital stock aggregate (an input from the viewpoint of the current period) using current-period capital stock prices as weights in the aggregation procedure; (*b*) a (depreciated) following-period capital stock aggregate (an output from the viewpoint of the current period) using discounted expected following-period capital stock prices as weights; (*c*) a current-period investment aggregate (an output) using current-period investment goods prices as weights in the aggregation procedure; and (*d*) a capital aggregate that is an aggregate of (*a*) and (*b*) where capital as an input and capital as an output are oppositely signed in the index number formula that is used. This "Hicksian" capital services aggregate[60] should be comparable to the Jorgenson-Griliches capital aggregate discussed earlier, except that the assumption of constant evaporation rates is not required.

Although the Hicksian model of producer behavior and the corresponding capital aggregates are more appealing than the constant evaporation rate model and the Jorgenson-Griliches capital aggregate, there is a major problem in implementing the Hicksian model; that is, the necessary data do not exist at present. Data on the market value and condition of the firm's beginning of the period holdings of durables are either: (*a*) nonexistent, (*b*) extremely aggregated, or (*c*) conventionally determined according to depreciation rules used for tax purposes.[61]

I will conclude this section by considering a problem Usher raised in his introduction: Should expenditures on maintenance and repair be lumped with expenditures on capital goods? If we have enough data (and we are willing to make the necessary imputations) to implement the approach to capital aggregation based on the Hicksian model of producer behavior defined by (108), then maintenance and repair should not be lumped with capital expenditures. A similar conclusion should hold if the model is based on the constant evaporation rate model (106), since we would expect maintenance and repair expenditures to change δ, the depreciation rate on the existing capital stock.

8.4.2 Special Problems in the Aggregation of Capital

In the previous section I may have left the impression that from a theoretical point of view constructing a capital aggregate is no more difficult than constructing a labor aggregate.[62] In the present section, I will readjust this impression by cataloging some of the practical difficulties and sources of controversy that occur when researchers attempt to construct capital aggregates that are suitable for estimating production functions or for estimating total factor productivity.

Producer's Expectations of Future Prices

Whether we construct a capital services aggregate using the constant evaporation rate model (equation 106) or the variable evaporation rate model (equation 108), it is necessary to estimate the producer's *expectations* about next period's capital stock prices (recall the price Q_{t+1} in eq. 105 and the expected prices $p_z/(1+r)$ in eq. 108). These expected prices are generally unobservable, and thus reasonable analysts could differ widely on how to estimate them. For example, Christensen and Jorgenson (1969, 1970) assume that producers perfectly anticipate next period's stock prices, whereas Woodland (1972, 1975) and many others[63] assume that producers expect current stock prices to prevail in the following period (static expectations or Hicks's [1946, p. 205] unitary elasticity of expectations). A third alternative (followed by Epstein 1977) is to use a forecasting model to predict next period's asset prices based on past information about the asset prices. It seems clear that the first two methods (perfect anticipations and static expectations) for forming expected prices are not generally correct, while the third alternative requires extensive econometric modeling expertise, which individual producers and accountants may not possess.

Another related difficulty must be mentioned at this point. We have been constructing aggregates under the assumption that producers are maximizing profits subject to their technological constraints, assuming that they are facing *known* prices for selling their outputs and buying their inputs. We have been assuming that there was no *uncertainty* involved in the individual producer's profit maximization problems, and thus that their attitude toward risk and uncertainty was irrelevant. However, since future-period prices of capital goods are not known with certainty, it is clear that our underlying profit maximization models (e.g., eq. 106 or 108) must be modified to incorporate producers' attitudes toward risk. This leads to a great number of complications[64] whose implications for the construction of aggregates have not been fully worked out.

The problem of modeling uncertainty is related to the problem of modeling the formation of expectations, in the sense that neither problem would exist (at least in theory) if there were sufficient future and insurance markets, for then the appropriate prices could be observed in the market. However, in the absence of these markets, the analyst who wishes to construct a capital services aggregate will be forced to make an *imputation* or assumption about future expected prices.

Jorgenson and Griliches (1967) have been criticized (cf. Denison 1969, pp. 6–12; and Daly 1972, pp. 49–50) for including capital gains terms in their rental price formulas for capital services, which they use as weights in order to aggregate different components of the capital stock into a capital aggregate. However, from our rather narrow viewpoint, which concentrates on the measurement of capital in the context of production function estimation and the measurement of total factor productivity, it seems clear that the capital gains term belongs in the rental price formula—what is not as clear is the validity of the Jorgenson-Griliches perfect anticipations assumption.

Interest Rates

The rental price formula (105) and the profit maximization problems (106) and (108) in the previous section all involve an interest rate r. Which r should be used? If the firm is a net borrower, then r should be the marginal cost of borrowing an additional dollar for one period, while if the firm is a net lender, then r should be the one-period interest rate it receives on its last loan. In practice, r is taken to be either (a) an exogenous bond rate that may or may not apply to the firm under consideration, or (b) an internal rate of return. I tend to use the first alternative, while Woodland (1972, 1975) and Jorgenson and his coworkers[65] use the second. As usual, neither alternative appears to be correct from a theoretical a priori point of view; so, again, reasonable analysts could differ on which r to use in order to construct a capital aggregate.[66]

Note that the appropriate interest rate is a *nominal* (not "real") rate of interest: any (anticipated) inflation should be taken into account by the (anticipated) capital gains term in the user cost formula (105).

Depreciation Rates

The user-cost formula in the previous section involved a depreciation rate δ. I have already commented that, in theory, the assumption of an exogenous evaporation rate δ is not warranted; but suppose that data limitations forced us to estimate a constant δ, or perhaps a series of δs, $\{\delta_s\}$, say, where δ_s would be the one-period evaporation rate applicable to a certain component of the capital stock that was s periods old. The depreciation rates $\{\delta_s\}$ are used not only in constructing rental prices in the Jorgenson-Griliches framework, but also by other analysts in constructing capital stock series from deflated investment series (cf. Kendrick 1961, 1976; Denison 1974).

What depreciation rates $\{\delta_s\}$ are to be used, and how are they to be constructed? There is considerable controversy in this area, much of it being very ably reviewed by Creamer (1972, pp. 62–68). Two relatively extreme positions can be discerned in the literature, one used by Jorgenson and his co-workers (constant evaporation; i.e., $\delta_s = \delta$ for all s) and the other used by Denison and Kendrick (one-horse-shay depreciation; i.e., $\delta_s = 0$ for all s, except $s = T$ when $\delta_T = 1$). Actually, Denison's depreciation assumptions are not quite as extreme as one-horse-shay depreciation, as the following quotations indicate:[67]

> It is not assumed that all of the investment in a category made in a particular year disappears from the gross stock simultaneously, after expiration of the average service life. Instead more realistically, retirements are dispersed around the average service life. The Winfrey S-3 distribution is used to obtain this dispersion. [Denison 1974, pp. 53–54]

> To introduce an allowance for rising maintenance expense and deterioration of capital services with the passage of time, I have adopted the following expedient. To measure input of structures and

> equipment I have used a weighted average of indexes of the gross stock and net stock based on straight-line depreciation, with the gross stock weighted three and the net stock one. [Denison 1974, p. 55]

On the other hand, for productivity comparisons, Kendrick prefers to use one-horse-shay depreciation "in order to relate real product to the comparable real capital stock estimates on a gross basis rather than a net basis" (Kendrick 1974, p. 20), but he also constructs estimates of a net capital stock using double declining-balance depreciation.

Two questions arise with respect to assumptions made about the depreciation, deterioration, and length of lives of components of the capital stock: (*a*) Do the assumptions make much difference empirically? and (*b*) What is the empirical evidence on the appropriateness of the various assumptions?

The answer to the first question appears to be an emphatic yes. Capital stocks constructed on the basis of different depreciation assumptions can differ considerably.[68]

Some negative empirical evidence on the validity of the constant evaporation rate form of depreciation (or declining-balance or geometric depreciation, as it is sometimes called) is reviewed by Feldstein and Rothschild (1974). Hulten and Wykoff (1977) utilize the theoretical model developed by Hall (1968) to estimate economic depreciation for various types of structures used in the United States manufacturing sector. They found that, in most cases, a constant geometric rate of depreciation could approximate the "true" rate of depreciation rather well, with the exception of the earliest years of the asset's life.

Overall, one can only conclude that empirical information on depreciation rates and lengths of lives of assets is scanty, and I can only echo the recommendations of others that governments devote more resources in order to obtain more information.

Treatment of Indirect Taxes

Indirect taxes in a national income accounting framework are generally defined as an amalgam of taxes on outputs produced by firms (sales taxes and various excise taxes) plus taxes on various inputs (including customs duties, real and property taxes, social insurance levies, and sometimes universal pension plan levies). There has been some controversy over where these taxes should be allocated when constructing a capital aggregate:

> The treatment of indirect taxes, property taxes, and corporate profits taxes can affect the income share of the capital-land category, and also the distribution to assets within that category. The choices of national income at market prices or factor costs for weights is in-

> fluenced by this question. This question had come earlier in a review of Solow's book, *Capital Theory and the Rate of Return*, as Solow had included both indirect taxes and corporate profits taxes in estimating the share of income to property. [Daly 1972, p. 49]

From the point of view that underlies the user-cost formula (eq. 105) (which was based on the assumption that producers competitively maximize profits subject to their technological constraints), the treatment of indirect taxes seems clear: indirect taxes (such as property taxes) that fall on durable inputs *should* be included in the user-cost formula for that input, other indirect taxes (such as Social Security payments) that fall on variable inputs *should* be added to the market price of those inputs, whereas indirect taxes (such as sales taxes) that fall on outputs should *not* be added to the market price of these outputs. The conceptually correct prices in the Jorgenson-Griliches framework are the output prices that reflect the revenue actually received by the firm and the input prices that reflect the actual costs paid by the firm for the use of the inputs involved in the production process.

Thus customs duties and tariffs *should* be added to the prices of various imported goods used by firms, but sales taxes imposed on the outputs of a firm as they are sold to households (or other nonbusiness sectors) should *not* be added to the firm's selling prices. But how should we treat (intermediate good) sales taxes imposed on the outputs of a firm (1, say) as the goods are sold to another firm (2, say)? Obviously the tax should not be added to the selling price of firm 1, but it should be added to the selling price of firm 1 when the good is treated as an input into firm 2.[69]

The Form of Business Organization

The user-cost formula (105) developed in the previous section implicitly assumed that the firm was an incorporated firm and thus faced the appropriate corporate tax rate. However, if the firm is not incorporated, then the appropriate tax rate is the owner's *personal* (marginal) tax rate, which will generally differ from the corporate rate. Thus Christenson and Jorgenson (1969) construct rental prices for the components of real capital input, disaggregated by class of asset and by legal form of organization. This appears to be a worthwhile methodological innovation, although reasonable analysts may find fault with some of the specific details of the Christensen-Jorgenson construction.

Weighting the Components of Capital

To construct real capital input, Kendrick (1972, p. 101) and Denison (1974, p. 51) favor weighting components of the captial stock by the components share of property income[70] while Kendrick (1976) simply adds constant-dollar components of the capital stock. None of these weights appear to agree with the user-cost weights we would obtain using formula (105), except in very restrictive circumstances: using the notation of the previous section, the suggested weights that would replace the Jorgenson-Griliches user cost (105) are rQ_t and Q_t respectively, where Q_t is the current period asset price and r is a (gross or net of depreciation) rate of return.

Since rQ_t and Q_t do not have a depreciation term, these weights tend to be much smaller than the Jorgenson-Griliches user-cost weights, and thus measures of real factor input (see section 8.5 below) using Jorgenson-Griliches weights tend to grow faster than Kendrick-Denison measures of real factor input, since capital typically has grown faster than labor during the past century. Thus the question of which weights to use when constructing a capital services aggregate is not empirically unimportant (cf. Denison 1969 and Jorgenson and Griliches 1972).

From the viewpoint of our restrictive theoretical model of production, it seems clear that the Jorgenson-Griliches weights are to be preferred over the Kendrick-Denison weights.[71]

Leased versus Owned Capital in the National Accounts

There is a problem in using national income accounting data to estimate sectoral capital stocks that must be mentioned here.[72] The problem is that all rented or leased components of a firm's capital stock appear as a primary input in the finance, insurance, and real estate sector and as an intermediate input in the firm's sector. This creates problems when sectoral "value added" production functions are estimated, since the sectoral capital services input will be too low.[73] Thus it would be helpful if official accounts were to provide a breakdown on which sector actually used the services of a leased component of the capital stock.

The Domain of Definition of Capital

Another fundamental problem in constructing a capital aggregate that we have not yet faced is the issue Usher raised in his introduction to this volume: Should capital be defined as an aggregate of produced means of production or as an aggregate of produced and nonproduced (natural resource) factors of production? Obviously, this is a definitional matter that could be decided either way. However, if we opt for the first definition of a capital aggregate and are interested in estimating aggregate production functions or explaining productivity change, then it is *essential* that we construct an aggregate for the noncapital, nonlabor, nonproduced primary inputs (such as land and natural resources), since omitting this latter aggregate ("land") will bias estimates of aggregate production functions as well as estimates of total factor productivity. This point has some applications for the current system of national income accounting in most countries, where land (and natural resources in general) is given a very minor role,[74] partly because of data limitations, but partly because researchers have focused for the most part on reproducible capital and neglected the contribution of nonreproducible resources.

A second problem with the definition of capital is that some national income accounting systems do not include inventories and goods in process as components of the capital stock. The essence of a capital good seems to be that it is a produced good, part of which lasts longer than the period under consideration. Thus, in agriculture, inventories of farm animals and feed and seed are generally large and should be included as components of the capital stock. This neglect of inventories seriously biases downward the contribution of capital in most industries. This point is well recognized, and Denison (1974), Kendrick (1976), and Christensen, Cummings, and Jorgenson (1976) all include inventory stocks as components of their capital stocks. However, when estimating production functions, many researchers (e.g., Woodland 1975) omit inventories as components of their capital stock series.

A final, related definitional problem has been raised by Creamer:

> Capital input is typically restricted to some combination of tangible assets although every analyst knows that an enterprise requires financial assets, (cash and accounts receivable) as well as tangible assets in order to function. . . . However, financial assets lead a double life—one entity's claim is another entity's obligation. Thus, at the level of aggregation of the national economy financial claims and obligations cancel each other, except for the net balance of international claims which have been a relatively small part of U.S. stock of capital. If this is the reason for the exclusion of financial capital, it constitutes, in my view, still another argument in favor of a disaggregative approach. [Creamer 1972, p. 60]

An individual firm will generally hold an "inventory" of financial capital (or working capital, as it is sometimes called) during our Hicksian period, and the cost of holding this "inventory" is just as real a cost to the firm as a payment to labor. Aggregating across firms in the private business sector of an economy will not generally cancel out these financial claims: they will cancel only if we include households, governments, and the rest of the world in the aggregate.

That financial capital is similar to physical inventories in some sense (both represent a real cost to a firm and to the private business sector of an economy) suggests that financial capital be treated like any other durable input in our accounting framework, and that financial capital should be included in any capital services aggregate, particularly since Creamer (1972, p. 60) suggests that there may have been substantial productivity gains in the use of financial capital since 1929, at least in

United States manufacturing. Unfortunately, we *cannot* insert financial capital into our production function as just another argument: the amount of financial capital a firm will require to produce a given output, given a vector of physical inputs, is simply not a well-defined quantity.

One way of proceeding would be to exclude financial capital as an input into the firm's "physical" production function, but to include financial capital as an input (along with labor, office space, etc.) into a "transactions" technology that would have the responsibility for *selling* the physical outputs the firm produces in its "plant" and *purchasing* the inputs the "plant" requires. Normal credit arrangements and payment procedures could be worked into the transactions technology.

We have drifted into the domain of monetary theory, and the reader is referred to Fischer (1974) and Nagatani (1978) for further references and suggestions. At this point we can only conclude that Creamer has pointed out a serious conceptual omission from most capital aggregates and that it is not immediately clear how we can insert financial capital into a capital aggregate using our naive production model.

Another definitional issue with respect to the scope of a capital aggregate has been raised by Christensen and Jorgenson's[75] inclusion of the stock of household consumer durables in their recent estimates of the capital stock. Creamer makes the following comments on their procedure:

> This is a puzzling addition. . . . It is certainly inconsistent with the underlying definition of a capital good—one that is used to produce other goods and services. Moreover, it is inconsistent with their own (Christensen-Jorgenson) statement that "the main analytical use of the production account is in the study of producer behavior. Revenue and outlay must be measured from the producer's point of view." . . . Moreover, the inclusion of consumer's durables in the capital stock understates aggregate total factor productivity since the methodology of estimates is such that this sector makes no contribution to productivity.[76] [Creamer 1972, p. 61]

Some further comments seem warranted. Christensen and Jorgenson have included in the private production sector of an economy the "process" that converts household stocks of consumer durables into service flows. Kendrick (1976) has also added consumer durables to the capital stock, justifying the procedure as follows: "This is merely an extension of the treatment presently accorded owner-occupied residential structures and may be justified by the argument cited above—that shifts in sector ownership patterns should not affect investment, capital, or the associated income estimates" (Kendrick 1976, pp. 5–6).

Thus, for some purposes the inclusion of consumer durables in a capital aggregate can be justified.[77] However, since this discussion of capital aggregation is in the context of the estimation of total factor productivity and production function estimation for the business sector of an economy that is oriented toward maximizing private profit, I would not recommend including consumer durables in a capital aggregate, since a household's conversion of durable stocks into flows is not generally a genuine business activity.

To conclude our discussion about the domain of definition of a capital aggregate, let us examine Kendrick's (1976) rather comprehensive definition of the capital stock. Kendrick includes the following items in his capital stock estimates for the United States:

1. *business nonhuman tangibles*, consisting of structures, land, natural resources, machinery and other durable equipment, and inventory stocks used in the private business sector;
2. *household nonhuman tangibles*, consisting of household residential real estate, automobiles, other durable goods, and household inventories;
3. *government nonhuman tangibles*, consisting of government structures, machinery and equipment, and public capital (such as highway construction);
4. *human tangibles*, which are defined as outlays required to produce mature human beings (rearing costs);
5. *research and development expenditures*;
6. *education and training expenses*;
7. *health and safety expenditures* (one-half of all outlays for health and safety, which reduce mortality and disability are taken as representing investment); and
8. *mobility payments*, which includes portions of unemployment insurance benefits paid, job search and hiring expenses, and moving expenses.[78]

From our narrow viewpoint, concerned with productivity and production function estimation for the private business sector of an economy, we would not recommend the inclusion of any of Kendrick's capital stock components beyond item (1), business nonhuman tangibles, since the main effect of the investments listed in items (3) to (8) is to change the prices and possibly the qualities of the inputs a private firm utilizes; but these price changes are quite consistent with our model of producer behavior and do not require any special treatment. However, we can discern three possible exceptions to the general statement made above.

First, if portions of the government capital stock (item 3 above) are leased to private firms, these rentals could be treated as intermediate inputs into the private sector. It would also make sense to aggregate these rentals of government capital together with the corresponding privately owned components of the business capital stock.

