NOTES

PITFALLS OF RESCALING REGRESSION MODELS WITH BOX-COX TRANSFORMATIONS

Marcel G. Dagenais and Jean-Marie Dufour*

Abstract—To facilitate ML estimation for Box-Cox models, several authors have suggested dividing the dependent variable by its sample geometric mean. This paper points out previously unmentioned drawbacks of this "rescalling." First, the "rescaled" model is not actually equivalent to the untransformed one, so that the procedure involves more than a unit change. Second, there is no clear interpretation of the parameters after such rescaling. We suggest an interpretation but find that the usual formulas for standard errors and confidence intervals are not asymptotically valid. Only tests for zero coefficients are valid. Thirdly, we discuss the appropriate way of measuring elasticities in such models.

I. Introduction

In a paper published some time ago, Spitzer (1984) demonstrated the lack of invariance of asymptotic t-ratios for the linear coefficients of regression models with the Box-Cox transformation on the dependent variable. Further, he discussed the merits of using a scaled version of the model, in which the observations on the dependent variable are divided by their geometric mean. This "rescaling" technique was in fact originally suggested by Box and Cox (1964) and has since been recommended or used in numerous textbooks [see, for example, Abraham and Ledolter (1983, pp. 56–60), Atkinson (1987, sections 6.2 and 6.4), Box, Hunter and Hunter (1978, p. 239), Cook and Weisberg (1982, p. 62), Davidson and MacKinnon (1993, pp. 486–488), Draper and Smith (1981, section 5.3), Judge et al. (1985, section 20.5) and Maddala (1977, section 13–11)] and articles [e.g., Blackley, Follain and Ondrich (1984), Collins (1991), John and Draper (1980), Lindsey (1972), Logothetis (1990), Rasmussen and Zuehlke (1990), Seaks and Vine (1990), Spitzer (1982), Zarembka (1974)].

This approach, however, raises a number of important questions that deserve further discussion. Even though some authors have issued warnings on the importance of taking into account the uncertainty of the Box-Cox transformation parameter (which is typically estimated) when making inferences in such models,[1] none of the above-cited authors has apparently pointed out the fact that "rescaling" of the observations by their geometric mean involves a very similar problem *even if the uncertainty of the estimation of the transformation exponent is taken into account* (using, for example, the corrections suggested by Spitzer (1982, 1984)). In particular, applications of Box-Cox models, with the dependent variable scaled by its geometric mean, have appeared in the literature, where the implications of such random "scaling" on statistical inference have been totally ignored; see, for example, Blackley, Follain and Ondrich (1990), Collins (1991), Logothetis (1990), and Rassmussen and Zuehlke (1984).

In this paper, we discuss some of the intricate implications of rescaling by the geometric mean for statistical inference in Box-Cox models. After describing the model in section II, we show in section III that rescaling the dependent variable with its sample geometric mean (without taking into account the random nature of the scaling factor) leads one to use the wrong likelihood function: the estimated model is not the same as in the original specification, and dividing the dependent variable by the sample geometric mean modifies the model in a basic way and thus involves much more than a mere unit change. Then, in sections IV and V, under an (additional) convergence assumption, we interpret the sample geometric mean as an estimator of a supplementary parameter and use this extended parametrization to interpret hypothesis tests in the "rescaled" model. In particular, regression coefficient standard errors obtained from the rescaled model are not generally valid, so that these standard errors cannot be used to build confidence sets (for either the original or the rescaled model). These standard errors are, however, valid for testing whether individual regression coefficients (other than the inter-

Received for publication April 24, 1992. Revision accepted for publication December 17, 1993.

*Université de Montréal.

This work was supported by the Government of Québec (Fonds FCAR), the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. The authors thank two anonymous referees for several useful comments.

[1] See Bickel and Doksum (1981), Davidson and MacKinnon (1993, p. 487), Judge et al. (1985, p. 842), Maddala (1977, p. 316) and Spitzer (1982, 1984). Here, we shall take for granted this point of view. For further discussion, see Box and Cox (1982), and Hinkley and Runger (1984, with comments).

cept) are zero. In section VI, we discuss the evaluation of elasticities in Box-Cox models. We conclude in section VII.

