Pork: Marketing

1981

ERROR SPECIFICATION IN TRANSFORMED MODELS

David Smallwood and James Blaylock*

ABSTRACT

Autocorrelated residuals and a heteroskedastic error variance are both jointly and separately incorporated into a Box-Cox transformed model. A quarterly demand equation for pork is used for purposes of illustration. Results indicate that both functional form and error term specification play a crucial role in hypothesis testing and estimation of elasticities.

*The authors are economists with USDA/ESS/NED/FE

LUS

Error Term Specification in Transformed Models

By

David Smallwood and Jim Blaylock*

An important issue in many applied demand analyses employing single equation demand models is the choice of the functional form to use in estimating consumer response parameters. Functional forms have been traditionally chosen on the basis of several criteria: 1) ease of estimation; 2) fit to the data; 3) simplicity of interpretation and; 4) compatibility with classical demand theory. The forms most often chosen include the linear, the double-log, the semi-log, inverse, and log-inverse models. These models are linear in the unknown parameters and, hence, may be conveniently estimated via ordinary least squares procedures. Rapidly declining computer costs and the availability of advanced econometric software have helped expand the choice set of functions to include many non-linear specifications which were previously excluded by criteria (1) and (3). Several recent studies have employed a monotonic transformation proposed by Box and Cox as a method to improve model fit and as a statistical tool to select the "appropriate" functional form (Chang, Hassan and Johnson, and Kulshreshtha). While the Box-Cox transformation (BCT) is a potentially useful device for discriminating among alternative functional forms, in most cases it has been applied without careful consideration of the underlying assumptions embodied in the statistical model.

Potential misuse of the BCT can be linked to two restrictive assumptions concerning the underlying error distribution. The first is the assumption of non-autocorrelated residuals. Savin and White have indicated that tests for functional form and autocorrelation need to be jointly considered. That is, using the BCT to test for functional specification can yield erroneous conclusions if autocorrelation is not simultaneously taken into account. The second consideration is the problem

---

* The authors are economists with USDA/ESS/NED/FE. The views expressed herein are not necessarily those of ESS or USDA. Paper presented at the annual AAEA meetings held at Clemson University on July 25-29, 1981.

of a heteroskedastic error variance. Zarembka has shown in a special case that if the error variance is heteroskedastic, then the estimator of the BCT parameter on the dependent variable is biased in the direction which would compensate for heteroskedasticity. As a result elasticity estimates and other parameters of interest are also biased. Furthermore, Gaudry and Dagenais have indicated that it is possible to begin with a nonlinear model with a homoskedastic error variance and have heteroskedasticity induced into the Box-Cox model via the estimation process.[1] Consequently, it is necessary to simultaneously test for functional form and error term specification.

The objective of this paper is to investigate the role that function specification, autocorrelation, and heteroskedasticity play in the estimation of a single equation demand function for pork.[2] We examine each of these factors independently and in combination, using a structured statistical model, to ascertain the impact of each factor on the estimated demand equation and on subsequent price and income elasticities.

GENERAL MODEL DEVELOPMENT

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

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

$$= \ln(W) \quad , \quad \lambda_w \to 0$$

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

$$(1) \qquad Y_t^{(\lambda_y)} = \alpha + \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_k)} + u_t \quad , (t = 1,\ldots,N)$$

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, B_k, k=1, 2,..., K are coefficients on the transformed exogenous variables, λ_y and λ_{kt}, k=1, 2,..., K are BCT parameters, and α is a constant. The BCT model provides a convenient framework for allowing both increased model flexibility and a means for discriminating among many of the commonly used classical functions.

Specification of the error structure for u_t is required for estimation of equation (1). With few exceptions, the error term u_t in the Box-Cox models is assumed to be independently, identically, and normally distributed with mean zero and constant variance for all t. However, imposition of these conditions is unnecessarily restrictive and may lead to biased parameter estimators if the underlying error structure violates these assumptions. Furthermore, Zarembka demonstrates that the Box-Cox transformation parameter on the dependent variable will be biased in the direction which will tend to stabilize the error variance, and hence, the predicted errors cannot be used to test for heteroskedasticity _ex post_. A solution to this problem is to allow more flexibility in the error structure and estimate it simultaneously with the functional form.

