Box-Cox Transformations and Error Term Specification in Demand Models

James Blaylock and David Smallwood

This paper analyzes the influence of error-term specification and functional form on a quarterly demand model for beef. The Box-Cox transformation is used to generalize the functional form while the equation error term is postulated to be both heteroskedastic and autoregressive. Results indicated that both functional form and error-term specification can play a major role in elasticity estimation, elasticity behavior, and hypothesis testing.

The monotonic transformation introduced by Box and Cox has become a popular tool for both discriminating among alternative functional forms and providing added flexibility in model specification. Most empirical analyses employing the Box-Cox transformation (BCT) have assumed that the model error term is homoskedastic and nonautoregressive. However, more recent evidence suggests that error specification is at least as important as functional form when the transformation is applied to the dependent variable. Savin and White have shown that estimating BCT models without specifying an autocorrelated error structure can yield both inefficient estimators and erroneous results of hypothesis tests. Zarembka has shown that if the true underlying error term is heteroskedastic, then the maximum likelihood estimator of the dependent variable's BCT parameter is biased in the direction which tends to make the estimated error term more homoskedastic. This can result in inconsistent estimators for all model parameters. One way to avoid this limitation is to relax the homoskedastic assumption of the model and estimate simultaneously the functional form and error-term specification. Subsequent statistical tests can then be performed to test for the significance of heteroskedasticity. This approach is in accordance with Zarembka's suggestion that it is important to ascertain the robustness of parameter estimates to heteroskedasticity.

Several studies have independently examined the roles of autocorrelation and heteroskedasticity in Box-Cox type models. These include Savin and White, Lahiri and Egy, Blaylock *et al.*, and Blaylock and Smallwood (1982). However, we are unaware of any empirical studies which have considered these problems jointly. Our objective is to investigate the influence of functional specification, autocorrelation, and heteroskedasticity on the parameters of a quarterly demand function for beef.[1] We examine these factors, using a struc-

James Blaylock and David Smallwood are Economists with the U.S. Department of Agriculture, Economic Research Service, Washington, D.C.

The authors wish to thank the referees for many helpful suggestions.

[1] Other studies which have used the transformation of variables technique to analyze meat demand include Chang, Hassan and Johnson, and Kulshreshtha. Following these studies, we assume that commodity price is exogenous. In addition, single equation demand models should be interpreted as approximations to a more complete system of equations.

Western Journal of Agricultural Economics, 8(1), 1983, pp. 68–75

tured statistical model, to ascertain their effects on the estimated demand equation and subsequent price and income elasticities.

General Model Development

The Box-Cox transformation for any positive variable W is defined as

$$W^{(\lambda_w)} = (W^{\lambda_w} - 1)/\lambda_w, \qquad \lambda_w \neq 0$$
$$= \ln W, \qquad \lambda_w \to 0 \quad (1)$$

where λ_w is a parameter to be estimated. A desirable property of the BCT is that, with the addition of a single parameter, λ_w, one obtains a general class of power transformations, including several that are frequently used in empirical analyses. For example, if $\lambda_w = 1$ one obtains the linear transformation, if $\lambda_w = 0$ one obtains the logarithmic transformation, and if $\lambda_w = -1$ one obtains the inverse transformation. The BCT is typically employed in econometric models of the form

$$Y_t^{(\lambda_y)} = \alpha + \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_k)} + u_t \qquad (2)$$

for each observation t, where u_t is the equation error term, Y_t is an endogenous variable, X_{kt} (k = 1, 2, . . . , K) are exogenous variables, β_k (k = 1, 2, . . . , K) are coefficients on the transformed exogenous variables, λ_y and λ_k (k = 1, 2, . . . , K) are BCT parameters, and α is a constant. Hence, the Box-Cox model provides a convenient framework for allowing both increased model flexibility and a means for discriminating among many of the commonly used "classical" functions.[2]

Specification of the error structure for u_t is required for estimation of equation (2). With few exceptions, the error term in Box-Cox models is assumed to be independently, identically, and normally distributed, with $E(u_t) = 0$ and $E(u_t^2) = \sigma^2$ for all t. However, imposition of these conditions is unnecessarily restrictive and may lead to inconsistent parameter estimators if the underlying error structure violates these assumptions.[3] Furthermore, given Zarembka's result that the estimator of the Box-Cox transformation parameter on the dependent variable is biased in the direction which tends to stabilize the error variance, the predicted errors cannot be used *ex post facto* to test for heteroskedasticity. A solution to this problem is to allow more flexibility in the specification of the error structure.

