The Normalized Quadratic Expenditure Function

W. Erwin Diewert,[1] University of British Columbia
Kevin J. Fox, University of New South Wales

January 8, 2009

Discussion Paper 09-04,
Department of Economics,
University of British Columbia,
#997-1873 East Mall,
Vancouver, BC V6T 1Z1 Canada.
email: diewert@econ.ubc.ca

Abstract

A concise introduction to the Normalized Quadratic expenditure or cost function is provided so that the interested reader will have the necessary information to understand and use this functional form. The Normalized Quadratic is an attractive functional form for use in empirical applications as correct curvature can be imposed in a parsimonious way without losing the desirable property of flexibility. We believe it is unique in this regard. Topics covered included the problem of cardinalizing utility, the modeling of nonhomothetic preferences, the use of spline functions to achieve greater flexibility and the use of a "semiflexible" approach to make it feasible to estimate systems of equations with a large number of commodities.

JEL Classification: D110

Key words: Normalized Quadratic, expenditure function, flexible functional forms, elasticities.

---

[1] The authors gratefully acknowledge the financial support from an Australian Research Council Discovery Grant (DP0559033) and the Social Sciences and Humanities Council of Canada.

1. Introduction

In this chapter, we will study the Normalized Quadratic expenditure or cost function and to a lesser extent, the Generalized Leontief cost function. Both of these functional forms are flexible; i.e., they can approximate arbitrary twice continuously differentiable functions in the appropriate class of functions to the second order at an arbitrary point of approximation. Thus they are very useful in applications where it is necessary to estimate elasticities of demand, since these flexible functional forms can approximate arbitrary differentiable demand functions to the first order. We will make extensive use of duality theory[2] in this chapter in order to obtain systems of demand functions that are consistent with economic theory but yet can be estimated by using linear regression techniques or "slightly" nonlinear regressions. Since many problems in applied economics depend on obtaining accurate estimates of elasticities, this topic is of considerable importance for the applied economist.

A producer's cost function is the solution to the problem of minimizing the cost of producing a given output target given input prices that are fixed to the producer. A consumer's expenditure function is the solution to the problem of minimizing the expenditure required to achieve a target level of utility given commodity prices that are fixed to the consumer. It turns out that these two problems are isomorphic to each other so up to a certain point, they can be studied using the same framework. In the end however, the consumer's expenditure minimization problem will prove to be more difficult to "solve" from an applied point of view. Initially we will make use of the similarity in these two minimization problems, because the econometric issues in the production context are not as complex as they turn out to be in the consumer context. Thus in sections 3 and 4, we will approach the econometrics of the consumer's problem by laying out solutions to the producer's cost minimization problem from an econometric point of view. Then in subsequent sections, when we study the consumer's expenditure minimization problem, we will find it relatively easy to adapt the previous producer oriented material.

Section 2 below starts off by giving a formal definition of a flexible functional form for a production or utility function and a cost or expenditure function. Basically, flexible functional forms are functional forms that have a second order approximation property so that elasticities of demand are not a priori restricted by using a flexible functional form. Sections 3 and 4 give two examples of flexible functional forms for cost functions: the Generalized Leontief cost function, and the Normalized Quadratic cost function. The Normalized Quadratic functional form is our preferred functional form, because convexity or concavity restrictions can be imposed on this functional form in a parsimonious way without destroying the flexibility of the functional form. We do not know of any other flexible functional form that has this property.[3]

Section 5 shows how cost functions can be applied to the problems involved in estimating systems of consumer demand functions that are consistent with utility maximizing behavior. Sections 5.1 and 5.2 apply the general strategy to the problem of estimating homothetic Generalized Leontief and Normalized Quadratic preferences. Section 6 notes a problem with the algebra presented in section 5 and provides a solution to the problem. The problem is that when we econometrically estimate preferences, we have to somehow cardinalize utility and section 6 discusses possible solutions to this cardinalization problem.

Section 7 draws on the previous sections and shows how flexible functional forms that are dual to nonhomothetic preferences can be estimated. The Generalized Leontief and Normalized Quadratic models studied earlier that were adequate to model homothetic preferences are modified (by the addition of some new parameters) to deal with nonhomothetic preferences in a flexible manner.

Section 8 shows how the use of linear spline functions can be added to the models presented in section 7 in order to better approximate arbitrary Engel curves.

A functional form requires approximately $N^2/2$ free parameters in order to be flexible if there are N commodities in the demand system. Thus if N is 10, we require roughly 50 free parameters, which can be handled in a time series context, but if the number of commodities is 100, we require 5,000 parameters, which is difficult to handle in a flexible functional form context. However, in section 9, we discuss semiflexible functional forms, which can be used to approximate flexible functional forms in situations where the number of commodities is large. Section 10 concludes.

---

[2] See Diewert (1974a) for a review of duality theory.
[3] For a comparison of the Normalized Quadratic functional form with other flexible functional forms, see Diewert and Wales (1993).

2. The Definition of a Flexible Functional Form

It is convenient to define the concept of a flexible functional form in two contexts: one where the underlying aggregator function[4] f (a production function or a utility function) is linearly homogeneous and another where the function f is not necessarily linearly homogeneous.

In the production function context, f is regarded as a production function, while in the utility context, f is regarded as a utility function. In the production function context, we have $y = f(x_1,x_2,...,x_N) = f(x)$ where $y \geq 0$ denotes the output produced by the nonnegative input vector $x \geq 0_N$.[5] In the consumer context, we replace the output level y by the utility level u and the vector x is now interpreted as a vector of commodity demands.

A *flexible functional form*[6] f is a functional form that has enough parameters in it so that f can approximate an arbitrary twice continuously differentiable function f^* to the second order at an arbitrary point x^* in the domain of definition of f and f^*. Thus f must have enough free parameters in order to satisfy the following $1+N+N^2$ equations:[7]

(1) $f(x^*) = f^*(x^*)$; (1 equation)

(2) $\nabla f(x^*) = \nabla f^*(x^*)$; (N equations)

(3) $\nabla^2 f(x^*) = \nabla^2 f^*(x^*)$; (N^2 equations).

Of course, since both f and f^* are assumed to be twice continuously differentiable, we do not have to satisfy all N^2 equations in (3) since Young's Theorem implies that $\partial^2 f(x^*)/\partial x_i \partial x_j = \partial^2 f(x^*)/\partial x_j \partial x_i$ and $\partial^2 f^*(x^*)/\partial x_i \partial x_j = \partial^2 f^*(x^*)/\partial x_j \partial x_i$ for all i and j. Thus the matrices of second order partial derivatives $\nabla^2 f(x^*)$ and $\nabla^2 f^*(x^*)$ are both symmetric matrices and so there are only $N(N+1)/2$

[4] Diewert (1976; 115) introduced this terminology.

[5] Notation: 0_N denotes a vector of N zeros. Then $x \geq 0_N$ means that each input is used in either zero or positive quantities.

[6] This terminology was introduced by Diewert (1976; 115).

[7] Notation: $\nabla f(x)$ denotes the (column) vector of first order partial derivatives of f evaluated at x, $[\partial f(x)/\partial x_1,..., \partial f(x)/\partial x_N]^T$, where the superscript ^{T} denotes transposition and $\nabla^2 f(x)$ denotes the N by N matrix of second order partial derivatives of f evaluated at the point x. The ij[th] element of $\nabla^2 f(x)$ will be denoted by either $\partial^2 f(x)/\partial x_i \partial x_j$ or $f_{ij}(x)$.

independent equations to be satisfied in the restrictions (3). Thus a general flexible functional form must have at least 1+N+N(N+1)/2 free parameters.

The simplest example of a flexible functional form is the following *quadratic function*:

(4) $f(x) \equiv a_0 + a^T x + (1/2)x^T Ax$; $A = A^T$

where a_0 is a scalar parameter, $a^T \equiv [a_1,...,a_N]$ is a vector of parameters and $A \equiv [a_{ij}]$ is a symmetric matrix of parameters. Thus the f defined by (4) has 1+N+N(N+1)/2 parameters. To show that this f is flexible, we need to choose a_0, a and A to satisfy equations (1)-(3). Upon noting that $\nabla f(x) = a + Ax$ and $\nabla^2 f(x) = Ax$, equations (1)-(3) become the following equations:

(5) $a_0 + a^T x^* + (1/2)x^{*T}Ax^* = f^*(x^*)$;

(6) $a + Ax^* = \nabla f^*(x^*)$;

(7) $A = \nabla^2 f^*(x^*)$.