Second, the existence of certain "free" government-provided goods such as highways creates some conceptual difficulties with our basic profit maximization problems listed in section 8.4.1. For example, consider the first profit maximization problem in section 8.4.1, modified to allow for the existence of government public goods:

(109)
$$\max_{\substack{w.r.t.\\ Y,W,B,K}} (1-u_t)\{p_yY - w_WW - w_BB - p_tK\}$$
$$\text{subject to } Y = f(W,B,K,R_1,R_2),$$

where all variables have been defined in section 8.4.1 except $R_1 \equiv$ number of miles of nonfreeway road utilized by the firm and $R_2 \equiv$ number of miles of freeway used by the firm. If f is differentiable and an interior solution to the profit maximization problem exists, Y^*, W^*, B^*, K^*, say, then the solution will satisfy the following first-order necessary conditions for (109):

(110) $$p_y \frac{\partial f}{\partial w}(W^*,B^*,K^*,R_1,R_2) = w_W$$

(111) $$p_y \frac{\partial f}{\partial B}(W^*,B^*,K^*,R_1,R_2) = w_B$$

(112) $$p_y \frac{\partial f}{\partial K}(W^*,B^*,K^*,R_1,R_2) = p_t.$$

The value of the marginal products of R_1 and R_2 can be defined as

(113) $$p_y \frac{\partial f}{\partial R_1}(W^*,B^*,K^*,R_1,R_2) \equiv p_1 \geq 0 \text{ and}$$

(114) $$p_y \frac{\partial f}{\partial R_2}(W^*,B^*,K^*,R_1,R_2) \equiv p_2 \geq 0,$$

respectively. Define the optimal output as $Y^* \equiv f(W^*,B^*,K^*,R_1,R_2)$. If the production function exhibits constant returns to scale in all five inputs, then Euler's theorem on homogeneous functions implies

(115) $$p_yY^* = w_WW^* + w_BB^* + p_tK^* + p_1R_1 + p_2R_2.$$

If p_1 or p_2 are positive, then the firm will capture the positive marginal products of the two free inputs R_1 and R_2, thus making excess profits. However, if there is free entry into the industry, new firms will enter the industry and the price of the output, p_y, will tend to fall. In fact, if the production function $f(W,B,K,R_1,R_2)$, with R_1 and R_2 held fixed, exhibits initially increasing returns to scale and eventual decreasing returns to scale in W,B,K,[79] then for industry equilibrium the price of output will be close to the average cost of production; that is, the following equation should (almost) hold for an individual firm in the industry:

$$(116) \qquad p_y Y^* = w_W W^* + w_B B^* + p_t K^*.$$

Equations (115) and (116) imply that the shadow prices of the two types of road, p_1 and p_2, should be close to zero.

With this highly simplified model in mind, we can return to our discussion of whether to include certain government capital goods such as highways in a capital aggregate. The answer (from the viewpoint of the economic theory of production) appears to be yes, except that the price weights for these government capital stock components will be shadow prices whose magnitude will not generally be known. However, if we assume competitive producer behavior with free entry into each industry using the public capital goods, then the price weights should be close to zero. In this case government capital goods would not show up in a capital aggregate constructed according to the index number formulas discussed in section 8.2.[80]

The third possible exception to my general statement that items (2) through (8) of Kendrick's capital aggregate should not be included in a capital aggregate based on my naive economic model of producer behavior is item (5), research and development expenditures. However, I am unable to make any concrete recommendations on just how research and development expenditures should be treated when forming a capital aggregate: it depends on how R&D enters the underlying economic model upon which we base our aggregation procedures.[81]

The Time Period

This discussion of capital aggregation based on the Jorgenson-Griliches economic model of producer behavior has thus far proceeded under the assumption that all components of the capital stock are freely variable during the period under consideration. Obviously, as we shorten our Hicksian period from, say, a decade to a week, an increasing number of inputs will become *fixed* rather than variable, and in these cases (observable) market prices should be replaced by (unobservable) shadow prices, which equal the value of the marginal products of the fixed inputs. Since these shadow prices are not generally observable, it will not generally be possible to construct capital aggregates based on our model of producer behavior when the time period becomes short enough to cause components of the capital stock to become fixed.

In view of this, one might think that capital stock aggregates based on annual data would be "better" than ones based on weekly data. This is not the case, however: the annual model of producer behavior assumes that all inputs are freely variable and that the prices the producer faces remain constant throughout the year when neither assumption is actually satisfied in practice, and thus neither annual nor weekly aggregates constructed on the basis of market data will be precisely equal to the "correct" aggregates, constructed using the appropriate shadow prices.

In section 8.7 I will discuss how to build up "annual" aggregates from "weekly" aggregates, assuming that the "weekly" aggregates have been constructed correctly from the viewpoint of our economic model of producer behavior.

Choice of Index Number Formula

> Unfortunately, there is no unique solution to the index number problem. Fisher's "ideal" index number, which employs a geometric mean of the weights from both periods in a binary comparison, is neat, but it has no more fundamental economic rationale than using either first or last period weights! [Kendrick 1972, p. 95]

In section 8.2.4 I argued that there was a strong economic rationale for using Fisher's ideal index number formula, since it is a *superlative* index number formula; that is, it corresponds to a *flexible* functional form, for the underlying production function. Moreover, we indicated that all superlative index number formulas approximate each other to the second order if changes in prices and quantities between the two periods are small, while the more commonly used Paasche or Laspeyres formulas approximate superlative indexes to the first order only.

Given that the economic justification for using a superlative index number formula seems fairly strong, should we use the fixed-base method for forming a capital aggregate, or should we compare each period with the immediately preceding period—that is, use the chain principle? In section 8.2.4 I argued for the use of the chain principle, since, if it is used, price and quantity changes should be small, and all superlative index number formulas should generate virtually the same aggregate series, so that the choice of a specific superlative index number formula becomes empirically irrelevant.

To conclude this section, let me note that the construction of a capital aggregate is fraught with both theoretical and empirical difficulties, even taking it as given that we wish to construct an aggregate that would be used in the context of production function or productivity estimation. It appears to me that the major conceptual problems are in the determination of producer's expectations about future prices and how to deal with the resulting uncertainty, while the major practical difficulties are in the estimation of depreciation rates.

With the above difficulties firmly in mind, let us turn now to a closely related topic: how to construct estimates of total factor productivity in the context of our naive model of producer behavior.

8.5 Capital and the Measurement of Technical Progress

8.5.1 The Measurement of Total Factor Productivity with a Separable Technology

For a brief but useful survey of the literature on the measurement of economic growth and total factor productivity, see Christensen, Cummings, and Jorgenson (1976).

Jorgenson and Griliches (1967, 1972, pp. 83–84) advocated the use of the Törnqvist quantity index number formula Q_0 (recall section 8.2.3 of this chapter) and the corresponding implicit Törnqvist price index P_0 in the context of the measurement of total factor productivity. In this section I repeat Diewert's (1976, pp. 124–27) justification for their procedure.

Jorgenson and Griliches (1972) use the index number formula $Q_0(p^0,p^1;x^0,x^1)$ defined in section 8.2.3 not only to form an index of real input, but also to form an index of real output. Just as the aggregation of inputs into a composite input rests on the duality between unit cost and homogeneous production functions, the aggregation of outputs into a composite output can be based on the duality between unit revenue and homogeneous factor requirements functions.[82] I will briefly outline this latter duality.

Suppose that a producer is producing M outputs, $(y_1,y_2,\ldots,y_M) \equiv y$, and the technology of the producer can be described by a *factor requirements function*, g, where $g(y)$ = the minimum amount of aggregate input required to produce the vector of outputs y.[83] The producer's *unit* (aggregate input) *revenue function*[84] is defined for each price vector $p \geqq 0_M$ by

$$(117) \qquad r(p) \equiv \max_y\{p\cdot y : g(y) \leqq 1, y \geqq 0_M\}.$$

Thus given a factor requirements function g, (117) may be used to define a unit revenue function. On the other hand, given a unit revenue function $r(p)$ that is a positive, linearly homogeneous, convex function for $p \gg 0_M$, a factor requirements function g^* consistent with r may be defined for $y \gg 0_M$ by[85]

$$(118) \qquad \begin{aligned} g^*(y) &\equiv \min_\lambda\{\lambda : p\cdot y \leqq r(p)\lambda \text{ for every } p \geqq 0_M\} \\ &= \min_\lambda \{\lambda : 1 \leqq r(p)\lambda \text{ for every } p \geqq 0_M \text{ such that } p\cdot y \\ &= 1\} = 1/\max_p\{r(p) : p\cdot y = 1, p \geqq 0_M\}. \end{aligned}$$

As usual, the translog functional form may be used to provide a second-order approximation to an arbitrary twice-differentiable factor requirements function. Thus, assume that g is defined (at least over the relevant range of ys) by

$$\ln g(y^r) \equiv a_0 + \sum_{m=1}^{M} a_m \ln y^r_m + \tfrac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} c_{jk} \ln y^r_j \ln y^r_k, \text{ for } r = 0, 1, \tag{119}$$

where $\sum_{m=1}^{M} a_m = 1$, $c_{jk} = c_{kj}$, $\sum_{k=1}^{M} c_{jk} = 0$, for $j = 1,2, \ldots, M$.

Now assume that $y^r \rangle\rangle 0_M$ is a solution to the aggregate input minimization problem $\min_y\{g(y) : p^r \cdot y = p^r \cdot y^r, y \geqq 0_M\}$, where $p^r \rangle\rangle 0_M$ for $r = 0,1$, and g is the translog function defined by (119). Then the first-order necessary conditions for the minimization problems along with the linear homogeneity of g yield the relations $p^r/p^r \cdot y^r = \nabla g(y^r)/g(y^r)$, for $r = 0,1$, and using these two relations in lemma (59) applied to (119) yields

$$g(y^1)/g(y^0) = Q_0(p^0, p^1; y^0, y^1) \tag{120}$$

where Q_0 is the Törnqvist quantity index defined by (38).

Thus the Törnqvist formula can again be used to aggregate quantities consistently, provided the underlying aggregator function is homogeneous translog.

Similarly, if the revenue function $r(p)$ is translog over the relevant range of data and if the producer is in fact maximizing revenue, then we can show that $r(p^1)/r(p^0) = P_0(p^0,p^1; y^0,y^1)$, the Törnqvist price index.

Using the material above, we may now justify the Jorgenson and Griliches (1972) method of measuring technical progress. Assume that the production possibilities efficient set can be represented as the set of outputs y and inputs x such that

$$g(y) = f(x), \tag{121}$$

where g is the homogeneous translog factor requirements function defined by (119), and f is the homogeneous translog production function defined in section 8.2.4. Let $p^r \rangle\rangle 0_M$, $w^r \rangle\rangle 0_N$, $r = 0, 1$ be vectors of output and input prices during periods 0 and 1, and assume that $y^0 \rangle\rangle 0_M$ and $x^0 \rangle\rangle 0_N$ is a solution to the period 0 profit-maximization problem,

$$\max_{y,x}\{p^0 \cdot y - w^0 \cdot x : g(y) = f(x)\}. \tag{122}$$

Suppose "technical progress" occurs between periods 0 and 1, which we assume to be a parallel outward shift of the "isoquants" of the aggregator function f; that is, we assume that the equation that defines the efficient set of outputs and inputs in period 1 is $g(y) = (1 + \tau)f(x)$, where τ represents the amount of "technical progress" if $\tau > 0$ or "technical regress" if $\tau < 0$. Finally, assume that $y^1 \rangle\rangle 0_M$ and $x^1 \rangle\rangle 0_N$ is a solution to the period 1 profit-maximization problem,

$$(123) \qquad \max_{y\,x}\{p^1\cdot y - w^1\cdot x : g(y) = (1+\tau)f(x)\}.$$

Thus we have $g(y^0) = f(x^0)$ and $g(y^1) = (1+\tau)f(x^1)$. It is easy to see that $y^r \rangle\rangle 0_M$ is a solution to the aggregate input minimization problem $\min_y\{g(y) : p^r\cdot y = p^r\cdot y^r, y \geqq 0_M\}$, for $r = 0,1$, and thus (120) holds. Similarly, $x^r \rangle\rangle 0_N$ is a solution to the aggregator maximization problem $\max_x\{f(x) : w^r\cdot x = w^r\cdot x^r, x \geqq 0_N\}$, for $r = 0,1$, and thus (43) holds. Substituting (43) and (120) into the identity $g(y^1)/g(y^0) = (1+\tau)f(x^1)/f(x^0)$ yields the following expression for $(1+\tau)$ in terms of observable prices and quantities:

$$(124) \qquad (1+\tau) = \sum_{m=1}^{M} [y^1_m/y^0_m]^{\frac{1}{2}\left[\frac{p^1_m\, y^1_m}{p^1\cdot y^1} + \frac{p^0_m\, y^0_m}{p^0\cdot y^0}\right]} \Big/ \sum_{n=1}^{N} [x^1_n/x^0_n]^{\frac{1}{2}\left[\frac{w^1_n\, x^1_n}{w^1\cdot x^1} + \frac{w^0_n\, x^0_n}{w^0\cdot x^0}\right]}$$

$$= Q_0(p^0,p^1,y^0,y^1)/Q_0(w^0,w^1,x^0,x^1).$$

Thus the Jorgenson-Griliches method of measuring technical progress can be justified if: (a) the economy's production possibilities set can be represented by a *separable* transformation surface defined by $g(y) = f(x)$, where the input aggregator function f and the output aggregator function g are both homogeneous translog functions; (b) producers are maximizing profits; and (c) technical progress takes place in the *"neutral"* manner postulated above.[86]

Since the separability assumption $g(y) = f(x)$ is somewhat restrictive from an a priori theoretical point of view, it would be of some interest to devise a measure of technical progress that did not depend on this separability assumption. This can be done, as we shall see in the next section.

8.5.2 The Measurement of Total Factor Productivity in the General Case

Before analyzing a general M outputs, N inputs case, we warm up with the one output, N inputs case.[87]

Suppose the technology of the producer can be represented by the following (time modified) translog production function f:

$$(125) \qquad \ln f(x,t) \equiv \alpha_0 + \sum_{n=1}^{N} \alpha_n \ln x_n + \tfrac{1}{2}\sum_{i=1}^{N}\sum_{h=1}^{N} \alpha_{ih} \ln x_i \ln x_h + \beta_0\, t + \sum_{n=1}^{N} \beta_n\, t \ln x_n + \gamma t^2,$$

where $y = f(x,t)$ is output produced during period t, and $x \equiv (x_1, x_2, \ldots, x_N)$ is a vector of inputs used by the firm during period t. We note that f defined by (125) can provide a second-order approximation to an arbitrary twice continuously differentiable function of x and t. With the following restrictions on the parameters,

$$(126)\qquad \sum_{n=1}^{N} \alpha_n = 1,\ \alpha_{ih} = \alpha_{hi},\ \sum_{i=1}^{N} \alpha_{ih} = 0 \text{ for } h = 1, 2, \ldots, N \text{ and } \sum_{n=1}^{N} \beta_n = 0,$$

f defined by (125) is linearly homogeneous in x, and the resulting function can provide a second-order approximation to an arbitrary twice continuously differentiable function of (x,t) that is linearly homogeneous in x (see Woodland 1976).

We interpret t as representing the effects of technological change. As t changes, the production function f shifts in the manner postulated by equation (125) above. Our present goal is to show how the *impact effect on output of technological change*, $\tau(x,t) \equiv \partial \ln f(x,t)/\partial t$,[88] can be estimated using only observable price and quantity data.

Assuming that f exhibits constant returns to scale (i.e., that the restrictions [126] above are satisfied), then application of the quadratic approximation lemma (59) or its consequence (64) to f defined by (125) and (126) yields the following identity:

$$(127)\qquad \begin{aligned}&\ln f(x^1, t^1) - \ln f(x^0,t^0) = \\ &\tfrac{1}{2}[\hat{x}^1 \nabla_x \ln f(x^1,t^1) + \hat{x}^0 \nabla_x \ln f(x^0,t^0)] \\ &\cdot[\ln x^1 - \ln x^0] + \tfrac{1}{2}\left[\frac{\partial \ln}{\partial t} f(x^1,t^1) + \frac{\partial \ln}{\partial t} f(x^0,t^0)\right] \\ &\cdot[t^1 - t^0],\end{aligned}$$

where the notation is the same as in section 8.2.6. If we now add the assumption that the producer faces the input price vectors $w^0 \gg 0_N$, $w^1 \gg 0_N$ during periods 0, 1 and that he competitively minimizes costs, then we can derive the usual identities (recall equation 23):

$$(128)\qquad \nabla_x \ln f(x^0,t^0) = w^0/w^0\cdot x^0;\ \nabla_x \ln f(x^1,t^1) = w^1/w^1\cdot x^1.$$

Substituting (128) into (127) yields the equation

$$(129)\qquad \begin{aligned}\ln y^1 - \ln y^0 = &\sum_{n=1}^{N} [s^1_n + s^0_n] \ln [x^1_n/x^0_n] \\ &+ \tfrac{1}{2}\left[\frac{\partial \ln}{\partial t} f(x^1,t^1)\right. \\ &\left.+ \frac{\partial \ln}{\partial t} f(x^0,t^0))\right] [t^1 - t^0],\end{aligned}$$

where $y^r \equiv f(x^r,t^r)$ and $s^r_n \equiv w^r_n x^r_n / w^r \cdot x^r$ for $r = 0,1$ and $n = 1,2, \ldots, N$. Equation (129) can be rearranged and exponentiated to yield the following exact relationship:

$$\text{(130)} \qquad \exp \left\{ \tfrac{1}{2} \left[\frac{\partial \ln}{\partial t} f(x^1,t^1) + \frac{\partial \ln}{\partial t} f(x^0,t^0) \right] [t^1 - t^0] \right\} = \frac{y^1}{y^0} / Q_0(w^0,w^1,x^0,x^1),$$

where $Q_0(w^0,w^1,x^0,x^1)$ is the Törnqvist quantity index in inputs. The left-hand side of (130) represents a theoretical expression for the cumulative effects of technical progress while the right-hand side of (130) can be calculated using observable data. If we define $\tau^r \equiv \partial \ln f(x^r,t^r)/dt$ for $r = 0,1$, then (130) can be rewritten as

$$\text{(131)} \qquad e^{1/2[\tau^1 + \tau^0][t^1 - t^0]} = [y^1/y^0]/Q_0(w^0,w^1,x^0,x^1).$$

The expression (131) simplifies further if we make the additional assumption that $f(x,t) = e^{\beta_0 t} f(x,0)$, which is a strong form of Hicks's neutral technological change.[89] This assumption is equivalent to the additional restrictions on the parameters

$$\text{(132)} \qquad \beta_n = 0,\ n = 1, 2, \ldots, N \text{ and } \gamma = 0.$$

With assumption (132), (131) can be rewritten as

$$\text{(133)} \qquad e^{\beta_0 [t^1 - t^0]} = [y^1/y^0]/Q_0(w^0,w^1,x^0,x^1),$$

where $\beta_0 \equiv \partial \ln f(x,t)/\partial t$ can be interpreted as a constant impact effect of technological change.