II The Scaled Model

The simple model considered by Spitzer (1984) to illustrate his approach is the following:

$$Y^{(\lambda)} = X\beta + \epsilon \tag{1}$$

where $Y^{(\lambda)}$ is a $T \times 1$ vector with typical element $y_t^{(\lambda)}, t = 1, \ldots, T$:

$$\begin{aligned} y_t^{(\lambda)} &= \left(y_t^{\lambda} - 1\right)/\lambda, \quad \text{if } \lambda \neq 0, \\ &= \ln(y_t), \qquad \text{if } \lambda = 0, \end{aligned} \tag{2}$$

where y_t is a positive random variable, X is a $T \times K$ matrix of (fixed or strictly exogenous) explanatory variables with a unit vector as its first column, β is a $K \times 1$ vector of parameters (with β_1 the constant term), ϵ is a $T \times 1$ vector of disturbances which is assumed to follow an (approximately) $N(0, \sigma_\epsilon^2 I_T)$ distribution. Given that y_t must be positive, a necessary condition to assume that ϵ is approximately normal, when $\lambda > 0$, is that for each possible observation in the admissible range of X_t (the t^{th} row of X), we have $(1/\lambda + X_t\beta)/\sigma_\epsilon \gg 1$, for all t (see Draper and Cox (1969)). For $\lambda < 0$, the condition becomes $(1/\lambda + X_t\beta)/\sigma_\epsilon \ll -1$.

Box and Cox (1964) and Spitzer (1984) have noted that a scaling trick which simplifies the computations for estimating the parameters of the model consists in defining

$$w_t = y_t/g, \qquad t = 1, \ldots, T \tag{3}$$

where

$$g = \left[\prod_{t=1}^{T} y_t\right]^{1/T}$$

is the sample geometric mean of the y_t's. From (1), we have

$$W^{(\lambda)} = X\gamma + v \tag{4}$$

where $W^{(\lambda)} = (w_1^{(\lambda)}, \ldots, w_T^{(\lambda)})', v = \epsilon/g^{\lambda}$, and

$$\begin{aligned} \gamma_1 &= (\beta_1 + 1/\lambda)/g^{\lambda} - 1/\lambda, \\ \gamma_k &= \beta_k/g^{\lambda}, \quad k = 2, \ldots, K. \end{aligned} \tag{5}$$

Then, writing the log-likelihood function of the transformed model *as if* the elements of v were independent $N(0, \sigma_v^2)$, where $\sigma_v^2 = \sigma^2/g^{2\lambda}$ and g is taken as fixed, the values of λ and γ that maximize this function are readily found, since the logarithm of the Jacobian vanishes. Further, the maximum likelihood estimates of the elements of β are easily obtained from equation (5).

It must be noted, however, that the above procedure is only a trick to facilitate computations. Clearly, the *sample* geometric mean g is a *random* variable and, since $\epsilon \sim N(0, \sigma_\epsilon^2 I_T)$, it is easy to see that $v = \epsilon/g^{\lambda}$ does not follow a normal distribution. Hence, equation (6), appearing in Spitzer (1984),

$$L(\gamma, \lambda; X, W) = -(T/2)\ln\left(\hat{\sigma}_v^2\right) \tag{6}$$

with $\hat{\sigma}_v^2 = v'v/T$, is not the concentrated log-likelihood of W, where W is a $T \times 1$ vector with typical element w_t.[2] Since $\prod_{t=1}^{T} w_t = 1$, the true density of the elements of W is necessarily degenerate. In addition, the elements of γ are *random* coefficients since they are all functions of the *stochastic* variable g. It will be of interest here to derive the correct likelihood function.