To satisfy these equations, choose $A \equiv \nabla^2 f^*(x^*)$ (and A will be a symmetric matrix since f^* is assumed to be twice continuously differentiable); $a \equiv \nabla f^*(x^*) - Ax^*$ and finally, choose $a_0 \equiv f^*(x^*) - [\, a^T x^* + (1/2)x^{*T}Ax^*]$.

In many applications, we want to find a flexible functional form f that is also linearly homogeneous. For example, in production theory, if the minimum average cost plant size is small relative to the size of the market, then we can approximate the industry technology by means of a constant returns to scale production function. As another example, in the pure theory of international trade, we often assume that consumer preferences are *homothetic*[8]; i.e., the consumer's utility function can be represented by g[f(x)] where f is linearly homogeneous and g is a monotonically increasing and continuous function of one variable. In this case, we can represent the consumer's preferences equally well by the linearly homogeneous utility function f(x).

[8] Shephard (1953) introduced this term.

Recall the definition for f to be linearly homogeneous:[9]

(8) $f(\lambda x) = \lambda f(x)$ for all scalars $\lambda \geq 0$ and vectors $x \geq 0_N$.

If in addition, f is twice continuously differentiable, then Euler's Theorem on homogeneous functions and Young's Theorem from calculus imply the following restrictions on the first and second order partial derivatives of f:

(9) $x^T\nabla f(x) = f(x)$; (1 restriction)

(10) $\nabla^2 f(x)x = 0_N$; (N restrictions)

(11) $\nabla^2 f(x) = [\nabla^2 f(x)]^T$ (N(N−1)/2 restrictions).

The restrictions given by (9) and (10) are implied by Euler's Theorem and the symmetry restrictions (11) are implied by Young's Theorem.

If the aggregator function f is linearly homogeneous, then the corresponding *cost function* C(y,p) in the production context or *expenditure function* C(u,p) in the consumer context has the following structure: for $u > 0$ and $p >> 0_N$,

(12) $C(u,p) \equiv \min_x \{p^Tx : f(x) \geq u\}$

$= \min_x \{p^Tx : f(x) = u\}$ if f is continuous and increasing in the components of x

$= \min_x \{p^Tx : (1/u)f(x) = 1\}$

$= \min_x \{p^Tx : f(\{1/u\}x) = 1\}$ using the linear homogeneity of f

$= \min_{x/u} \{yp^T(x/u) : f(x/u) = 1\}$

$= u \min_z \{p^Tz : f(z) = 1\}$ letting $z \equiv x/u$

$= uC(1,p)$

$= uc(p)$

[9] Notation: 0_N is an N dimensional vector of zeros; $x \geq 0_N$ means each component of x is nonnegative; $x >> 0_N$ means each component of x is strictly positive and $x > 0_N$ means $x \geq 0_N$ but $x \neq 0_N$.

where we define the *unit cost function* c(p) as C(1,p), the minimum cost of producing one unit of output (in the production context) or utility (in the consumer context).

It is straightforward to show that C(1,p) and c(p) must be nondecreasing, linearly homogeneous and concave in the components of the price vector p; see for example Diewert (1974).[10]

As indicated above, linearly homogeneous functions primal aggregator functions f arise naturally in a variety of economic applications. Moreover, even if we allow our production function or utility function f to be a general nonhomogeneous function, it is often of interest to allow f to have the capability to be flexible in the class of linearly homogeneous functions.

Consider what happens to the general quadratic function f defined by (4) if we attempt to specialize it to become a linearly homogeneous flexible functional form. In order to make it homogeneous of degree one, we must set $a_0 = 0$ and $A = 0_{NxN}$ and the resulting functional form collapses down to the following linear function:

(13) $f(x) = a^T x$.

But the f defined by (13) is not a flexible linearly homogeneous functional form! Thus finding flexible linearly homogeneous functional forms is not completely straightforward in the case where the aggregator function is restricted to be linearly homogeneous.

Let us determine the minimal number of free parameters that a flexible linearly homogeneous functional form must have. If both f and f^* are linearly homogeneous (and twice continuously differentiable), then both functions will satisfy the restrictions (9)-(11). In view of these restrictions, it can be seen that instead of f having to satisfy all $1+N+N^2$ of the equations (1)-(3), f need only satisfy the following $N+N(N-1)/2 = N(N+1)/2$ equations:

(14) $\nabla f(x^*) = \nabla f^*(x^*)$; (N equations)

(15) $f_{ij}(x^*) = f^*_{ij}(x^*)$ for $1 \le i < j \le N$ (N(N−1)/2 equations)

where $f_{ij}(x^*) \equiv \partial^2 f(x^*)/\partial x_i \partial x_j$. Note that equations (15) are the equations in the upper triangle of the matrix equation (3) above. If the upper triangle equations in (3) are satisfied, then by Young's Theorem, the lower triangle equations will also be satisfied if equations (15) are satisfied. The main diagonal equations in (3) will also be satisfied if equations (15) are satisfied: the diagonal elements $f_{ii}(x^*)$ are determined by the restrictions $\nabla^2 f(x^*)x^* = 0_N$ and the $f^*_{ii}(x^*)$ are determined by the restrictions $\nabla^2 f^*(x^*)x^* = 0_N$.

Thus in order for f(x) or the dual unit cost function c(p) to be a flexible linearly homogeneous functional form, it must have at least $N + N(N-1)/2 = N(N+1)/2$ free parameters. If it has exactly this number of free parameters, then we say that f is a *parsimonious flexible functional form*.

In the following sections, we shall give some examples of parsimonious flexible functional forms for unit cost functions. Thus we look for linearly homogeneous functions c(p) that can satisfy the following N(N+1)/2 equations:

(16) $\nabla c(p^*) = \nabla c^*(p^*)$; (N equations)

(17) $c_{ij}(p^*) = c^*_{ij}(p^*)$ for $1 \le i < j \le N$ (N(N−1)/2 equations).

Why is it important that functional forms used in applied economics be flexible? From Shephard's (1953; 11) Lemma, the producer's system of cost minimizing input demand functions, x(y,p), is equal to the vector of first order partial derivatives of the cost function with respect to input prices, $\nabla_p C(y,p)$. Thus the matrix of *first order input demand price derivatives* $\nabla_p x(y,p)$ is equal to the matrix of second order partial derivatives with respect to input prices, $\nabla^2_{pp} C(y,p)$. Hence, if the functional form for C is *not* flexible, *price elasticities of input demand will be a priori restricted in some arbitrary way*. Of course, a similar comment applies in the consumer context. The consumer's system of Hicksian demand functions,[11] x(u,p), is equal to the vector of first order

[10] The underlying aggregator function f(x) need only be positive for strictly positive x and continuous from above in order to obtain these regularity conditions on the cost function or the unit cost function in the case where f is linearly homogenous.

[11] See Hicks (1946; 311-331).

partial derivatives of the cost function with respect to commodity prices, $\nabla_p C(u,p)$ and the matrix of first derivatives of these Hicksian demand functions with respect to commodity prices is equal to $\nabla^2_{pp} C(u,p)$. Hence, if the functional form for C is *not* flexible, *Hicksian price elasticities of demand will be a priori restricted in some arbitrary way*. Many practical problems in applied economics depend crucially on estimates of elasticities and hence it is usually not appropriate to use estimates of elasticities that are restricted in some arbitrary manner.

In the following two sections, we will exhibit some examples of flexible functional forms for unit cost functions. Econometric strategies for estimating the unknown parameters in these functional forms will be illustrated in the production function context; i.e., we will show how a system of estimating equations can be obtained where input demands are the dependent variables and input prices and output are the independent variables. It turns out that we cannot apply the same methods to the estimation of a consumer's system of Hicksian demand functions because unlike output y, utility u is not observable. In section 5 below, we will indicate how this problem can be overcome and we will show how the analysis in the following two sections can be adapted to the consumer context.

3. The Generalized Leontief Unit Cost Function.

Define the *generalized Leontief unit cost function* c(p) as follows: [12]

$$(18)\quad c(p_1,\ldots,p_N) \equiv \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}\, ; \qquad b_{ij} = b_{ji} \text{ for } 1 \le i < j \le N.$$

Thus c is a quadratic form in the square roots of input prices and has N(N+1)/2 b_{ij} parameters.

We need to determine whether the unit cost function c(p) defined by (18) is flexible; i.e., whether we can choose the b_{ij} so as to satisfy equations (16) and (17). Upon differentiating (18), equations (16) and (17) become the following equations:

$$(19)\quad c_i(p^*) = \sum_{j=1}^{N} b_{ij}\, (p_i^*)^{-(1/2)}(p_j^*)^{1/2} = c_i^*(p^*)\, ; \qquad i = 1,\ldots,N;$$

[12] This functional form was introduced by Diewert (1971).