We make the assumption that if u_t is heteroskedastic it can be adequately decribed by a form suggested by Gaudry and Dagenais:

(2) $$u_t = \{\psi^2 \exp(\delta_o + \delta_1 Z_t^{(\lambda z)})\}^{\frac{1}{2}} v_t$$

where Z is an exogenous variable used to explain the heteroskedasticity, δ_o is a constant, λ_z is a BCT parameter, and v_t is a random disturbance term distributed with mean zero and constance variance. It follows that

(3) $$E(u_t^2) = \omega_{tt} = \psi^2 \exp[\delta_o + \delta_1 Z_t^{(\lambda z)}]$$

$$= \phi^2 \exp[\delta_1 Z_t^{(\lambda z)}]$$

where $\phi^2 = \psi^2 \exp\{\delta_0\}$. Many of the traditional empirical specifications of heteroskedasticity are special cases of equation (3) and are shown below.

Heteroskedastic Error Specifications

Restriction	Functional form	Description
none	$\omega_{tt} = \phi^2 \exp[\delta_1 Z^{(\lambda z)}]$	general
$\lambda_z = 0$	$\omega_{tt} = \phi^2 Z_t^{\delta_1}$	(Park,1966)
$\lambda_z = 0,\ \delta_1 = 1$	$\omega_{tt} = \phi^2 Z_t$	univariate
$\delta_1 = 0$	$\omega_{tt} = \phi^2$	homoskedastic

In order to correct for the presence of autocorrelated residuals in the above model, we assume that the v_t's follow a stationary first-order autocorrelation process of the form

(4) $$v_t = \rho v_{t-1} + w_t, \qquad |\rho| < 1$$

where w_t is assumed to be a normally, independently, and identically distributed random error term with mean zero and constant variance.

Pork Model

The regression model we propose for estimating the quarterly demand for pork can be written as follows:

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

where the sample period (t = First quarter 1960 through Fourth quarter 1979) was chosen for expository purposes only and the following definitions apply:

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

X_{1t} : retail price index of pork (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 beef and veal 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.

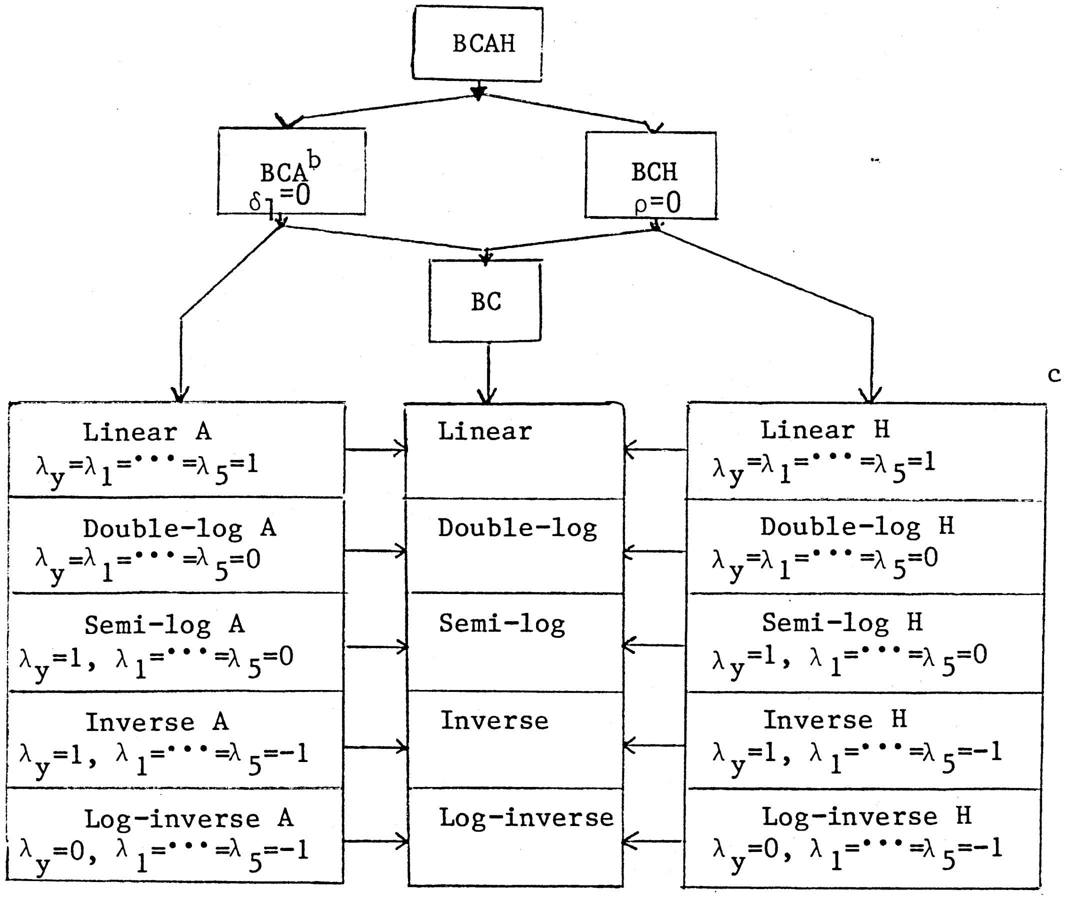
Prices and income were divided by the CPI to impose homogeneity of degree zero on the demand function. The dummy variables enter linearly for ease of estimation and simplicity of interpretation. The stochastic error term u_t in equation (5) is specified in equations (2) - (4) and the price of pork is selected as the Z variable to stabilize the error variance.[3] The model is estimated via a non-linear maximum likelihood procedure (Liem).