Consider now the multiple output, multiple input case. Recall the definition of the firm's variable profit function in sections 8.2.1 and 8.2.6:

$$\text{(134)} \qquad \Pi(x^r,p^r,t^r) \equiv \max_y \{p^r \cdot y : (x^r,y) \in S^{t^r}\},$$

where S^{t^r} is the firm's production possibilities set at time t^r, $(x^r,y) \equiv (x^r_1, \ldots, x^r_N, y_1, \ldots, y_M)$ is a feasible vector of inputs and outputs for the firm at time t^r, and $p \gg 0_M$ is a vector of output prices at time t^r.[90]

Recall that the variable profit function can provide a complete description of the technology of a firm under certain conditions. Now assume that the firm's variable profit function is the following (time modified) translog function:[91]

$$\text{(135)} \qquad \ln \Pi^*(x^r,p^r,t^r) \equiv \alpha_0 + \sum_{n=1}^{N} \alpha_n \ln x^r_n + \tfrac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \alpha_{ih} \ln x^r_i \ln x^r_h + \beta_0 t^r + \sum_{n=1}^{N} \beta_n t^r \ln x^r_n + \sum_{m=1}^{M}$$

$$\delta_m \ln p^r{}_m + \tfrac{1}{2} \sum_{m=1}^{M} \sum_{k=1}^{M} \delta_{mk} \ln p^r{}_m \ln p^r{}_k +$$

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \epsilon_{nm} \ln x^r{}_n \ln p^r{}_m + \sum_{m=1}^{M} \epsilon_m t^r \ln p^r{}_m + \gamma(t^r)^2,$$

where $x^r \equiv (x^r{}_1, \ldots, x^r{}_N) \rangle\rangle 0_N$ is the vector of inputs utilized by the firm at time t^r, and $p^r \equiv (p^r, \ldots, p^r{}_M) \rangle\rangle 0_M$ is the vector of output (and intermediate input) prices that the firm is facing at time t^r, $r = 0,1$. The parameters on the right-hand side of (135) satisfy the following restrictions (which ensure that $\Pi(x,p,t)$ is linearly homogeneous in p):

(136) $$\sum_{m=1}^{M} \delta_m = 1, \delta_{mk} = \delta_{km}, \sum_{m=1}^{M} \delta_{mk} = 0 \text{ for } k = 1, \ldots, M$$

$$\sum_{m=1}^{M} \epsilon_{nm} = 0 \text{ for } n = 1, 2, \ldots, N \text{ and } \sum_{m=1}^{M} \epsilon_m = 0.$$

We will also assume that the firm's production is subject to constant returns to scale so that the following restrictions are also satisfied:

(137) $$\sum_{n=1}^{N} \alpha_n = 1, \alpha_{ih} = \alpha_{hi}, \sum_{i=1}^{N} \alpha_{ih} = 0, h = 1, 2, \ldots, N,$$

$$\sum_{n=1}^{N} \beta_n = 0 \text{ and } \sum_{n=1}^{N} \epsilon_{nm} = 0 \text{ for } m = 1, 2, \ldots, M.$$

If the producer is (variable) profit maximizing at time t^r, $r = 0,1$, where y^r denotes the profit-maximizing vector of outputs (and intermediate inputs) and the producer is also cost-minimizing at time t^r, where $w^r \rangle\rangle 0_N$ denotes the vector of input prices that the producer is facing at time t^r, then it can be shown[92] that the following equations hold:

(138) $$y^r/p^r \cdot y^r = \nabla_p \ln \Pi^*(x^r,p^r,t^r);$$

$$w^r/w^r \cdot x^r = \nabla_x \ln \Pi^*(x^r,p^r,t^r); \quad r = 0,1.$$

Note that the right-hand side of (135) is quadratic in the variables $\ln x_n$, $\ln p_m$, and t. Thus we can apply the quadratic approximation lemma (59) to (135) and obtain the following equality, which is analogous to (127) above:

(139) $$\ln \Pi^*(x^1,p^1,t^1) - \ln \Pi^*(x^0,p^0,t^0) = \tfrac{1}{2} [$$

$$\hat{x}^0 \nabla_x \Pi^*(x^0,p^0,t^0) + \hat{x}^1 \nabla_x \ln \Pi^*(x^1,p^1,t^1)]$$

$$\cdot[\ln x^1 - \ln x^0] + \tfrac{1}{2}[\hat{p}^1 \nabla_p \ln \Pi^*(x^1,p^1,t^1) + \hat{p}^0$$

$$\nabla_p \Pi^*(x^0,p^0,t^0)]\cdot[\ln p^1 - \ln p^0]$$

$$+ \tfrac{1}{2}\left[\frac{\partial \ln \Pi^*}{\partial t}(x^1,p^1,t^1) + \frac{\partial \ln \Pi^*}{\partial t}(x^0,p^0,t^0)\right][t^1 - t^0].$$

Now define the *impact effect on real value added of technological change* as $\tau^*(x,p,t) \equiv \partial \ln \Pi^*(x,p,t)/\partial t$. This simply a convenient way of summarizing the percentage change in real value added due to a small increment of time. In particular, define $\tau^{*1} \equiv \partial \ln \Pi^*(x^1,p^1,t^1)\partial t$ and $\tau^{*0} \equiv \partial \ln \Pi^*(x^0,p^0,t^0)/\partial t$. Substitution of these definitions plus the relations (138) plus the identities $\Pi^*(x^1,p^1,t^1) = p^1 \cdot y^1$ and $\Pi^*(x^0,p^0,t^0) = p^0 \cdot y^0$ into (139) yields

$$(140) \qquad \tfrac{1}{2}[\tau^{*1} + \tau^{*0}]\,[t^1 - t^0] = \ln\left[\frac{p^1 \cdot y^1}{p^0 \cdot y^0}\right] - \sum_{m=1}^{M} \tfrac{1}{2}\left[\frac{p^1{}_m\, y^1{}_m}{p^1 \cdot y^1} + \frac{p^0{}_m\, y^0{}_m}{p^0 \cdot y^0}\right] \ln\left[\frac{p^1{}_m}{p^0{}_m}\right] - \sum_{n=1}^{N} \tfrac{1}{2}\left[\frac{w^1{}_n\, x^1{}_n}{w^1 \cdot x^1} + \frac{w^0{}_n\, x^0{}_n}{w^0 \cdot x^0}\right] \ln\left[\frac{x^1{}_m}{x^0{}_m}\right].$$

After exponentiating both sides of (140), we get

$$(141) \qquad e^{1/2[\tau^{*1} + \tau^{*0}]\,[t^1 - t^0]} = \tilde{Q}_0(p^0,p^1,y^0,y^1)/Q_0(w^0,w^1,x^0,x^1),$$

an *implicit* Törnqvist index of outputs divided by the Törnqvist index of inputs. Thus the right-hand side of (141) is almost identical[93] to the right-hand side of (124), and the Jorgenson and Griliches (1967, 1972) measure of technical progress can be (approximately) justified in the context of a general (not necessarily separable) technology.

Finally, suppose Π^* satisfies the additional restrictions:

$$(142) \qquad \beta_n = 0,\ n = 1,2,\ldots,N,\ \epsilon_m = 0,\ m = 1,2,\ldots,M \text{ and } \gamma = 0.$$

Then $\Pi^*(x,p,t) = e^{\beta_0 t}\,\Pi^*(x,p,0) = \Pi^*(xe^{\beta_0 t},p,0)$; that is, technical change is of the primary factor augmenting strongly Hicks's neutral variety. Then $\tau^{*1} = \tau^{*0} = \beta_0$ and (141) becomes

$$(143) \qquad e^{\beta_0[t^1 - t^0]} = \tilde{Q}_0(p^0,p^1,y^0,y^1)/Q_0(w^0,w^1,x^0,x^1).$$

Usher [1974, p. 278] has criticized the use of the continuous time Divisia index (recall section 8.2.3) in the measurement of total factor productivity. I conclude this section by evaluating my measure of the residual (141) in the light of Usher's objections.

Usher's (1974, pp. 277–82) first objection to the Divisia index is that it will not give the correct answer unless the technology is homothetic[94] and technical change affects the technology in a Hicks neutral manner. In my model the technology is restricted to be homothetic, since I have imposed constant returns to scale on my technology by the restrictions (137) above, and thus this part of Usher's objection applies also to my model. However, we do not require Hicks neutral technological change in my model: isoquants are allowed to twist owing to technical change.

Usher's (1974, p. 278) second objection to the Divisia index is that it is defined using time as a continuous variable but has to be computed using some sort of discrete time approximation, and that this approximation will introduce errors that will possibly cumulate over time. My method of calculating the residual seems free from this defect of the Divisia index, since (141) can be evaluated as an exact equality using discrete time data. However, in reality my method is not entirely free from this criticism, since it is unlikely that my modified translog variable profit function Π^* defined by (135) could provide a very accurate approximation to the actual technology we are modeling for very long periods of time.[95]

Usher's (1974, p. 278) third criticism of the Divisia index also applies to my formula (141): the formula depends on the assumption that there is competitive price-taking behavior on the part of producers. However, the assumption of competitive behavior can readily be relaxed in theory: if there is monopolistic pricing behavior on the part of a producer, all we have to do is replace the observed w, p prices that occur in (141) with the appropriate marginal prices.[96] In practice this is extremely difficult to do.

Usher's (1974, p. 288) final criticism of the Divisia index methodology is more subtle than the criticisms above and deserves to be extensively quoted:

> The 1965 graduate is equally productive in some occupations, more productive in others, and he possesses skills that were unknown in 1940 because they depend on technology developed in the intervening period. The point I am making is that the relative wage of college graduates has been preserved because, and only because, technical advance has brought forth new skills and has made it profitable for people to acquire these skills, so that what we measure as labour input contains a very large component of technical change. Inputs with the same name are not the same inputs at different periods of time. . . . These considerations suggest that the use of the Divisia index coupled with the practice of treating factors of production with identical names as though they were identical factors of production may be leading us to attribute a disproportionate share of observed economic growth to the mere replication of factors of production, and may conceal the vital role of invention.

Obviously the above criticism applies with equal force to my formula (141). Of course, one method of attenuating the force of Usher's criticism would be to treat changed inputs as new inputs. This leads us to consider the new goods problem, a problem that will be considered in section 8.6.

Before studying the new goods problem, we will study one additional issue in the measurement of total factor productivity: the problem of defining an aggregate over sectors (or producers) measure of technical change.

8.5.3 Sectoral Estimates of Total Factor Productivity versus Economywide Measures

Domar (1961), in a classic paper,[97] raised the issue of working out a method of measuring technical progress that would be invariant to the degree of aggregation and integration of processes (at the firm level), firms, industries (aggregates of firms), and sectors (aggregates of industries): "We should be free to take the economy apart, to aggregate one industry with another, to integrate final products with their inputs, and to reassemble the economy once more and possibly over different time units without affecting the magnitude of the Residual. The latter's rate of growth should, therefore, be invariant to the degree of aggregation and integration and to the choice of time unit, be it a year or a decade" (Domar 1961, pp. 713–14).

Suppose we have two time periods, J sectors (or processes, or firms, or industries), and that the constant returns to scale technology of each sector can be represented by a variable profit function (which can be interpreted as a value-added function)[98] Π^j, where

$$(144)\qquad \Pi^j(x^{rj}, p^r, t^r) \equiv \max_y\{p^r \cdot y : (y, x^{rj}) \in S^{jt^r}\}$$
$$= p^r \cdot y^{rj} = w^{rj} \cdot x^{rj}, \quad r = 0, 1;$$
$$j = 1, 2, \ldots, J,$$

where S^{jt^r} is the jth sector's production possibilities set at time t^r, $x^{rj} \equiv (x^{rj}_1, \ldots, x^{rj}_{N_j}) \rangle\rangle 0_{N_j}$ is an N-dimensional vector of primary inputs used by sector j during period r, $w^{rj} \equiv (w_1{}^{rj}, \ldots, w^{rj}_{N_j}) \rangle\rangle 0_{N_j}$ is the corresponding vector of primary input prices the jth sector faces during period r, $p^r \equiv (p^r_1, \ldots, p^r_M) \rangle\rangle 0_M$ is the vector of positive final product (and intermediate input) prices all sectors face during period r, and $y^{rj}_1, \ldots, y^{rj}_M)$ is the vector of outputs produced (and intermediate inputs used) by the jth sector during period r. As usual, if $y^{rj}_m > 0$, then the jth sector is producing the mth good during period r while if $y^{rj}_m < 0$, then the mth good is being utilized as an input by the jth sector during period r. Thus the components of y^{rj} are not restricted in sign but $p^r \cdot y^{rj} > 0$, since the value of outputs (minus the value of intermediate inputs used), $p^r \cdot y^{rj}$, equals the value of primary inputs used by the jth sector during period r, $w^{rj} \cdot x^{rj} > 0$. Note also that the primary inputs need not be the same across sectors, but that each sector faces the same output (and intermediate input) prices.

Define the aggregate real value function Π as

$$\Pi(x^{r1}, \ldots, x^{rJ}, p^r, t^r) \equiv \max_{y^1, \ldots, y^J}\{p^r \cdot \left(\sum_{j=1}^{J} y^j\right): \tag{145}$$

$$(y^j, x^{rj}) \in S^{jt^r}, j = 1, \ldots, J\} = \sum_{j=1}^{J}$$

$$\max_{y^j}\{p^r \cdot y^j : (y^j, x^{rj}) \in S^{jt^r}\}$$

$$= \sum_{j=1}^{J} \Pi^j(x^{rj}, p^r, t^r) \tag{146}$$

$$= p^r \cdot \left(\sum_{j=1}^{J} y^{rj}\right).$$

Thus aggregate value added is equal to the sum of sectoral value added. Define the aggregate net output vector in period r as

$$y^r \equiv \sum_{j=1}^{J} y^{rj}, \quad r = 0,1, \tag{147}$$

and define the "aggregate" vectors of inputs and input prices as

$$x^r \equiv (x^{r1}, x^{r2}, \ldots, x^{rJ}), \quad r = 0,1 \tag{148}$$

$$w^r \equiv (w^{r1}, w^{r2}, \ldots, w^{rJ}), \quad r = 0,1. \tag{149}$$

Finally, define the *sectoral technical change impact coefficients* during period r (assuming differentiability of the Π^j with respect to time) as

$$\tau^{rj} \equiv \partial \ln \Pi^j(x^{rj}, p^r, t^r)/\partial t, \quad r = 0,1; j = 1,2, \ldots, J, \tag{150}$$

and define the *aggregate technical change impact coefficients* during period r as

$$\tau^r \equiv \partial \ln \Pi(x^r, p^r, t^r)/\partial t, \quad r = 0,1. \tag{151}$$

Using (146) above, it is easy to show that the following relationship between the sectoral coefficients τ^{rj} and the aggregate technical change impact coefficient τ^r holds:

$$\tau^r = \sum_{j=1}^{J} \frac{\partial \ln \Pi^j}{\partial t}(x^{rj}, p^r, t^r) \frac{\Pi^j(x^{rj}, p^r, t^r)}{\Pi(x^r, p^r, t^r)} \tag{152}$$

$$= \sum_{j=1}^{J} \tau^{rj} s^{rj}, r = 0,1,$$

using the definitions (150) and defining the sectoral value added shares as $s^{rj} \equiv \Pi^j(x^{rj}, p^r, t^r)/\Pi(x^r, p^r, t^r) = p^r \cdot y^{rj}/p^r \cdot y^r$.

Recall that in section 8.5.2 I indicated that, under certain conditions, an arithmetic average of the impact coefficients $\frac{1}{2}[\tau^0 + \tau^1]$ could be calculated using only observable price and quantity data. Using (152), we see that the following relationship holds between the sectoral and aggregate average technical change coefficients:

$$(153) \qquad \tfrac{1}{2}[\tau^0 + \tau^1] = \sum_{j=1}^{J} \tfrac{1}{2}[\tau^{0j}s^{0j} + \tau^{1j}s^{1j}].$$

Now suppose that for each sector the technology can be adequately represented by a sectoral translog variable profit function Π^{j*} similar to the function defined by (135), (136), and (137). Then for each sector we can derive an identity similar to (141):

$$(154) \qquad e^{1/2[\tau^{0j}+\tau^{1j}][t^1-t^0]} = \tilde{Q}_0(p^0,p^1,y^{0j},y^{1j})/ Q_0(w^{0j},w^{1j},x^{0j},x^{1j}) \text{ for } j = 1, 2, \ldots, J.$$

Unfortunately, the relations (154) do not enable us to calculate the terms $\tfrac{1}{2}[\tau^{0j}s^{0j} + \tau^{1j}s^{1j}]$, which are needed to calculate the average of the aggregate technical change impact coefficients $\tfrac{1}{2}[\tau^0 + \tau^1]$ via formula (153). However, if we assume that the sectoral translog variable profit functions Π^{j*} satisfy the additional restrictions similar to (142) (so that technical change is strongly Hicks neutral in each sector), then we can show, as in section 8.5.2, that

$$(155) \qquad \tau^{0j} = \tau^{1j} \equiv \tau^j \text{ for } j = 1, 2, \ldots, J,$$

and the relations (154) can be rewritten as

$$(156) \qquad e^{\tau^j[t^1-t^0]} = \tilde{Q}_0(p^0,p^1,y^{0j},y^{1j})/Q_0(w^{0j},w^{1j},x^{0j},x^{1j}), \quad j = 1, \ldots, J,$$

which means that the (constant) sectoral technical change coefficients can be calculated using observable price and quantity data for the two periods. Using (153), (155), and (156), it can be seen that the *correct average aggregate technical change impact coefficient* $\tfrac{1}{2}[\tau^0 + \tau^1]$ can be calculated from observable data using the following equations:

$$(157) \qquad e^{1/2[\tau^0+\tau^1][t^1-t^0]} = e^{\sum_{j=1}^{J} \tau^j \, 1/2[s^{0j}+s^{1j}][t^1-t^0]}$$
$$= \prod_{j=1}^{J} [e^{\tau^j[t^1-t^0]}]^{1/2[s^{0j}+s^{1j}]}$$
$$= \prod_{j=1}^{J} [\tilde{Q}_0(p^0,p^1,y^{0j},y^{1j})/Q_0(w^{0j},w^{1j},x^{0j},x^{1j})]^{1/2[s^{0j}+s^{1j}]}$$
$$(158) \qquad \equiv f(p^0,p^1,y^0,y^1,w^0,w^1,x^0,x^1).$$

Thus the correct aggregate measure of technical change, $e^{1/2[\tau^0+\tau^1][t^1-t^0]}$, is equal to a *geometric average* of the sectoral measures of technical change, $e^{\tau_j[t^1-t^0]}$, with weights equal to the sector's average share of value added, $\frac{1}{2}[s^{0j}+s^{1j}]$ (which sum to unity).

Suppose now that we (incorrectly) assume that the technology of the "economy" (i.e., the aggregate over sectors technology) was represented by a translog variable profit function $\Pi^*(x^r,p^r,t^r)$ similar to that defined by (135), (136), and (137) (except that here $x^r \equiv (x^{r1}, x^{r2}, \ldots, x^{rJ})$ and each $x^{rj} \rangle\rangle 0_{N_j}$ is a vector and we calculated the following *incorrect average aggregate technical change impact coefficient* $\frac{1}{2}[\tau^{*0} + \tau^{*1}]$ from observable data using the following equation that corresponds to formula (141):

$$(159) \qquad e^{1/2[\tau^{*0}+\tau^{*1}][t^1-t^0]} \equiv \tilde{Q}_0\left(p^0,p^1,\sum_{j=1}^{J} y^{0j},\sum_{j=1}^{J} y^{1j}\right)/Q_0(w^0,w^1,x^0,x^1)$$

$$(160) \qquad \equiv g(p^0,p^1,y^0,y^1,w^0,w^1,x^0,x^1).$$

The formula (159) is incorrect because we are assuming that each of the sectoral technologies is *precisely* representable by a translog variable profit function satisfying the appropriate restrictions, and thus the aggregate technology is not precisely representable by a translog variable profit function. However, the aggregate technology could be *approximated* to the second order by an aggregate translog variable profit function. Thus we would hope that the two estimates of average aggregate technical progress defined by (157) and (159) would give approximately the same answer when applied to empirical data. This hope turns out to be justified, as the following theorem indicates.

(161) *Theorem*: The functions f and g, defined by (158) and (160), respectively, differentially approximate each other to the second order[99] at any point where $p^0 = p^1$, $w^0 = w^1$, $x^0 = x^1$ and $y^0 = y^1$.

The proof of this theorem is a very tedious series of computations that can be simplified using tricks similar to those used in Diewert (1978*b*).

8.6 The New Goods Problem

8.6.1 New Goods and Index Number Formulas

One of the problems that has troubled index number theorists and practioners is constructing price and quantity indexes that are comparable over a period when new commodities are being introduced into the economy. For example, how can one construct meaningful price and quantity indexes of capital during a period of time when new capital goods are constantly appearing on (and disappearing from) the market?

My solution to the problem is quite conventional:[100] assume that the consumer (or producer) is consistently trying to solve the aggregator maximization subject to an expenditure constraint problem $\max_x\{f(x) : p^r \cdot x \leq p^r \cdot x^r, x \geq 0_N\}$ for $r = 0,1$ except that in period 0, when some goods are not available, the aggregator maximization problem has additional constraints imposed upon it that set the components of $x = (x_1, x_2, \ldots, x_N)$ that correspond to unavailable goods equal to zero.

For the sake of definiteness, let us suppose the first good is the new good that is introduced into the economy at some stage.

Obviously, if a given price and quantity are always zero, then when calculating a price or quantity index the zero good can simply be omitted from the computations. However, if a good is at a zero level during period 0 and nonzero during period 1, it is clear that the Törnqvist quantity index number formula Q_0 cannot be used, since the logarithm of zero is minus infinity. The Fisher quantity index Q_2 is well defined even if a subset of prices and quantities is zero, but we shall show below that it is not in general correct (from the viewpoint of the theory of exact index numbers) to use Q_2 without some modification.