III The Likelihood Function

The likelihood associated with the model where the dependent variable has been scaled by its sample geometric mean can be expressed in terms of the joint density of, say, $w_1, w_2, \ldots, w_{T-1}$ and g. This likelihood could be denoted by $l(\beta, \lambda, \sigma_\epsilon; w_1, \ldots, w_{T-1}, g)$, thus indicating that the *fixed parameters* are λ, σ_ϵ and the elements of β (not γ, λ and σ_v) and that the stochastic variables considered in the scaled model are $w_1, \ldots, w_{T-1}$, and g. The joint density of $w_1, \ldots, w_{T-1}$ may then be obtained by integrating out g. As for the joint density of $w_1, \ldots, w_T$, it is identical to that of $w_1, \ldots, w_{T-1}$, since $\prod_{t=1}^{T} w_t = 1$ with probability 1.

Specifically, the log-likelihood function associated with the y_t's can be written as

$$\begin{aligned} L(\beta, \sigma_\epsilon, \lambda; y_1, \ldots, y_T) &= -(T/2)\ln(2\pi) - T\ln(\sigma_\epsilon) \\ &\quad - \left(Y^{(\lambda)} - X\beta\right)'\left(Y^{(\lambda)} - X\beta\right)/\left(2\sigma_\epsilon^2\right) \\ &\quad + \ln(J_1) \end{aligned} \tag{7}$$

where J_1 is the Jacobian of the change of variables from $Y^{(\lambda)}$ to Y:

$$\ln(J_1) = (\lambda - 1)\sum_{t=1}^{T} \ln(y_t) = T(\lambda - 1)\ln(g). \tag{8}$$

Then the log-likelihood associated with the model expressed in terms of the scaled variables $w_1, \ldots, w_{T-1}$ and the sample geometric mean g is

$$\begin{aligned} L(\beta, \sigma_\epsilon, \lambda; w_1, \ldots, w_{T-1}, g) &= L(\beta, \sigma_\epsilon, \lambda; y_1, \ldots, y_T) + \ln(J_2) \end{aligned} \tag{9}$$

[2] For a similar remark about a somewhat different data transformation in the Box-Cox model, see Doksum (1984) and Bickel (1984). Note that a constant term is missing in equation (6), but this has no consequence on our discussion.

where J_2 is the Jacobian associated with the change of variables from $(Y_1, \ldots, Y_T)$ to $(w_1, \ldots, w_{T-1}, g)$:

$$\ln(J_2) = \ln(T) + (T-1)\ln(g) - \sum_{t=1}^{T-1} \ln(w_t). \tag{10}$$

We can therefore write

$$\begin{aligned} L(\beta, \sigma_\epsilon, \lambda; w_1, \ldots, w_{T-1}, g) &= -(T/2)\ln(2\pi) - T\ln(\sigma_\epsilon) \\ &\quad -(Y^{(\lambda)} - X\beta)'(Y^{(\lambda)} - X\beta)/(2\sigma_\epsilon^2) \\ &\quad + \ln(T) + (T\lambda - 1)\ln(g) - \sum_{t=1}^{T-1} \ln(w_t), \end{aligned} \tag{11}$$

where the y_t's contained in $Y^{(\lambda)}$ may be expressed in terms of $w_1, \ldots, w_{T-1}$ and g as

$$y_t = g w_t, \qquad t = 1, \ldots, T-1, \qquad y_T = g \prod_{t=1}^{T-1} w_t^{-1}. \tag{12}$$

Finally, the probability density of $(w_1, \ldots, w_{T-1})$ can be obtained from $l(\beta, \sigma, \lambda; w_1, \ldots, w_{T-1}, g)$ by integrating out g:

$$\begin{aligned} &l(\beta, \sigma, \lambda; w_1, \ldots, w_{T-1}) \\ &= T(2\pi)^{-T/2}\sigma_\epsilon^{-T} \prod_{t=1}^{T-1} w_t^{-1} \int_0^\infty [\, g^{(T\lambda - 1)} \\ &\quad \times \exp\{-(Y^{(\lambda)} - X\beta)'(Y^{(\lambda)} - X\beta)/(2\sigma_\epsilon^2)\}]\, dg. \end{aligned} \tag{13}$$