We assume in this analysis that the error distribution can be adequately described by structures suggested by Gaudry and Dagenais (heteroskedasticity) and Savin and White (autocorrelation).

We make the assumption that if u_t is heteroskedastic, it can be expressed as

$$u_t = [f(Z_t)]^{1/2} v_t \qquad (3)$$

where Z is an exogenous variable used to explain the heteroskedasticity. We also assume $E(v_t) = 0$, $E(v_t^2) = \psi^2$ and $E(v_t Z_t) = 0$. The functional form we select for $f(Z_t)$ is

$$u_t = [\exp\{\delta Z_t^{(\lambda_z)}\}]^{1/2} v_t \qquad (4)$$

where a BCT (λ_z) has been applied to the variable which is postulated to stabilize the error variance. It follows that

$$E(u_t^2) = \omega_t = \psi^2\{\delta Z_t^{(\lambda_z)}\}. \qquad (5)$$

Many of the traditional empirical specifications of heteroskedasticity are special cases of equation (5). For example, Park's specification

$$\omega_t = \psi^2 Z_t^{\delta}$$

is obtained from equation (5) by setting

[2] The "classical" functions include linear, double-log, semi-log, inverse, and the log-inverse.

[3] If the value of the transformation parameter on the dependent variable is known a priori, then least squares estimators can be used to obtain consistent estimators of the remaining model parameters. It is for this reason that estimators of the "classical" forms are consistent in the presence of heteroskedasticity while estimators obtained from direct estimation of the transformation parameters are inconsistent if the error variance is heteroskedastic.

$\lambda_z = 0$. The classical univariate form

$$\omega_t = \psi^2 Z_t$$

is obtained by setting $\lambda_z = 0$ and $\delta = 1$. The homoskedastic case is obtained from expression (5) by setting $\delta = 0$.

To allow and test subsequently for the presence of autocorrelated residuals in the model, we assume that the v_t's follow a stationary first-order autoregressive process of the form

$$v_t = \rho v_{t-1} + w_t \qquad (6)$$

where w_t is assumed to be a normally, independently, and identically distributed random-error term with zero mean and a constant variance σ^2.[4] It should be noted that the error term in Box-Cox models cannot be strictly normal because the transformation can be applied only to positive variables. However, Draper and Cox have shown that estimates of the transformation parameters are robust to nonnormality if the error distribution is reasonably symmetric.

Beef Model

The regression model we propose for estimating the quarterly demand for beef can be written as

$$Y_t^{(\lambda_y)} = \alpha + \sum_{k=1}^{5} \beta_k X_{kt}^{(\lambda_k)} + \sum_{r=1}^{3} \gamma_r D_r + u_t \qquad (7)$$

where the sample period (t = first quarter, 1960, through fourth quarter, 1979) and variable specifications were chosen only for expository purposes and the following definitions apply:

Y_t : per capita consumption of beef and veal (Source: Livestock and Meat Situation)

X_{1t} : retail price index of beef and veal (Source: Livestock and Meat Situation) divided by the Consumer Price Index (CPI, 1967 = 100, source: Bureau of Labor Statistics)

X_{2t} : Retail price index of pork divided by CPI (Source: Livestock and Meat Situation)

X_{3t} : retail price index of poultry divided by CPI (Source: Poultry and Egg Situation)

X_{4t} : retail price index of fish divided by CPI (Source: Bureau of Labor Statistics)

X_{5t} : index of per capita disposable income divided by CPI (Source: Survey of Current Business)

D_1, D_2, D_3: seasonal dummies for the second, third, and fourth quarters of the calendar year, respectively.