(20) $c_{ij}(p^*) = (1/2)\, b_{ij}\, (p_i^*)^{-(1/2)}(p_j^*)^{-(1/2)} = c_{ij}^*(p^*)$; $\quad 1 \le i < j \le N$.

Use equations (20) to determine the b_{ij} for $1 \le i < j \le N$. Then use equations (19) to solve for the b_{ii} for $i = 1,...,N$. This proves that the c(p) defined by (18) is flexible. Since it has only N(N+1)/2 parameters, it is also parsimonious.

In a production study where there is only one output and N inputs and if the assumption of competitive cost minimization is justified, then the ith input demand x_i is equal to $\partial C(y,p)/\partial p_i$ using Shephard's Lemma and this derivative is equal to $y\partial c(p)/\partial p_i$ in the case where the production function f is linearly homogeneous, where c is the dual unit cost function. Thus given period t data on input demands, x_i^t, input prices, p_i^t and on output produced, y^t, then the unknown parameters in (18) can be estimated by using the following N estimating equations:[13]

(21) $x_i^t/y^t = \sum_{j=1}^N b_{ij}\, (p_j^t/p_i^t)^{1/2} + e_i^t$; $\quad i = 1,...,N,$

where the e_i^t are stochastic error terms for $i = 1,...,N$.[14]

Note that the b_{ij} in equation i should equal the b_{ji} in equation j. These *cross equation symmetry restrictions* can be imposed in the estimation procedure or we could *test* for their validity.

After estimating the b_{ij}, it is necessary to check whether $\nabla^2 c(p^t)$ is *negative semidefinite* at each data point p^t.[15] Thus it will be necessary to calculate the second order derivatives of c at each data point. Differentiating the c(p) defined by (18) yields the following formulae for the derivatives:

(22) $c_{ij}(p^t) = (1/2)\, b_{ij}\, (p_i^t p_j^t)^{-(1/2)}$ $\quad$ for $i \ne j$;

---

[13] We divided the inputs by the output level here because this will typically reduce heteroskedasticity.
[14] The error terms could be due to a variety of causes including: (i) errors in cost minimization; (ii) errors in the measurement of x_i^t/y^t; (iii) errors in the measurement of the input prices p_i^t and (iv) errors due to functional form approximation error; i.e., the true cost function may not be adequately represented by our assumed functional form. All of these problems may lead to the error terms being correlated with the independent variables in the system of regression equations, leading to biased estimates. We will not deal with possible econometric remedies for these econometric estimation problems in this chapter.
[15] A necessary and sufficient condition for a twice continuously differentiable c(p) to be concave over a convex set S is that $\nabla^2 c(p)$ be negative semidefinite for all p belonging to S.

$$c_{ii}(p^t) = -(1/2)\sum_{k\neq i,k=1}^{N} b_{ik}\,(p_i^t)^{-(3/2)}(p_k^t)^{(1/2)}; \qquad \text{for } i = 1,\ldots,N.$$

Note that the b_{ii} do not appear in the formulae (22) for the second order partial derivatives of the generalized Leontief unit cost function. Note also if all $b_{ij} = 0$ for $i \neq j$, then the functional form defined by (18) collapses down to the no substitution Leontief (1941) functional form.[16] Under these restrictions, the input demand functions defined by (21) collapse down to the following system of equations:

$$(23)\quad x_i^t/y^t = b_{ii} + e_i^t\,; \qquad i = 1,\ldots,N.$$

Thus input demands are not affected by changes in input prices if the producer's cost function has the Leontief functional form.

Experience with the Generalized Leontief unit cost function has shown that if the number of inputs is greater than four or so (or the number of commodities is greater than four in the consumer context), then the estimated unit cost function is often *not* locally concave for prices in the data set. Thus the concavity (or curvature) conditions that *must* be satisfied by a cost function fail and the resulting estimated elasticities cannot be used in practical applied economic problems. This failure of the curvature conditions can be avoided by restricting all of the off diagonal b_{ij} to be nonnegative.[17] However, if we impose nonnegativity on our estimated b_{ij}, then we rule out complementarity, which is a severe a priori restriction on elasticities of demand if the number of inputs or commodities is greater than two.[18]

If we are lucky, our estimated Generalized Leontief unit cost function will satisfy the concavity conditions, at least locally around the data in our sample, and all is well. But frequently, we will not be lucky and so we need to turn to flexible functional forms where the correct curvature conditions can be imposed without destroying the flexibility of the functional form. The

---

[16] This functional form was actually used by Walras (1954; 243); the first edition of this book was published in 1874.

[17] In a nonlinear regression, these restrictions can easily be imposed by setting each $b_{ij} = (a_{ij})^2$ for $i\neq j$.

[18] The translog functional form suffers from a similar problem: unrestricted translog estimates frequently fail the local concavity in prices conditions and if concavity is imposed, then the flexibility of the functional form is destroyed. See Diewert and Wales (1987) for a discussion of these problems. The translog functional form is due to Christensen, Jorgenson and Lau (1971) (1975).

normalized quadratic functional form is just such a parsimonious flexible functional form and we turn to a discussion of it in the following section.

4. The Normalized Quadratic Unit Cost Function.

The normalized quadratic unit cost function c(p) is defined as follows for $p >> 0_N$:[19]

(24) $c(p) \equiv b^T p + (1/2)p^T Bp/\alpha^T p$

where $b^T \equiv [b_1,...,b_N]$ and $\alpha^T \equiv [\alpha_1,...,\alpha_N]$ are parameter vectors and $B \equiv [b_{ij}]$ is a matrix of parameters. The vector α and the matrix B satisfy the following restrictions:

(25) $\alpha > 0_N$;
(26) $B = B^T$; i.e., the matrix B is symmetric;
(27) $Bp^* = 0_N$ for some $p^* >> 0_N$.

In most empirical applications, the vector of nonnegative but nonzero parameters α is fixed a priori. The two most frequent a priori choices for α are $\alpha \equiv 1_N$, a vector of ones or $\alpha \equiv (1/T) \sum_{t=1}^{T} x^t$, the sample mean of the observed input vectors in the producer context or the sample mean of the observed commodity vectors in the consumer context. The two most frequent choices for the reference price vector p^* are $p^* \equiv 1_N$ or $p^* \equiv p^t$ for some period t; i.e., in this second choice, we simply set p^* equal to the observed period t price vector.

Assuming that α has been predetermined, there are N unknown parameters in the b vector and N(N−1)/2 unknown parameters in the B matrix, taking into account the symmetry restrictions (26) and the N linear restrictions in (27). Note that the c(p) defined by (24) is linearly homogeneous in the components of the input price vector p.

[19] This functional form was introduced by Diewert and Wales (1987; 53) where it was called the Symmetric Generalized McFadden functional form. Additional material on this functional form and applications can be found in Diewert and Wales (1988a, 1988b, 1992, 1993), Kohli (1993, 1994, 1998) and Fox (1996, 1998).

Another possible way of defining the normalized quadratic unit cost function is as follows:

(28) $c(p) \equiv (1/2)\, p^T A p/\alpha^T p$

where the parameter matrix A is symmetric; i.e., $A = A^T \equiv [a_{ij}]$ and $\alpha > 0_N$ as before. Assuming that the vector of parameters α has been predetermined, the c(p) defined by (28) has N(N+1)/2 unknown a_{ij} parameters.

Comparing (24) with (28), it can be seen that (28) has dropped the b vector but has also dropped the N linear constraints (27). It can be shown that the model defined by (24) is a special case of the model defined by (28). To show this, given (24), define the matrix A in terms of B, b and α as follows:

(29) $A \equiv B + [b\alpha^T + \alpha b^T]$.

Substituting (29) into (28), (28) becomes:

$$
\begin{aligned}
(30)\ c(p) &= (1/2)p^T\{B + [b\alpha^T + \alpha b^T]\}p/\alpha^T p \\
&= (1/2)p^T Bp/\alpha^T p + (1/2)\, p^T[b\alpha^T + \alpha b^T]p/\alpha^T p \\
&= (1/2)p^T Bp/\alpha^T p + (1/2)\{p^T b\alpha^T p + p^T \alpha b^T p\}/\alpha^T p \\
&= (1/2)p^T Bp/\alpha^T p + (1/2)\{2p^T b\alpha^T p\}/\alpha^T p \\
&= (1/2)p^T Bp/\alpha^T p + p^T b
\end{aligned}
$$

which is the same functional form as (24). However, we prefer to work with the model (24) rather than with the seemingly more general model (28) for three reasons:

- The c(p) defined by (28) clearly contains the no substitution Leontief functional form as a special case (simply set $B = 0_{N\times N}$);
- The estimating equations that correspond to (24) will contain constant terms and
- It is easier to establish the flexibility property for (24) than for (28).