Suppose that the nonzero prices and quantities in period 0 are $p^0{}_2, p^0{}_3, \ldots, p^0{}_N$ and $x^0{}_2, x^0{}_3, \ldots, x^0{}_N$, while the nonzero prices and quantities in period 1 are $p^1 \equiv (p^1{}_1, p^1{}_2, \ldots, p^1{}_N)$ and $x^1 \equiv (x^1{}_1, x^1{}_2, \ldots, x^1{}_N)$ respectively. We suppose that the quantity of good 1 in period 0 is $x^0{}_1 = 0$. In some circumstances we will often incorrectly assume that the price of good 1 in period 0 is also zero. However, when a new good enters the domain of our model during period 1, we should attempt to estimate the *reservation price* of the new good for the previous period that would rationalize the zero demand for the new good of the previous period.

Thus the theoretically correct procedure would be to form an estimate of the (reservation) price of good 1 in period 0, $p^0{}_1 > 0$, say, and then apply our usual index number formulas (P_2 and Q_2, say), using $p^0 \equiv (p^0{}_1, p^0{}_2, \ldots, p^0{}_N)$, $x^0 \equiv (0, x^0{}_2, \ldots, x^0{}_N)$ and the period 1 price and quantity vectors, p^1 and x^1. Let us denote the theoretically correct Fisher price index in the usual manner as:

$$(162) \qquad P_2(p^0,p^1;x^0,x^1) = [(p^1 \cdot x^0 p^1 \cdot x^1)/ (p^0 \cdot x^0 p^0 \cdot x^1)]^{1/2}.$$

Note that the theoretically correct index depends on the empirically unobservable price $p^0{}_1 > 0$. If we incorrectly set $p^0{}_1 = 0$ and substitute the resulting price vector into (162), we obtain the following incorrect Fisher price index (recall that $x^0{}_1 = 0$):

(163) $$P^{**}_2(p^0,p^1;x^0,x^1) \equiv P_2([0,p^0_2,\ldots,p^0_N],p^1; [0,x^0_2,\ldots,x^0_N],x^1) = \left[\frac{p^1\cdot x^0\, p^1\cdot x^1}{p^0\cdot x^0(p^0\cdot x^1 - p^0_1 x^1_1)}\right]^{1/2}.$$

Another theoretically incorrect Fisher price index could be obtained by simply ignoring good 1 for both periods; that is, set $p^0_1 = p^1_1 = x^0_1 = x^1_1 = 0$ and substitute the resulting price and quantity vectors into (162) to obtain the following index:

(164) $$\begin{aligned} P^*_2(p^0,p^1;x^0,x^1) &\equiv \\ P_2([0,p^0_2,\ldots,p^0_N],\ &[0,p^1_2,\ldots,p^N_2]; \\ [0,x^0_2,\ldots,x^0_N],\ &[0,x^1_2,\ldots,x^1_N]) \\ &= \left[\frac{p^1\cdot x^0(p^1\cdot x^1 - p^1_1 x^1_1)}{p^0\cdot x^0(p^0\cdot x^1 - p^0_1 x^1_1)}\right]^{1/2}. \end{aligned}$$

The virtue of the incorrect indexes P^*_2 and P^{**}_2 is that they may be calculated without a knowledge of the empirically unobservable $p^0_1 > 0$.

We now evaluate the *bias* in each of the incorrect index number formulas; that is, we take the ratio of (163) to (162) and the ratio of (164) to (162):

(165) $$\frac{P^{**}_2(p^0,p^1;x^0,x^1)}{P_2(p^0,p^1;x^0,x^1)} = [1/(1-s^{01}_1)]^{1/2} > 1$$

(166) $$\frac{P^*_2(p^0,p^1;x^0,x^1)}{P_2(p^0,p^1;x^0,x^1)} = \left[\frac{1-s^{11}_1}{1-s^{01}_1}\right]^{1/2},$$

where $s^{01}_1 \equiv p^0_1\, x^1_1/p^0\cdot x^1$, the share of good 1 using period 0 prices and period 1 quantities, and

$s^{11}_1 \equiv p^1_1\, x^1_1/p^1\cdot x^1$, the share of good 1 using period 1 prices and period 1 quantities.

Several points immediately become apparent. (*a*) The Fisher price index P_2^{**} that incorrectly sets the price of good 1 equal to zero for period 0 is always biased upward. (*b*) The Fisher price index P^*_2 that incorrectly ignores the existence of good 1 for both periods need not be biased. The bias will be zero if $p^1_1/p^0_1 = p^1\cdot x^1/p^0\cdot x^1 = P_P(p^0, p^1; x^0,x^1)$; that is, the bias will be zero if the relative change in the price of good 1 over the two periods is equal to the general change in prices as measured by a Paasche price index. In general we would expect that the relative price of good 1 would be higher in period 0 when good 1 is not yet being demanded; that is, we would expect that $p^1_1/p^0_1 \leq p^1\cdot x^1/p^0\cdot x^1$, in which case

$$(167) \qquad P_2(p^0,p^1;x^0,x^1) \leq P^*_2(p^0,p^1;x^0,x^1) < P^{**}_2(p^0,p^1;x^0,x^1).$$

Thus, in general P^*_2 will probably have a upward bias, while P^{**}_2 will definitely have a larger upward bias. Hence, in empirical applications where nothing is known about the magnitude of the reservation price, $p^0{}_1$, I would recommend the use of P^*_2 rather than P^{**}_2 (which is sometimes used).

The quantity index corresponding to P^*_2 is defined by deflating the actual expenditure ratio (remember $x^0{}_1 = 0$) by P^*_2:

$$(168) \qquad Q^*_2(p^0,p^1;x^0,x^1) \equiv \left[p^1 \cdot x^1 \Big/ \left(\sum_{i=2}^{N} p^0{}_i x^0{}_i\right)\right] \Big/ P^*_2(p^0,p^1;x^0,x^1) = [p^1 \cdot x^1/p^0 \cdot x^0]/P^*_2(p^0,p^1;x^0,x^1).$$

Thus the quantity index Q^*_2 can be calculated without a knowledge of $p^0{}_1$. Note that P^*_2 and Q^*_2 are consistent with the weak factor reversal test.

In the following section we will consider a method for obtaining empirical estimates of demand reservation prices.

8.6.2 A Simple Econometric Approach to the New Goods Problem

Let us consider a slightly more general situation than the model of the previous section. We now suppose that only the first K goods to be aggregated are available in period 0 where $1 < K < N$ and that N goods are available in period t, $t = 1, 2, \ldots, T$. Then the period 0 aggregator maximization problem is

$$(169) \qquad \max_{x_1,x_2,\ldots,x_K} \{f(x_1,x_2,\ldots,x_K,0,\ldots,0): \sum_{k=1}^{K} p^0{}_k x_k \leq Y^0, x_k \geq 0, k = 1,2,\ldots,K\},$$

where $Y^0 > 0$ is period 0 expenditure, and the period t aggregator maximization problems are

$$(170) \qquad \max_{x_1,x_2,\ldots,x_N} \{f(x_1,x_2,\ldots,x_N): \sum_{k=1}^{N} p^t{}_k x_k \leq Y^t, x_k \geq 0, k = 1,2,\ldots,N\}; t = 1,\ldots,T,$$

where $Y^t > 0$ is period t expenditure. Denote a solution to the period 0 aggregation maximization problem by the K-dimensional vector $\tilde{x}^0 \equiv (x^0{}_1,x^0{}_2,\ldots,x^0{}_K)$ and define the N-dimensional vector $x^0 \equiv (x^0{}_1,x^0{}_2,\ldots,x^0{}_K,0,\ldots,0) = (\tilde{x}^0,0_{N-K})$ and similarly denote period 0 prices by the K-dimensional vector $\tilde{p}^0 \equiv (p^0{}_1,p^0{}_2,\ldots,p^0{}_K)$. Denote a solution to the period t aggregation maximization problem by the N-dimensional vector $x^t \equiv (x^t{}_1, x^t{}_2, \ldots, x^t{}_N)$ and define the period t price vector as $p^t \equiv (p^t{}_1, p^t{}_2, \ldots, p^t{}_N) \gg 0_N$.

The "demand reservation prices" $p^0{}_{k+1}, \ldots, p^0{}_N$ are defined to be prices that "rationalize" the consumer's or producer's choice of x^0 in period 0, assuming that the new goods were available in period 0; that is, $(p^0{}_{K+1}, p^0{}_{K+2}, \ldots, p^0{}_N)$ is a set of period 0 demand reservation (or shadow) prices if x^0, a solution to (169), is also a solution to:

$$(171) \qquad \max_x\{f(x) : p^0 \cdot x \leq p^0 \cdot x^0, x \geq 0_N\},$$

where $p^0 \equiv (\tilde{p}^0, p^0{}_{k+1}, p^0{}_{k+2}, \ldots, p^0{}_N)$.

If f were differentiable at x^0 and we knew the functional form for f, p^0 could be defined by using the first-order conditions for the constrained maximization problem (171) (after eliminating the Lagrange multiplier); that is, $p^0/p^0 \cdot x^0 = \nabla_x f(x^0)/x^0 \cdot \nabla f(x^0)$ or, since $p^0 \cdot x^0 = \tilde{p}^0 \cdot \tilde{x}^0$,

$$(172) \qquad p^0 = (\tilde{p}^0 \cdot \tilde{x}^0) \nabla_x f(x^0)/x^0 \cdot \nabla f(x^0).$$

Thus, if the functional form for the aggregator function f were known, formula (172) could be used to estimate the "shadow" price components $p^0{}_{k+1}, \ldots, p^0{}_N$ of $p^0 \equiv (\tilde{p}^0, p^0{}_{k+1}, \ldots, p^0{}_N) \equiv (p^0{}_1, \ldots, p^0{}_k, p^0{}_{k+1}, \ldots, p^0{}_N)$.

Now assume $f = f_r$ for some $r > 0$[101] where the quadratic means of order r aggregator functions f_r were defined in section 8.2.4:

$$(173) \qquad f_r(x) \equiv \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}\, x^{r}{}_i{}^{/2} x^{r}{}_j{}^{/2\,1/r}; \; a_{ij} = a_{ji}.$$

Assuming that all components of $\tilde{x}^0$ are nonzero, the Kuhn-Tucker conditions for the period 0 aggregator maximization problem (169) imply:

$$(174) \qquad \tilde{p}^0/\tilde{p}^0 \cdot \tilde{x}^0 = \nabla_{\tilde{x}} f_r(\tilde{x}^0, 0_{N-K})/f_r(\tilde{x}^0, 0_{N-K}).$$

Similarly, assuming that all components of x^1 are nonzero,[102] the Kuhn-Tucker conditions for the period t aggregator maximization problems (170) imply:

$$(175) \qquad p^t/p^t \cdot x^t = \nabla_x f_r(x^t)/f_r(x^t), \quad t = 1, 2, \ldots, T.$$

Make the base period normalization:

$$(176) \qquad f_r(\tilde{x}^0, 0_{N-K}) = 1.$$

Now regard the system of equations defined by (174), (175), and (176) as a system of equations in the unknown a_{ij} parameters occurring in f_r defined by (173); that is, we are back to *method I* (recall section 8.2.2) for the determination of an aggregator function. The equations (174)–(176) are particularly simple if $r = 1$ or $r = 2$. Once the parameters a_{ij} that occur in the definition of f_r have been statistically determined, estimated demand reservation prices for period 0 can be calculated using formula (172). However, in order to estimate econometrically the $N(N+1)/2$ a_{ij} parameters,[103] we will require that the number of observations $T+1$ be large relative to the number of goods N. In many practical situations, this condition is unlikely to be met.

8.6.3 The Hedonic Approach to the New Goods Problem

Many capital goods (e.g., trucks) came in so many varieties that it is possible to think of the good as being indexed by varying amounts $x_1,x_2,\ldots,x_N$ of N continuous characteristics. For example, in Griliches's (1961) classic work,[104] automobiles were indexed by the continuous variables horsepower, weight, and length in addition to some discrete variables. In this section I will attempt to provide a theoretical justification for Griliches's hedonic price index approach.

Let us suppose that the producers of "trucks" can produce a truck indexed by the vector of characteristics $x \equiv (x_1,x_2,\ldots,x_N)$ in period r at a price $P_r(x)$ equal to the minimum cost of production:

$$(177) \qquad P_r(x) \equiv C(w^r,x),$$

where C is a "truck" producer's *joint cost function*, w^r is a vector of input prices the truck producer faces during period r, and x is the vector of characteristics that indexes the truck. It can be shown[105] that, under reasonable assumptions on the technology, the joint cost function C will be nondecreasing, linearly homogeneous, and concave in the input prices w, and, assuming that we are measuring characteristics so that more of a characteristic increases the cost of a truck, then C will be nondecreasing and concave in the vector x (assuming that the underlying technology is convex). In addition, we make the not-so-reasonable assumption that the technology is subject to constant returns to scale, so that C is linearly homogeneous with respect to x as well as w.

Another producer who uses "trucks" as an input into his productive process will want to solve the following profit maximization problem:

$$(178) \qquad \max_{u_0,u,x}\{p^r \cdot u + P_r(x)u_0 : u_1 = t(\tilde{u}, u_0 x)\},$$

where u_0 is the number of "trucks" with characteristics x purchased[106] during period r at price $P_r(x)$, $u \equiv (u_1,\tilde{u}) \equiv (u_1,u_2,u_3,\ldots,u_M)$ is a vector of nontruck outputs (indexed positively) and inputs (indexed negatively) produced and utilized by the producer, $p^r \equiv (p^r{}_1,p^r{}_2,\ldots,p^r{}_M) \gg 0_M$ is the vector of nontruck prices facing the producer during period r, and t is the producer's transformation function.[107] Note that we are assuming that characteristics enter the producer's transformation function as $u_0 x = (u_0 x_1, u_0 x_2, \ldots, u_0 x_N)$, the number of "trucks" purchased times the per truck vector of characteristics. This is not an innocuous assumption. Now define the total characteristics purchased vector y as

$$(179) \qquad y \equiv u_0 x$$

and substitute (177) and (179) into (178). Making use of the linear homogeneity in x property of C, $C(w^r,x)u_0 = C(w^r,u_0x)$, (178) becomes

$$(180) \qquad \max_{u,y}\{p^r \cdot u + C(w^r,y) : u_1 = t(\tilde{u},y), u \equiv (u_1,\tilde{u})\}.$$

If t is differentiable with respect to its arguments and C is differentiable with respect to the components of y, then a solution u^r, y^r to (180) will satisfy the following conditions:

$$(181) \qquad p^r \equiv (p^r_1, p^r_2, \ldots, p^r_M) = \lambda^r(1, \partial t(\tilde{u}^r,y^r)/\partial u_2, \ldots, \partial t(\tilde{u}^r,y^r/\partial u_M),$$

$$(182) \qquad \nabla_y C(w^r,y^r) = \lambda^r \nabla_y t(\tilde{u}^r,y^r), \quad r = 1, 2, \ldots, T,$$

where $\nabla_y C(w^r,y^r) \equiv (\partial C(w^r,y^r)/\partial y_1, \ldots, \partial C(w^r,y^r)/\partial y_M)$, $\nabla_y t(\tilde{u}^r,y^r) \equiv (\partial t(\tilde{u}^r,y^r)/\partial y_1, \ldots, \partial t(\tilde{u}^r,y^r)/\partial y_M)$, $u^r \equiv (u^r_1, \tilde{u}^r)$, and λ^r is the Lagrange multiplier for the constrained maximization problem (180).

From (182) it can be seen that the partial derivative $\partial C(w^r,y^r)/\partial y_n$ can be interpreted as the price of one unit of the nth characteristic in period r, P^r_n; that is, define

$$(183) \qquad P^r \equiv \nabla_y C(w^r,y^r),$$

where$P^r \equiv (P^r_1,P^r_2, \ldots, P^r_N)$ is a vector of characteristic prices during period r. The constant returns to scale property of C in y implies that

$$(184) \qquad P^r \cdot y^r = C(w^r,y^r) = C(w^r,x^r)u^r_0.$$

Thus, if econometric estimates of the "truck" producer's joint cost function are available and if we can observe a purchasing firm's input of "trucks" u^r_0 with characteristics x^r during period r, then we can calculate the characteristic prices P^r using (183), and we can decompose the purchasing firm's expenditure on "trucks," $P_r(x)u^r_0$, into a price component P^r and a quantity component y^r. At this point, standard index number formulas can be used to form a "truck" aggregate for the purchasing firm.

A further useful specification of this model is possible. Suppose the truck-producing technology is separable[108] so that the joint cost function C decomposes in the following manner:

$$(185) \qquad C(w,x) = c(w)g(x).$$

The effect of the additional assumption (185) is that a combination of cross-sectional and time-series analysis can be used to estimate the parameters of C; that is, we can econometrically estimate the parameters that occur in the following equation:

$$P_r(x^{rj}) = a_r g(x^{rj}), \quad r = 1, 2, \ldots, T; j = 1, 2, \ldots, J, \tag{186}$$

where $a_r \equiv c(w^r)$ and $P_r(x^{rj}) \equiv C(w^r, x^{rj})$ is the price of a "truck" with characteristics x^{rj} purchased by the jth firm during period r.

If, in addition, the function g in (185) can be approximated by the translog functional form, then (186) can be rewritten as (after taking logarithms of both sides):

$$\begin{aligned} \ln P_r(x^{rj}) &= \ln a_r + \ln \alpha_0 + \sum_{n=1}^{N} \alpha_n \ln x^{rj}{}_n \\ &+ \tfrac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_{nm} \ln x^{rj}{}_n \ln x^{rj}{}_m \\ & r = 1,2, \ldots, T; j = 1,2, \ldots, J, \end{aligned} \tag{187}$$

where $\sum_{n=1}^{N} \alpha_n = 1$, $\alpha_{nm} = \alpha_{mn}$ and $\sum_{n=1}^{N} \alpha_{nm} = 0$ for $m = 1,2, \ldots, N$. If we further specify that $\alpha_{nm} = 0$ for all n,m, then the model defined by (187) becomes very close to Griliches's (1961) classic hedonic prices model. However, if $\alpha_{nm} = 0$ for all n,m and $\alpha_n > 0$ with $\sum_{n=1}^{N} \alpha_n = 1$, then the function g[109] reduces to the Cobb-Douglas function that is concave in x instead of being convex in x. Thus the Griliches model cannot be obtained as a special case of our model (187), which is based on the assumption that the "truck" producer's technology is separable. However, if we assumed that the *consumers* of "trucks" all used a concave, linearly homogeneous, weakly separable aggregator function f in the characteristics x in order to form a "truck" aggregate $f(x)$, then instead of (186) we would obtain the model $P_r(x^{rj}) = a_r f(x^{rj})$, where a_r can be interpreted as the price of the "truck" aggregate during period r. If we further specify f to be the translog aggregator function, we would again obtain the system of equations (187), but now the Cobb-Douglas case is perfectly consistent with this second model of producer behavior.

There are many difficulties with these theoretical treatments of the new goods problem in the context of a continuous characteristics model. However, in certain industries it should be possible to modify these models into empirically useful techniques.[110]

We have discussed the problems of aggregation over goods and aggregation over producers, but we have not yet discussed the problem of aggregation over time.

8.7 Aggregation over Time: The Problem of Seasonality

All the analysis thus far has been based on the implicit assumption that the time under consideration is a year, or a decade, or some period where seasonal influences are absent. The questions I address in this section are: (*a*) how should "monthly" (or weekly) indexes be constructed (*b*) how should "annual" indexes be related to the monthly indexes? These questions are relevant to the problem of forming capital aggregates as well as other aggregates.

Zarnowitz (1961) looks at the problem of constructing seasonal index in an excellent paper. I extend his analysis a bit by utilizing some of the results discussed earlier in the present paper.