EMPIRICAL RESULTS

Each of the "classical" functional forms, i.e., the linear, double-log, semi-log, inverse, and log-inverse, were estimated both with and without the assumption of first-order autocorrelation. Variations of the general model that were estimated and the relationships among them are reported in Figure 1. Estimated parameters, asymptotic standard errors, and the maximum values of the alternative log-likelihood functions are presented in Table 1.

To statistically compare the fit of the alternative models maximum likelihood ratio tests are used. This test is appropriate since many of the models are nested. The maximum likelihood ratio test statistic is defined as -2 times the logarithm of the ratio of the restricted to the unrestricted likelihood function.

Figure 1. Nesting Relationships Among Models[a].



---

[a] (BC) : Box-Cox model with a homoskedastic-nonautoregressive error term,

(BCA) : Box-Cox model with a homoskedastic-first-order autoregressive error term,

(BCH) : Box-Cox model with a heteroskedastic-nonautoregressive error term and,

(BCAH) : Box-Cox model with a heteroskedastic-first-order autoregressive error term.

[b] Restrictions on the parameters are denoted below the model designation. Restrictions are cummulative in the direction of the diagram arrows.

[c] These models are not estimated.

Table 1. Parameter Estimates

		Independent Variables								Transformation Parameters						Auto-Correlation Parameter	Heteroskedastic Parameters		
Model	α	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	γ_3	λ	λ_1	λ_2	λ_3	λ_4	λ_5	ρ	δ_1	λ_Z	-LL
Linear	1.542 (0.148)	-1.382 (0.107)	0.655 (0.103)	0.159 (0.141)	-0.187 (0.081)	-0.043 (1.413)	-0.097 (0.018)	-0.087 (0.019)	0.112 (0.019)	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	-	223.74
Linear A	1.484 (0.274)	-1.347 (0.171)	0.763 (0.145)	0.130 (0.219)	-0.144 (0.134)	-1.107 (2.220)	-0.095 (0.019)	-0.088 (0.016)	0.114 (0.016)	1.0	1.0	1.0	1.0	1.0	1.0	0.448 (0.152)	0.0	-	230.90
Double-log	0.626 (0.172)	-0.874 (0.066)	0.406 (0.061)	0.087 (0.082)	-0.138 (0.050)	0.024 (0.074)	-0.056 (0.010)	-0.051 (0.010)	0.060 (0.010)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	233.77
Double-log A	0.477 (0.285)	-0.841 (0.104)	0.470 (0.079)	0.039 (0.117)	-0.235 (0.090)	-0.039 (0.122)	-0.055 (0.010)	-0.051 (0.009)	0.060 (0.008)	0.0	0.0	0.0	0.0	0.0	0.0	0.486 (0.151)	0.0	-	243.07
Semi-log	0.847 (0.296)	-1.480 (0.113)	0.663 (0.104)	0.171 (0.141)	-0.208 (0.085)	0.031 (0.128)	-0.097 (0.018)	-0.087 (0.018)	0.111 (0.018)	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	234.01
Semi-log A	0.695 (0.446)	-1.448 (0.170)	0.762 (0.137)	0.135 (0.197)	-0.193 (0.142)	-0.032 (0.192)	-0.096 (0.018)	-0.088 (0.016)	0.112 (0.016)	1.0	0.0	0.0	0.0	0.0	0.0	0.416 (0.148)	0.0	-	240.58
Inverse	0.792 (0.108)	-1.532 (0.123)	0.675 (0.110)	0.151 (0.144)	-0.228 (0.101)	0.002 (0.012)	-0.096) (0.018)	-0.086 (0.018)	0.109 (0.019)	1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	-	231.52
