An Explicit Variance Formula for the Box-Cox Functional Form Estimator

Zhenlin Yang* and Tilak Abeysinghe

Department of Economics
National University of Singapore
1 Arts Link, Singapore 117570

February 15, 2002

Abstract

Although the Box-Cox transformation provides a flexible functional form for regression models, its applicability is often hampered by the difficulty of choosing an appropriate value for the Box-Cox parameter. This paper presents an explicit variance formula for the Box-Cox estimator of the functional form, from which the analytical behavior of the estimator and its precision can be assessed.

Keywords: Box-Cox transformation, transformation estimator, precision, export function

JEL classification: C10

*Corresponding author: Address as above; Tel.: (65) 874 3958; Fax: (65) 775 2646; email: ecsyzl@nus.edu.sg

1 Introduction

Economic variables typically have skewed distributions with heteroscedastic errors and require instantaneous transformations before they are used in linear regression models. Applied econometricians are so attuned to the logarithamic transformation that they hardly bother to look for other possible transformations. Obviously one could make a strong case for the log transformation on the ground that it is simple, it provides constant elasticities and log first differences approximate growth rates. It is often the case, however, that regressions based on two cross sections of data, ten or twenty years apart, provide different elasticity estimates, which indicates that a functional form that allows for changing elasticity over time is more plausible than a constant elasticity formulation. The Box-Cox transformation provides non-constant elasticities over time when the Box-Cox parameter (λ) is not zero. Furthermore, recent publications by Franses and McAleer (1997) and Franses and Koop (1998) on unit roots under general transformations is likely to heighten the interest again in the Box-Cox family of transformations (Box and Cox, 1964).

The Box-Cox transformation provides a flexible functional form given by

$$h(Y_t) = \beta_0 + \beta_1 X_{t1}(\lambda_1) + \cdots + \beta_k X_{tk}(\lambda_k) + \epsilon_t, \quad t = 1, \cdots, T. \tag{1}$$

where $X(\lambda)$ is defined by the power transformation

$$X(\lambda) = \begin{cases} (X^\lambda - 1)/\lambda, & \lambda \neq 0, \\ \log X, & \lambda = 0, \end{cases} \tag{2}$$

and $h(Y)$ is a general transformation that renders normal or near normal responses.[1] An important special case of model (1) is that all transformations are the same:

$$Y_t(\lambda) = \beta_0 + \beta_1 X_{t1}(\lambda) + \cdots + \beta_k X_{tk}(\lambda) + \epsilon_t, \quad t = 1, \cdots, T, \tag{3}$$

with the standard linear form given by ($\lambda = 1$) and the log-linear form by ($\lambda = 0$). The general form given in (1) is often considered too cumbersome than necessary (Green, 2000,

[1]Zarembka (1968) seems to be the first one to apply this transformation in an econometric model. In Zarembka's model for money demand, both endogenous (response) and exogenous variables are transformed, whereas in the original Box-Cox model only the response is transformed. For some review articles and for more references see, Collins (1991), Sakia (1992), Kim and Hill (1995), and Kemp (1996).

Section 10.5). Our study concentrates only on model (3) where some exogenous variables such as time trend and dummies may enter the model untransformed.

Despite the long history of the Box-Cox transformation, having to use an estimated value for λ remains an unattractive feature of the procedure for many econometricians. This is because, at least partially, the analytical behavior of the transformation estimator remains unclear. As a result it is not clear with what precision a given data set can be transformed. For example, it is certainly more reliable to use a transformation estimate of 0.25 with a standard error of 0.01 than an estimate of 0.25 with a standard error of 0.2. The objective of this study is to provide an explicit expression for the variance of the transformation estimator so that the properties of this estomator can be examined and its precision can easily be assessed for a given data set. Moreover, the expression allows the easy construction of tests and confidence intervals for λ. The accuracy of the explicit expression is assessed through a Monte Carlo simulation. The final section provides an illustrative example by modeling demand for Singapore's exports.

