Empirical Economics (1994) 19:639–645

EMPIRICAL ECONOMICS

Using the Quadratic Box-Cox for Flexible Functional Form Selection and Unconditional Variance Computation

FERMIN S. ORNELAS,[1] C. RICHARD SHUMWAY AND TEOFILO OZUNA, JR.

Department of Agricultural Economics, Texas A & M University, College Station, Texas, 77843-2124, USA

Summary: A quadratic Box-Cox methodology is presented for choice of flexible functional form that includes consistent computation of variance estimates. Empirical viability of the procedure is investigated by specifying a dual profit function using highly aggregated U.S. agricultural data. Conditional and unconditional variance estimates for the parameters are compared and contrasted. Likelihood ratio tests are utilized to discriminate among the generalized Leontief, normalized quadratic, translog, and square-rooted quadratic functional forms. Statistical results indicate that the square-rooted quadratic is the preferred choice of functional form for these data, followed by the normalized quadratic.

JEL Classification System-Numbers: C12, C13, C52, Q11

1 Introduction

Quadratic Box-Cox models (QBCM) have become increasingly popular for choosing among nested locally flexible functional forms (LFFF) such as the generalized Leontief, quadratic, and translog (e.g., Berndt and Khaled, 1979; Appelbaum, 1979; Blackley et al., 1984; Dagenais et al., 1987; Rasmussen and Zuehlke, 1990). Grid search or gradient methods have generally been employed to obtain parameter estimates. However, econometric evidence (e.g., Seaks and Layson, 1983; Spitzer, 1982a, 1984) suggests that standard applications (e.g., SHAZAM, SAS) of both methods may render inconsistent *t*-statistics of the Box-Cox parameter estimates.

In this paper, a QBCM is programmed for a dual profit function of agricultural production in order to choose among alternative nested LFFFs. In order to obtain consistent estimates of the standard errors, it computes the variance

[1] Fermin S. Ornelas is an econometrician, American Express Travel Co., Phoenix, Arizona. C. Richard Shumway is a professor and Teofilo Ozuna, Jr. is an associate professor in the Department of Agricultural Economics, Texas A & M University, College Station, Texas.

Appreciation is extended to Eldon Ball for access to his data set for U.S. agriculture and to Carl Shafer and Steve Fuller for helpful comments on an earlier draft.

Texas Agricultural Experiment Station Technical Article No. 25892. This material is based on work partially supported by the U.S. Department of Agriculture under Agreement No. 58-3AEM-8-00104.

0377–7332/94/4/639–645 $2.50

of the parameters from the Hessian matrix. Emphasis is given to the fact that choosing a LFFF is seldom an independent research objective. Most often, precise estimation of individual parameters and variances is also a concern. Therefore, correct application of likelihood ratio tests and accurate computation of all parameter variances are essential for correctly determining the precision of parameter estimates.

The paper is organized as follows: Procedures for more accurately computing variances of QBCM parameter estimates are developed in the next section. They are followed in turn by a description of the data and estimation details, a discussion of the empirical findings, and the conclusions.

2 Methodology

A functional form is considered locally flexible if it can reflect any combination of economic effects at a particular point. However, LFFFs are usually well behaved only over a limited range of points (Despotakis, 1986). Because no single functional form is unequivocally superior with respect to all theoretical and empirical criteria, this study focuses on parametric testing to discriminate among LFFFs nested within the generalized Box-Cox. Among the second-order Taylor expansions (LFFFs) nested within the QBCM as specified in this paper are the translog (TL), generalized Leontief (GL), normalized quadratic (NQ), and square-rooted quadratic (SRQ) functional forms. Statistical discrimination among them is possible by testing for the validity of alternative restrictions on the parameter transformation of the QBCM.

For a dual profit function, let the QBCM be defined as follows:

$$Y(\delta) = \alpha_0 + \alpha' X(\lambda) + .5X(\lambda)'\beta X(\lambda) + \varepsilon \ , \tag{1}$$

where X is a vector of output and variable input normalized prices and fixed input quantities; Y is the level of normalized profits; δ and λ are power transformations; α_0, α, and β are conformable parameters (scalar, vector, and matrix, respectively) to be estimated; and ε is a column vector of random error terms. Profit and prices are normalized by an output or variable input price, or a linear combinaton of them. This normalization maintains linear homogeneity of the profit function in prices regardless of functional form.