Inverse A	0.743 (0.170)	-1.501 (0.186)	0.772 (0.145)	0.123 (0.181)	-0.220 (0.170)	-0.002 (0.018)	-0.095 (0.018)	-0.086 (0.017)	0.111 (0.016)	1.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.420 (0.152)	0.0	-	238.28
Log-inverse	0.580 (0.065)	-0.893 (0.074)	0.415 (0.066)	0.064 (0.086)	-0.157 (0.060)	0.001 (0.007)	-0.055 (0.011)	-0.051 (0.011)	0.059 (0.011)	0.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.0	0.0	-	229.14
Log-inverse A	0.528 (0.116)	-0.859 (0.121)	0.480 (0.087)	0.020 (0.112)	-0.151 (0.112)	-0.004 (0.012)	-0.054 (0.010)	-0.051 (0.010)	0.059 (0.009)	0.0	-1.0	-1.0	-1.0	-1.0	-1.0	0.501 (0.148)	0.0	-	239.26
BC	0.569 (0.104)	-0.750 (0.274)	0.374 (0.193)	-0.039 (0.107)	0.003 (0.008)	-0.163 (0.125)	-0.053 (0.021)	-0.052 (0.020)	0.059 (0.024)	-0.063 (0.639)	1.208 (1.057)	5.909 (2.593)	2.890 (2.315)	18.638 (7.810)	13.015 (3.506)	0.0	0.0	-	246.00
BCA	0.558 (0.087)	-0.709 (0.244)	0.409 (0.208)	-0.014 (0.123)	0.002 (0.008)	-0.137 (0.133)	-0.052 (0.018)	-0.051 (0.017)	0.058 (0.020)	-0.126 (0.546)	1.456 (0.930)	4.400 (2.827)	3.560 (4.279)	19.892 (12.075)	13.418 (5.574)	0.377 (0.158)	0.0	-	251.26
BCH	0.513 (0.120)	-0.678 (0.324)	0.387 (0.235)	0.005 (0.047)	0.005 (0.013)	-0.103 (0.102)	-0.045 (0.023)	-0.042 (0.022)	0.053 (0.028)	-0.387 (0.852)	0.993 (1.088)	3.883 (3.091)	-19.303 (56.90)	17.117 (8.391)	14.796 (4.343)	0.0	1.382 (2.812)	6.451 (13.046)	248.66
BCAH	0.473 (0.105)	-0.570 (0.293)	0.383 (0.238)	0.004 (0.062)	0.005 (0.020)	-0.900 (0.118)	-0.038 (0.019)	-0.036 (0.018)	0.045 (0.023)	-0.675 (0.842)	0.902 (1.291)	2.346 (3.101)	-19.682 (109.4)	16.701 (14.95)	14.410 (6.649)	0.411 (0.162)	0.994 (2.143)	10.293 (11.05)	255.04

Note: Values in parenthesis are asymptotic standard errors. LL is the maximum value of the log-likelihood function.

This test statistic is asymptotically distributed under the null hypothesis as a chi-square random variable with the degrees of freedom corresponding to the number of independent parametric restrictions placed on the unrestricted model. Comparison of the alternative nested models involves comparing the calculated test statistics with the tabulated values of the chi-square variable at the 0.05 significance level with the appropriate degrees of freedom. The calculated values of the chi-square variable for the alternative models are not presented but can be easily derived from Table 1.

Results indicate that the autocorrelated versions of the "classical" functions are a statistical improvement over the nonautocorrelated forms in all cases. The BC, BCH, and BCAH models are a statistically significant improvement over the non-autocorrelated "classical" forms in all cases as are the BCA and BCAH over the autocorrelated versions of the "classical" forms. Test statistics indicate that the BCA, BCH, and BCAH models are a statistical improvement over the BC model. Both the BCA and BCH models are rejected in favor of the more general BCAH model.