2 Variance of the Transformation Estimator

Assume that there exists some λ value such that the error process of model (3) is iid $N(0, \sigma^2)$. Then, the log likelihood function is:

$$\ell(\beta, \sigma^2, \lambda) \propto -\frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{Y_t(\lambda) - \sum_{j=0}^{k}\beta_j X_{tj}(\lambda)\right\}^2 + \sum_{t=1}^{T}\log J_t(\lambda) \tag{4}$$

where $X_{t0}(\lambda) = 1$, and $J_t(\lambda) = |\partial Y_t(\lambda)/\partial Y_t|$ is the Jacobian of the transformation from Y_t to $Y_t(\lambda)$. Writing model (3) in matrix form: $\mathbf{Y}(\lambda) = \mathbf{X}(\lambda)\beta + \epsilon$, and denoting the geometric mean of the Ys by $\tilde{Y}$, the maximum likelihood estimators (MLE) can be written as:

$$\begin{aligned} \hat{\beta}(\hat{\lambda}) &= [\mathbf{X}(\hat{\lambda})'\mathbf{X}(\hat{\lambda})]^{-1}\mathbf{X}(\hat{\lambda})'\mathbf{Y}(\hat{\lambda}), \\ \hat{\sigma}^2(\hat{\lambda}) &= \frac{1}{T}\|\mathbf{M}(\hat{\lambda})\mathbf{Y}(\hat{\lambda})\|^2, \\ \hat{\lambda} &= \arg\min_{\ell} \tilde{Y}^{-\ell}\|\mathbf{M}(\ell)\mathbf{Y}(\ell)\|, \end{aligned} \tag{5}$$

where $\mathbf{M}(\lambda) = \mathbf{I}_T - \mathbf{X}(\lambda)[\mathbf{X}(\lambda)'\mathbf{X}(\lambda)]^{-1}\mathbf{X}(\lambda)'$, and $\mathbf{I}_T$ is a $T \times T$ identity matrix.

As stated earlier, it is important to find an explicit expression for the variance of $\hat{\lambda}$ (Box and Cox, 1982; Yang, 1999). Let this variance be denoted by $\tau^2(\beta, \sigma^2, \lambda)$. The following conventions are followed. For a sequence $\{a_t, t = 1, \cdots T\}$, $\mathbf{a}$ is the corresponding column vector, $\bar{a}$ is the average, $\mathbf{a}^2 = \{a_t^2\}_{T\times 1}$, $\mathbf{a}\#\mathbf{b} = \{a_t b_t\}_{T\times 1}$ for another vector $\mathbf{b}$ of the same length, and $\mathbf{a} + c = \{a_t + c\}_{T\times 1}$ where c is a scalar.

Theorem. *Let* $\theta_t = \sigma\lambda/(1 + \lambda\mu_t(\lambda))$ *and* $\theta_0 = max|\theta_t|$. *For small* θ_0 *and large* T,

$$\tau^2(\beta, \sigma^2, 0) \approx \frac{1}{\frac{1}{\sigma^2}\|\mathbf{M}(0)\delta\|^2 + 2\|\mu(0) - \bar{\mu}(0)\|^2 + \frac{3}{2}T\sigma^2}, \tag{6}$$

$$\tau^2(\beta, \sigma^2, \lambda) \approx \frac{1}{\frac{1}{\sigma^2}\|\mathbf{M}(\lambda)\delta\|^2 + \frac{1}{\lambda^2}\left[2\|\phi - \bar{\phi}\|^2 - 4(\phi - \bar{\phi})'(\theta^2 - \bar{\theta^2}) + \frac{3}{2}\|\theta\|^2\right]}, \tag{7}$$

where $\mu(\lambda) = \mathbf{X}(\lambda)\beta$, $\phi = \log(1 + \lambda\mu(\lambda))$, $\delta = \frac{1}{2}[\mu^2(0) + \sigma^2] - \dot{\mathbf{X}}(0)\beta$ *for* $\lambda = 0$, $\delta = \frac{1}{\lambda^2}(1 + \lambda\mu(\lambda))\#\phi + \frac{\sigma}{2\lambda}\theta - \dot{\mathbf{X}}(\lambda)\beta$ *for* $\lambda \neq 0$, *and* $\dot{\mathbf{X}}(\lambda) = d\mathbf{X}(\lambda)/d\lambda$.