The first and second order partial derivatives of the normalized quadratic unit cost function defined by (24) are given by:

(31) $\nabla c(p) = b + (\alpha^T p)^{-1} Bp - (1/2)(\alpha^T p)^{-2} p^T Bp\alpha$;
(32) $\nabla^2 c(p) = (\alpha^T p)^{-1} B - (\alpha^T p)^{-2} Bp\alpha^T - (\alpha^T p)^{-2} \alpha p^T B + (\alpha^T p)^{-3}\, p^T Bp\alpha\alpha^T$.

We now prove that the c(p) defined by (24)-(27) (with $\alpha > 0_N$) predetermined) is a flexible functional form at the point p^*. Using the restrictions (27), $Bp^* = 0_N$, we have $p^{*T}Bp = p^{*T}0_N = 0$. Thus evaluating (31) and (32) at $p = p^*$ yields the following equations:

(33) $\nabla c(p^*) = b$;
(34) $\nabla^2 c(p^*) = (\alpha^T p^*)^{-1} B$.

We need to satisfy equations (16) and (17) above to show that the c(p) defined by (24)-(27) is flexible at p^*. Using (33), we can satisfy equations (16) if we choose b as follows:

(35) $b \equiv \nabla c^*(p^*)$.

Using (34), we can satisfy equations (17) by choosing B as follows:

(36) $B \equiv (\alpha^T p^*)\, \nabla^2 c^*(p^*)$.

Since $\nabla^2 c^*(p^*)$ is a symmetric matrix, B will also be a symmetric matrix and so the symmetry restrictions (26) will be satisfied for the B defined by (36). Moreover, since $c^*(p)$ is assumed to be a linearly homogeneous function, Euler's Theorem implies that

(37) $\nabla^2 c^*(p^*)p^* = 0_N$.

Equations (36) and (37) imply that the B defined by (36) satisfies the linear restrictions (27). This completes the proof of the flexibility property for the normalized quadratic unit cost function.

It is convenient to define the vector of *normalized input prices*, $v^T \equiv [v_1,...,v_N]$ as follows:

(38) $v \equiv (p^T\alpha)^{-1}p$.

In the production function context, the system of input demand functions x(y,p) that corresponds to the normalized quadratic unit cost function c(p) defined by (24) can be obtained using Shephard's Lemma in the usual way:

(39) $x(y,p) = y\nabla c(p)$.

Using (39) and (31) evaluated at the period t data, we obtain the following system *of estimating equations*:

(40) $x^t/y^t = b + Bv^t - (1/2)v^{tT}Bv^t\alpha + e^t$; $\quad t = 1,...,T$

where x^t is the observed period t input vector, y^t is the period t output, $v^t \equiv p^t/\alpha^Tp^t$ is the vector of period t normalized input prices and $e^t \equiv [e_1{}^t,...,e_N{}^t]^T$ is a vector of stochastic error terms. Equations (40) can be used in order to statistically estimate the parameters in the b vector and the B matrix. Note that equations (40) are linear in the unknown parameters. Note also that the symmetry restrictions (26) can be imposed in (40) (using standard econometric packages) or their validity can be tested.

Once estimates for b and B have been obtained (denote these estimates by b^* and B^* respectively), then equations (40) can be used in order to generate a period t vector of fitted input demands, x^{t*} say:

(41) $x^{t*} \equiv y^t[b^* + B^*v^t - (1/2)v^{tT}B^*v^t\alpha]$; $\quad t = 1,...,T.$

Equations (32) and (39) may be used in order to calculate the matrix of period t *estimated input price derivatives*, $\nabla_p x(y^t,p^t) = \nabla^2_{pp}C(y^t,p^t)$. Our point estimate for the matrix $\nabla^2_{pp}C(y^t,p^t)$ is:

(42) $[C_{ij}^{t*}] \equiv y^t[(\alpha^T p^t)^{-1}B^* - (\alpha^T p^t)^{-2}B^* p^t \alpha^T - (\alpha^T p^t)^{-2}\alpha p^{tT}B^* + (\alpha^T p^t)^{-3}\, p^{tT}Bp^t\alpha\alpha^T]$; $t = 1,...,T$.

Equations (41) and (42) may be used in order to obtain estimates for the matrix *of period t input demand price elasticities*, $[E_{ij}^t]$:

(43) $E_{ij}^t \equiv \partial \ln x_i(y^t,p^t)/\partial \ln p_j = p_j^t\, C_{ij}^{t*}/x_i^{t*}$; $i, j = 1,...,N$; $t = 1,...,T$

where x_i^{t*} is the ith component of the vector of fitted demands x^{t*} defined by (41).

There is one important additional topic that we have to cover in our discussion of the normalized quadratic functional form: what conditions on b and B are necessary and sufficient to ensure that c(p) defined by (24)-(27) is concave in the components of the price vector p?

The function c(p) will be concave in p if and only if $\nabla^2 c(p)$ is a negative semidefinite matrix for each p in the domain of definition of c. Evaluating (32) at $p = p^*$ and using the restrictions (27) yields:

(44) $\nabla^2 c(p^*) = (\alpha^T p^*)^{-1}B$.

Since $\alpha > 0_N$ and $p^* >> 0_N$, $\alpha^T p^* > 0$. Thus in order for c(p) to be a concave function of p, the following *necessary condition* must be satisfied:

(45) B is a negative semidefinite matrix.

We now show that the necessary condition (45) is also *sufficient* to imply that c(p) is concave over the set of p such that $p >> 0_N$. Unfortunately, the proof is somewhat involved.[20]

Let $p >> 0_N$. We assume that B is negative semidefinite and we want to show that $\nabla^2 c(p)$ is negative semidefinite or equivalently, that $-\nabla^2 c(p)$ is positive semidefinite. Thus for any vector z, we want to show that $-z^T\nabla^2 c(p)z \geq 0$. Using (32), this inequality is equivalent to:

$$(46)\ -(\alpha^Tp)^{-1}\, z^TBz + (\alpha^Tp)^{-2}\, z^TBp\alpha^Tz + (\alpha^Tp)^{-2}\, z^T\alpha p^TBz - (\alpha^Tp)^{-3}\, p^TBpz^T\alpha\alpha^Tz \geq 0 \qquad \text{or}$$

$$(47)\ -(\alpha^Tp)^{-1}\, z^TBz - (\alpha^Tp)^{-3}\, p^TBp(\alpha^Tz)^2 \geq -2(\alpha^Tp)^{-2}\, z^TBp\alpha^Tz \qquad \text{using } B = B^T.$$

Define $A \equiv - B$. Since B is symmetric and negative semidefinite by assumption, A is symmetric and positive semidefinite. Thus there exists an orthonormal matrix U such that

$$(48)\ U^TAU = \Lambda\ ;$$
$$(49)\ U^TU = I_N$$

where I_N is the N by N identity matrix and Λ is a diagonal matrix with the nonnegative eigenvalues of A, λ_i, i = 1,...,N, running down the main diagonal. Now premultiply both sides of (48) by U and postmultiply both sides by U^T. Using (49), $U^T = U^{-1}$, and the transformed equation (48) becomes the following equation:

$$\begin{aligned}(50)\ A &= U\Lambda U^T \\ &= U\Lambda^{1/2}\,\Lambda^{1/2}\,U^T \\ &= U\Lambda^{1/2}\,U^TU\,\Lambda^{1/2}\,U^T \qquad \text{since } U^TU = I_N \\ &= S\,S\end{aligned}$$

where $\Lambda^{1/2}$ is the diagonal matrix that has the nonegative square roots $\lambda_i^{1/2}$ of the eigenvalues of A running down the main diagonal and the symmetric square root of A matrix S is defined as

$$(51)\ S \equiv U\Lambda^{1/2}\,U^T.$$

If we replace $-$ B in (47) with A, the inequality that we want to establish becomes

$$(52)\ 2(\alpha^Tp)^{-1}\, z^TAp\alpha^Tz \leq z^TAz + (\alpha^Tp)^{-2}\, p^TAp(\alpha^Tz)^2$$

[20] The proof is due to Diewert and Wales (1987).

where we have also multiplied both sides of (47) by the positive number $\alpha^T p$ in order to derive (51) from (47).

Recall the Cauchy-Schwarz inequality for two vectors, x and y:

(53) $x^T y \le (x^T x)^{1/2} (y^T y)^{1/2}$.