Suppose x^{rm} is a solution to the following producer (or consumer) aggregator maximization problem[111] in year r and month m:

$$(188) \qquad \max_x\{f(y^{rm},x) : p^{rm}\cdot x \leq p^{rm}\cdot x^{rm}, x \geq 0_N\},$$

where $p^{rm} \equiv (p^{rm}{}_1, p^{rm}{}_2, \ldots, p^{rm}{}_N) \rangle\rangle 0_N$ is the vector of goods prices facing the producer (or consumer) during year r, and month m, $y^{rm} \equiv (y^{rm}{}_1, y^{rm}{}_2, \ldots, y^{rm}{}_M) \rangle\rangle 0_M$ is a vector of variables that expresses weather and seasonal taste variations in year r and month m, and f is the producer's production function (or the consumer's subutility function) that is a function of both the goods vector $x \equiv (x_1,x_2, \ldots, x_N)$ and the vector of seasonal variables y^{rm}. Suppose further that the aggregator function f can be closely approximated by a linearly homogeneous in x translog aggregator function $f^*(y,x)$ where f^* is defined exactly in the same manner as D^* was defined equation (85). Then we can prove the following result in exactly the same manner as equation (70) or theorem (84) was proved:

$$(189) \qquad \textit{Theorem}: Q_0\,(p^{rm}, p^{sn}, x^{rm}, x^{sn}) = f^*(y^*, x^{sn})/f^*(y^*, x^{rm}),$$

where Q_0 is the Törnqvist quantity index and the vector of average seasonal variables y^* is defined by $y^* \equiv (y^*{}_1, y^*{}_2, \ldots, y^*{}_M)$ where $y^*{}_j \equiv (y^{rm}{}_j\, y^{sm}{}_j)^{1/2}, j = 1, 2, \ldots, M$.

The proof of this theorem rests on the assumptions of: (*a*) optimizing behavior, (*b*) the translog functional form for the aggregator function, and (*3*) the quadratic approximation lemma (59).

We can now attack the questions that were posed at the beginning of this section. First, should the monthly indexes be computed using the chain principle across months within a year, or should we construct twelve separate monthly indexes, chaining the twelve indexes across years? Thus we could calculate the Törnqvist indexes $Q_0(p^{rm}, p^{r,m+1}, x^{rm},x^{r,m+1})$, or the twelve monthly Törnqvist indexes $Q_0(p^{rm}, p^{r+1,m}, x^{rm}, x^{r+1,m}), m = 1,2, \ldots, 12$. In view of theorem (189) above, it seems that the latter procedure of constructing twelve monthly indexes would be better in normal circumstances, where we would expect the seasonal variables y^{rm} to repeat themselves every twelve months. Thus, if $y^{rm} = y^{r+1,m}$, then theorem (189) tells us that we can calculate precisely the ratio we are interested in, $f^*(y^{r+1,m}, x^{r+1,m})/f^*(y^{rm},x^{rm})$, by evaluating the Törnqvist index, $Q_0(p^{rm}, p^{r+1,m}, x^{rm}, x^{r+1,m})$, using observable price and quantity data. This conclusion agrees with that reached by Hofsten and Zarnowitz, as the following quotations indicate.

> This difficulty is especially obvious if seasonal fluctuations are considered. It is unnatural to accept an index which may after a year give a result different from 1, if the prices have returned to their initial values. . . . Yearly links should then be used. [Hofsten 1952, p. 27]

> Since 1887, when Marshall first advanced the chain system and Edgeworth seconded it, many students of index numbers have come to look upon the chain index as the standard statistical solution to changing weights. But careful consideration must be given to the question of how well chain indexes can be applied to the seasonal weight changes with whose specific features they were surely not designed to cope.
>
> It is easy to demonstrate that a chain index with varying weights does not fulfill the test of proportionality (or identity). . . . Thus, on the identity test, the indexes for the same seasons should be equal, too, but they are so only for the fixed-base, not for the chain, formulae. [Zarnowitz 1961, p. 235]

Second, given that we are going to construct twelve monthly indexes, how should these indexes be related during the base year, $r = 0$? A reasonable procedure would be to use the Törnqvist quantity index formula to construct the following eleven numbers, which could be used to compare the levels of the twelve monthly indexes during the base year: $Q_0(p^{0m}, p^{0,m+1}, x^{0m}, x^{0,m+1}), m = 1,2, \ldots, 11$. Theorem (189) can then be used to provide an economic interpretation of the resulting indexes. We should also note at this point (as does Zarnowitz 1961, p. 244) that the problem of disappearing goods giving rise to zero prices and quantities is particularly acute when we deal with seasonal indexes, and the reader is reminded of the discussion of the new goods problem in sections 8.6.1 and 8.6.2. The techniques discussed there can also be used in the present context. Summarizing the discussion thus far, I have recommended that the chain principle across months during a base year be used to construct monthly indexes for the base year, and then the chain principle across years be used to construct twelve separate monthly indexes. This procedure is of course not invariant to the choice of the base year, but in practice we would expect deviations from circularity[112] to be rather small.

Finally, there is the question of how an annual index should be related to the monthly indexes. In theory, the most appropriate way of forming annual aggregates would be to treat each good in each month as a separate argument in the index number formula; for example, we should compute $Q_0([p^{0,1}, \ldots, p^{0,12}], [p^{1,1}, \ldots, p^{1,12}], [x^{0,1}], \ldots, x^{0,12}], [x^{1,1}, \ldots, x^{1,12}])$ as representing the ratio of the aggregate in year 1 to year 0. However, we could apply the results on two-stage aggregation outlined in section 8.2.5 to conclude that a close approximation to the above aggregate ratio can be obtained by either (*a*) constructing monthly indexes and then aggregating these indexes over the year, or (*b*) constructing annual indexes for each good and then aggregating over goods. The index number formula Q_0 (or any other superlative quantity index) is to be used whenever an aggregate is calculated in the above two-stage procedures.

8.8 Concluding Comments

It is necessary to reemphasize that this discussion of capital aggregation (and aggregation in general) has taken place in the context of production function and total factor productivity estimation, where we have consistently assumed that producers are competitively profit-maximizing or cost-minimizing or both. However, I have noted that the assumption of competitive or price taking behavior can be easily relaxed in theory: simply replace observed prices with the appropriate marginal or shadow prices.[113] In practice, the assumption of competitive behavior will probably be required for some time yet in order to construct aggregates.

Given the rather narrow competitive optimizing framework, I have discussed two methods for justifying aggregation over goods such as components of the capital stock: (*a*) price proportionality[114] or Hicks's aggregation theorem (section 8.2.1), and (*b*) homogeneous weak separability (section 8.2.2). We have discussed a number of methods for justifying aggregation over sectors, including: (*c*) the method that assumes that all producers face the same prices with all goods (except possibly one) being freely variable during the period under consideration (section 8.3.2) and (*d*) a method due to Gorman (1968*a*) and Fisher (1965) that assumes some goods are fixed but the functional forms for producer's production functions are restricted in a certain manner (section 8.3.3).

In actual practice, we do not expect any of the above justifications for aggregation to hold *exactly*; however, we can hope that both methods (a) and (c) hold *approximately* at least, so that, if aggregates are used in actual applications, there is some hope that microeconomic theory will be at least approximately relevant.

I have argued that *superlative* index number formulas (recall section 8.2.4) should be used when aggregating over goods, assuming that there is a homogeneous weakly separable aggregator function defined over the goods in the aggregate, since superlative index number formulas correspond to *flexible* functional forms for aggregator functions. If the prices of the goods to be aggregated move proportionally, then the use of a superlative index number formula will also lead to the construction of an aggregate that is consistent with Hicks's aggregation theorem, even if there does not exist a homogeneous weakly separable aggregator function defined over the micro goods to be aggregated. Thus the use of a superlative index number formula is consistent with both of the general methods above for justifying aggregation over goods, and thus my first specific recommendation is that *superlative indexes* be used to construct aggregates whenever possible.

My second specific recommendation is that the *chain* principle be used (rather than a fixed base) whenever possible. Theoretical and practical reasons for this recommendation are scattered throughout the chapter and will not be reviewed here.

My third recommendation is that *rental prices* be used to weight the components of the capital stock when constructing a capital aggregate suitable for the measurement of productivity and the estimation of production functions. These rental prices should involve depreciation rates, taxes, and expectations of capital gains, although the last item presents some conceptual and practical difficulties (cf. section 8.4). However, rental prices for capital stock components need not be constructed if one employs the Hicksian view of production, which regards depreciated capital as a separate output.

My fourth specific recommendation is that *new goods* be treated in the manner outlined in section 8.6.1 when resources do not permit the implementation of the theoretically more refined techniques outlined in sections 8.6.2 and 8.6.3.

My fifth specific recommendation is that *seasonal series* be constructed in the manner outlined in section 8.7; that is, roughly speaking, "seasonal weights" must be estimated and utilized in the construction of seasonal series.

My final recommendation is that serious consideration be given to *revising the system of national accounts* used in most Western countries. The basic problem with the current system is that it is not very well suited to estimating production functions or systems of consumer demand and labor supply functions: prices that producers face are not generally distinguished from prices consumers pay. In fact, on primary input markets (labor, capital, land, and natural resources), it is often difficult to determine prices or quantities separately at all: total payments to labor, total payments to capital (including land and natural resources), and certain payments to governments (direct and indirect taxes) are distinguished in the current system of accounts, but there is no systematic decomposition of these highly aggregated payments into detailed price and quantity components for each type of labor, capital, land, and so on.

Finally, it is useful to contrast this chapter on the aggregation of capital with the excellent chapter by Murray Brown in this volume (chap. 7). Our discussions of the theoretical conditions allowing for the construction of capital aggregates have been very similar and we have reached broadly consistent conclusions—no small accomplishment considering that our papers were written completely independently. Some differences in emphasis remain—Brown's chapter has a somewhat broader theoretical coverage (his excellent discussion of the Cambridge controversies and of the general equilibrium approach to aggregation is entirely missing in my chapter), whereas mine has placed a greater emphasis on index number problems. However, taken together, perhaps the two offer a fairly comprehensive survey of the current state of aggregation theory, with particular emphasis on the problems of capital aggregation.

Appendix: Proofs of Theorems

Proof of (6)

$$\Pi(w,p) \equiv \max_{x,y}\{w^T x + p^T y : (x,y) \in S\}$$

$$= w^T x^* + p^T y^* \quad \text{by assumption} \tag{A1}$$

$$= \max_x\{w^T x + p^T y^* : (x,y^*) \in S\}$$

$$= \Pi^*(w,y^*) + p^T y^* \quad \text{by the definition of } \Pi^*$$

$$= \Pi^*(p_0\alpha, y^*) + p^T y^* \quad \text{using (3)}$$

$$= p_0\Pi^*(\alpha,y^*) + p^T y^* \quad \text{by a homogeneity property of } \Pi^*$$

$$\equiv p_0\, y^*_0 + p^T y^* \quad \text{defining } y^*_0 \equiv \Pi^*(\alpha,y^*) \tag{A2}$$

$$\leq \max_{y_0\, y}\{p_0 y_0 + p^T y : (y_0,y) \in S_\alpha\} \quad \text{since } (y^*_0,y^*) \in S_\alpha$$

$$\equiv \hat{\Pi}(p_0,p).$$

Suppose $\hat{\Pi}(w,p) < \Pi(p_0,p) = p_0\bar{y}_0 + p^T\bar{y}$, where $(\bar{y}_0,\bar{y}) \in S_\alpha$. Then $\bar{y}_0 = \Pi^*(\alpha,\bar{y}) = \alpha^T\bar{x}$ for some $\bar{x}$ such that $(\bar{x},\bar{y}) \in S$

$$\begin{aligned}\therefore \Pi(w,p) &\equiv \max_{x,y}\{p_0\alpha^T x + p^T y : (x,y) \in S\}\\ &\geq p_0\alpha^T\bar{x} + p^T\bar{y}\\ &= p_0\bar{y}_0 + p^T\bar{y}\\ &> \Pi(w,p),\end{aligned}$$

which is a contradiction, and thus our supposition is false and thus $\Pi(w,p) = \hat{\Pi}(p_0,p)$. Note that (A1) = (A2) implies that $p_0 y^*_0 = w^T x^*$, which is equivalent to (7).

Proof of (84)

The producer's technology S can be completely described by means of a transformation function t (see Diewert 1973*a*): $t(y_2,y_3,\ldots,y_M, x_1,x_2,\ldots,x_N) \equiv \max_{y_1}\{y_1 : (y_1,y_2,\ldots,y_M,x_1,\ldots,x_N) \in S\}$. Furthermore, it is easy to see that if the producer has minimized the cost $w{\cdot}x$ of producing a given vector of outputs $(y_1,y_2,\ldots,y_N)$, then under the usual monotonicity assumptions, the producer will also be producing the maximal amount of output 1 given that he must also produce $y_2,\ldots,y_M$ and is subject to an expenditure constraint on inputs. Thus we assume that x^0 is a solution to $\max_x\{t^*(y^0{}_2,y^0{}_3,\ldots,y^0{}_M,x) : w^0{\cdot}x = w^0{\cdot}x^0, x \geq 0_N\} = y^0{}_1$ and that x^1 is a solution to $\max_x\{t^*(y^1{}_2, y^1{}_3,\ldots,y^1{}_M, x) : w^1{\cdot}x = w^1{\cdot}x^1\} = y^1{}_1$, where t^* is the firm's transformation function that corresponds to the translog distance function D^*. The proof of the rest of the theorem is virtually identical to the proof of theorem (2.17) in Diewert (1976, pp. 139–40), except that the transformation function $t^*(y_2,y_3,\ldots,y_M, x)$ replaces the utility function $f(x)$, and y_1 replaces the utility level u.

Proof of (97)

Let $x^{1*}, x^{2*},\ldots,x^{M*}$ be a solution to the maximization problem (91). Using our assumed regularity conditions on Π^m, (91) becomes a concave programming problem, and we may apply the saddle-point theorem of Karlin (1959, p. 201) and Uzawa (1958) to obtain the existence of shadow prices $w^* \geq 0_N$, such that $x^{1*}, x^{2*},\ldots,x^{M*}$ is a solution to the following unconstrained maximization problem:

$$\text{(A3)} \qquad \max_{x^1,\ldots,x^M}\left\{\sum_{m=1}^{M}\Pi^m(p,x^m,z^m) - w^*{\cdot}\left(\sum_{m=1}^{M}x^m\right)\right\}.$$

In fact, our strong monotonicity assumptions on Π^m (along with the concavity assumptions) imply that $w^* \gg 0_N$ and that

(A4) $$\sum_{m=1}^{M} x^{m*} = x.$$

Now rewrite (A3) as

$$\max_{x^1,\ldots,x^M} \left\{ \sum_{m=1}^{M} \Pi^m(p,x^m,z^m) - w^{*}\cdot\left(\sum_{m=1}^{M} x^m\right)\right\}$$

$$= \sum_{m=1}^{M} \max_{x^m} \{\Pi^m(p,x^m,z^m) - w^{*}\cdot x^m\}$$

(A5) $$\equiv \sum_{m=1}^{M} \Pi^{*m}(p,w^{*},z^m) \qquad \text{using definitions (94).}$$

$$= \max_{x,x^1,\ldots,x^M} \left\{ \sum_{m=1}^{M} \Pi^m(p,x^m,z^m) - w^{*}\cdot x : \sum_{m=1}^{M} x^m = x \right\},$$

which is equivalent to (A3) upon defining $x \equiv \sum_{m=1}^{M} x^m$

$$= \max_{x,x^1,\ldots,x^M} \left\{ \sum_{m=1}^{M} \Pi^m(p,x^m,z^m) - w^{*}\cdot x : \sum_{m=1}^{M} x^m \leq x \right\}$$

using the monotonicity properties of Π^m

$$= \max_x \{\Pi(p,x,z^1,\ldots,z^M) - w^{*}\cdot x\}$$

upon optimizing *w.r.t.* $x^1,\ldots,x^N$ and using the definition of Π, (92)

(A6) $$= \Pi^{*}(p,w^{*},z^1,\ldots,z^M)$$

using the definition of Π^{*}, (95).

Thus (A5) = (A6), the desired result.

Notes

1. Notation: $w \rangle\rangle 0_N$ means that each component of the N-dimensional vector $w \equiv (w_1,w_2,\ldots,w_N)$ is positive where 0_N is a vector of zeros; $w \geq 0_N$ means that each component is nonnegative; $w > 0_N$ means $w \geq 0_N$ but $w \neq 0_N$; $w^T x \equiv \sum_{n=1}^{N} w_n x_n = w\cdot x$ is the inner product of the vectors w and x.

2. The only regularity condition we need impose on S is that a solution to the profit maximization problem (1) exist for the set of prices (w,p) under con-

sideration. This existence will generally be assured if S is a closed nonempty set with an appropriate property of boundedness from above. See McFadden (1978) or Diewert (1973*a*) on this point.

3. Gorman (1968*a*) uses this terminology.

4. McFadden (1978) and Lau (1976) use this terminology.

5. Diewert (1973*a*, 1974*b*) uses this terminology.

6. See Gorman (1968*a*), Diewert (1973*a*), and Lau (1976). A property of Π^* that we require below is the linear homogeneity of Π^* in w; that is, for every $\lambda > 0$, $w \rangle\rangle 0_N$, we have $\Pi^*(\lambda w, y) = \lambda \Pi^*(w,y)$.

7. S is convex if and only if for every scalar such that $0 \leq \lambda \leq 1$ and $(x^1,y^1) \epsilon S$, $(x^2,y^2) \epsilon S$, we have $(\lambda x^1 + (1-\lambda)x^2, \lambda y^1 + (1-\lambda)y^2) \epsilon S$.

8. Let $0 \leq \lambda \leq 1$, $(y^1{}_0, y^1) \epsilon S_\alpha$ and $(y^2{}_0, y^2) \epsilon S_\alpha$. Then $y^1{}_0 \leq \Pi^*(\alpha,y^1)$ and $y^2{}_0 \leq \Pi^*(\alpha,y^2)$. Convexity of S implies that $\Pi^*(\alpha,y)$ is a concave function of y (cf. Gorman 1968*a*) so $\Pi^*(\alpha, \lambda y^1 + (1-\lambda)y^2) \geq \lambda\Pi^*(\alpha,y^1) + (1-\lambda)\Pi^*(\alpha,y^2) \geq \lambda y^1{}_0 + (1-\lambda)y^2{}_0$, which implies $(\lambda y^1{}_0 + (1-\lambda)y^2{}_0,$
$\lambda y^1 + (1-\lambda)y^2) \epsilon S_\alpha$.

9. That is, S is a cone; if $(x,y) \epsilon S$ and $\lambda \geq 0$, then $(\lambda x, \lambda y) \epsilon S$.

10. If S is a cone, then $\Pi^*(\lambda\alpha, y) = \lambda\Pi^*(\alpha,y)$ for every $\lambda \geq 0$, and the proof follows readily.

11. See Diewert (1974*b*, p. 139). It is an obvious modification of the translog function introduced by Jorgenson and Lau (1970).

12. $E\epsilon^r = \frac{1}{2} \sum_{i=1}^{N} \sum_{h=1}^{N} \beta_{ih}\, \sigma_{ih}$, which does not depend on r. If $N = 2$, then the regularity conditions (9) on β_{ih} along with the symmetry conditions $\beta_{ih} = \beta_{hi}$ imply that $\beta = \beta_{11} = -\beta_{12} = -\beta_{21} = \beta_{22}$, so that in this case $E\epsilon^r = \frac{1}{2}\beta[\sigma_{11} - 2\sigma_{12} + \sigma_{22}]$. Note that positive semidefiniteness of the variance covariance matrix $[\sigma_{ij}]$ implies that $\sigma_{11} - 2\sigma_{12} + \sigma_{22} \geq 0$, so that the sign of the bias $E\epsilon^r$ is determined by the sign of β. For a general N, we could expect the bias to be small, since we would not expect a systematic correlation between β_{ih} and σ_{ih}.

13. The assumption of time independence is somewhat unrealistic: if the relative price of the first good is higher than usual during a given period, we would expect this condition to persist for a number of subsequent periods. Thus autocorrelation is to be expected when estimating the parameters of an equation like (11).

14. This terminology follows Geary and Morishima (1973). The concept of weak separability is due to Sono (1961) and Leontief (1947). Note that Shephard actually considered the problem of simultaneously aggregating x and z into two aggregates.