The error of approximation to the variance $\tau^2(\beta, \sigma^2, \lambda)$ is of order $O_p(T^{-3/2})$ for $\lambda = 0$, and $O_p(T^{-3/2}) + O_p(\theta_0^3)$ for $\lambda \neq 0$. The proof of the theorem is given in the Appendix.

The expressions (6) and (7) readily reveal the analytical behavior of the transformation estimator. It can easily be seen that the variability of $\hat{\lambda}$ is governed by the interplay of three factors. First, the model structure as captured by the $\mathbf{M}\delta$ terms in (6) and (7) determines the magnitude of the variance. The larger is this term, the smaller is the variability of $\hat{\lambda}$ and more precisely can λ be estimated. Second, the variability of explanatory variables as captured by the second terms in the denominators of (6) and (7) help reduce the variance of $\hat{\lambda}$. Again, the larger the variability the smaller is the variance of $\hat{\lambda}$. Third, the magnitude of error variance affects the variance of $\hat{\lambda}$ in a non-uniform way. Smaller σ enlarges the $\mathbf{M}\delta$ term in (6) and (7) and reduces the variability of $\hat{\lambda}$. When σ increases, $Var(\hat{\lambda})$ first increases and then start to decrease because the last terms in the denominators of (6) and (7) become dominant.

In addition to understanding the analytical behavior of $\hat{\lambda}$, the expressions (6) and (7) provide ready estimates of the variance of $\hat{\lambda}$, allowing us to assess the precision of the transformation estimator for a given data set. Once the MLEs $\hat{\beta}, \hat{\sigma}^2$ and $\hat{\lambda}$ are obtained, they can be substituted into $\tau(\beta, \sigma^2, \lambda)$ to give an estimated standard error of $\hat{\lambda}$ that can then be used to construct tests and confidence intervals for λ.

3 Monte Carlo Evaluation

The variance formula given in the above theorem is an asymptotic expression. In this section we examine the small sample accuracy of this formula through a Monte Carlo exercise. We compare the value of $\tau(\beta, \sigma^2, \lambda)$ with the corresponding simulated standard deviation of $\hat{\lambda}$. For this we use the simple model:

$$Y_t(\lambda) = \beta_0 + \beta_1 X_t(\lambda) + \epsilon_t, \quad t = 1, \cdots, T. \tag{8}$$

We consider several sets of parameter values, two sample sizes, and select $\log X$ values uniformly from the interval (0, 5). The simulation process basically involves generating a sample of T values from $N(0, \sigma^2)$, converting them to $Y_t(\lambda)$ through model (8), inverting $Y_t(\lambda)$ to obtain Y_t, and finally calculating $\hat{\lambda}$ by minimizing $\tilde{Y}^{-\lambda}\|M(\lambda)\mathbf{Y}(\lambda)\|$. For each parameter setting, 10,000 replicates of $\hat{\lambda}$ are obtained and the standard deviation of these 10,000 values gives a Monte Carlo estimate of $sd(\hat{\lambda})$. Table 1 summarizes the results.

Table 1 here

For many economic applications the λ required to render normal errors is likely to be small and a small λ usually comes with a small σ. Therefore, the results in Table 1, though somewhat limited, are informative for many practical cases. A comparison of simulated and formula-based standard deviations shows that our variance formula is very accurate even in small samples, hence the variance estimator it provides should also be very accurate. As σ increases the formula seems to underestimate $Var(\hat{\lambda})$ slightly. The table also shows that the magnitude of the slope coefficients also affects the standard deviation of $\hat{\lambda}$. We also computed (not reported for brevity) the standard deviations by increasing the range of X and observed that $Var(\hat{\lambda})$ drops substantially as the range of X increases.