Prices and income are divided by the CPI to approximate homogeneity of degree zero in the demand function. The dummy variables enter linearly for ease of estimation.[5]

Estimation

The demand equation defined in expression (7), together with the structure of the error term given in expressions (4) and (6), is estimated via nonlinear maximum likelihood procedures.[6] Briefly, the estimation procedure is outlined as follows. First, note that the Jacobian of the transformation from $Y_t^{(\lambda_y)}$ to the Y_t's is given by

[4] The models were originally estimated under the assumption of first- and fourth-order autocorrelation. The fourth-order autocorrelation coefficient was not found to be significant, but the data did indicate a significant first-order autocorrelation.

[5] The effects of dummy variables on the original dependent variable are difficult to interpret. The interested reader is referred to Blaylock and Smallwood (1983).

[6] A general discussion of the derivation of likelihood functions and the role of the Jacobian can be found in Theil or Kmenta.

TABLE 1. Estimated Transformation, Autocorrelation, and Heteroskedastic Parameters.

Model	Transformation Parameters						Autocorrelation	Heteroskedasticity		
	λ_y	λ_1	λ_2	λ_3	λ_4	λ_5	ρ	δ	λ_z	$-LL^a$
BC	1.73	1.73	1.73	1.73	1.73	1.73	0.0	0.0		193.3
BCA	−0.99	−0.99	−0.99	−0.99	−0.99	−0.99	0.67	0.0		213.1
BCAH	0.60	0.60	0.60	0.60	0.60	0.60	0.57	3.41	7.57	215.8
BCG	−0.71	1.47	17.59	4.71	6.32	−0.30	0.0	0.0		217.1
BCGA	−0.31	2.55	−8.10	2.42	6.97	−5.91	0.64	0.0		223.0
BCGAH	0.24	2.58	−6.70	3.68	9.48	−4.64	0.63	1.56	7.83	223.8

[a] Value of the log-likelihood function.

$$|J| = |\det(\partial Y_t^{(\lambda_y)}/\partial Y_t)| = \sum_{t=1}^{N} Y_t^{\lambda_y - 1}. \quad (8)$$

The likelihood function corresponding to the beef model is written as[7]

$$L = \prod_{t=2}^{N} (1/\sqrt{\sigma^2 2\pi})[Y_t^{\lambda_y-1}/\sqrt{f(Z_t)}] \cdot \exp\{-(1/2\sigma^2)[(S_t/\sqrt{f(Z_t)}) - \rho(S_{t-1}/\sqrt{f(Z_{t-1})})]^2\}. \quad (9)$$

where

$$S_t = Y_t^{(\lambda_y)} - \alpha - \sum_{k=1}^{5} \beta_k X_{kt}^{(\lambda_k)} - \sum_{r=1}^{3} \gamma_r D_r \quad (10)$$

and

$$f(Z_t) = \exp\{\delta Z_t^{(\lambda_z)}\}. \quad (11)$$

The log-likelihood function corresponding to equation (9) is written as

$$LL = [(N-1)/2]\ln(\sigma^2) + (\lambda_y - 1)\sum_{t=2}^{N} \ln(Y_t) - \tfrac{1}{2}\sum_{t=2}^{N} \ln[f(Z_t)] - (1/2\sigma^2)\sum_{t=2}^{N} [S_t/\sqrt{f(Z_t)} - \rho(S_{t-1}/\sqrt{f(Z_{t-1})})]^2 + c \quad (12)$$

where c is a constant.

The log-likelihood function is then concentrated on σ^2, α, and the β's in order to reduce the number of parameters to be estimated directly by a nonlinear optimizing algorithm.[8]

Empirical Results

The linear, double-log, semi-log, inverse, and log-inverse functions have been shown to be special cases of the general Box-Cox model (e.g., see Savin and White). The linear and double-log forms are special cases of the Box-Cox model when both the dependent and independent variables are transformed identically. Each of the "classical" type forms was estimated both with and without the assumption of first-order autocorrelation. The following variations of the Box-Cox model were estimated:

- BC: Model with the same BCT applied to both dependent and independent variables. The designations BCA and BCAH will be used to denote the autoregressive and heteroskedastic-autoregressive versions, respectively;
- BCG: Model with all the variables transformed differently. The definitions of BCGA and BCGAH are as above.

[7] It is recognized that the effect of the first observation is lost. However, to recapture its effect would impose a severe computational burden.

[8] A computer program written by Liem was used in this analysis. The program uses the Fletcher-Powell algorithm.

TABLE 2. Summary of Nested Specifications.