15. See Diewert (1974*b*, p. 112). Similar formulas have been derived by Chipman (1970) and Samuelson (1972).

16.
$$\begin{aligned}
p^{*T}x^* + w^{*T}z^* &= C^*(u^*; p^*, w^*)\\
&= \min_{x,z}\{p^{*T}x + w^{*T}z : f^*(x,z) \geq u^*\}\\
&= \min_{x,z}\{p^{*T}x + w^{*T}z : \hat{f}[f(x),z] \geq u^*\} \text{ using (14)}\\
&= \min_{x,z,y}\{p^{*T}x + w^{*T}z : \hat{f}[y,z] \geq u^*, y = f(x)\} \text{ adding an additional variable and equation}\\
&= \min_{z,y}\{c(p^*)y + w^{*T}z : \hat{f}[y,z] \geq u^*\} \text{ upon minimizing with respect to } x \text{ using (17)}\\
&= \hat{C}(u^*; c(p^*), w^*) \text{ using definition (16)}\\
&= c(p^*)f(x^*) + w^{*T}z^* \text{ since } x^*, z^* \text{ is a solution to the first cost-minimization problem.}
\end{aligned}$$

17. *Proof*: $x^{rT}\nabla f(x^r) = f(x^r)$ by Euler's theorem on homogeneous functions.

18. *Proof*: divide (26) by $p^{rT}x^r = c(p^r)f(x^r)$.

19. If all $\gamma_{jk} = 0$, then we are in the Cobb-Douglas case and the two aggregator functions coincide.

20. Or, alternatively, assume that x^r is a solution to the cost-minimization problem (17): $\min_x\{p^{rT}x : f(x) \geq f(x^r)\}$.

21. Solow (1957) presented his exposition in terms of a single output and two inputs. His argument is readily extended to the case of N inputs, as is done in Richter (1966), Star (1974), and Star and Hall (1976), for example.

22. Note that we do not have to estimate the unknown parameters of the production function.

23. See, for example, the empirical examples worked out in Fisher (1922).

24. See Konyus and Byushgens (1926), Afriat (1972), Pollak (1971), and Samuelson and Swamy (1974).

25. See Shephard (1953) or Diewert (1974*b*), or recall the material presented in section 8.2.2.

26. Actually, we require (43) to hold only for an empirically relevant subset of positive prices and outputs.

27. This theory is perhaps more clearly presented in Shephard (1953) and Solow (1955–56).

28. See also Afriat (1972), Pollak (1971), Samuelson and Swamy (1974), and Diewert (1976).

29. The term is due to Lau (1974): f is a second-order differential approximation to f^* at the point x^0 if and only if $f(x^0) = f^*(x^0)$, $\nabla f(x^0) = \nabla f^*(x^0)$ and $\nabla^2 f(x^0) = \nabla^2 f^*(x^0)$; that is, the levels of the functions coincide as well as their first- and second-order partial derivatives evaluated at x^0. Note that $\nabla f(x^0)$ is the vector of first-order partial derivatives of f evaluated at x^0, while $\nabla^2 f(x^0)$ is the matrix of second-order partial derivatives.

30. Note that ∇Q_2 stands for the vector of first-order partial derivatives of Q_2 with respect to all 4N arguments, etc.

31. The Laspeyres and Paasche quantity indexes give the same answer as the three "better" indexes only to the first order; that is,
$\nabla^2 Q_L(p,p;x,x) \neq \nabla^2 Q_P(p,p;x,x) \neq \nabla^2 Q_2(p,p;x,x)$.

32. The functional forms f_r and c_r were studied by Denny (1974).

33. The proofs of theorems (50) and (51) do not rest on any assumption of optimizing behavior: they are simply theorems in numerical analysis rather than economics.

34. When the chain principle is used, the limited empirical evidence in Diewert (1978*b*) suggests that even the Paasche and Laspeyres indexes give virtually the same answer as the superlative indexes.

35. See Shephard (1953, pp. 61–71), Solow (1955–56), Gorman (1959), Blackorby et al. (1970), and Geary and Morishima (1973, pp. 100–103).

36. We consider only the Vartia I indexes, since they are indexes that have the property of consistency in aggregation. Sato (1976*a*) showed that the Vartia II indexes were exact for a CES aggregator function.

37. In fact, any pseudosuperlative index could be used. It is shown in Diewert (1978*b*) that any twice continuously differentiable symmetric mean of the Paasche and Laspeyres price indexes is a pseudosuperlative price index; for example, $\frac{1}{2} P_L(p^0,p^1;x^0,x^1) + \frac{1}{2} P_P(p^0,p^1;x^0,x^1)$ is pseudosuperlative.

38. Actually, some of the components of x^r can be outputs instead of inputs, in which case the corresponding components of w^r are indexed with a minus sign.

39. See McFadden (1978) for the properties of these functions.

40. The translog joint cost function is defined analogously to the translog variable profit function defined in section 8.2.1. However, since the logarithm of a negative number is not defined, we must renormalize the components of $y \equiv (y_1, \ldots, y_M)$ so that each $y_m > 0$. If y_m is an intermediate input, then we renormalize the corresponding price p_m to be minus the initial positive price.

41. Diewert (1978*b*) shows that $\tilde{Q}_0$ differentially approximates Q_0 to the second order if the partial derivatives are evaluated at equal price and quantity vectors. Thus, normally, $\tilde{Q}_0(w^0,w^1,x^0,x^1)$ will be numerically close to $Q_0(w^0,w^1, x^0,x^1)$. Similarly, $\tilde{P}_0$ will normally be close to P_0.

42. We make the same sign conventions as were noted in note 40. Thus, if there are intermediate inputs, the corresponding components of p^0 and p^1 will be negative. However, $Q_0(p^0,p^1,y^0,y^1)$ can still be calculated in the usual way even if some components of p^0 and p^1 are negative.

43. We now revert back to our original sign conventions: output and input prices are all positive, as are inputs x, but components of y can be negative if the corresponding good is an (intermediate) input. Note that even if some components of y^0 or y^1 are negative, we can still calculate P_0 using the usual formula.

44. There is also a close correspondence with consumer surplus concepts. Define the consumer's cost or expenditure function m as $m[f(x),p] \equiv \min_x\{p \cdot x : f(x) \geq f(x)\}$, where f is the consumer's utility function and p is a vector of commodity rental prices that the consumer faces. The (Laspeyres) Allen (1949) quantity index is defined as $Q_A(x^0,x^1,p^0) \equiv m[f(x^1),p^0]/m[f(x^0),p^1]$, while the (Laspeyres) Konyus (1939) cost of living index is defined as $P_K(p^0, p^1, x^0) \equiv m[f(x^0),p^1]/m[f(x^0),p^0]$. The consumer surplus concepts use arithmetic differences rather than ratios. Thus Hicks's (1946, pp. 40–41) *compensating variation* in income can be defined as $m[f(x^0),p^1] - m[f(x^0),p^0]$, while Hicks's (1946, p. 331) *equivalent variation* in income can be defined as $m[f(x^1),p^0] - m[f(x^0),p^0]$.

45. For formal proofs of duality theorems between distance functions and production or utility functions, see Shephard (1970), Hanoch (1978), McFadden (1978), Rockafellar (1970) and Blackorby, Primont, and Russell (1978).

46. See Gorman (1970), McFadden (1978), Hanoch (1975), and Blackorby and Russell (1976) for discussions on the separability properties of distance functions.

47. As usual, we have to change our sign conventions with respect to the components of y: assume that $y \equiv (y_1, y_2, \ldots, y_M) \rangle\rangle 0_M$, but if the mth good is actually an intermediate input, then the corresponding price p_m is taken to be negative. Note that we have not restricted $D^*[y,x]$ to be homogeneous of degree -1 in the components of y, which would be the case if production were subject to constant returns to scale. Finally, we note that the translog distance function can provide a second-order approximation to an arbitrary twice-differentiable distance function.

48. The sign conventions of note 47 are operative here also.

49. Actually, Klein considered the problem of simultaneously aggregating over commodities as well as sectors.

50. For an alternative proof, see Green (1964). Nataf (1948) and Green assumed the f^m were twice differentiable. These regularity conditions were relaxed by Gorman (1968*b*) to continuity and by Pokropp (1972) to monotonicity conditions alone.

51. A strongly separable production function can provide only a first-order approximation to a general production function.

52. If the individual production sets are S^m, $m = 1,2,\ldots,M$, then the aggregate technology S can be defined as the sum of the individual production sets; that is, $S \equiv \{z : z = \sum_{m=1}^{M} z_m$, where $z^m \in S^m$ for $m = 1,2,\ldots,M\}$. If the individual producer profit functions are defined as $\Pi^m(p) \equiv \max_{z^m}\{p \cdot z^m : z^m \in S^m\}$, then $\Pi(p) \equiv \max_z\{p \cdot z : z \in S\} = \sum_{m=1}^{M} \Pi^m(p)$; that is, the aggregate profit function equals the sum of the individual profit functions. There does not appear to be a simple characterization of the aggregate transformation function in terms of the individual transformation or production functions.

53. In the Solow problem, K (the number of outputs) is taken to be 1, and p_1 (the price of the output) is taken to be 1 also.

54. Concavity of $\Pi^m(p,x^m,z^m)$ in x^m is implied by convexity of S^m but is a considerably weaker restriction on the technology than convexity of S^m.

55. This approach was pioneered by Cornwall (1973).

56. This result can be traced back to Hotelling (1932). For modern proofs, see Gorman (1968*a*), McFadden (1978), or Diewert (1973*a*).

57. The rental price formula defined by equation (105) is similar to those derived by Jorgenson and Griliches (1967), except that they derive their formulas on the basis of a continuous-time optimization problem (as opposed to my discrete-time "Hicksian" period formulation), and thus the term $(1 + r)$ is missing from the denominator of their formulas. However, my formula (105) has the property that if $\delta = 1$ (i.e., the good is actually a nondurable), then the rental price equals the purchase price of the good plus associated tax payments (cf. eq. 104).

58. For example, consider the controversy between Jorgenson and his coworkers (Jorgenson and Griliches 1967, 1972; Christensen and Jorgenson 1969, 1970; Christensen, Cummings, and Jorgenson 1976) and Denison 1969, 1974, and Kendrick 1961, 1972, 1976.

59. The tax system that the producer is facing is not explicitly modeled in (108) but is implicit in the definitions of the prices p_z, p_y, and p_x.

60. Generally, if interest rates are positive, this capital aggregate would be an aggregate input. However, if maintenance and renovation expenditures were particularly large for the firm under consideration, it is possible for the capital aggregate to be a net output.

61. Epstein (1977, chap. 7) shows how the data limitations can be overcome in theory by an explicit econometric model.

62. From a theoretical point of view, the "Hicksian" capital services aggregate appears to be more appropriate than the Jorgenson-Griliches aggregate. However, from an empirical point of view, the "Hicksian" aggregate is more difficult to construct.

63. Kendrick (1972, p. 37) notes that official national income accounts in the United States and generally elsewhere exclude capital gains and losses, a comment that also applies to private accounting practices. With the recent upsurge of worldwide inflation, it has become more difficult to ignore capital gains, and a literature on accounting for inflation has sprung up (cf. Shoven and Bulow 1975 and the discussion following their paper).

64. See the analysis and references to the literature in Epstein (1977).

65. See Jorgenson and Griliches (1967), Christensen and Jorgenson (1969, 1970), and Christensen, Cummings, and Jorgenson (1976).

66. The former alternative seems more plausible to me: ex post internal rates of return seem to be too volatile to be an adequate approximation to the firm's actual borrowing or lending rates.

67. Winfrey distributions are widely used in the construction of capital stock aggregates. However, the following quotation from Creamer (1972, p. 68) indicates that their empirical foundation is not strong: "An examination of Winfrey's report discloses that the empirical basis of his distribution is his analysis of a sample of equipment retirements that are heavily weighted with railroad ties, trestles and power generating equipment. Moreover, these retirements occurred over the period 1869 and 1934. Clearly, this is an area that calls for new research."

68. See the examples tabulated in Tice (1967) and Creamer (1972).

69. Taxes on intermediate goods could be treated as follows in a model that aggregated firms 1 and 2: for each intermediate tax, break the corresponding commodity up into two commodities, one of which would be the "untaxed" market where firm 1 would sell its output q at price p. The government is thought of as buying the untaxed commodity at price p and then selling it back to firm 2 on the "taxed" market price $p(1+t)$. Aggregating over the two firms would yield an aggregate "untaxed" output of $q > 0$ selling at price p and an aggregate "taxed" input of $-q$ at price $p(1+t)$, while aggregation over the other goods could proceed as outlined in section 8.3.2. If t remained constant over time, the taxed and untaxed commodities could be aggregated using Hicks's aggregation theorem.

70. Denison (1974) favors estimating property income *net* of depreciation, whereas Kendrick (1972, p. 102) favors the *net* concept if one is interested in real product from a welfare standpoint, but the *gross* concept for production and productivity analysis.

71. The most preferred alternative would be to construct the "Hicksian" measure of capital services mentioned in the previous section.

72. See Denny and Sawyer (1976) for references to the theoretical national income accounting literature, plus a review of current Canadian accounting practices from the viewpoint of a neoclassical as opposed to a Keynesian approach to macroeconomic theory.

73. This would not be a problem if the capital services were included in sectoral estimates of intermediate inputs and gross output sectoral production functions were estimated.

74. Typically, rents to land are included in the accounting system, but asset prices and quantities of the different types of land do not appear.

75. See Christensen and Jorgenson (1969) and Christensen, Cummings, and Jorgenson (1976).

76. Creamer attributes this last point to Denison.

77. Under the appropriate assumptions, we can use the results of section 8.3.2 to justify the aggregation over the private producer sector and the household "processes" that create services out of consumer durable stocks.

78. Kendrick's (1976, p. 15) justification for including this item is given in the following quotation: "The costs of transfering resources are a form of investment, for investment in mobility results in an increase in the future income stream beyond what incomes would be if the shifts were not made."

79. Alternative assumptions giving the same result (item 8) are: f is linearly homogeneous, nondecreasing, and concave in all five inputs and the marginal products $\partial f(0,0,0,R_1,R_2)/\partial W$, $\partial f(0,0,0,R_1,R_2)/\partial K$ exist and are finite. In this case the industry will be made up of many tiny firms each earning tiny excess profits.

80. This does not mean that government public capital is unimportant (consider the massive United States interstate highway system): it means that in the

context of production function estimation, public capital goods should not be aggregated together with private capital but should appear as a separate inputs into the production function. However, in the context of productivity measurement, it does no harm to omit public capital goods, provided the shadow prices are zero, as we shall see in section 8.5.

81. Alternative economic models of R&D have recently been very ably surveyed by Woodland (1976).

82. See Diewert (1974*a*), Fisher and Shell (1972), and Samuelson and Swamy (1974) on this topic.

83. Assume g is defined for $y \geq 0_M$, and has the following properties: (*i*) $g(y) > 0$ for $y \rangle\rangle 0_M$ (positivity) (*ii*) $g(\lambda y) = \lambda g(y)$ for $\lambda \geq 0$, $y \geq 0$ (linear homogeneity), and (*iii*) $g(\lambda y^1 + (1-\lambda)y^2) \leq \lambda g(y^1) + (1-\lambda)g(y^2)$ for $0 \leq \lambda \leq 1, y^1 \geq 0_M, y^2 \geq 0_M$ (convexity).

84. If g satisfies the three properties listed in note 83, then r also has those three properties.

85. The proof is analogous to the proof of the Samuelson-Shephard duality theorem presented in Diewert (1974*a*); alternatively, see Samuelson and Swamy (1974).

86. A more complete exposition of the material presented in this section, with some additional material on the theory of partial Divisia indexes, can be found in Chinloy (1974).

87. The analysis for this case has been independently developed by Christensen, Cummings, and Jorgenson (1976), who cite Jorgenson and Lau (1979) as their source.

88. Ohta (1974) calls τ the primal rate of technical progress, and he shows that it is equal to the dual rate of technical progress λ defined as $\lambda \equiv -\partial \ln C(y,w,t)/\partial t$ if f exhibits constant returns to scale, where $C(y,w,t) \equiv \min_x\{w \cdot x : y = f(x,t)\}$ is the producer's total cost function.

89. See Blackorby, Lovell, and Thursby (1976) for a discussion of the different definitions of neutral technological change.

90. Make the same sign conventions as were made in the first part of section 8.2.6.

91. As usual, $\Pi^*(x,p,t)$ can provide a second-order approximation to an arbitrary twice continuously differentiable $\Pi(x,p,t)$. A special case of the time-modified translog variable profit function Π^* has been considered by Berndt and Wood (1975): in their model, x is a scalar output and y is a vector of inputs, so that, with appropriate sign changes, Π^* becomes a cost function. They also show under what conditions such a functional form can be consistent with factor-augmenting technical change.

92. See Diewert (1974*b*, pp. 137–40).

93. In Diewert (1978*b*) it is shown that $\tilde{Q}_0(p^0,p^1,y^0,y^1)$ and $Q_0(p^0,p^1,y^0,y^1)$ approximate each other to the second order at any point where $p^0 = p^1$ and $y^0 = y^1$.

94. See Shephard (1953, p. 41). A production function f is homothetic if there exists a monotonically increasing function of one variable g such that $g[f(\lambda x)] = \lambda g[f(x)]$ for every $\lambda > 0$, $x \geq 0_N$; that is, $g[f]$ is linearly homogeneous.

95. The use of the chain principle should minimize this type of error: for any two consecutive time periods, t^r, t^{r+1}, we could approximate accurately the shifting technology of the sector by a Π^{*r} defined by (135), whose parameters depend on r.

96. This point was made by Frisch (1936) forty years ago. For more details, see Diewert (1974*a*, p. 155) and Appelbaum (1979).

97. Other papers on the subject include Star (1974) and Hulten (1978).

98. Recall section 8.2.6.

99. Recall the discussion on this concept in section 8.2.5.

100. See Fisher and Shell (1972, p. 101) or Hofsten (1952, p. 97).

101. We require $r > 0$ so that $f_r(x^0, 0_{N-K})$ is well defined. Thus the translog functional form cannot be used as an aggregator function in this section.

102. If any components of x^0 or x^t are zero, then drop the corresponding equations from (174) or (175). For $r = 2$, we can drop the requirement that all components of the x vectors x^0 and x^t be nonnegative. We require only that $p^0 \cdot x^0 > 0$ and $p^t \cdot x^t > 0$. Thus we can deal with the case where f_2 is a transformation function. A negative component of the x vector indicates an output, a positive component indicates an input.

103. Actually, the base period normalization (176) implies that the number of independent parameters is $(N(N+1)/2) - 1$.

104. This paper is reprinted in Griliches (1971), where several other papers and an extensive bibliography on the hedonic approach to the quality adjustment problem will be found. Additional empirical work can be found in Gordon (1977), King (1976), and Ohta and Griliches (1976).

105. See McFadden (1978) for properties of joint cost functions.

106. If "trucks" are a durable input, then $P_r(x)$ should be the user cost of a "truck" with characteristics x during period r rather than the purchase price. Recall the discussion of rental price formulas in section 8.4.1.

107. See Diewert (1973*a*) for a discussion of the properties of such transformation functions. We do not require constant returns for the technology described by t.

108. See Blackorby, Primont, and Russell (1978) for a comprehensive discussion of separability.

109. Such functions are sometimes called input requirements functions. See Diewert (1974*a*) for a discussion of their properties.

110. An excellent discussion of many of the theoretical and practical difficulties associated with "hedonic" techniques can be found in Triplett (1975, 1976).

111. Recall the discussion about homogeneous weak separability in section 8.2.2. Note that we do not assume that the seasonal variables y^{rm} enter the objective function f in a separable way. Gersovitz and MacKinnon (1977) argue that in this case it is extremely difficult to construct deseasonalized p^{rm} and x^{rm} series that will be consistent with an underlying (nonseasonal) economic model. Thus, in the case of nonseparable seasonal variables interacting with economic variables, they suggest that it may often be appropriate to estimate econometrically completely separate models, one for each season, rather than attempting to estimate econometrically one model using data "seasonally adjusted" by conventional methods.

112. See Fisher (1922) for a discussion of the circular test and some empirical evidence that Q_2 satisfies circularity rather well. Since Q_0 is very close to Q_2 in most empirical situations, we would expect the same conclusion to hold for Q_0.

113. Appelbaum (1979) and Appelbaum and Kohli (1979) have applied this theoretical technique due originally to Frisch (1936).

114. Brown (this volume, chap. 7) calls this the commodity aggregation approach.

References

Afriat, S. N. 1972. The theory of international comparisons of real income and prices. In *International comparisons of prices and outputs*, ed. D. J. Daly, pp. 13–69. New York: National Bureau of Economic Research.

Allen, R. G. D. 1949. The economic theory of index numbers. *Economica*, n.s., 16:197–302.