4 An Application

This section presents very briefly an illustrative application by modeling the demand for Singapore's exports. Here the demand for Singapore's non-oil domestic real exports (X) is assumed to depend on exogenously determined export price level (P_x), and price (P_w) and real income (Y_w) levels of the importing countries.[2]

We estimated the model with a lag dependent variable as an additional explanatory variable. The resulting estimate for λ is 0.1523 with a standard error of 0.0712. These estimates reject the hypothesis $\lambda = 0$ at the 5 percent level. The likelihood ratio test leads to the same conclusion. We estimated the model for both $\lambda = 0$ (log-linear) and $\lambda = 0.15$. Both specifications fit the data very well, the coefficient estimates are statistically significant and have the expected signs, the models pass the diagnostics (except for normality) available in PCGIVE and provide similar elasticities at the average values of the variables. Both specifications fail the normality test due to excess kurtosis. The coefficient of the lagged-dependent variable is around 0.8 for both cases.

Although it is difficult to choose between the two specifications based on standard diagnostics, the (short-run) elasticity estimates from the non-log model given in Figure 1 show the importance of obtaining a good estimate for λ. Unlike the log-linear model which provides constant elaticities, the non-log specification shows highly plausible trends in income and price elaticities. Figure 1 shows that elasticity with respect to the key determinant of Singapore's exports (Y_w) has dropped by about 26 percent over the last two decades. This is to be expected for economies that move from a fast-growing stage to a more mature stage.

Figure 1 here.

[2] X is in millions of Sin dollars in 1995 prices, and P_x, P_w, and Y_w are indices with 1995 as the base. Y_w is an export-weighted (geometric) average of real GDP of ASEAN4 (Indonesia, Malaysia, Phillipines, Thailand), NIE3 (Hong Kong, South Korea, Taiwan), China, Japan, USA, and the rest of OECD as a group. P_w is similarly constructed from WPI or PPI depending on the data availability. Both P_x and P_w are in Sin dollars. The data series are over the period 1978Q1-1999Q4. The data set can be obtained from the authors.

5 Conclusion

The derived variance estimator shows that the precision of the Box-Cox functional-form determination depends on the model structure, the variability of the explanatory variables and the error variance. Monte Carlo results show that the variance estimator provides reliable estimates in small samples. The illustrative example shows why it is important to pay attention to the functional form though the log-linear model may appear to fit the data well.

Appendix: Proof of the Theorem

Let $\phi = (\beta', \sigma^2)'$. For large T, $\tau^2(\beta, \sigma^2, \lambda) \approx (I_{\lambda\lambda} - I_{\lambda\phi} I_{\phi\phi}^{-1} I_{\phi\lambda})^{-1}$. The elements of the expected information matrix are:

$$
\begin{aligned}
&I_{\beta\beta} = \tfrac{1}{\sigma^2}\mathbf{X}'(\lambda)\mathbf{X}(\lambda), && I_{\sigma^2\sigma^2} = \tfrac{T}{2\sigma^4}, && I_{\lambda\lambda} = \tfrac{1}{\sigma^2} E[\dot{e}'(\lambda,\beta)\dot{e}(\lambda,\beta) + e'(\lambda,\beta)\ddot{e}(\lambda,\beta)],\\
&I_{\beta\lambda} = -\tfrac{1}{\sigma^2}[\mathbf{X}'(\lambda)E[\dot{e}(\lambda,\beta)], && I_{\beta\sigma^2} = 0, && I_{\sigma^2\lambda} = -\tfrac{1}{\sigma^4}E[e'(\lambda,\beta)\dot{e}(\lambda,\beta)],
\end{aligned}
$$

where $e(\lambda, \beta) = \mathbf{Y}(\lambda) - \mathbf{X}(\lambda)\beta \equiv \epsilon$, $\dot{e} = \partial e/\partial\lambda$, and $\ddot{e} = \partial^2 e/\partial\lambda^2$. Using the expected information given above, one obtains, after some alegbra,