$Y(\delta)$ and $X(\lambda)$ represent the transformations: $Y(\delta)$ is $(Y^{2\delta} - 1)/2\delta$ if $\delta \neq 0$ or $\ln Y$ if $\delta = 0$, and $X(\lambda)$ is $(X^{\lambda} - 1)/\lambda$ if $\lambda \neq 0$ or $\ln X$ if $\lambda = 0$. By utilizing l'Hopital's rule, the transformations are continuous around zero. Therefore, as $\delta = \lambda \rightarrow 0$ the QBCM becomes the TL functional form:

$$\ln Y = a_0 + a' \ln X + .5 \ln X' A \ln X + \varepsilon_1 \ . \tag{2}$$

At $\delta = \lambda = .5$ the QBCM becomes the GL functional form:

$$Y = b_0 + b'X^{.5} + .5(X^{.5})'BX^{.5} + \varepsilon_2 \ . \tag{3}$$

At $\delta = .5$ and $\lambda = 1$ it is the NQ functional form:

$$Y = c_0 + c'X + .5X'CX + \varepsilon_3 \ . \tag{4}$$

At $\delta = \lambda = 1$ it is the SRQ functional form:

$$Y = [d_0 + 2d'X + X'DX]^{.5} + \varepsilon_4 \ . \tag{5}$$

Assuming that there exists some δ and λ for which the random term for equation (1) is approximately normally distributed with mean zero and variance σ^2, the concentrated log-likelihood function can be written as:

$$L(\theta; X, Y) = K - T/2 \ln \sigma^2 + (2\delta - 1) \sum_{i=1}^{T} \ln Y_i \ , \tag{6}$$

where $\theta = (\alpha, \beta, \delta, \lambda)$, K is a constant, and T is the number of observations. From this log-likelihood function one can derive the first and second-order conditions to be utilized in estimation of the asymptotic variance-covariance matrix, $\{-\partial^2 L/\partial\theta\partial\theta'\}^{-1}$. Computation of this matrix permits estimation of the unconditional parameter variances.

To estimate the QBCM, re-specify equation (1) as follows:

$$Y(\delta) = X(\lambda)\beta + \varepsilon \ , \tag{7}$$

where $X(\lambda)$ is the design matrix conformed as

$$X(\lambda) = \begin{bmatrix} 1 & X_{11}(\lambda) & \dots & X_{1n}(\lambda) & X_{11}(\lambda)X_{11}(\lambda) & \dots & X_{11}(\lambda)X_{1n}(\lambda) & X_{12}(\lambda)X_{12}(\lambda) & \dots & X_{1n}(\lambda)X_{1n}(\lambda) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & X_{T1}(\lambda) & \dots & X_{Tn}(\lambda) & X_{T1}(\lambda)X_{T1}(\lambda) & \dots & X_{T1}(\lambda)X_{Tn}(\lambda) & X_{T2}(\lambda)X_{T2}(\lambda) & \dots & X_{Tn}(\lambda)X_{Tn}(\lambda) \end{bmatrix} . \tag{8}$$

ε and $Y(\delta)$ are column vectors of order $(T \times 1)$, $X(\lambda)$ is a $T \times (1 + (n + 1) \cdot (n + 2)/2)$ matrix of independent variables, and β is a $(1 + (n + 1)(n + 2)/2) \times 1$ column vector of parameters.

To maximize the concentrated log-likelihood function (6) let $\varepsilon_\lambda = \partial\varepsilon/\partial\lambda$, $\varepsilon_{\lambda\lambda} = \partial^2\varepsilon/\partial\lambda^2$, $\varepsilon_\delta = \partial\varepsilon/\partial\delta$, $\varepsilon_{\delta\delta} = \partial^2\varepsilon/\partial\delta^2$, and $X_\lambda = \partial X(\lambda)/\partial\lambda$. Hence, the required first-order conditions are:

$$\partial L/\partial\beta = X(\lambda)'\varepsilon/\sigma^2 = 0 \ , \tag{9}$$

$$\partial L/\partial\delta = 2 \sum_{i=1}^{T} \ln Y_i - \varepsilon'\varepsilon_\delta/\sigma^2 = 0 \ , \tag{10}$$

$$\partial L/\partial\lambda = -\varepsilon'\varepsilon_\lambda/\sigma^2 = 0 \ . \tag{11}$$

For these conditions to hold, the second-order conditions require that the Hessian matrix for the unconditional variances,

$$\partial^2 L/\partial\theta\partial\theta' = -1/\sigma^2 \begin{bmatrix} X(\lambda)'X(\lambda) & X(\lambda)'\varepsilon_\delta & -(X_\lambda'\varepsilon_\lambda + X(\lambda)\varepsilon_{\lambda\lambda}) \\ & (\varepsilon'\varepsilon_{\delta\delta} + \varepsilon_\delta'\varepsilon_\delta) - 4T^{-1}\varepsilon'\varepsilon_\delta \sum_{i=1}^{T} \ln Y_i & \varepsilon_\delta'\varepsilon_\lambda \\ \text{symmetric} & & -(\varepsilon'\varepsilon_{\lambda\lambda} + \varepsilon_\lambda'\varepsilon_\lambda) \end{bmatrix} . \tag{12}$$

be negative definite. The inverse of the negative of this matrix is the estimated covariance matrix of $\theta = (\beta, \delta, \lambda)$. Its dimensions are $(3 + n(n + 3)/2) \times (3 + n(n + 3)/2)$.