Appelbaum, E. 1979. Testing price taking behavior. *Journal of Econometrics* 9:283–94.

Appelbaum, E., and Kohli, U. R. 1979. Canada–United States trade: Test for the small-open-economy hypothesis. *Canadian Journal of Economics* 12:1–14.

Archibald, R. B. 1977. On the theory of industrial price measurement: Output price indexes. *Annals of Economic and Social Measurement* 6:57–72.

Arrow, K. J. 1974. The measurement of real value added. In *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, ed. P. David and M. Reder. New York: Academic Press.

Bergson (Burk), A. 1936. Real income, expenditure proportionality and Frisch's "New methods of measuring marginal Utility." *Review of Economic Studies* 4:33–52.

———. 1961. *National income of the Soviet Union since 1928*. Cambridge: Harvard University Press.

Berndt, E. R., and Christensen, L. R. 1973. The translog function and the substitution of equipment, structures and labor in U.S. manufacturing, 1929–68. *Journal of Econometrics* 1:81–113.

Berndt, E. R., and Wood, D. O. 1975. Technical change, tax policy, and the derived demand for energy. Mimeographed. Department of Economics, University of British Columbia.

Blackorby, C.; Lady, G.; Nissen, D.; and Russell, R. R. 1970. Homothetic separability and consumer budgeting. *Econometrica* 38:468–72.

Blackorby, C.; Lovell, D.; and Thursby, M. 1976. Extended Hicks neutral technical progress. *Economic Journal* 86:845–52.

Blackorby, C.; Primont, D.; and Russell, R. R. 1978. *Duality separability and functional structure: Theory and economic applications*. New York: American Elsevier.

Blackorby, C., and Russell, R. R. 1976. Conjugate implicit separability. Submitted to the *Review of Economic Studies*.

———. 1978. Indices and subindices of the cost of living and the standard of living. *International Economic Review* 19:229–40.

Bliss, C. J. 1975. *Capital theory and the distribution of income*. Amsterdam: North-Holland.

Bowley, A. L. 1928. Notes on index numbers. *Economic Journal* 38: 216–37.

Bruno, M. 1978. Duality, intermediate inputs and value-added. In *Produciton economics: A dual approach to theory and applications,* vol. 2, ed. M. Fuss and D. McFadden. Amsterdam: North-Holland.

Chinloy, P. 1974. Issues in the measurement of labor input. Ph.D. diss., Harvard University.

Chipman, J. S. 1970. Lectures on the mathematical foundations of international trade theory. I. Duality of cost and production functions. Mimeographed. Vienna: Institute of Advanced Studies.

Christensen, L. R.; Cummings, D.; and Jorgenson, D. W. 1976. Economic growth, 1947–1973: An international comparison. Harvard Institute of Economic Research Discussion Paper 521, Harvard University.

Christensen, L. R., and Jorgenson, D. W. 1969. The measurement of U.S. real capital input, 1929–1967. *Review of Income and Wealth,* ser. 15, no. 4, pp. 293–320.

———. 1970. U.S. real product and real factor input, 1929–1967. *Review of Income and Wealth* 16, no. 1:19–50.

Christensen, L. R.; Jorgenson, D. W.; and Lau, L. J. 1971. Conjugate duality and the transcendental logarithmic production function. *Econometrica* 39, no. 4:255–56.

Cornwall, R. R. 1973. A note on using profit functions to aggregate production functions. *International Economic Review* 14:511–19.

Creamer, D. 1972. Measuring capital input for total factor productivity analysis: Comments by a sometime estimator. *Review of Income and Wealth,* 18, pp. 55–78.

Daly, D. J. 1972. Combining inputs to secure a measure of total factor input. *Review of Income and Wealth,* ser. 18, pp. 27–53.

Denison, E. F. 1969. Some major issues in productivity analysis: An examination of estimates by Jorgenson and Griliches. *Survey of Current Business* 49, no. 5, part 2:1–27.

———. 1974. *Accounting for United States economic growth 1929–1969.* Washington, D.C.: Brookings Institution.

Denny, M. 1974. The relationship between functional forms for the production system. *Canadian Journal of Economics* 7:21–31.

Denny, M. G. S., and Sawyer, J. A. 1976. Revising the national accounts. *Canadian Journal of Economics* 9:720–32.

Diewert, W. E. 1971. An application of the Shephard duality theorem: A generalized Leontief production function. *Journal of Political Economy* 79:481–507.

———. 1973*a*. Functional forms for profit and transformation functions. *Journal of Economic Theory* 6:284–316.

———. 1973*b*. Afriat and revealed preference theory. *Review of Economic Studies* 40:419–26.

———. 1974*a*. Functional forms for revenue and factor requirements functions. *International Economic Review* 15:119–30.

———. 1974*b*. Applications of duality theory. In *Frontiers of quantitative economics*, vol. 2, ed. M. D. Intriligator and D. A. Kendrick. Amsterdam: North-Holland.

———. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4:115–45.

———. 1978*a*. Hicks' aggregation theorem and the existence of a real value added function. In *Production economics: A dual approach to theory and applications*, ed. M. Fuss and D. McFadden, pp. 17–51. Amsterdam: North Holland.

———. 1978*b*. Superlative index numbers and consistency in aggregation. *Econometrica* 46:883–900.

Divisia, F. 1926. *L'indice monétaire et la théorie de la monnaie*. Paris: Société Anonyme du Recueil Sirey.

Domar, E. D. 1961. On the measurement of technological change. *Economic Journal* 71:709–29.

Epstein, L. E. 1977. Essays in the economics of uncertainty. Ph.D. diss. University of British Columbia, Vancouver.

Feldstein, M. S., and Rothschild, M. 1974. Towards an economic theory of replacement investment. *Econometrica* 42:393–424.

Fischer, S. 1974. Money and the production function. *Economic Inquiry* 12:517–33.

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., and Shell, K. *The economic theory of price indices*. New York: Academic Press.

Fisher, I. 1922. *The making of index numbers*. Boston: Houghton Mifflin.

Frisch, R. 1936. Annual survey of general economic theory: The problem of index numbers. *Econometrica* 4:1–39.

Fuss, M., and McFadden, D. eds. 1978. *Production economics: A dual approach to theory and applications*. Amsterdam: North-Holland.

Geary, P. T., and Morishima, M. 1973. Demand and supply under separability. In *Theory of demand: Real and monetary*, ed. M. Morishima and others. Oxford: Clarendon Press.

Gersovitz, M., and MacKinnon, J. G. 1977. Seasonality in regression: Economic theory and econometric practice. Discussion Paper 257, Queen's University, Kingston, Canada.

Gordon, R. J. 1977. *The measurement of durable goods prices*. New York: National Bureau of Economic Research. Forthcoming.

Gorman, W. M. 1953. Community preference fields. *Econometrica* 21:63–80.

———. 1959. Separable utility and aggregation. *Econometrica* 27:469–81.

———. 1968*a*. Measuring the quantities of fixed factors. In *Value, capital and growth: Papers in honour of Sir John Hicks*, ed. J. N. Wolfe, pp. 141–72. Chicago: Aldine.

———. 1968*b*. The structure of utility functions. *Review of Economic Studies* 35:367–90.

———. 1970. Quasi-separable preferences, costs and technologies. Mimeographed. London School of Economics and Political Science.

Green, H. A. J. 1964. *Aggregation in economic analysis: An introductory survey*. Princeton, N.J.: Princeton University Press.

Griliches, Z. 1961. Hedonic price indexes for automobiles: An econometric analysis of quality change. In *The price statistics of the federal government*. New York: National Bureau of Economic Research.

———, *Price indexes and quality change: Studies in new methods of measurement*. Cambridge, Mass.: Harvard University Press.

Hall, R. 1968. Technical change and capital from the point of view of the dual. *Review of Economic Studies* 35:35–46.

Hanoch, G. 1975. Production or demand models with direct or indirect implicit additivity. *Econometrica* 43:395–420.

———. 1978. Symmetric duality and polar production functions. In *Production economics: A dual approach to theory and applications*, ed. M. Fuss and D. McFadden, 1:113–31. Amsterdam: North Holland.

Hardy, G. H.; Littlewood, J. E.; and Polya, G. 1934. *Inequalities*. Cambridge: Cambridge University Press.

Hicks, John R. 1946. *Value and capital*. 2d ed. Oxford: Clarendon Press.

———. 1965. *Capital and growth*. Oxford: Oxford University Press.

Hofsten, E. v. 1952. *Price indexes and quality changes*. London: George Allen and Unwin.

Hotelling, H. 1932. Edgeworth's taxation paradox and the nature of demand and supply functions. *Journal of Political Economy* 40:577–616.

———. 1935. Demand functions with limited budgets. *Econometrica* 3:66–78.

Houthakker, H. S. 1955–56. The Pareto distribution and the Cobb-Douglas production function in activity analysis. *Review of Economic Studies* 23:27–31.

Hulten, C. R. 1973. Divisia index numbers. *Econometrica* 41:1017–26.

———. 1978. Growth accounting with intermediate inputs. *Review of Economic Studies* 45:511–18.

Hulten, C. R., and Wykoff, F. C. Economic depreciation and the taxation of structures in U.S. manufacturing industries: An empirical analysis. Discussion Paper 77–02, Department of Economics, University of British Columbia.

Johansen, L. 1959. Substitution vs. fixed production coefficients in the theory of economic growth: A synthesis. *Econometrica* 27:157–76.

———. 1972. *Production functions.* Amsterdam: North-Holland.

Jorgenson, D. W. 1965. Anticipations and investment behavior. In *The Brookings Quarterly econometric model of the United States*, ed. James S. Duesenberry and others, pp. 35–94. Chicago: Rand McNally.

———. 1966. The embodiment hypothesis. *Journal of Political Economy* 74:1–17.

Jorgenson, D. W., and Griliches, Z. 1967. The explanation of productivity change. *Review of Economic Studies* 34:249–83.

———. 1972. Issues in growth accounting: A reply to Edward F. Denison. *Survey of Current Business* 52, no. 5, part 2: 65–94.

Jorgenson, D. W., and Lau, L. J. 1970. The transcendental logarithmic utility function and demand analysis. Department of Economics, Harvard University.

———. 1979. *Duality and technology.* Amsterdam: North-Holland. Forthcoming.

Karlin, S. 1959. *Mathematical methods and theory in games, programming and economics.* vol. 1, Palo Alto: Addison-Wesley.

Kendrick, J. W. 1961. *Productivity trends in the United States.* Princeton: Princeton University Press.

———. 1972. *Economic accounts and their uses.* New York: McGraw-Hill.

———. 1976. *The formation and stocks of total capital.* New York: National Bureau of Economic Research.

Khang, C. 1971. An isovalue locus involving intermediate goods and its applications to the pure theory of international trade. *Journal of International Economies* 1:315–25.

King, A. T. 1976. The demand for housing: Integrating the roles of journey-to-work, neighborhood quality and prices. In *Household production and consumption*, ed. N. E. Terleckyj, pp. 451–83. New York: National Bureau of Economic Research.

Klein, L. R. 1946*a*. Macroeconomics and the theory of rational behavior. *Econometrica* 14:93–108.

———. 1946*b*. Remarks on the theory of aggregation. *Econometrica* 14:303–12.

Kloek, T. 1967. On quadratic approximations of cost of living and real income index numbers. Report 6710, Econometric Institute, Netherlands School of Economics. Mimeographed.

Könus [Konyus], A. A. 1939. The problem of the true index of the cost of living. *Econometrica* 7:10–29. (Translation of a paper first published in *Ekonomicheskii Byulleten Konyunkturnovo Instituta* 3 [1924]:64–71.)

Konyus, A. A., and Byushgens, S. S. 1926. K probleme pokupatelnoi cili deneg. *Voprosi Konyunkturi* 2:151–72.

Lau, L. J. 1974. Comments on applications of duality theory. In *Frontiers of quantitative economics*, ed. M. D. Intriligator and D. A. Kendrick, 2:176–99. Amsterdam: North-Holland.

———. 1976. A characterization of the normalized restricted profit function. *Journal of Economic Theory* 12:131–63.

———. 1979. On exact index numbers. *Review of Economics and Statistics* 61:73–82.

Leontief, W. W. 1947. Introduction to a theory of the internal structure of functional relationships. *Econometrica* 15:361–73.

McFadden, D. 1978. Cost, revenue and profit functions. In *Production economics: a dual approach to theory and applications*. eds. M. Fuss and D. McFadden, 1:3–109. Amsterdam-North-Holland.

Malinvaud, E. 1953. Capital accumulation and the efficient allocation of resources. *Econometrica* 21:233–68.

Malmquist, S. 1953. Index numbers and indifference surfaces. *Trabajos de Estadistica* 4:209–42.

May, K. 1946. The aggregation problem for a one-industry model. *Econometrica* 14:285–98.

Moorsteen, R. H. 1961. On measuring productive potential and relative efficiency. *Quarterly Journal of Economics* 75:451–67.

Morishima, M. 1969. *Theory of economic growth*. London: Oxford University Press.

Nagatani, K. 1978. *Monetary theory*, Amsterdam: North-Holland.

Nataf, A. 1948. Sur la possibilité de construction de certains macromodèles. *Econometrica* 16:232–44.

Neumann, J. von. 1945–46. A model of general economic equilibrium. *Review of Economic Studies* 12:1–9.

Ohta, M. 1974. A note on duality between production and cost functions: Rate of returns to scale and rate of technical progress. *Economic Studies Quarterly* 25:63–65.

Ohta, M., and Griliches, Z. 1976. Automobile prices revisited: Extensions of the hedonic hypothesis. In *Household production and consumption*, ed. N. E. Terleckyj, pp. 325–90 New York: National Bureau of Economic Research.

Parkan, C. 1975. Nonparametric index numbers and tests for the consistency of consumer data. Department of Manpower and Immigration, Ottawa, Research Projects Group.

Pokropp, F. 1972. A note on the problem of aggregation. *Review of Economic Studies* 39:221–30.

Pollak, R. A. 1971. The theory of the cost of living index. Research Discussion Paper no. 11, Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.

Pu, S. S. 1946. A note on macroeconomics. *Econometrica* 14:299–302.

Richter, M. K. 1966. Invariance axioms and economic indexes. *Econometrica* 34:739–55.

Rockafellar, R. T. 1970. *Convex analysis*. Princeton, N.J.: Princeton University Press.

Ruggles, R. 1967. Price indexes and international price comparisons. In *Ten economic studies in the tradition of Irving Fisher*, ed. W. Fellner and others, pp. 171–205. New York: John Wiley

Samuelson, P. A. 1953–54. Prices of factors and goods in general equilibrium. *Review of Economic Studies* 21:1–20.

———. 1972. Unification theorem for the two basic dualities of homothetic demand theory. *Proceedings of the National Academy of Sciences, U.S.A.* 69:2673–74.

Samuelson, P. A., and Swamy, S. 1974. Invariant economic index numbers and canonical duality: Survey and synthesis. *American Economic Review* 64:566–93.

Sato, K. 1975. *Production functions and aggregation*. Amsterdam: North-Holland.

———. 1976*a*. The ideal log-change index number. *Review of Economics and Statistics* 58:223–28.

———. 1976*b*. The meaning and measurement of the real value added index. *Review of Economics and Statistics* 58:434–42.

Schworm, W. 1977. *User cost and investment theory*. Ph.D. diss., University of Washington, Seattle.

Shephard, R. W. 1953. *Cost and production functions*. Princeton: Princeton University Press.

———. 1970. *Theory of cost and production functions*. Princeton: Princeton University Press.

Shoven, J. B., and Bulow, J. I. 1975. Inflation accounting and nonfinancial corporate profits: Physical assets. *Brookings Papers on Economic Activity* pp. 557–98.

Solow, R. M. 1955–56. The production function and the theory of capital. *Review of Economic Studies* 23:101–8.

———. 1957. Technical change and the aggregate production function. *Review of Economics and Statistics* 39:312–20.

———. 1964. Capital, labor and income in manufacturing. In *The behavior of income shares*, pp. 101–28. Studies in Income and Wealth no. 26, Princeton: Princeton University Press.

Sono, M. 1961. The effect of price changes on the demand and supply of separable goods. *International Economic Review* 2:239–75.

Star, S. 1974. Accounting for the growth of output. *American Economic Review* 64:123–35.

Star, S., and Hall, R. E. 1976. An approximate divisia index of total factor productivity. *Econometrica* 44:257–64.

Stigum, B. P. 1967. On certain problems of aggregation. *International Economic Review* 8:349–67.

———. 1968. On a property of concave functions. *Review of Economic Studies* 35:413–16.

Taubman, P., and Wilkinson, M. 1970. User cost, capital utilization and investment theory. *International Economic Review* 11:209–15.

Theil, H. 1967. *Economics and information theory*. Amsterdam: North-Holland.

———. 1968. On the geometry and the numerical approximation of cost of living and real income indices. *De Economist* 116:677–89.

Tice, H. S. 1967. Depreciation, obsolescence, and the measurement of the aggregate capital stock of the United States, 1900–1962. *Review of Income and Wealth* 13:119–54.

Törnqvist, L. 1936. The Bank of Finland's consumption price index. *Bank of Finland Monthly Bulletin*, no. 10, pp. 1–8.

Triplett, J. E. 1975. The measurement of inflation: A survey of research on the accuracy of price indexes. In *Analysis of Inflation*, ed. P. H. Earl, pp. 19–82 Lexington, Mass.: D. C. Heath.

———. 1976. Consumer demand and characteristics of consumption goods. In *Household production and consumption*, ed. N.'E. Terleckyj, pp. 305–24. New York: National Bureau of Economic Research.

Usher, D. 1974. The suitability of the Divisia index for the measurement of economic aggregates. *Review of Income and Wealth* 20:273–88.

———. 1958. The Kuhn Tucker theorem in concave programming. In *Studies in linear and nonlinear programming*, ed. K. J. Arrow, L. Hurwicz, and H. Uzawa. Stanford: Stanford University Press.

Uzawa, H. 1964. Duality principles in the theory of cost and production. *International Economic Review* 5:216–20.

Vartia, Y. O. 1974. Relative changes and economic indices. Licensiate thesis in Statistics, University of Helsinki.

———. 1976. Ideal log change index numbers. *Scandanavian Journal of Statistics* 3:121–26.

Wold, H. 1944. A synthesis of pure demand analysis. *Skandinavisk Aktuarietidskift* 27:69–120.

Wold, H., in association with L. Juréen. 1953. *Demand analysis*. New York: John Wiley.

Woodland, A. D. 1972. The construction of price and quantity components of inputs for Canadian industries, 1927–1969. Department of Manpower and Immigration, Ottawa.
———. 1975. Substitution of structures, equipment and labor in Canadian production. *International Economic Review* 16:171–87.
———. 1976. Modelling the production sector of an economy: A selective survey and analysis. Discussion Paper 76-21, Department of Economics, University of British Columbia.
Zarnowitz, V. 1961. Index numbers and the seasonality of quantities and prices. In *The price statistics of the federal government.* New York: National Bureau of Economic Research.

Comment Michael Denny

The analysis of capital aggregation given at this conference may appear technically complex, but it is possible to summarize the results in a simpler, though not rigorous, manner. Before turning to the details of Professor Diewert's paper, let me consider some implications of the basic results that underlie and motivate both his paper and Professor Brown's. As theorists, both Brown and Diewert are arguing that the conditions required for aggregation are stringent. So stringent that perhaps we should not publish or use aggregated data in the unassuming manner that is our current practice.

We are all familiar with the necessity to aggregate quantities of different goods or services. It is impossible to imagine economic data without aggregation: thus we should seriously consider the losses involved in our current techniques. Fundamentally, aggregation in practice involves weighting the elements according to some formula that produces an aggregate of the elements. One must remember that it is *not* the voluminous literature on index numbers that is relevant. Our authors are asking a prior question: When may we aggregate by any method and not suffer losses because of the information suppressed by aggregation? The aggregate provides less information, and in the loose framework that I am currently using there is a loss that will result in errors. Consider a specific example in which two types of capital, K_1 and K_2, and labor, L, are used to produce output (fig. C8.1). We can characterize this process abstractly as a production function, $Q = f(K_1,$

Michael Denny is associated with the Institute for Policy Analysis and the Department of Political Economy, University of Toronto.