$$
\begin{aligned}
\tau^{-2}(\beta, \sigma^2, \lambda) &= I_{\lambda\lambda} - I_{\lambda\psi} I_{\psi\psi}^{-1} I_{\psi\lambda} = I_{\lambda\lambda} - I_{\lambda\beta} I_{\beta\beta}^{-1} I_{\beta\lambda} - I_{\lambda\sigma^2} I_{\sigma^2\sigma^2}^{-1} I_{\sigma^2\lambda}\\
&= \frac{1}{\sigma^2}\left\{ E[\epsilon'\ddot{\mathbf{Y}}(\lambda)] + \sum_{t=1}^{T} Var[\dot{Y}_t(\lambda)] - \frac{2}{T\sigma^2}[E(\epsilon'\dot{\mathbf{Y}}(\lambda))]^2 + E[\dot{e}'(\lambda,\beta)]\mathbf{M}(\lambda)E[\dot{e}(\lambda,\beta)] \right\}
\end{aligned}
$$

where $\dot{\mathbf{Y}}(\lambda)$ and $\ddot{\mathbf{Y}}(\lambda)$ are, respectively, the first and second order partial derivatives of $\mathbf{Y}(\lambda)$ with respect to λ and are given by:

$$
\begin{aligned}
\dot{Y}_t(\lambda) &= \begin{cases} \frac{1}{\lambda}[1 + \lambda Y_t(\lambda)]\log Y_t - \frac{1}{\lambda}Y_t(\lambda), & \lambda \neq 0,\\ \frac{1}{2}(\log Y_t)^2, & \lambda = 0, \end{cases}\\
\ddot{Y}_t(\lambda) &= \begin{cases} \dot{Y}_t(\lambda)(\log Y_t - \frac{1}{\lambda}) - \frac{1}{\lambda^2}[\log Y_t - Y_t(\lambda)], & \lambda \neq 0,\\ \frac{1}{3}(\log Y_t)^3, & \lambda = 0. \end{cases}
\end{aligned}
$$

When $\lambda = 0$, we have

$$
\begin{aligned}
&E[\epsilon_t \dot{Y}_t(0)] = \sigma^2 \mu_t(0), && E[\epsilon_t \ddot{Y}_t(0)] = \sigma^4 + \sigma^2 \mu_t^2(0)\\
&Var[\dot{Y}_t(0)] = \sigma^2\mu_t^2(0) + \tfrac{1}{2}\sigma^4; && E[\dot{e}_t(0,\beta)] = \tfrac{1}{2}[\mu_t^2(0) + \sigma^2] - \dot{X}_t'(0)\beta.
\end{aligned}
$$

Putting everything together and simplifying gives (6). When $\lambda \neq 0$, an approximation to $\log Y_t$ is necessary. The following approximation

$$\log Y_t \approx \frac{1}{\lambda}(\phi_t + \theta_t z_t - \frac{1}{2}\theta_t^2 z_t^2 + \frac{1}{6}\theta_t^3 z_t^3), \ \ z_t \sim N(0,1)$$

is developed based on a third-order Taylor expansion and the consideration the θ_t is small so that its accuracy is guaranteed. Using this approximation, we have,

$$\begin{aligned}
\dot{Y}_t(\lambda) &= \frac{1}{\lambda^2}(1+\lambda\mu_t)\phi_t - \frac{1}{\lambda}\mu_t + \frac{\sigma}{\lambda}\phi_t z_t + \frac{\sigma}{2\lambda}\theta_t z_t^2 - \frac{\sigma}{3\lambda}\theta_t^2 z_t^3 + O(\theta_t^3) \\
Var[\dot{Y}_t(\lambda)] &= \frac{\sigma^2}{\lambda^2}(\frac{\theta_t^2}{2} + \phi_t^2 - 2\phi_t\theta_t^2) + O(\theta_t^3) \\
E[\epsilon_t \dot{Y}_t(\lambda)] &= \frac{\sigma^2}{\lambda}(\phi_t - \theta_t^2) + O(\theta_t^3) \\
E[\epsilon_t \ddot{Y}_t(\lambda)] &= \frac{\sigma^2}{\lambda^2}(\phi_t^2 - 2\phi_t\theta_t^2 + \theta_t^2) + O(\theta_t^3)
\end{aligned}$$