Clearly, this likelihod function is quite different from the usual Box-Cox likelihood function (where g would be taken as nonrandom). As mentioned above, on the hypersurface defined by $\prod_{t=1}^{T} w_t = 1$, we have:

$$l(\beta, \sigma, \lambda; w_1, \ldots, w_T) = l(\beta, \sigma, \lambda; w_1, \ldots, w_{T-1}). \tag{14}$$

IV. A Large-sample Interpretation of Geometric Mean Rescaling

It appears that an alternative way of interpreting the Box-Cox-Spitzer approach of rescaling by the sample geometric mean might avoid the above complications. It would consist in considering that the original model has been scaled not by the stochastic geometric mean g but by a *constant term* that takes on the specific value $\Xi = \text{plim}\ g$, where plim refers to the probability limit as $T \to \infty$.[3] One would then obtain:

$$z_t^{(\lambda)} = X_t\psi + \zeta_t, \qquad t = 1, \ldots, T, \tag{15}$$

where $z_t = y_t/\Xi$, $\psi_k = \beta_k/\Xi^\lambda$ for $k = 2, \ldots, K$, $\psi_1 = (\beta_1 + 1/\lambda)/\Xi^\lambda - (1/\lambda)$ and $\zeta_t = \epsilon_t/\Xi^\lambda$.

If Ξ were known, a consistent estimator of ψ could be obtained by maximizing the likelihood function associated with the z_t's. Given the properties of ML estimators, the ML estimates of the ψ_k's would satisfy the following equations:

$$\begin{aligned} &\hat{\psi}_1 = \left(\hat{\beta}_1 + 1/\hat{\lambda}\right)/\Xi^{\hat{\lambda}} - \left(1/\hat{\lambda}\right), \\ &\hat{\psi}_k = \hat{\beta}_k/\Xi^{\hat{\lambda}}, \; k = 2, \ldots, K. \end{aligned} \tag{16}$$

Note that the ML estimate of λ would be the same for the original model and the model scaled by Ξ. Now, the $\hat{\gamma}$ estimators proposed by Spitzer (1984) could be seen as alternative computable consistent estimates of the ψ_k's, since

$$\text{plim}\ \hat{\gamma}_1 = \text{plim}\left[\left(\hat{\beta}_1 + 1/\hat{\lambda}\right)/g^{\hat{\lambda}} - \left(1/\hat{\lambda}\right)\right] = \psi_1, \tag{17}$$

$$\text{plim}\ \hat{\gamma}_k = \text{plim}\left(\hat{\beta}_k/g^{\hat{\lambda}}\right) = \psi_k, \qquad k = 2, \ldots, K. \tag{18}$$

This interpretation of $\hat{\gamma}_k$ as a computable estimator of β_k/Ξ^λ avoids the difficulties encountered in section II and preserves the meaning of the tests discussed by Spitzer, since each null hypothesis $\beta_k = 0$ is equivalent to $\psi_k = 0$, for $2 \le k \le K$.[4] However, adopting this alternative viewpoint also has its drawbacks, as pointed out in the next section.

V. The Asymptotic Covariance Matrix Estimator

If one adopts the interpretation discussed in the previous section, the expression shown in equation (12) of Spitzer's (1984) paper may be considered as an estimator of the asymptotic covariance matrix of the elements of $\hat{\psi}$. However, given that g is a random variable, is this estimator appropriate as an estimator of the asymptotic covariance of $\hat{\gamma}$?

There is no evidence that the elements of $\sqrt{T}(\hat{\gamma} - \psi)$ will, in general, have the same asymptotic variances as the corresponding elements of $\sqrt{T}(\hat{\psi} - \psi)$. This would be the case (Theil (1971, p. 370)) if we had

$$\text{plim}\left[\sqrt{T}(\hat{\gamma} - \psi) - \sqrt{T}(\hat{\psi} - \psi)\right] = 0. \tag{19}$$

This does not appear to be the case, in general, except

[3] Clearly, we assume here that plim(g) exists. Whether Ξ coincides with the actual geometric mean of the population or not is not relevant for our purpose.