Now we are ready to establish the inequality (52). Using (50), we have:

(54) $(\alpha^T p)^{-1} z^T A p \alpha^T z = (\alpha^T p)^{-1} z^T S S p \alpha^T z$

$\le (z^T S S^T z)^{1/2} ([\alpha^T p]^{-2} [\alpha^T z]^2 p^T S^T S p)^{1/2}$

using (53) with $x^T \equiv z^T S$ and $y \equiv (\alpha^T p)^{-1} (\alpha^T z) Sp$

$= (z^T S S z)^{1/2} ([\alpha^T p]^{-2} [\alpha^T z]^2 p^T S S p)^{1/2}$ using $S = S^T$

$= (z^T A z)^{1/2} ([\alpha^T p]^{-2} [\alpha^T z]^2 p^T A p)^{1/2}$ using (50), $A = SS$

$\le (1/2)(z^T A z) + (1/2)[\alpha^T p]^{-2} [\alpha^T z]^2 (p^T A p)$

using the nonnegativity of $z^T Az$, $p^T Ap$ and $\alpha^T z$, the positivity of $\alpha^T z$ and the Theorem of the Arithmetic and Geometric Mean.

The inequality (54) is equivalent to the desired inequality (52).

Thus the normalized quadratic unit cost function defined by (24)-(27) will be concave over the set of positive prices if and only if the symmetric matrix B is negative semidefinite. Thus after econometric estimates of the elements of B have been obtained using the system of estimating equations (40), we need only check that the resulting estimated B matrix is negative semidefinite.

However, suppose that the estimated B matrix is *not* negative semidefinite. How can one reestimate the model, impose negative semidefiniteness on B, but without destroying the flexibility of the normalized quadratic functional form?

The desired imposition of negative semidefiniteness can be accomplished using a technique due to Wiley, Schmidt and Bramble (1973): simply replace the matrix B by

(55) $B \equiv -AA^T$

where A is an N by N lower triangular matrix; i.e., $a_{ij} = 0$ if $i < j$.[21]

We also need to take into account the restrictions (59), $Bp^* = 0_N$. These restrictions on B can be imposed if we impose the following restrictions on A:

(56) $A^Tp^* = 0_N$.

To show how this curvature imposition technique works, let $p^* = 1_N$ and consider the case N = 2. In this case, we have:

$$A \equiv \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \text{ and } A^T = \begin{bmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{bmatrix}.$$

The restrictions (56) become: $A^T 1_2 = \begin{bmatrix} a_{11} + a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

and hence we must have $a_{21} = -a_{11}$ and $a_{22} = 0$. Thus in this case,

$$(57)\ B \equiv -AA^T = -\begin{bmatrix} a_{11} & 0 \\ -a_{11} & 0 \end{bmatrix}\begin{bmatrix} a_{11} & -a_{11} \\ 0 & 0 \end{bmatrix} = -\begin{bmatrix} a_{11}^2 & -a_{11}^2 \\ -a_{11}^2 & a_{11}^2 \end{bmatrix} = a_{11}^2\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}.$$

---

[21] Since $z^TAA^Tz = (A^Tz)^T(A^Tz) = y^Ty \geq 0$ for all vectors z, AA^T is positive semidefinite and hence $-AA^T$ is negative semidefinite. Diewert and Wales (1987; 53) showed that any positive semidefinite matrix can be written as AA^T where A is lower triangular. Hence, it is not restrictive to reparameterize an arbitrary negative semidefinite matrix B as $-AA^T$.

Equations (57) show how the elements of the B matrix can be defined in terms of the single parameter, a_{11}^2. Note that with this reparameterization of the B matrix, it will be necessary to use nonlinear regression techniques rather than modifications of linear regression techniques. This turns out to be the cost of imposing the correct curvature conditions on the unit cost function.

In the following sections, we indicate how the functional forms described in sections 3 and 4 in the producer context can be adapted to estimate consumer preferences.

5. The Estimation of Consumer Preferences: The General Framework

It would seem that the producer cost function framework described in the previous two sections can be readily adapted to the problem of estimating consumer preferences: simply replace output y by utility u, reinterpret the production function f as a utility function, reinterpret the input vector x as a vector of commodity demands and reinterpret the vector of input prices p as a vector of commodity prices. If the cost function is differentiable with respect to the components of the commodity price vector p, then Shephard's (1953; 11) Lemma applies and the consumer's system of Hicksian commodity demand functions as functions of the chosen utility level u and the commodity price vector p, x(u,p), is equal to the vector of first order partial derivatives of the cost or expenditure function C(u,p) with respect to the components of p:

(58) $x(u,p) = \nabla_p C(u,p)$.

Thus, initially, it seems that we can adapt the theory of cost and production functions used in sections 3-4 above in a very straightforward way and estimate consumer preferences in exactly the same way that we estimated cost functions that were dual to production functions. Thus we need only replace period t output, y^t, by period t utility, u^t, in the estimating equations (21) (for the generalized Leontief cost function) and (40) (for the normalized quadratic cost function) and reinterpret the resulting equations. However, there is a major problem: the period t output level y^t is an *observable* variable but the period t utility level u^t is *not observable*!

However, this problem can be solved. We need only equate the cost function C(u,p) to the consumer's *observable expenditure* in the period under consideration, Y say, and solve the resulting equation for u as a function of Y and p. Thus u = g(Y,p) is the solution to:

(59) $C(u,p) = Y$

and the resulting solution function g(Y,p) is the *consumer's indirect utility function*. Now replace the u in the system of Hicksian demand functions (58) by g(Y,p) and we obtain the consumer's system of (observable) market demand functions:

(60) $x = \nabla_p C(g(Y,p),p)$.

We will conclude this section by showing how the above general framework can be implemented for the Generalized Leontief and Normalized Quadratic models explained in sections 3 and 4 above in the production context. In the remainder of this section, we will assume that the consumer's preferences can be represented by a homothetic utility function and so without loss of generality, we will assume that the consumer's utility function is a linearly homogeneous function. In the following section, we will indicate how the models in this section can be generalized to deal with nonhomothetic preferences.

5.1 The Generalized Leontief Expenditure Function for Homothetic Preferences.

We illustrate the above procedure for the generalized Leontief cost function defined in section 3 above. For this functional form, equation (59) becomes:

(61) $u \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2} = Y$; $\qquad (b_{ij} = b_{ji}$ for all i and j)

and the u solution to this equation is:

(62) $u = g(Y,p) = Y/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}]$.

Substituting (62) into (60) leads to the following system of market demand functions:

(63) $x_i = [\sum_{j=1}^{N} b_{ij}\, (p_j/p_i)^{1/2}]\; Y/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{1/2}\, p_j^{1/2}]$; $i = 1,\ldots,N$.

Evaluating (63) at the period t data and adding a stochastic error term e_i^t to equation i in (63) for $i = 1,\ldots,N$ leads to the following system of estimating equations:[22]

(64) $x_i^t = [\sum_{j=1}^{N} b_{ij}\, (p_j^t/p_i^t)^{1/2}]\; Y^t/[\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, (p_i^t)^{1/2}\, (p_j^t)^{1/2}] + e_i^t$; $t = 1,\ldots,T$; $i = 1,\ldots,N$.

5.2 The Normalized Quadratic Expenditure Function for Homothetic Preferences.

We can also illustrate the above procedure for the normalized quadratic cost function defined in section 4 above. For this functional form, equation (59) becomes:

(65) $u[b^Tp + (1/2)(\alpha^Tp)^{-1}p^TBp] = Y$

and the u solution to this equation is:

(66) $u = g(Y,p) = Y/[b^Tp + (1/2)(\alpha^Tp)^{-1}p^TBp]$.

Substituting (66) into (58) leads to the following system of market demand functions:

(67) $x = [b + Bv - (1/2)v^TBv\alpha][(\alpha^Tp)^{-1}Y]/[b^Tv + (1/2)v^TBv]$

[22] Since Y^t will typically equal $\sum_{i=1}^{N} p_i^t x_i^t$, it can be verified that the errors in (97) for any period t cannot be independently distributed since they must satisfy the restriction $\sum_{i=1}^{N} p_i^t e_i^t = 0$ for each t; see (103) below. It is also necessary to impose a normalization on the b_{ij} since the right hand side of each equation in (7) is homogeneous of degree 0 in the b_{ij}. We will deal with the normalization problem in section 7 below.

where $v \equiv (\alpha^T p)^{-1} p = p/\alpha^T p$ is the vector of normalized prices. Evaluating (67) at the period t data and adding a vector of stochastic error terms e^t to the resulting equations leads to the following system of estimating equations:

$$(68)\ x^t = [b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}Y^t]/[b^T v^t + (1/2)v^{tT}Bv^t] + e^t\ ; \qquad t = 1,...,T$$

where $v^t \equiv p^t/\alpha^T p^t$ for t = 1,...,T.