General Specification	Restricted Version(s)
Autocorrelated "classical" forms	"Classical" forms [1]
BC	Linear [1], Double-log [1]
BCA	Linear [2], Double-log [2], Autocorrelated linear [1], Autocorrelated double-log [1], BC [1]
BCAH	Linear [3], Double-log [3], Autocorrelated linear [2], Autocorrelated double-log [2], BC [2], BCA [1]
BCG	"Classical" forms [6], BC [5]
BCGA	"Classical" forms [7], Autocorrelated "classical" forms [6], BC [6], BCA [5], BCG [1]
BCGAH	"Classical" forms [8], Autocorrelated "classical" forms [7], BC [7], BCA [6], BCAH [5], BCG [2], BCGA [1]

Note: Values in brackets refer to the number of independent parametric restrictions placed on the general specification used to obtain the restricted form(s).

TABLE 3. Summary of Chi-square Test Results.

Unrestricted Specification	Accepted Restricted Specification(s)[a]
Autocorrelated "classical" forms	None
BC	Double-log
BCA	Linear A[b], Double-log A
BCAH	Double-log A, BCA
BCG	None
BCGA	None
BCGAH	BCGA

[a] Accepted at the 0.05 significance level.
[b] Denotes that the function was estimated with first-order autocorrelation.

The transformation parameters, autocorrelation coefficients, parameters of the heteroskedastic structure, and the maximum values of the log-likelihood functions for the alternative Box-Cox models are presented in Table 1.[9] Income was found to be the most appropriate variable to stabilize the error variance as measured by increases in the log-likelihood function. Thus, Z_t in the above notation is X_{5t}.

To compare statistically the fit of the various functions, the maximum likelihood ratio test is used. This test is appropriate since many of the models are nested. A model is considered to be a nested version of another if the simpler of the two models can be obtained by appropriately restricting the coefficients of the more general model. For example, the "classical" forms are nested or special cases of their respective autocorrelated forms when $\rho = 0$; the linear and double-log functions are nested versions of the BC model when the transformation parameter is restricted to 1 and 0, respectively; and the BC model is a special case of the BCAH model when $\rho = 0$ and $\delta = 0$. A summary of the nested specifications is given in Table 2.

The likelihood ratio method was used to test the null hypothesis that no significant difference existed between the models which are nested versions of each other. The likelihood ratio is defined as

$$\phi = L(R)/L(UR) \qquad (13)$$

where L(R) is the restricted likelihood function, which corresponds to the simpler case of the more general function represented by the likelihood function L(UR). It can be shown that minus two times the logarithm of the likelihood ratio is asymptotically distributed as a chi-square random variable with the degrees of freedom corresponding to the number of independent parametric restrictions placed on the unrestricted model used to obtain the special case (Theil). Statistical analysis of the nested models involves comparing the calculated test statistics with the tabulated values of the chi-square

[9] The residuals from all models were analyzed via a Box-Jenkins procedure. The residuals appeared to behave as white noise after correction for first-order autocorrelation.

variable at the 0.05 significance level (with the appropriate degrees of freedom).

A summary of the chi-square test results is presented in Table 3. Results indicated that the autocorrelated "classical" functions were a significant statistical improvement vis-à-vis their nonautocorrelated counterparts. The BC model was not statistically different from the double-log form while the BCA model was a statistical improvement over the linear, double-log, and BC functions. The autocorrelated double-log form and the BCA were the only restricted versions of the BCAH model that were accepted. The generalized models (i.e., BCG, BCGA, and BCGAH) were a statistical improvement over all their respective nested specifications, except the BCGA model, which the chi-square test failed to reject as significantly different from the BCGAH equation. In the case of the beef equations, correcting for nonspherical residuals as well as further generalizing the Box-Cox model by adding additional BCT parameters has significantly altered the hypothesis test results.

In some cases, the change in the magnitude of the transformation parameters was substantial. For example, the BCT on the dependent variable changed from 1.73 for the BC model to 0.24 for the BCGAH model. There was even more variation in the transformation parameters associated with the independent variables. The BCT on income changed from −0.30 for the BCG to −4.64 for the BCGAH model. The autocorrelation coefficient, ρ, appeared relatively stable as did the parameters associated with the heteroskedastic structure.