[4] A somewhat different interpretation suggested by Berndt, Showalter and Wooldridge (1993) would be to assume that the data is generated not by the original equation (1) as is usually done, but directly by equation (15) and that the observed Y's are then obtained by multiplying the z_t' by an *extra parameter* ν, which corresponds to the population geometric mean. The likelihood function used by Spitzer would clearly not be applicable to this extended model either, since it does not incorporate the extra parameter ν.

when $\beta_k = 0$ for $2 \le k \le K$. Indeed, for $2 \le k \le K$, we have

$$\begin{aligned}\sqrt{T}(\hat{\gamma}_k - \psi_k) - \sqrt{T}(\hat{\psi}_k - \psi_k) &= \sqrt{T}(\hat{\gamma}_k - \hat{\psi}_k) \\ &= \sqrt{T}\left(\frac{\hat{\beta}_k}{g^{\hat{\lambda}}} - \frac{\hat{\beta}_k}{\Xi^{\hat{\lambda}}}\right) = \hat{\beta}_k\sqrt{T}\left[\frac{1}{g^{\hat{\lambda}}} - \frac{1}{\Xi^{\hat{\lambda}}}\right].\end{aligned} \tag{20}$$

Since $\hat{\beta}_k\sqrt{T}(1/g^{\hat{\lambda}} - 1/\Xi^{\hat{\lambda}})$ converges in distribution, as $T \to \infty$, to a nondegenerate *random variable*, $\sqrt{T}(\hat{\gamma}_k - \psi_k)$ will not *in general* have the same asymptotic distribution as $\sqrt{T}(\hat{\psi}_k - \psi_k)$. This would be the case, however, under the null hypothesis H_0: $\beta_k = \psi_k = 0$, since the probability limit of $\hat{\beta}_k\sqrt{T}(1/g^{\hat{\lambda}} - 1/\Xi^{\hat{\lambda}})$ then vanishes.

Thus, the t-test proposed by Spitzer for H_0: $\gamma_k = 0$, as well as the corresponding test for the original model, are both *asymptotically* valid. In fact, it can be shown that *under* H_0, as $T \to \infty$ and $\psi_k, \gamma_k \to 0$, the value of the *asymptotic* t-statistic for $\sqrt{T}(\hat{\psi}_k - \psi_k)$ does not change with the scaling factor; the proof of this result is available in Dagenais and Dufour (1993). However, it must be noted that asymptotically valid confidence intervals could not be established for γ_k, using the variance estimator employed by Spitzer (1984) in the denominator of the right-hand side of his equation (15).

Even if the proposed t-test is *asymptotically* valid for testing $\gamma_k = \psi_k = 0$, it does remain, however, as demonstrated by Spitzer, that *in finite samples* the value of the t-statistic can be altered arbitrarily by changing the scaling factor and that invariant tests should therefore be preferred (for some proposals, see Dagenais and Dufour (1991)).

VI. Elasticities

It is often of interest to evaluate elasticities from an econometric model. To be meaningful, such quantities must correspond to a well-defined population parameter (i.e., a nonstochastic characteristic of the model). In nonlinear models of the form

$$f(y_t, X_t, \epsilon_t) = 0, \qquad t = 1, \ldots, T, \tag{21}$$

where X_t is fixed and ϵ_t is a random disturbance, it is natural to define the elasticity of y_t with respect to x_{kt}, given X_t, as:

$$\theta_k = \frac{\partial E(y_t|X_t)}{\partial x_{kt}} \cdot \frac{x_{kt}}{E(y_t|X_t)}. \tag{22}$$