Maximization of equation (6) can be accomplished by using nonlinear estimation (e.g., Newton-Raphson or modified Newton nonlinear procedures with the dependent variable scaled by the geometric mean [Zarembka, 1974; Spitzer, 1982b]) or iterative OLS. In iterative OLS, a bi-dimensional grid search for the power transformation can be implemented to obtain $\beta^* = (X(\lambda)'X(\lambda))^{-1}X(\lambda)'Y(\delta)$. Variances and t-statistics of the parameter estimates not conditional on λ or δ are calculated for both methods using matrix (12).

3 Data and Estimation

The procedure outlined in the previous section was applied to (1) using Ball's (1988) annual U.S. agricultural data for the period 1948–79. Independent variables included three aggregate expected output prices (for grains, other crops, and livestock and fluid milk), one fixed-input quantity (for self-employed labor), and one aggregate variable-input price (for all other inputs). Time was included as a proxy for disembodied technical change.

Linear homogeneity of the profit function in prices was maintained by normalizing profit and output and variable input prices by grain price. Because of high collinearity, the squared terms on the fixed input and time variables were deleted from the model specification as were the interaction terms between other crops and time and between self-employed labor and time. Thus, the second-order Taylor expansion is complete in the prices of output and variable inputs but is incomplete in the fixed input quantity and in time.

Computer programs (available on request) were written using SAS PROC MATRIX to estimate the QBCM in its quadratic form with different power transformations for the response and explanatory variables and to compute unconditional variances. The QBCM and four alternative LFFFs (TL, GL, NQ, and SRQ) were estimated with iterative OLS. A bi-dimensional grid search for both power transformations was performed using equation (1); δ and λ were iterated between 0 and 2.5 in increments of .01. The QBCM was also estimated with a Newton-Raphson program and obtained similar results.

To discriminate among the four LFFFs, a likelihood ratio test was utilized. Each functional form test required that the power transformations, δ and λ, satisfy the relevant restrictions. The null hypotheses were $\delta = \lambda = 0$ for the TL, $\delta = \lambda = .5$ for the GL, $\delta = .5$ and $\lambda = 1$ for the NQ, and $\delta = \lambda = 1$ for the SRQ. The alternative hypothesis in each case allowed the unrestricted parameters of the power transformations to take on values that maximized equation (6).

4 Empirical Results

Table 1 reports the statistical results for choice of functional form. The square-rooted quadratic and the normalized quadratic failed to be rejected at a .05 significance level. At a significance level of .01, only the TL would be rejected. Since the square-rooted quadratic had the lowest value for the test statistic, it may be regarded as the preferred choice and is followed by the normalized quadratic.

Table 2 presents unrestricted QBCM estimates along with their conditional (on λ and δ) and unconditional standard errors. All estimated first-order parameter terms ($\alpha_1 - \alpha_5$) have the expected signs (i.e., positive for output prices, negative for variable-input price, positive for fixed-input quantity, and positive for time). Estimated second-order terms (β_{ij}) are also consistent with the expectation that the profit function is convex in prices at the point of approximation. Conditional standard errors for both sets of estimates were computed using a maximum likelihood estimated variance.

At a 5 percent significance level, nearly half the parameters were judged statistically significant using the inappropriate conditional standard errors. Although caution must still be exercised in interpreting any standard errors since some statistics are not invariant to scale of the regressors (Spitzer, 1984), only δ and λ were found to be statistically significant when the unconditional standard errors were used. Thus, the parameter estimates are less precise than the conditional variances would suggest. These results agree with the findings of Blackley et al. (1984) and Spitzer (1982b) for the linear generalized Box-Cox. Some previous studies using the Box-Cox transformation for selection of functional form may have incurred a serious statistical fault utilizing grid search and/or approxima-

Table 1. Log-likelihood results

Locally Flexible Functional Form	Transformation (δ, λ)	Log-Likelihood Function Value	Likelihood Ratio Test[a]
Translog	(0, 0)	−27.9276	16.50
Generalized Leontief	(.5, .5)	−22.8537	6.35
Normalized Quadratic	(.5, 1)	−21.9977	4.64
Square-rooted Quadratic	(1, 1)	−21.0033	2.65
Quadratic Generalized Box-Cox	(.88, 2.19)	−19.6744	