In general, the transformation parameters indicate whether the underlying variable has resulted in a convex or concave transformation. If the transformation parameter is greater than one, this indicates that the transformed variable is a convex function of the original variable, and if it is less than one, the function is concave. The Box-Cox model can also be interpreted as a linear functional form with variational parameters (Hassan and Johnson). This interpretation indicates how the linear approximation must adjust to explain the sample data.

It is also interesting to note that only the BC, BCAH and BCGAH models satisfy the mean convergence criterion outlined by Huang and Grawe. This criterion is satisfied if the BCT on the dependent variable lies outside the interval $-1 <$ BCT < 0. Satisfaction of this criterion is important to ensure that the conditional expectation of the untransformed dependent variable (that is, conditional on the deterministic part of the model) exists. This is an important issue if one attempts to calculate elasticities based on the expected value of the dependent variable. In Box-Cox models, this expected value is not necessarily given by the deterministic part of the model.

In summary, the general BCGAH appears to be the most satisfactory quarterly beef model based on chi-square tests and the mean convergence criterion. This lends supporting evidence to the contention that functional and error-term specification need to be considered simultaneously. Further evidence is provided by analyzing the estimated elasticities from the alternative models.

Elasticities, evaluated at the sample means, for the alternative models are presented in Table 4. The own-price elasticity appears to be relatively robust as to both application of the BCT parameters and error-term specification. The income elasticities, however, varied widely among the models. For example, the income elasticity estimated from the parameters of the BCGAH model was 0.44 compared to 0.89 for the BCAH specification and 0.80 for the autocorrelated linear function. The "classical" forms appeared to overestimate the beef income elasticity compared to the elasticity from the BCGAH model. This is important if the elasticities are to

TABLE 4. Estimated Elasticities.[a]

Model	Own Price	Income
Linear	−0.50	0.83
Linear A[b]	−0.47	0.80
Double-log	−0.49	0.86
Double-log A	−0.50	0.74
Semi-log	−0.51	0.77
Semi-log A	−0.52	0.65
Inverse	−0.51	0.71
Inverse A	−0.52	0.63
Log-inverse	−0.48	0.79
Log-inverse A	−0.50	0.71
BC	−0.49	0.98
BCA	−0.50	0.71
BCAH	−0.51	0.89
BCG	−0.52	0.87
BCGA	−0.46	0.37
BCGAH	−0.47	0.44

[a] Elasticities were evaluated at the sample means.
[b] Denotes autoregressive model.

be used for policy analysis or projecting demand because the elasticities from the "classical" forms would severely overestimate the effect of income changes on beef demand.

Another criterion for selecting among the alternative beef equations is the behavior of the elasticities as beef prices or income changes. Chang argues that "the income elasticity for a specific food, like meat, generally should be falling rather than rising" as consumption rises in response to increased income. This requires that the first derivative of the income elasticity (E_I) with respect to income be negative. This condition can be expressed for the BCG models as

$$(\lambda_5 - \lambda_y E_I) < 0$$

and as

$$\lambda_y(1 - E_I) < 0$$

for the BC models. The only quarterly beef models that satisfied these conditions were the semi-log, inverse, log-inverse, BCA, BCGA, and BCGAH models.

It may also be reasonable to expect that the price elasticity (E_p) at a given income level will rise in absolute value as beef prices increase. This condition can be expressed for the BCG and BC models as

$$(\lambda_1 - \lambda_y E_p) > 0 \text{ and } \lambda_y(1 - E_p) > 0,$$

respectively. This condition is satisfied only by the semi-log, linear, BC, BCAH, and all BCG models.

It is also reasonable to assume that the price elasticity will fall in absolute value at a given price as income and consumption increase. Likewise, it is reasonable that the income elasticity will rise at a given income level as beef prices rise and consumption falls. It can be shown (see Gemmill) that these conditions on elasticity behavior can be satisfied only if the transformation parameter on the dependent variable is greater than zero. Among the various models examined, only the BC, BCAH, BCGAH, linear, inverse, and semi-log models satisfied these conditions.