In practice, period t "income" Y^t is defined to be period t expenditure, $p^{tT}x^t = \sum_{i=1}^N p_i^t x_i^t$; i.e., we have:

$$(69)\ Y^t = p^{tT}x^t = \sum_{i=1}^N p_i^t x_i^t\ ; \qquad t = 1,...,T.$$

However, the identities (69) create some econometric difficulties: namely, we cannot assume that all of the error terms, e_i^t, in each period are independently distributed. Thus if we premultiply both sides of equation i for period t in (64) by p_i^t and sum over i, we obtain the following identity using (69):

$$(70)\ Y^t = Y^t + \sum_{i=1}^N p_i^t e_i^t\ ; \qquad t = 1,...,T$$

which in turn implies that the period t error terms e_i^t satisfy the following exact identity:

$$(71)\ \sum_{i=1}^N p_i^t e_i^t = 0\ ; \qquad t = 1,...,T.$$

In a similar fashion, premultiplying both sides of the period t equation in (68) by p^{tT}, we obtain the following equations:

$$(72)\ p^{tT}x^t = p^{tT}[b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}Y^t]/[b^T v^t + (1/2)v^{tT}Bv^t] + p^{tT}e^t\ ; \qquad t = 1,...,T \text{ or}$$
$$Y^t = p^{tT}\ \alpha^T p^t(\alpha^T p^t)^{-1}[b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}Y^t]/[b^T v^t + (1/2)v^{tT}Bv^t] + p^{tT}e^t \qquad \text{or}$$
$$Y^t = v^{tT}\ \alpha^T p^t[b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}Y^t]/[b^T v^t + (1/2)v^{tT}Bv^t] + p^{tT}e^t \qquad \text{or}$$
$$Y^t = v^{tT}\ [b + Bv^t - (1/2)v^{tT}Bv^t\alpha][Y^t]/[b^T v^t + (1/2)v^{tT}Bv^t] + p^{tT}e^t \qquad \text{or}$$

$$Y^t = [b^Tv^t + (1/2)v^{tT}Bv^t][Y^t]/[b^Tv^t + (1/2)v^{tT}Bv^t] + p^{tT}e^t \qquad \text{or}$$
$$Y^t = Y^t + p^{tT}e^t$$

which in turn implies that the period t error term vector e^t satisfies the following exact identity, (71).

Thus for both the generalized Leontief and the normalized quadratic cost function models the period t error vectors satisfy an exact identity and hence in both models, we must drop one estimating equation; i.e., we must drop one of the estimating equations in (64) and one of the estimating equations in (68). Thus there are some substantial differences between the cost function models in the producer context and in the consumer context.

6. The Problem of Cardinalizing Utility.

There is another significant difference between the producer models discussed in sections 3 and 4 and the consumer models discussed in section 5. Looking at (64), it can be seen that the right hand side explanatory variables are *homogeneous of degree 0* in the b_{ij} coefficients. Thus the regression will not be able to determine the *scale* of the b_{ij} parameters. Similarly, by looking at the right hand side of (68), it can be seen that the right hand side explanatory variables are *homogeneous of degree 0* in the components of the b vector and the B matrix. Thus the regression will not be able to determine the *scale* of the parameters in b and B. This indeterminacy means that we require at least one additional restriction or normalization on the parameters of each of these models. Basically, what we have to do is *cardinalize* our measure of utility in some way.

There are two simple ways of cardinalizing utility:[23]

- Pick a strictly positive reference quantity vector $x^* >> 0_N$. Let the period t consumption vector x^t be on the indifference surface $I(x^t) \equiv \{x: f(x) = f(x^t)\}$. Let $\lambda^t x^*$ be on the $I(x^t)$ indifference curve. Then measure period t utility as λ^t.

[23] The two methods are equivalent in the case of homothetic preferences.

- Pick a positive reference price vector $p^* >> 0_N$. Then normalize the consumer's cost function C(u,p) so that it has the following property:

(73) $C(u,p^*) = u$ for all $u > 0$.

The meaning of (73) is that if the consumer faces the reference price vector p^*, then his or her utility will be equal to his or her "income" or expenditure on commodities at those reference prices. Thus if relative prices never changed, the consumer's utility is proportional to the size of the observed budget set. This serves to cardinalize utility for all consumption vectors. Samuelson (1974) called this type of cardinalization of utility, *money metric utility*.[24]

We will follow the money metric method of scaling utility. For the generalized Leontief model, (73) implies the following normalization of the b_{ij}:

(74) $\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}\, p_i^{*1/2}\, p_j^{*1/2} = 1.$

For the normalized quadratic model, (73) implies the following normalization of the components of the b vector and the B matrix:

(75) $b^T p^* + (1/2)p^{*T}Bp^*/\alpha^T p^* = 1.$

If we choose the reference vector p^* in (73) to be the same as the reference vector p^* which occurred in (27), then $Bp^* = 0_N$ and the cardinalization restriction (75) becomes:

(76) $b^T p^* = 1.$

The Generalized Leontief and Normalized Quadratic models for estimating consumer preferences that we have considered thus far assume that preferences are homothetic. Unfortunately, empirical evidence indicates that consumer preferences are far from being homothetic. Hence, in the

[24] The basic idea can be traced back to Hicks (1941-42).

following section, we indicate how the material in this section can be generalized to accommodate nonhomothetic preferences.

7. Modeling Nonhomothetic Preferences.

Since empirical evidence (and common sense) indicates that consumer preferences are not homothetic, we need to generalize our functional forms in order to accommodate nonhomothetic preferences.

Let $C^*(u,p)$ be an arbitrary twice continuously differentiable cost function that satisfies money metric scaling at the positive reference price vector $p^* >> 0_N$; i.e., C^* satisfies:

(77) $C^*(u,p^*) = u$ for all $u > 0$.

Let c(p) be a flexible unit cost function. Then Diewert (1980; 597) showed that the following functional form could approximate C^* to the second order at (u^*,p^*) where $u^* > 0$:

(78) $C(u,p) \equiv a^Tp + uc(p)$

where the vector of parameters a can be chosen to satisfy the following restriction:

(79) $a^Tp^* = 0$.

In order for the C(u,p) to satisfy the money metric utility scaling counterpart to (77),[25] we also require that the parameters of the unit cost function satisfy the following restriction:

(80) $c(p^*) = 1$.

In order to derive the system of market demand functions that corresponds to the cost junction defined by (78), we again set C(u,p) equal to "income" Y and solve for the u = g(Y,p) solution:

[25] This counterpart is $C(u,p^*) = a^Tp^* + uc(p^*) = u$ for all $u > 0$.

(81) $u = [Y - a^T p]/c(p)$.

The *system of Hicksian demand functions* that corresponds to the cost function defined by (78) is as usual obtained using Shephard's Lemma:

(82) $x(u,p) \equiv \nabla_p C(u,p) = a + u\nabla_p c(p)$.

Now replace u in the right hand side of (82) by the right hand side of (81) and we obtain the *consumer's system of market demand functions*:

(83) $x(Y,p) = a + \nabla_p c(p)[Y - a^T p]/c(p)$.

Letting $c(p) \equiv \sum_{i=1}^N \sum_{j=1}^N b_{ij}\, p_i^{1/2}\, p_j^{1/2}$ be the generalized Leontief unit cost function, the system of market demand functions (84) becomes, after adding stochastic error terms:

(84) $x_i^t = a_i + \{[\sum_{j=1}^N b_{ij}\, (p_j^t/p_i^t)^{1/2}][Y^t - \sum_{k=1}^N a_k p_k^t]/[\sum_{i=1}^N \sum_{j=1}^N b_{ij}\, (p_i^t)^{1/2}\, (p_j^t)^{1/2}]\} + e_i^t$;
$t = 1,...,T$; $i = 1,...,N$.

One of the a_i needs to be eliminated from the estimating equations (84) using the restriction $a^T p^* = 0$ and one of the b_{ij} needs to be eliminated using the restriction $c(p^*) = 1$ in order to obtain the final system of estimating equations. Note also, if period t "income" Y^t is equal to period t expenditure on the commodities, $p^{tT}x^t$, then as before, we can only use N−1 of the N equations in (84) as estimating equations. Note that nonlinear regression techniques have to be used to estimate the unknown parameters in (84).