The change in the magnitude of the transformation parameters was substantial in some cases. For example, the BCT on the dependent variable changed from -0.063 in the BC model to -0.675 in the BCAH specification. The autocorrelation coefficient (ρ) was significant in all the autoregressive models but the parameters associated with the analytic form of the heteroskedasticity (i.e. δ_1 and λ_z) were not significant as indicated by individual t-tests. The joint interaction of the two parameters did, however, produce a significant statistical improvement as was noted above. This would appear to indicate that likelihood ratio tests are a more appropriate vehicle than t-statistics for testing homoskedasticity of the error variance (at least with the heteroskedastic form postulated in this paper) for these types of nonlinear models.

In summary, the general BCAH Model is a significant statistical improvement as indicated by chi-square tests over all the simpler model specifications. This lends supporting evidence to the contention that the traditional specification of the error term in the simple BC model is incorrect. Further evidence is provided by analyzing the elasticities generated from the alternative models.

The dummy variables for seasons are found to be statistically significant and display the same general pattern in each of the estimated models. Examination of the coefficients reveals that, other variables being the same, pork demand is lower in the second and third quarters and higher in the fourth than in the base period (first quarter). Partial differentiation of equation (5) reveals that the magnitude of the seasonality variables on consumption depends on the coefficients of both the dummy variables as well as the BCT parameter on the dependent variable. For example, if $\lambda_y = 1$ the seasonal effect on demand relative to the base period is equal to the coefficient on the dummy variable. Similarly for $\lambda_y = 0$, the seasonal effect is proportional to the level of consumption, with the coefficient of the dummy variable being the constant of proportionality. For $\lambda_y < 0$, as was found in the unconstrained models, the seasonal effect on demand increases more than proportionately to the level of consumption.

The own price, cross-price, and income elasticities for the alternative models are presented in Table 2. The BCH and BCAH specifications were the only models for which all elasticities were of the expected sign. That is, the income elasticity is negative (researchers have generally found pork to have a negative income elasticity, e.g. Kulshreshtha), the own price elasticity is negative, and all cross-price elasasticities are positive. The nonautoregressive and autoregressive "classical" forms indicate that fish is a complement of pork (i.e., the cross elasticity is negative). The BC and BCA models indicate that the cross-price elasticity between poultry and pork is negative. In addition, all the "classical" forms generate very

Table 2. Estimated Elasticities

Model 1/	Cross Price Elasticities				
	Pork	Beef	Poultry	Fish	Income
Linear	-0.817	0.393	0.091	-0.121	-0.003
	(0.182)	(0.174)	(0.246)	(0.126)	(23.753)
Linear A	-0.797	0.458	0.075	-0.093	-0.067
	(0.291)	(0.243)	(0.380)	(0.210)	(37.383)
Double-log	-0.874	0.406	0.087	-0.138	0.024
	(0.066)	(0.061)	(0.082)	(0.050)	(0.074)
Double-log A	-0.841	0.470	0.039	-0.125	-0.039
	(0.104)	(0.079)	(0.117)	(0.090)	(0.112)
Semi-log	-0.852	0.382	0.099	-0.120	0.018
	(0.065)	(0.060)	(0.081)	(0.049)	(0.073)
Semi-log A	-0.833	0.438	0.078	-0.111	-0.019
	(0.097)	(0.079)	(0.113)	(0.081)	(0.110)
Inverse	-0.863	0.374	0.086	-0.118	0.014
	(0.069)	(0.061)	(0.082)	(0.052)	(0.065)
Inverse A	-0.846	0.428	0.070	-0.114	-0.013
	(0.104)	(0.080)	(0.104)	(0.087)	(0.101)
Log-inverse	-0.874	0.400	0.064	-0.141	0.013
	(0.072)	(0.063)	(0.086)	(0.054)	(0.067)
Log-inverse A	-0.841	0.462	0.020	-0.135	-0.038
	(0.118)	(0.084)	(0.119)	(0.101)	(0.119)
BC	-0.797	0.485	-0.040	0.025	-0.279
	(0.066)	(0.105)	(0.109)	(0.044)	(0.156)
BCA	-0.784	0.518	-0.015	0.019	-0.248
	(0.090)	(0.126)	(0.131)	(0.048)	(0.192)
BCH	-0.857	0.556	0.006	0.040	-0.227
	(0.062)	(0.116)	(0.057)	(0.067)	(0.151)
BCAH	-0.843	0.607	0.006	0.041	-0.228
	(0.093)	(0.135)	(0.089)	(0.104)	(0.202)

Note: Elasticities are evaluated at the sample means. Estimates of the standard errors are in parentheses.