Consequently, only the semi-log and the BCGAH functions satisfy all of the above conditions for elasticity behavior believed to characterize the demand for beef. Hence, based on the chi-square tests and satisfaction of both the mean convergence criterion and elasticity conditions, the BCGAH model appears to be the best function of those considered for analyzing the quarterly demand for beef.

Conclusions

This paper has demonstrated, via a quarterly demand model for beef, several key points concerning the Box-Cox functional form and its error-term specification. We provided a specific example to show that some *a priori* assumptions concerning the distribution of the error term can lead to inconsistent parameter estimators. Our results revealed that in quarterly beef models the specification of the error term can be at least as important as the functional form for hypothesis testing and elasticity estimation. Generalizing to other applications of Box-Cox models, the following conclusions appear applicable for researchers employing flexible functional forms:

1. Autocorrelated residuals, which are likely to occur when time-series data are used, should be corrected in the Box-Cox models.
2. The analytic form of heteroskedasticity should be estimated simultaneously with the nonstochastic (i.e., fixed) part of the model.
3. Transforming the dependent and independent variables identically as opposed to using a different transformation parameter on all model variables can play a crucial role in hypothesis testing and elasticity estimation.
4. The transformation parameters can vary widely in magnitude; thus care should be taken not to overly restrict the parameter range if the grid-search method is employed for estimation.

We recommend that results from models employing the Box-Cox transformation be viewed with skepticism unless proper analysis of the error term is conducted.

References

Blaylock, J., L. Salathe, and R. Green. "A Note on the Box-Cox Transformation Under Heteroskedasticity," *Western Journal of Agricultural Economics*, 4(1980): 129–35.

Blaylock, J. and D. Smallwood. "Analysis of Income and Food Expenditure Distributions: A Flexible Approach," *Review of Economics and Statistics*, 64(1982): 104–09.

Blaylock, J. and D. Smallwood. "Interpretating the Effects of Binary Variables in Transformed Models," *Economics Letters,* 1983.

Box, G. E. P. and D. R. Cox. "An Analysis of Transformations," *Journal of the Royal Statistical Society*, 26(1964): 211–43.

Box, G. E. P. and G. M. Jenkins. *Time-Series Analysis: Forecasting and Control,* Holden-Day, San Francisco, 1976.

Chang, H. S. "Functional Forms and the Demand for Meat in the United States," *Review of Economics and Statistics*, 59(1977): 355–59.

Draper, N. R. and D. R. Cox. "On Distributions and their Transformation to Normality," *Journal of the Royal Statistical Society*, 31(1969): 472–76.

Gaudry, M. and M. Dagenais. "Heteroscedasticity and the Use of Box-Cox Transformations," *Economics Letters*, 2(1979): 225–29.

Gemmill, G. "Using the Box-Cox Form for Estimating Demand: A Comment," *Review of Economics and Statistics*, 62(1980): 147–48.

Hassan, Z. and S. Johnson. "The Demand for Meat in Canada: An Application of the Transformation of the Variables," *Canadian Journal of Agricultural Economics*, 27(1979): 1–12.

Huang, C. and O. Grawe. "Functional Forms and the Demand for Meat in the United States: A Comment," *Review of Economics and Statistics*, 62(1980): 144–46.

Kmenta, J. *Elements of Econometrics*, The Macmillan Company, New York, 1971.

Kulshreshtha, S. "Functional Form Specification in the Quarterly Demand for Red Meats in Canada," *Western Journal of Agricultural Economics*, 4(1979): 89–97.

Lahiri, K. and D. Egy. "Joint Estimation and Testing for Functional Form and Heteroskedasticity," *Journal of Econometrics*, 15(1981): 299–307.

Liem, T. C. *A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals*, University of Montreal, Centre de Recherche Sur les Transports, Publication No. 134, 1980.

Park, R. E. "Estimation with Heteroskedastic Error Terms," *Econometrica*, 4(1966): 888.

Savin, N. and K. White. "Estimation and Testing for Functional Form and Autocorrelation: A Simultaneous Approach," *Journal of Econometrics*, 8(1978): 1–12.

Theil, H. *Principles of Econometrics*, John Wiley & Sons, New York, 1971.

Zarembka, P. "Transformation of Variables in Econometrics," in P. Zarembka (Ed.), *Frontiers of Econometrics*, Academic Press, New York, 1974: 81–104.