When evaluated at the sample geometric mean of the elements of X (say $\dot{X}$), this yields:

$$\dot{\theta}_k = \frac{\partial E(y|X)}{\partial x_k} \cdot \frac{x_k}{E(y|X)}\bigg|_{X=\dot{X}}. \tag{23}$$

Spitzer (1984) suggests computing elasticities in a Box-Cox regression as

$$\theta_k^* = \left[\frac{\partial y^{(\lambda)}}{\partial x_k} \cdot \frac{x_k}{g^{\lambda}}\right]_{X=\dot{X}} \tag{24}$$

or, for the scaled model,

$$\Gamma_k = \left[\frac{\partial w^{(\lambda)}}{\partial x_k} \cdot x_k\right]_{X=\dot{X}} = \gamma_k \dot{x}_k. \tag{25}$$

However, it appears more meaningful, for most applications, to define the elasticity in terms of proportional changes of the variable of interest (namely, y_t in the case at hand) rather than in terms of proportional changes of a power function of this variable (such as $y_t^{(\lambda)}$). A more important drawback of defining the elasticity as in equations (24) or (25) is that this is a *random variable* which conveys little information unless a confidence interval is also supplied.[5] Clearly, the true elasticity defined by equation (23) is a fixed number, given $\dot{X}$. It will, in general, only depend on the unknown coefficients of the model. But to evaluate the elasticity, the unknown coefficients must be replaced by estimates.[6]

VII. Conclusion

We have tried, in this paper, to clarify a number of questions raised by Spitzer's (1984) paper. In particular, we have pointed out that the coefficients of Spitzer's scaled model are random coefficients and we have derived explicitly the likelihood function of the scaled model. We have also suggested a reinterpretation of Spitzer's approach, in which the coefficients of the scaled model would correspond to fixed parameters. We have then established that the variance estimators suggested by Spitzer for the parameter estimates of the scaled model are inconsistent, except under the null hypothesis of a zero coefficient. We have also pointed out that elasticities should be defined in terms of expected values.

Because of the pitfalls involved in the interpretation of the scaled model, we conclude that there is defi-

[5] Furthermore, if the elasticity is defined as in equation (25), testing hypotheses such as H_0: $\Gamma_k = 1$, where $\Gamma_k = \gamma_k \dot{x}_k$, would involve a contradiction, since γ_k is *stochastic*. A more meaningful null hypothesis would be $\psi_k \dot{x}_k = 1$.

[6] For examples of the estimation of such elasticities in regression models with Box-Cox transformations, see Dagenais, Gaudry and Liem (1987).

nitely no advantage from a statistical viewpoint, in using this type of approach. Nor do we believe that scaling each variable by its sample geometric mean would produce scale invariant t-statistics in regression models with Box-Cox transformations. Clearly, if all practitioners agree, when using regression models with Box-Cox transformations, to adopt a uniform scaling factor (such as the geometric or the arithmetic mean), the results of all experimenters using the same sample will be exactly similar, including the t-ratios of the linear coefficients. However, this uniformity of results, which would arise from the adoption of a *similar* scaling factor, does not render the statistics invariant to *changes* in measurement units.[7] Note that, as pointed out by a referee, researchers may be only interested in testing the linear model ($\lambda = 1$) against the loglinear model ($\lambda = 0$), rather than in performing a full nonlinear estimation of the model with continuous possible values of λ. Within such a framework, scaling by the geometric mean is not an issue since the asymptotic t-tests for λ are scale invariant. Such a procedure, however, would be affected by the biases associated with pretest estimators (Judge et al. (1985)).

REFERENCES

Abraham, B., and J. Ledolter, *Statistical Methods for Forecasting* (New York: Wiley, 1983).

Atkinson, A. C., *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis* (Oxford: Clarendon Press, 1985).

Berndt, E. R., M. H. Showalter and J. M. Wooldridge, "A Theoretical and Empirical Investigation of the Box-Cox Model and a Nonlinear Least Squares Alternative," *Econometric Reviews* 12 (1) (1993), 65–102.