Letting $c(p) \equiv b^T p + (1/2)(\alpha^T p)^{-1} p^T B p$ be the normalized quadratic unit cost function (with $b^T p^* = 1$ and $Bp^* = 0_N$), the system of market demand functions (83) becomes, after adding stochastic error terms:

(85) $x^t = a + \{[b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}][Y^t - a^T p^t]/[b^T v^t + (1/2)v^{tT}Bv^t]\} + e^t$; $t = 1,...,T$

where $v^t \equiv p^t/\alpha^T p^t$ for t = 1,...,T. Obviously, nonlinear regression techniques have to be used in order to estimate the unknown parameters in the systems of estimating equations (85). One of the a_i needs to be eliminated from the estimating equations (85) using the restriction $a^T p^* = 0$ and one of the b_i needs to be eliminated using the restriction $b^T p^* = 1$ in order to obtain the final system of estimating equations. However, if period t "income" Y^t is equal to period t expenditure on the commodities, $p^{tT} x^t$, then as before, we can only use N−1 of the N equations in (85) as estimating equations. If the estimated B matrix turns out to be *not* negative semidefinite, then we need to replace B by $-AA^T$ where A is a lower triangular matrix satisfying $A^T p^* = 0_N$. Obviously, the computer coding to set up the estimating equations for the normalized quadratic system is rather complex, particularly when B must be replaced by $-AA^T$ (but it does work).[26]

One final comment on the regularity properties of the normalized quadratic functional form. As indicated above, if we replace B by $-AA^T$, the normalized quadratic functional form will be globally concave and linearly homogeneous. But another regularity property that must be satisfied is monotonicity; i.e., cost functions must be nondecreasing in input (or commodity) prices. There is no guarantee that this monotonicity property will be globally satisfied but it will generally be satisfied in the sample region because of Shephard's Lemma, which equates positive demand vectors to first order derivatives of the cost function. Thus the estimated cost function is very likely to satisfy the monotonicity property (unless the fit in one or more equations is extremely poor).

8. The Use of Linear Spline Functions to Achieve Greater Flexibility.

Although the above model is flexible around the point (u^*,p^*), as we move away from this point, the normalized quadratic regression model defined by (85) may not fit the data very well. If the plots of the actual and fitted values using the normalized quadratic model have a zigzag appearance, then it may be worthwhile to try a *linear spline model*. We will indicate below how a two segment linear spline model can be implemented. For more details (and an extension to 3 segments instead of 2), see Diewert and Wales (1993; 81-85).

We redefine the normalized quadratic cost function C(u,p) as follows:

(86) $C(u, p) = a^Tp + u(1/2)(\alpha^Tp)^{-1}p^TBp + d(u, p)$

where a satisfies $a^Tp^* = 0$ and α and B satisfy the restrictions (25)-(27). The function d(u, p) is defined as follows:

(87) $d(u,p) \equiv ub^Tp$ for $0 \leq u \leq u^*$
$\equiv u^*b^Tp + (u - u^*)f^Tp$ for $u^* \leq u$.

where $b^T \equiv [b_1,...,b_N]$ and $f^T \equiv [f_1,...,f_N]$ parameter vectors to be estimated and u^* is *a break point level of utility* to be chosen by the investigator. The vectors b and f satisfy the restrictions:

(88) $b^Tp^* = 1$; $f^Tp^* = 1$.

How should one pick the break point u^*? Examine the plots of the regression model defined by (85) and look for an observation number where the plot changes from a zig to a zag. Suppose that this observation number is t^*. Now compute index numbers of utility,[27] using the price and quantity data in the sample and determine what level of utility corresponds to the chosen observation and set this level equal to u^*. This choice of u^* will work satisfactorily if the observations which precede the chosen observation have estimated indirect utilities which are equal to or less than u^* and the remaining observations have indirect utilities that are greater than u^*.

The estimating equations for the first t^* observations will still be given by (85); i.e., for the first t^* observations, our estimating equations are:

(89) $x^t = a + \{[b + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^Tp^t)^{-1}][Y^t - a^Tp^t]/[b^Tv^t + (1/2)v^{tT}Bv^t]\} + e^t$; $t = 1,...,t^*$

where as usual, $v^t \equiv p^t/\alpha^Tp^t$.

[26] For examples of the normalized quadratic cost function in action, see Diewert and Wales (1988a) (1988b) (1993).

In order to obtain the estimating equations for the last $T - t^*$ observations, we need to form the Hicksian demand functions and calculate the indirect utility function. If $t > t^*$, then the Hicksian demand functions that correspond to the functional form defined by (86) and (87) are:

$$(90)\ x(u,p) \equiv \nabla_p C(u,p) = a + u[(\alpha^T p)^{-1}Bp - (1/2)(\alpha^T p)^{-2}p^T Bp\alpha] + u^* b + (u - u^*)f$$
$$= a + u^* b - u^* f + u[f + (\alpha^T p)^{-1}Bp - (1/2)(\alpha^T p)^{-2}p^T Bp\alpha].$$

For $t > t^*$, the indirect utility function $u = g(Y,p)$ can be obtained by solving $C(u,p) = Y$. The solution is:

$$(91)\ u = [Y - a^T p - u^* b^T p + u^* f^T p]/[f^T p + (1/2)(\alpha^T p)^{-1}p^T Bp].$$

Now substitute (91) into (90) in order to obtain the consumer's market demand functions for periods $t > t^*$. After adding stochastic error terms, we obtain the following estimating equations:

$$(92)\ x^t = a + u^* b - u^* f$$
$$+ \{[f + Bv^t - (1/2)v^{tT}Bv^t\alpha][(\alpha^T p^t)^{-1}][Y^t - a^T p^t - u^* b^T p^t + u^* f^T p^t]/[f^T v^t + (1/2)v^{tT}Bv^t]\} + e^t$$
$$\text{for } t^* < t \leq T.$$

Although the estimating equations (92) look rather formidable, they can be programmed with a bit of effort. The most difficult part of implementing the above spline model is choosing the "right" observation at which the break point occurs. By adding additional linear segments, one can approximate an arbitrary pattern of income elasticities reasonably well. However, the disadvantage of the linear splines in utility setup is that income elasticities of demand will shift discontinuously as we move from one time period to the next time period that corresponds to a different spline

[27] A superlative index number formula should be used such as the Fisher (1922) ideal quantity index. See Diewert (1976) for other examples of superlative index number formulae.

segment. This discontinuity problem can be avoided by using quadratic splines; see Diewert and Wales (1993) for an example of this quadratic spline technique.[28]

As usual, if "income" Y^t in period t is equal to expenditure $p^{tT}x^t$, then we must drop one equation in the system of estimating equations (89) and (92). Finally, if the estimated B matrix is *not* negative semidefinite, then the model should be rerun, setting $B = -AA^T$, where A is lower triangular and satisfies the restrictions $A^Tp^* = 0_N$.

9. Semiflexible Functional Forms and the Normalized Quadratic Functional Form

In models where the number of commodities N is large (say greater than 20), it can be difficult to estimate all of the parameters in the B or A matrices in a single regression: there are simply too many parameters for a nonlinear econometric package to handle without "reasonable" starting values. Thus suppose that we impose curvature on our normalized quadratic model so that we replace B by $-AA^T$ where A is lower triangular. An effective way to estimate the A matrix is to estimate it one column at a time. Thus in the first stage of the nonlinear regression model, we use the estimating equations (85) with the A (and hence the B) matrix set equal to zero. Then at the next stage we use the estimates for the parameters which are not in the B matrix as starting values for the stage 2 nonlinear regression model with B set equal to AA^T where A is a rank 1 lower triangular matrix; i.e., at this second stage, A is set equal to:[29]

$$(93)\ A \equiv \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ a_{N1} & 0 & \ldots & 0 \end{bmatrix}.$$

[28] For applications of quadratic splines using normalized quadratic functional forms in the producer context, see Diewert and Wales (1992) and Fox (1998). The latter paper proposes the use of an algorithm to adaptively fit the spline function by endogenizing the choice of both the number and location of break points.

[29] We also need to use the restrictions (56) to express a_{11} in terms of $a_{21}, \ldots, a_{N1}$. Thus if p^* is a vector of ones, the a_{11} in (56) is replaced by $-a_{21} - a_{31} \ldots - a_{N1}$. If maximum likelihood estimation is used, then in the stage 2 nonlinear regression, the starting values for $a_{21}, \ldots, a_{N1}$ are taken to be 0's so the starting log likelihood for the stage 2 nonlinear regression will be equal to the final log likelihood of the stage 1 regression. This provides a check on the programming code used. A similar strategy should be used with the subsequent stage 3, 4 and so on regressions.