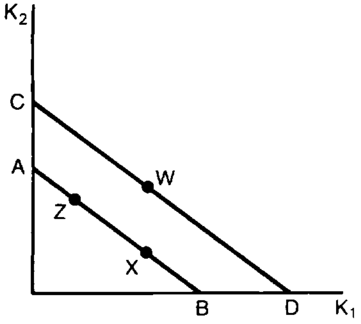

Fig. C8.1

K_2, L). Although I have chosen a production function, the argument would be similar for any other technical or behavioral function, for example, supply, demand, or cost function.

Suppose we consider the aggregation of the two types of capital. In general we will have, $K = g(K_1, K_2)$, where K is the aggregate quantity of capital and g is the function or rule that describes how we aggregate. In figure C8.1, the line "AB" represents a specific value of aggregate capital using a particular aggregation rule. Similarly, the line "CD" represents a greater value of aggregate capital using the same aggregation rule or formula. The loss of information is obvious. Any point Z, on line AB, represents particular quantities of the two capital services. Once we aggregate we can not distinguish Z from any other point X on the same line AB. We can distinguish Z (and X) from W on line CD.

We wish to know how this loss of information will affect our ability to investigate our production process. In figure C8.2, the line AB represents an aggregate quantity of capital. An isoquant, labeled (Q_o, L_o), has also been drawn tangent to AB at point W. The isoquant shows the alternative combinations of the two types of capital that can be used

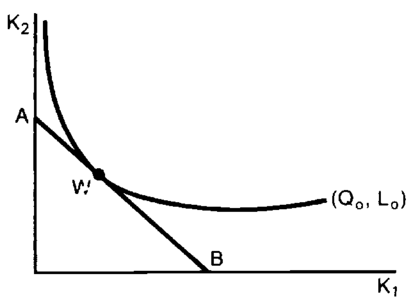

Fig. C8.2

to produce output *level*, Q_o, when labor is used at a *level*, L_o. Remember that we have to fix the level of labor; we will return to this in a moment.

Suppose we have data only on Q, L, and aggregate capital, K. Can we adequately acquire information about the disaggregated production technology from the aggregate capital, labor, and output data? From figure C8.2, we can state that at the point of tangency W, the quantity of aggregate capital represented by all the points on AB is an aggregate of the true disaggregated quantities of capital. Nontangency points on AB, while they present the same aggregate quantity of capital as W, must represent capital input combinations that lie on different isoquants.

Holding labor constant, the same aggregate quantity of capital will produce smaller and smaller quantities of output as we move away from W along AB in either direction. The use of aggregate capital will imply that the same quantities of aggregate capital and labor are capable of producing a wide variety of output levels. This is inconsistent with the production function that assumes that only one output level is associated with the efficient use of a given input bundle.

A very special linear aggregation function was used in figure C8.1. Consider bending the line AB so that it coincides precisely with isoquant (Q_o,L_o). Now this will mean that the single aggregate quantity of capital corresponds to all disaggregated input quantities that produce output Q_o in conjunction with a labor input, L_o. This seems hopeful, since now we have a measure of aggregate capital that corresponds to a particular unique input-output combination.

We have made a different but special assumption about the aggregation formula when we require that it correspond to the isoquant. However, this assertion is required if we are to eliminate aggregation errors.

Our special aggregator function that corresponds to a unique isoquant in figure C8.2 must be generalized to cover situations in which the level of output or labor do not equal (Q_o,L_o). The set of isoquants in figure C8.3 represents three different levels of output and the *same* quantity of labor input. However, these labels are not necessarily unique: any point on an isoquant could be consistent with a large number of output-labor combinations. For example, our initial capital combination, point A, could also produce output levels Q_1 and Q_2 for some levels of labor greater than L_0.

We have defined our aggregator function, $K = g(K_1,K_2)$, to depend on the disaggregated quantities of capital only. Thus when we plot the level sets of this function, we can label them as representing values of aggregate capital, independent of the level of Q and L. This is simply the condition required for *weak separability* of the production function, and in this case, we can write our production function as

$$Q = f(g(K_1,K_2),L). \tag{1}$$

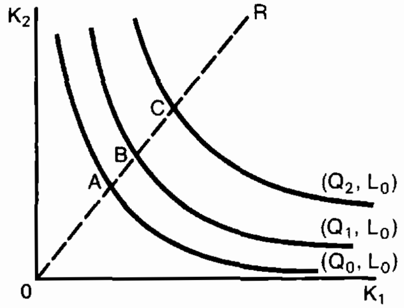

Fig. C8.3

Our aggregator function g is the first argument of the production function in this case. It turns out that the micro capital inputs, K_1 and K_2, must be weakly separable from all other inputs and output in order for a capital aggregate to exist; that is, we must be able to write the production function in the form (1) in order for a capital aggregate to exist (*unless* the rental prices of K_1 and K_2 happen to vary proportionally, in which case an aggregate can be constructed using Hicks's aggregation theorem even if the technology is not weakly separable in the micro capital inputs). Moreover, to actually *measure* the capital aggregate using just market data, we require an additional assumption in addition to (1): we require that the aggregator function g be *homothetic*, that is, we require that g be a monotonically increasing function of a linearly homogeneous function. In fact, it turns out that there is no real loss of generality in assuming that g is actually linearly homogeneous once we have made the initial assumption of homotheticity.

In figure C8.3, the ray OR through the origin cuts the isoquants at points A, B, and C. It was mentioned above that, to measure the capital aggregate, we required homotheticity in addition to weak separability. This property requires that the slopes of the isoquants at A, B, and C are equal. Why is this property or the (equivalent in the present context) property of linear homogeneity needed? The link between aggregation theory and index number theory requires the homogeneity property. Index numbers almost always have the property that they are linear homogeneous in their elements. If you double all the components, then you double the aggregate. If we are to have an aggregate quantity calculated by a rule that is consistent with an index number formula, then the isoquants in figure C8.3 must be homothetic (except when one uses a Malmquist index number formula, which does not require homotheticity). However, there are other reasons for assuming homotheticity. Consistent two-stage maximization will require this property, and this is the primary theoretical rationale.

The other major conditions that permit aggregation can be discussed in relation to the first case. Remember that from the point of view of a national statistics office, our first case implies that aggregation of capital must be done *separately* for each *different* production function. When the aggregation function was "bent" to match the isoquant in figure C8.2, this was done for a specific production technology. The unfortunate conclusion must be that not only must *each* production technology be homothetically weakly separable, but the aggregation must be done separately if the isoquants are different in any two technologies. Practically, this is impossible, and it may be that this problem is at least as important in practice as the assumption of separability.

For a single aggregate capital to be useful in two different production sectors, it turns out that each sector, i, must have a production function, $Q_i = f_i(g(K_1,K_2),L)$, $i = 1,2$. The idea of aggregating across sectors is only an extension of using the same aggregate in different sectors. Although the details will not be included here, the nonmathematical reader should be able to understand the following argument.

Consider the following special case of intersectoral aggregation. The disaggregated problem is to maximize the value of one output given the value of the other output and a fixed total amount of our three inputs. This problem will have a solution of the form $\emptyset\ (Q_1,Q_2,K_1,K_2,L) = 0$, where Q_i is the ith output and K_j and L are the simple aggregates of each input. Suppose we wish to have aggregate production technology of the form,

$$Q = H(Q_1,Q_2) = f(K_1,K_2,L), \tag{2}$$

or

$$Q = H(Q_1,Q_2) = F(K,L), \tag{3}$$

where $K = g(K_1,K_2)$ is an aggregate of the different types of capital.

Without rigorous proof we can link this type of problem to our earlier case. If we are to shift from the disaggregated function $\emptyset$ to the aggregation involved in (2) we are already severely constraining the technologies of the sectors. For (2) to be a valid representation of $\emptyset$, the production technologies for the individual sectors must have *almost* identical isoquants. The *almost* has to be put in because the isoquants can be numbered differently for each sector. Notice that in (2) we have not aggregated capital of *different types*. Even without capital aggregation, the aggregation of output will force the isoquants of the two sectors to be almost identical. If we now aggregate the different *types* of capital, the restrictive assumptions on the already similar isoquants of the two sectors will be increased. This movement from equation (2) to equation (3) is nothing more than the simultaneous application of our earlier argument to both sectors. Aggregation of outputs and capital requires a more complex and restrictive set of assumptions.

Should we throw up our hands? No, the requirements of both policy and science rule against that reaction. Economic theory without empirical confirmation will not be science. *Both* theorists and empirical investigators must accept the stringent conditions of aggregating. However, while more disaggregated data is desirable and is becoming available, the high costs of high-quality disaggregated data will preclude the elimination of aggregation. The notion of a *totally disaggregated* production function as a technical constraint is an abstraction. There is no room for an extensive catalog of possibilities, but I will state a rough guide. Both empirical and theoretical economists must continue attempts to reorient the theory to bring the "level" of abstraction closer to the "level" of observations.

At the most general level, one can approach Diewert's paper in the following manner. There is a large body of literature on aggregation and index numbers. What Diewert has done is to focus and link the powerful theoretical tools of duality theory and the recent work on flexible functional forms with this traditional literature. While there are no startling new results, the paper does integrate a scattered literature and provide some interesting insights on the interface between these areas.

Diewert has broken down the problems of aggregation into (*a*) aggregation over goods, (*b*) aggregation over sectors, and (*c*) aggregation over time. The aggregation of capital is initially treated as a special case of these types of aggregation. However, in the final three sections the special problems of technical progress, new goods, and seasonality are investigated. These are all problems that are intimately related to the special nature of capital goods and their production.

There are two sets of conditions that permit aggregation of goods. These are price proportionality and homogeneous weak separability (which is equivalent to the homothetic weak separability assumption we discussed above). The first condition states that if the prices of a group of goods varies proportionally, then it is possible to define an aggregate quantity of the goods. Provided the micro price proportionality holds, then the aggregate quantity can be used in place of the micro quantities. Many empirical price series seem to move with approximate proportionality. Diewert opens the investigation of an area that could have wide application. For a particular model of how prices deviate from proportionality, Diewert shows how the absence of strict price proportionality will affect the results. In this example as well as several others, Diewert does not clearly indicate the possibilities of generalizing his special case. It may be possible to aggregate with relatively small errors in a wide variety of situations if the very particular model can be expanded.

The second set of conditions, homogeneous weak separability, imposes restrictions on the production, demand, or utility function. Diewert concentrates on investigating the alternative methods of finding the most suitable aggregate under the assumption that homogeneous weak separability is acceptable. This problem provides the core of a very large section of the paper. If micro data are available and the second set of aggregation conditions is acceptable, then one can proceed by two methods. Using the micro data, the investigator may either estimate a functional form for an aggregator function or else choose an index number formula. Diewert defines the concepts of exact and superlative index numbers to provide a link between these two methods. If we choose an index number, then what assumptions about the underlying technology are we making? For a number of well-known index formulas, Diewert shows that they are equivalent to the use of a particular functional form. This work provides a link to the recent development of flexible functional forms. Essentially, the following proposition is being suggested. Flexible functional forms such as the Translog can approximate to the second order any functional form. Consequently, if we do not know the true functional form for an index, we should choose a *superlative* index, that is, one that is *exact* for a flexible functional form. This will ensure that we can approximate the true form, and we do not need to estimate the true function. This also suggests that the choice of a particular index formula from those that are superlative is not important. All the formulas will provide a second-order approximation to the true function, and it will not matter which formula is chosen.

If one uses one of the superlative indexes, then the following problem will arise. If one aggregates over some group of commodities and then uses the calculated aggregates in a second-state aggregation, will the results be consistent with single-stage aggregation? The answer in general is no. For consistency in two-stage aggregation, the function must be Cobb-Douglas. What Diewert does is to show that, provided one uses a superlative index or a Vartia index, the results will be approximately consistent. Basically, as one would expect, the underlying rationale is that if observations in adjacent periods are chained, then for small changes between adjacent periods, no problem will arise with multistage aggregates.

Let me provide one concrete example of the type of specific problem that is being considered. A *true index of inputs* between two periods $X(X^0,X^1;p^*)$ must equal the ratio of the variable profit functions in the two periods: $\pi(X^1,p^*)/\pi(X^0,p^*)$.

It is shown that if there are:

(*a*) constant returns to scale,

(*b*) profit-maximizing behavior with respect to inputs X and outputs Y for periods 0 and 1 given output price vectors p^0, p^1 and input price vectors w^0,w^1,

(*c*) a translog function for π, and

(*d*) the reference price p^* equals the geometric mean of p^0 and p^1,

then $Q_0(w^0,w^1,X^0 X^1)$ — a Törnqvist index of inputs will be correct.

An alternative approach to index numbers has been developed by Malmquist and extended by Pollak and by Blackorby and Russell. What Diewert is able to show is that the Törnqvist index defined above can be interpreted as a Malmquist index provided the distance function is translog and producers minimize with respect to the inputs. The point is that the Malmquist interpretation requires fewer restrictions. Neither constant returns to scale nor profit maximization is required.

Diewert has developed a very useful set of procedures for choosing an aggregation function. We must remember that he has accepted the weak separability conditions required for aggregation. While this may seem like a very weak second-best procedure, I believe this type of two-stage investigation will become very common. This topic is beyond the concern of this conference, but it is one of the links of index numbers and aggregation theory with a growing empirical literature.

The section on aggregation over sectors is brief and, as expected, the results do not suggest much optimism. I believe Diewert is correct in focusing on the possibilities of models, such as Johansen's, that attempt to link micro and macro observations. Further empirical investigations of these models is needed.

Capital aggregation suffers from all the problems of goods aggregation in general. While Diewert does suggest a more general intertemporal Hicksian model for a capital-using firm, he backs away from any serious suggestion of its implementation owing to the difficulties of obtaining the required data. This section should be extended to clarify the possibilities of measurement. If capital aggregation is viewed as equivalent to noncapital aggregation, as Diewert states in section 8.4.1, then the problems of aggregation over time must be solved. Better data on depreciation, discards, and used capital market prices could mean that capital aggregation is similar to noncapital aggregation. However, at present this is not true. The durability of capital creates difficulties in constructing a capital aggregate from an empirical point of view, *additional* to the theoretical difficulties inherent in constructing any kind of an aggregate.

The rest of the chapter turns to a number of special problems that are closely connected with capital goods. The latter are generally thought to be durable, heterogeneous in design, and subject to rapid design changes. This description may be biased toward problems with equipment, although structures are by no means homogeneous even if design changes may be less relevant.

The measurement of technical change has been fraught with all the problems of capital goods measurement. It has also been an emotional area where prior beliefs often determine one's evaluation of particular studies. Using the work of Jorgenson and Griliches as an example, several points are made. Provided all variables are measured correctly the Jorgenson-Griliches technique requires: (*a*) separability of outputs and inputs; (*b*) competitive profit-maximizing behavior; (*c*) neutral technical progress. Diewert develops a more general case in which (*a*) and (*c*) are weakened. His results, which are mildly surprising, are that the measure of technical change is approximately equal in both cases. The only difference is the use of an implicit rather than explicit index. This is encouraging.

In attempting to extend this result to technical progress over a number of sectors, some difficulties arise. To obtain an answer, further restrictions on technical change within a sector are required. They must all be strongly Hicks neutral. In this case, the geometric mean of the sectoral measures of technical change provides the correct answer.

Having assessed the present status of measures of technical change, Diewert turns to perhaps the most serious and frustrating problem of all. New goods are continually being developed, and our capability of analyzing problems is limited by the complexity of measurement and theory in the presence of new goods.

In examining the new-goods problem, two approaches are considered. An attempt to evaluate the errors associated with (*a*) setting the price of a good equal to zero in the period in which a new good is unavailable and (*b*) setting the price equal to zero in both periods. The first is always biased upward. The second is biased upward if the new good has a relative price change less than the relative price change of a Paasche price index of all goods. It is shown that the upward bias in (*a*) is smaller than in (*b*), which suggests a method for measuring prices not commonly used in these cases.

The second approach uses duality theory and flexible functional forms to suggest a possible method for estimating the "reservation" prices of the new good in the first period. I cannot explain the details here. However, I think it is clear as the author states that the data requirements for implementation are severe.

I would like to see some suggestions made from the floor over the course of this conference on how we proceed under conditions that I believe we can roughly agree upon. If our theoretical understanding of measurement and aggregation problems in capital requires vastly improved data, then how are we to proceed on either generating that data or evaluating the net benefits of alternative data collection systems? The distance in Canada and the United States between the designers and implementers of data-collection systems and the users in economics is large. While my appeal is not original, I believe we need to and can provide assistance in improving the data. I will go no further here, but I hope that some mechanism for more serious consultation will arise in both the United States and Canada.

Although he has not pursued it intensively enough, in my opinion, Diewert does begin the investigation of the problems of specifying and estimating a hedonic model in which goods have characteristics. Unfortunately, the problem of aggregating the qualities of characteristics is simplified by assuming that only one type of trade is purchased and that the total quantity of a given characteristic is simply the product of the number of trades, times the per trade quantity of that characteristic. Diewert is aware of this limitation, and within a limited space he does provide the beginnings of a useful model for estimating hedonic models.

I have omitted several topics such as aggregation of seasonal variables and vintage capital models. It is very difficult to provide an evaluation of a very long and detailed paper. I will restrict myself to some quite general remarks. The theory of index numbers has predominantly been developed in terms of homogeneous functions. This appears perfectly reasonable when you think of an aggregate quantity index as independent of any behavioral or technical function. However, if you pursue the links between economic theory, index numbers, and flexible forms, then this assumption becomes suspect. If you double a subset of the micro inputs of a production function, you need not expect the aggregate input to double unless you want the production technology to be homogeneous. The problem is that, though weak separability does not require homotheticity or homogeneity, consistent two-stage optimization and consequently consistent aggregation does require one of these assumptions. It would be pleasant to hope some work could be done on considering weaker forms of homotheticity and errors associated with approximate homotheticity. The latter, of course, should be contrasted with the possibility of approximate price proportionality.

I would like to emphasize a few points of danger. First, the detailed concrete results in many portions of the paper are derived using particular cases of flexible functional forms. I have urged Diewert to attempt to clarify the following issue. In what cases is it true that the results can be obtained for any or many flexible forms, and in what case are the results highly specific to a particular form?

Empirical studies in economics are slowly recognizing the necessity of a more direct recognition of the approximations involved in both the data and the functional forms. The pure theory of aggregation is never going to comfort the empirical economists. The gap must be closed with more explicit models of the errors of approximation and aggregation that are bound to arise. Diewert's paper is a contribution to this very broad question, although it is only a beginning.

Reply by Diewert

A brief response to a number of specific points raised by Professor Denny seems in order. Section 8.4 of my paper has been totally revised to reflect Denny's comment that "the durability of capital creates difficulties in constructing a capital aggregate from an empirical point of view, *additional* to the theoretical difficulties inherent in constructing any kind of an aggregate."

Second, Professor Denny astutely observes that the many concrete results in the paper have been derived under the assumption that the underlying functional form is translog, and he asks whether similar concrete results can be generated by using other flexible functional forms instead of the translog. My answer is that it may be possible, but I have not been able to do it. It appears to be difficult to obtain functional forms that are linearly homogeneous, flexible, and quadratic in logarithms so that the quadratic approximation lemma (59) yields the very useful identity (64), upon which my concrete results are built.

Third, Professor Denny notes that the theory of index numbers has predominantly been developed in terms of homogeneous functions, and he wonders to what extent this assumption could be relaxed. I have certainly made liberal use of the assumption of constant returns to scale in my paper. However, the reader should note that all of my results involving the Malmquist quantity index did *not* require the linear homogeneity assumption (but they *did* require the choice of a very specific reference vector). I further note that although I have assumed homogeneous weak separability in order to justify two-stage optimization and aggregation, the *theoretical* literature on two-stage budgeting and decentralization does not require homogeneous weak separability. The main theorems in this area are due to Gorman (1959) and the extensive literature on the subject is reviewed and extended by Blackorby, Primont, and Russell (1978, chap. 5). On the other hand, the index number implications of this literature have not yet been completely worked out, although Afriat's (1972) theory of marginal price indexes makes a start in this direction. This appears to be a fruitful area for further research, as Denny notes.