Bickel, P. J., "Comment on 'The Analysis of Transformed Data,'" *Journal of the American Statistical Association* (1984), 315.

Bickel, P. J., and K. A. Doksum, "An Analysis of Transformation Revisited," *Journal of the American Statistical Association* 76 (1981), 296–311.

Blackley, P., J. R. Follain and J. Ondrich, "Box-Cox Estimation of Hedonic Models: How Serious Is the Iterative OLS Variance Bias?," this REVIEW 66 (1984), 348–353.

Box, G. E. P., and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B 26 (1964), 211–243.

_____, "An Analysis of Transformation Revisited, Rebutted," *Journal of the American Statistical Association* 77 (1982), 209–216.

Box, G. E. P, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction. Design, Data Analysis and Model Building* (New York: Wiley, 1978).

Collins, S., "Prediction Techniques for Box-Cox Regression Models," *Journal of Business and Economic Statistics* 3 (35) (1991), 267–277.

Cook, R. D., and S. Weisberg, *Residuals and Influence in Regression* (New York: Chapman and Hall, 1982).

Dagenais, M. G., and J.-M. Dufour, "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica* (1991), 1601–1615.

_____, "Pitfalls of Rescaling Regression Models with Box-Cox Transformations," Working Paper, C.R.D.E., Université de Montréal (1993).

Dagenais, M. G., M. J. I. Gaudry and T. C. Liem, "Urban Travel Demand: The Impact of Box-Cox Transformations with Non-Spherical Residual Errors," *Transportation Research*, Series B (1987), 443–477.

Davidson, R., and J. G. MacKinnon, *Estimation and Inference in Econometrics* (Oxford: Oxford University Press, 1993).

Doksum, K. A., "Comment on the Analysis of Transformed Data," *Journal of the American Statistical Association* (1984), 316.

Draper, N. R., and D. R. Cox, "On Distributions and Their Transformation to Normality," *Journal of the Royal Statistical Society* B (1969), 472–476.

Draper, N. R., and H. Smith, *Applied Regression Analysis*, 2nd edition (New York: Wiley, 1981).

Hinkley, D. V., and G. Runger, "The Analysis of Transformed Data," *Journal of the American Statistical Association* 79 (1984), 302–324 (with comments).

John, J. A., and N. R. Draper, "An Alternative Family of Transformations," *Applied Statistics* 29 (1980), 190–197.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl and T.-C. Lee, *The Theory and Practice of Econometrics*, 2nd edition (New York: John Wiley and Sons, 1985).

Lindsey, J. K., "Fitting Response Surfaces with Power Transformations," *Applied Statistics* 21 (1972), 234–247.

Logothetis, N., "Box-Cox Transformations and the Taguchi Method," *Applied Statistics* 39 (1) (1990), 31–48.

Maddala, G. S., *Econometrics* (New York: McGraw-Hill, 1977).

Rasmussen, D. W., and T. W. Zuehlke, "On the Choice of Functional Form for Hedonic Price Functions," *Applied Economics* 22 (1990), 431–438.

Seaks, T. G., and D. P. Vines, "A Monte Carlo Evaluation of the Box-Cox Difference Transformation," this REVIEW 72 (1990), 506–510.

Spitzer, J. J., "A Primer on Box-Cox Estimation," this REVIEW 64 (1982), 307–313.

_____, "Variance Estimates in Models with the Box-Cox Transformation: Implications for Estimation and Hypothesis Testing," this REVIEW 66 (1984), 645–652.

Theil, H. *Principles of Econometrics* (New York: Wiley, 1971).

Zarembka, P., "Functional Forms in the Demand for Money," *Journal of the American Statistical Association* 77 (1968), 502–511.

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

[7] As pointed out earlier, scaling the dependent variable by its geometric mean may still be resorted to as an intermediate computational technique to obtain the maximum likelihood estimates of the $(\beta', \lambda)'$ vector of parameters. However, given the efficiency of the presently available hardware and software, the usefulness of this algorithmic device is no longer very significant.