The estimated parameters from this stage 2 nonlinear regression are then used as starting values in a stage 3 nonlinear regression that fills in column 2 of the lower triangular matrix A; i.e., in the stage 3 regression, A is set equal to the following rank 2 lower triangular matrix:[30]

$$(94)\ A \equiv \begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ a_{N1} & a_{N2} & \ldots & 0 \end{bmatrix}.$$

This procedure of gradually adding nonzero columns to the A matrix can be continued until the full number of N−1 nonzero columns have been added, provided that the number of time series observations T is large enough compared to N, the number of commodities in the model. However, in models where T is small relative to N, the above procedure of adding nonzero columns to A will have to be stopped well before the maximum number of N−1 nonzero columns has been added, due to the lack of degrees of freedom. Suppose that we stop the above procedure after $K < N-1$ nonzero columns have been added. Then Diewert and Wales (1988b; 330) call the resulting normalized quadratic functional form a *flexible of degree K* functional form or a *semiflexible functional form*. A flexible of degree K functional form for a cost function can approximate an arbitrary twice continuously differentiable functional form to the second order at some point, except the matrix of second order partial derivatives of the functional form with respect to prices is restricted to have maximum rank K instead of the maximum possible rank, N−1.

What is the cost of estimating a semiflexible functional form for a cost function instead of a fully flexible functional form? When we estimate a fully flexible functional form, we need the B matrix to be able to approximate an arbitrary negative semidefinite symmetric matrix B^* of rank N−1. This arbitrary B^* can be represented as a sum of N−1 rank one negative semidefinite matrices as we now show.

[30] The starting values for the stage 3 nonlinear regression for the elements in the first column of A are the final estimated values from the stage 2 nonlinear regression and the starting values for the elements in the second column of A are 0's. Again, if p^* is a vector of ones, the a_{22} in (222) is replaced by $-a_{32} - a_{42} \ldots - a_{N2}$.

Recall that any symmetric matrix can be diagonalized by means of an orthonormal transformation; i.e., there exists a matrix U equal to $[u^1,u^2,\ldots,u^N]$, where the u^n are the columns of U, such that:

$$(95)\ U^TBU = \Lambda \equiv \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_N \end{bmatrix}$$

where U satisfies

$$(96)\ U^TU = I_N$$

and Λ is a diagonal matrix with the nonpositive eigenvalues of B, the λ_n, running down the main diagonal. We order these eigenvalues starting with the biggest in magnitude and ending up with the smallest in magnitude (which is equal to 0):

$$(97)\ -\lambda_1 \geq -\lambda_2 \geq \ldots \geq -\lambda_{N-1} \geq -\lambda_N = 0.$$

Now premultiply both sides of (95) by U and post multiply both sides of (95) by U^T. Using (96), we find that:

$$\begin{aligned}(98)\ B &= U\Lambda U^T \\ &= [u^1\lambda_1, u^2\lambda_2, \ldots, u^N\lambda_N]\,[u^1, u^2, \ldots, u^N]^T \\ &= \textstyle\sum_{n=1}^{N} \lambda_n u^n u^{nT} \\ &= \textstyle\sum_{n=1}^{N-1} \lambda_n u^n u^{nT}\end{aligned}$$

where the last equality in (98) follows from the fact that $\lambda_N = 0$.

If we estimate a normalized quadratic that is flexible of degree K, then it turns out that the resulting $-AA^T$ matrix can approximate B defined by (98) as follows:

(99) $-AA^T = \sum_{n=1}^{K} \lambda_n u^n u^{nT}$.

Thus the cost of using a semiflexible functional form of degree K where K is less than N−1 is that we will miss out on the part of B that corresponds to the smallest in magnitude eigenvalues of B; i.e., our estimating $-AA^T$ will omit the negative semidefinite matrix $\sum_{n=K+1}^{N-1} \lambda_n u^n u^{nT}$, where these λ_n for n > K are smaller in magnitude than the first K eigenvalues. In many situations, this cost will be very small; i.e., as we go through the various stages of estimating A by adding an extra nonzero column to A at each stage, we can monitor the increase in the final log likelihood (if we use maximum likelihood estimation) and when the increase in stage k+1 over stage k is “small”, we can stop adding extra columns, secure in the knowledge that we are not underestimating the size of B by a large amount.

This semiflexible technique has not been widely applied but it would seem to offer some big advantages in estimating substitution matrices in situations where there are a large number of commodities in the model.[31]

10. Conclusion

This chapter has provided a concise introduction to the Normalized Quadratic expenditure or cost function. The interested reader will have the necessary information to understand and use the Normalized Quadratic form, with the references providing examples of applications in diverse contexts. The Normalized Quadratic is an attractive functional form for use in empirical applications as correct curvature can be imposed in a parsimonious way without losing the desirable property of flexibility. We believe it is unique in this regard.

We examined the use of the Normalized Quadratic functional form in both the producer and consumer contexts, highlighting the differences between the two contexts. Along the way, useful reference was made to the Generalized Leontief functional form, which was the first flexible functional form to be proposed. Other topics covered included the problem of cardinalizing utility, the modeling of nonhomothetic preferences, the use of spline functions to achieve greater flexibility and the use of a "semiflexible" approach to make it feasible to estimate systems of equations with a large number of commodities. These topics provide the reader with the tools to use this functional form in a wide range of applications.

References.

Christensen, L.R., Jorgenson, D.W. and Lau, L.J., 1971. Conjugate duality and the transcendental logarithmic production function. Econometrica 39, 255-256.

Christensen, L.R., Jorgenson, D.W. and Lau, L.J., 1975. Transcendental logarithmic utility functions. American Economic Review 65, 367-383.

Diewert, W.E., 1971. An application of the Shephard duality theorem: A generalized Leontief production function. Journal of Political Economy 79, 481-507.

Diewert, W.E., 1974. Applications of duality theory. In Intriligator, M.D. and Kendrick, D.A. (Eds.), Frontiers of Quantitative Economics, Vol. 2, Amsterdam: North-Holland. 106-171

Diewert, W.E., 1976. Exact and superlative index numbers. Journal of Econometrics 4, 114-145.

Diewert, W.E., 1980. Symmetry conditions for market demand functions. The Review of Economic Studies 47, 595-601.

Diewert, W.E. and Wales, T.J., 1987. Flexible functional forms and global curvature Conditions. Econometrica 55, 43-68.

Diewert, W.E. and Wales, T.J., 1988a. Normalized Quadratic systems of consumer demand functions. Journal of Business and Economic Statistics 6, 303-12.

Diewert, W.E. and Wales, T.J., 1988b. A Normalized Quadratic semiflexible functional form. Journal of Econometrics 37, 327-42.

Diewert, W.E. and Wales, T.J., 1992. Quadratic spline models for producer's supply and demand functions. International Economic Review 33, 705-722.

Diewert, W.E. and Wales, T.J., 1993. Linear and quadratic spline models for consumer demand functions. Canadian Journal of Economics 26, 77-106.

Fisher, I., 1922. The Making of Index Numbers. Houghton-Mifflin, Boston.

---

[31] Kohli (1994, 1998) has estimated both flexible and semi-flexible versions of Normalized Quadratic aggregator functions in the context of estimating trade elasticities. Diewert and Lawrence in some unpublished work have successfully estimated semiflexible models for profit functions for 40 to 45 commodities.

Fox, K.J., 1996. Specification of functional form and the estimation of technical progress. Applied Economics 28, 947-956.

Fox, K.J. 1998. Non-parametric estimation of technical progress. Journal of Productivity Analysis 10, 235-250.

Kohli, U., 1993. A symmetric Normalized Quadratic GNP function and the U.S. demand for imports and supply of exports. International Economic Review 34, 243-255.

Kohli, U., 1994. Canadian imports and exports by origin and destination: A semiflexible approach. Canadian Journal of Economics 27, 580-603.

Kohli, U., 1998. Semiflexibility vs. flexibility: U.S. imports and exports by commodity group. International Review of Economics and Finance 7, 1-21.

Hicks, J.R., 1941-42. Consumers' surplus and index numbers. The Review of Economic Studies 9, 126-137.

Hicks, J.R., 1946. Value and Capital. Second Edition, Oxford: Clarendon Press.

Leontief, W.W., 1941. The structure of the American economy 1919-1929. Cambridge, Massachusetts: Harvard University Press.

Samuelson, P.A., 1974. Complementarity—An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. Journal of Economic Literature 12, 1255-1289.

Shephard, R.W., 1953. Cost and Production Functions, Princeton: Princeton University Press.

Walras, L., 1954. Elements of Pure Economics, (a translation by W. Jaffé of the Edition Définitive (1926) of the Eléments d'économie pure, first edition published in 1874). Homewood, Illinois: Richard D. Irwin.

Wiley, D.E., Schmidt, W.H. and Bramble, W.J., 1973. Studies of a Class of Covariance Structure Models. Journal of the American Statistical Association 68, 317-323.