1/ The models designated with an A (e.g. Linear A) indicates that the functional form was estimated with first-order autocorrelation.

small negative income elasticities. On the other hand, the Box-Cox specifications have income elasticities ranging from -0.227 (BCH) to -0.279 (BC). A negative income elasticity would indicate that pork is regarded as an inferior good.

The own price elasticity for pork was relatively stable around a value of -.85 regardless of the model specification. The cross-price elasticity between beef and pork varied from 0.393 (linear) to 0.607 for the BCAH Model. An examination of Table 2 reveals that substantial differences exist in both sign and magnitude among the elasticities generated from the alternative models. Thus, *a priori* choice of functional and error term specification can influence both statistical fit and elasticity estimates.

Given that most empirical analyses have used the Box-Cox model with a homoskedastic-nonautoregressive error structure, (e.g., Chang, Kulshreshtha, and Hassan and Johnson) it appears useful to compare this model with the BCAH specification. The cross-price elasticity between poultry and pork was of a different sign for the two models and the beef cross-elasticity showed a 20 percent difference. The own price and income elasticities showed a 5 and 22 percentage difference, respectively, between the two models. This not only supports the desirablity of adjusting for autocorrelation as shown by Savin and White but also suggests that the analytic form of heteroskedasticity should be simultaneously estimated with the remainder of the model.

Conclusions

This paper has demonstrated, via a quarterly demand model for pork, several key points concerning the Box-Cox functional form and its error term specification. We emphasize the fact that the Box-Cox transformation parameter on the dependent variable can change the implied distribution of the error term which may then bias parameter estimators. Consequently, the error specification is at least as important as the functional form. Many facets of this paper can be generalized to other applications of Box-Cox models. The following conclusions appear applicable for researchers employing Box-Cox flexible functional forms:

1. Autocorrelated residuals, which are likely to occur when time-series data are used, should be corrected for in the Box-Cox models; and

2. The analytic form of heteroskedasticity should be simultaneously estimated with the nonstochastic (i.e., fixed) part of the model.

This paper also shows that the "classical" type functions, including those estimated under autocorrelation, can yield elasticities which are substantially different from those estimated from more general model specifications. The same is true for the simpler versions of the BCAH model. We would suggest that researchers using the Box-Cox transformation should consider estimating the error structure along with the nonstochastic part of the model.

Footnotes

1/ Zarembka (p.92) also indirectly implies this result. He found that the transformation parameter on the dependent variable was not robust to error specification in the demand for money (p. 96).

2/ The supply of pork is assumed to be independent of the current period price.

3/ Preliminary analysis found that the price of pork was the most appropriate variable for stabilizing the error variance as determined by increases in the estimated likelihood function.

4/ The estimated standard errors reported in this paper are only approximate.

References

Benus, J., J. Kmenta, and H. Shapiro. "The Dynamics of Household Budget Allocation to Food Expenditure." Rev. Econ. and Statist. 57(1976): 129-38.

Box, G.E.P., and D.R. Cox. "An Analysis of Transformations." J. Royal Statist. Soc. B. 26(1964): 211-52.

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

Chang, H.S. "Functional Forms and the Demand for Meat in the United States." Rev. Econ. and Statist. 59(1977): 355-59.

Fletcher, R., and M.J.D. Powell. "A Rapidly Convergent Descent Method for Minimization." Computer J. 6(1963): 163-68.

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

Hassan, Z.A., and S.R. Johnson. "The Demand for Meat in Canada: An Application of the Transformation of the Variables." Can. J. Agr. Econ. 27(1979): 1-12.

Kulshreshtha, S. "Functional Form Specification in the Quarterly Demand for Red Meats in Canada." West J. Agr. Econ. 4(1979): 89-97.

Liem, T.C. "A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals." Univ. of Montreal, Centre de Recherche sur les Transports, Pub. No. 134.

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

Savin, N.E., and K.J. White. "Estimation and Testing for Functional From and Autocorrelation: A Simultaneous Approach." J. Econometrics. 8(1978): 1-12.

Zarembka, Paul. "Transformation of Variables in Econometrics." Frontiers of Econometrics, pp. 81-104. New York: Academic Press, 1974.