Putting everything together and simplifying gives (7). Note that direct derivation of $E[\epsilon_t \ddot{Y}_t(\lambda)]$ gives the last term as $2\theta_t^2$. This is due to the approximation and checking with the case of $\lambda = 0$ (the case of exact expression) allows us to conclude that this term should be θ_t^2 instead. Monte Carlo simulation results given in Table 1 confirm this point.

References

Box, G.E.P. and Cox,D.R. 1964. An analysis of transformation (with discussion). Journal of Royal Statistical Society B. 26, 211-252.

Box, G.E.P. and Cox,D.R. 1982. An analysis of transformation revisited, rebutted. Journal of the American Statistical Association 77, 209-210.

Collins, S. 1991 Prediction Techniques for Box-Cox Regression Models. Journal of Business and Economic Statistics 9, 267-277.

Greene, W. H. 2000. *Econometric Analysis*, 4^{th} ed., Prentice-Hall, Inc.

Franses, P.H., Koop, G. 1998. On the sensitivity of unit root inference to nonlinear data transformations. Economics Letters 59, 7-15.

Franses, P.H., McAleer, M. 1998. Testing for unit roots and non-linear transformations. Journal of Time Series Analysis 19, 147-164.

Kemp, G. C. R. 1996. Scale equivariance and the Box-Cox transformation. Economics Letters 51, 1-6.

Kim, M. and Hill, R. C. 1995. Shrinkage estimation in nonlinear regression The Box-Cox transformation. Journal of Econometrics 66, 1-33.

Sakia, R. M. 1992. The Box-Cox transformation technique: a review. The Statistician 41, 169-178.

Yang, Z. 1999. Estimating transformation and its effect on Box-Cox T-ratio. Test 8, 167-190.

Zarembka, P. 1968. Functional form in the demand for money. Journal of the American Statistical Association 63, 502-511.

Table 1: Standard Deviations of $\hat{\lambda}$: Simulated (Col. 1) and Formula-based (Col. 2)

		$T = 25$				$T = 50$			
λ	σ	$\beta = (-5, -0.5)$		$\beta = (-5, -2)$		$\beta = (-5, -0.5)$		$\beta = (-5, -2)$	
-0.25	.01	.0073	.0073	.0015	.0015	.0051	.0051	.0010	.0010
	0.1	.0731	.0726	.0148	.0147	.0507	.0504	.0101	.0102
	0.5	.3480	.3099	.0718	.0708	.2259	.2159	.0493	.0491
	1.0	.3714	.3258	.1275	.1246	.2415	.2288	.0866	.0863
λ	σ	$\beta = (5, 0.5)$		$\beta = (5, 2)$		$\beta = (5, 0.5)$		$\beta = (5, 2)$	
0.0	.01	.0086	.0087	.0011	.0011	.0061	.0061	.0008	.0008
	0.1	.0814	.0788	.0107	.0105	.0566	.0556	.0074	.0074
	0.5	.1914	.1566	.0394	.0361	.1224	.1107	.0267	.0254
	1.0	.1493	.1242	.0498	.0431	.0962	.0878	.0325	.0304
	10.	.0187	.0163	.0177	.0155	.0124	.0115	.0118	.0110
λ	σ	$\beta = (5, 0.5)$		$\beta = (5, 2)$		$\beta = (5, 0.5)$		$\beta = (5, 2)$	
0.25	.01	.0035	.0035	.0094	.0093	.0025	.0025	.0066	.0066
	0.1	.0350	.0349	.0755	.0698	.0248	.0248	.0511	.0493
	0.5	.1643	.1460	.1185	.1017	.1096	.1038	.0768	.0719
	1.0	.2371	.1928	.1194	.1017	.1507	.1369	.0776	.0719