[a] Critical value of $LRT = -2(L(\delta^*, \lambda^*) - L(\delta, \lambda)) \sim \chi^2_{2,.05} = 5.99$, where δ^* and λ^* are the unrestricted power tranformations

Table 2. Parameter and standard error estimates from the iterative OLS estimation of the quadratic generalized Box-Cox

Parameter[a]	Iterative OLS Value	Conditional[b] Standard Error	Unconditional[c] Standard Error
α_0	201.69	24.27	267.16
α_1	1041.26	258.26	1747.46
α_2	4811.94	1478.68	8004.13
α_3	−2354.03	547.56	3885.21
α_4	194.35	114.45	368.59
α_5	3.40	5.37	7.96
β_{11}	917.77	1221.52	2146.59
β_{22}	96219.00	37567.70	162740.00
β_{33}	11630.60	3596.50	21343.2
β_{12}	7856.70	5186.18	13644.00
β_{13}	−1899.24	1822.54	4446.65
β_{14}	−350.96	194.56	608.29
β_{15}	13.49	38.29	53.93
β_{23}	−28806.50	10081.80	48727.70
β_{24}	−1568.27	1131.87	2783.58
β_{34}	932.76	374.50	1634.70
β_{35}	−41.44	48.82	95.39
δ	.89	n.a.	.2388
λ	2.19	n.a.	.2955

n.a. means not applicable

[a] Parameter codes: α_0-intercept; α_1-livestock and fluid milk price; α_2-other crops price; α_3-variable inputs price; α_4-fixed input quantity; α_5-time variable; β_{11}-square of livestock and fluid milk price; β_{22}-square of other crops price; β_{33}-square of variable inputs price; β_{ij} ($i = 1, 2, 3; j = 1, \ldots, 5; i \neq j$)-cross products (interaction terms) among prices, fixed input quantity, and time; δ-power transformtion for profit; λ-power transformation for the explanatory variables except time

[b] Conditional statistics were computed by maximizing equation (6) in a bi-dimensional grid search for δ and λ using iterative OLS

[c] Unconditional statistics were computed by expanding the OLS variance-covariance matrix to conform the full Hessian matrix (12)

tions of the Hessian matrix to compute (inappropriate conditional) variances of the parameter estimates because they were inconsistent estimators of the true variances.

5 Conclusions

Based on the developments of Spitzer (1982a, 1982b, 1984) for the linear generalized Box-Cox, a methodology has been presented to estimate the quadratic generalized Box-Cox and obtain unconditional variance estimates. The main concerns of this paper were selection of a LFFF, computation of unconditional variances, and empirical application to U.S. agriculture.

Empirical results indicated that (a) the popular translog functional form was strongly rejected for the profit function specification of U.S. agriculture, (b) unconditional variances were much larger than conditional variances for some parameters and (c) the power transformations were statistically significant. Current software packages (e.g., SAS, SHAZAM) require modification in order to conveniently handle different power transformations and compute unconditional standard errors. With convenient software, functional form tests could become common practice in exploratory analysis for model specification.

6 References

Appelbaum E (1979) On the choice of functional forms. International Economic Review 20:449–57

Ball VE (1988) Modelling supply response in a multiproduct framework. American Journal of Agricultural Economics 70:813–25

Berndt ER, Khaled MS (1979) Parametric productivity measurement and choice among flexible functional forms. Journal of Political Economy 87:1221–45

Blackley P, Follain JR, Ondrich J (1984) Box-Cox estimation of hedonic models: how serious is the iterative OLS variance bias? The Review of Economics and Statistics 66:348–53

Dagenais MG, Gaudry MJI, Liem TC (1987) Urban travel demand: the impact of Box-Cox transformations with nonspherical residual errors. Transportation Research 21B:443–47

Despotakis KA (1986) Economic performance of flexible functional forms. European Economic Review 30:1107–43

Rasmussen DW, Zuehlke TW (1990) On the choice of functional form for hedonic price functions. Applied Economics 22:431–38

Seaks TG, Layson SK (1983) Box-Cox estimation with standard econometric problems. The Review of Economics and Statistics 65:160–64

Spitzer JJ (1982a) A primer on the Box-Cox estimation. The Review of Economics and Statistics 64:307–13

Spitzer JJ (1982b) A fast and efficient algorithm for the estimation of parameters in models with the Box-Cox transformation. Journal of the American Statistical Association 77:760–66

Spitzer JJ (1984) Variance estimates in models with the Box-Cox transformation: implications for estimation and hypothesis testing. Review of Economics and Statistics 66:645–52

Zarembka P (1974) Frontiers in econometrics. New York Academic Press

First version received: June 1993
Final version received: March 1994