8289-

1990

THE BOX-COX METHODOLOGY:

A 26-YEAR MISTAKE

by

Octavio A. Ramirez

and

J. S. Shonkwiler*

February 18, 1990

Statistics

Submitted to the AAEA meetings

*Octavio A. Ramirez is a Graduate Assistant and J. S. Shonkwiler is a Professor in the Department of Food and Resource Economics at the University of Florida, Gainesville, 32611.

AAEA 1990.

THE BOX-COX METHODOLOGY:

A 26-YEAR MISTAKE

Abstract

The Box-Cox methodology has been extensively applied since it was first introduced in 1964. This study establishes that the function currently used as a basis for implementing the Box-Cox methodology is not a valid probability density function. The proper density function associated with the random variable underlying the Box-Cox methodology is derived. Simulation evidence documents that the original Box-Cox approach can result in considerable bias.
Keywords: Box-Cox, density function, random variable, bias.

#8289

THE BOX-COX METHODOLOGY: A 26-YEAR MISTAKE

Introduction

The Box-Cox methodology has been extensively applied since it was first introduced in 1964. A vast amount of research has been conducted attempting to evaluate the performance of this methodology under different empirical scenarios (i.e. Burbidge et al., Davidson, Spitzer, Magee, Mackinnon et al.). Research has also dealt with possible modifications of the Box-Cox methodology that accommodate negative values of the dependent variable (Bickel). Poirier (1978) pioneered the use of the Box-Cox transformation in limited dependent variable models. Furthermore, extensions of the Box-Cox methodology to accommodate simultaneous equation models have also been proposed (Spitzer, 1977).

The "density function" underlying the Box-Cox methodology, however, does not fulfill the requirements of a proper density function and it should not be used as a basis for maximum likelihood estimation. The problem results from the application by Box and Cox of an incorrect methodology in their derivation of the "density function". This study shows that the function proposed by Box and Cox is not a bona fide probability density function and it derives the correct probability density function underlying the Box-Cox transformation.

The Box-Cox Methodology

The Box-Cox methodology is based on transforming the random variable Y according to

(1) $(Y^{\lambda}-1)/\lambda = G$

where G is assumed to be a normal random variable with mean μ and variance σ^2. The task is then to derive the density function associated with the random variable Y.

(2) $Y = (\lambda G+1)^{1/\lambda}$.

Box and Cox use the standard transformation technique to obtain the following probability density function

(3) $f_Y^B(y) = (y^{\lambda-1})(2\pi\sigma^2)^{-1/2} \exp(-.5[([(y^{\lambda}-1)/\lambda]-\mu)/\sigma]^2)\ dy \quad 0<y<\infty$

Applying the standard transformation technique will yield the appropriate probability density function for the random variable Y if (Mood et al.):

(a) $Y = Q(G)$ defines a one-to-one transformation of G^d onto Y^d; where G^d and Y^d are the domains of the random variables G and Y respectively.

(b) The derivative of $G = Q^{-1}(Y)$ with respect to Y is continuous and nonzero for all Y belonging to Y^d; where $Q^{-1}(Y)$ is the inverse function of $Q(G)$.

The transformation defined by equation (2), however, does not satisfy these requirements for any value of λ between 0 and 1. If λ is such that $1/\lambda$ is an even integer, the domain of the random variable Y is restricted between zero and infinity, but equation (2) does not define a one-to-one transformation of G onto Y. If λ is such that $1/\lambda$ is an odd integer, equation (2) is a one-to-one transformation but $d(Q^{-1}(Y))/dY = Y^{\lambda-1}$ is not continuous (or even defined) for values of Y less than or equal to zero. If λ is such that $1/\lambda$ is not an integer, equation (2) is not defined for values of G less than zero.

Application of the standard Transformation Technique under these circumstances will not yield the correct density function associated with the random variable Y. In fact, it can be shown that the density function currently used as a basis for the Box-Cox methodology does not integrate out to one (see appendix 1):

$$(4) \quad \int_0^\infty f_Y^B(y) = \int_0^\infty f_G(g) < 1$$

where $f_G(g)$ is a normal density function .

This characteristic makes $f_Y^B(y)$ unacceptable as a density function. If equation (3) is used as the basis for maximum likelihood estimation, the underlying estimators will not have all of the desirable properties associated with maximum likelihood estimators since they are not maximum likelihood estimators. In general, they will be biased and inconsistent. Furthermore, the estimators of the standard errors will be meaningless.

The correct methodology

The proper density function associated with the random variable Y can be derived using a more general form of the **transformation technique** (Mood et al.). If λ is such that $1/\lambda$ is an even integer, equation (2) can be used as the basis to obtaining the correct probability density function for the random variable Y. Otherwise, a slight modification of equation (2) is required

(5) $Y = (|\lambda G+1|)^{1/\lambda}$

where $|\lambda G+1|$ is the absolute value of the normal random variable $\lambda G+1$. Since $\lambda G+1$ is a normal random variable, it takes on negative values with some probability greater that zero. It follows that taking the absolute value of $\lambda G+1$ is required for the transformation to be defined when $1/\lambda$ is not an integer.

Although this modification is, in concept, necessary to properly apply the more general form of the **transformation technique** when $1/\lambda$ is not an even integer; the probability density function~~s~~ resulting when such technique is applied, is exactly the same under both circumstances (see appendix 2):

$$(6) \quad f_Y^0(y)=(y^{\lambda-1})(2\pi\sigma^2)^{-1/2}\ (\exp(-.5[([(y^{\lambda}-1)/\lambda]-\mu)/\sigma]^2) + \exp(-.5[([-(y^{\lambda}+1)/\lambda]-\mu)/\sigma]^2))\ dy \qquad 0<y<\infty.$$

Equation (6) is the probability density function that should be used as a basis for the Box-Cox methodology. It can be shown that (see appendix 1)

$$(7) \quad \int_0^\infty f_Y^0(y) = \int_0^\infty f_G(g) + \int_{-\infty}^0 f_G(g) = 1$$

so that $f_Y^0(y)$ is a proper probability density function.

An example

In order to illustrate the problems associated with using equation (3) instead of the appropriate probability density function given by equation (6) as the basis for maximum likelihood estimation, a simple monte carlo experiment can be designed. For four different sets of parameters, 100 samples each of them containing 1000 observations from the probability space associated with the random variable Y where taken using a normal random variable generator and equation (5). Using equation (5) is equivalent to using equation (2) if $1/\lambda$ is and even integer. Then, parameter estimates were obtained using two different likelihood functions; one based on equation (6) and one based on equation (3).

Tables 1, 2, 3 and 4 present the results of the experiment. Notice that when equation (6) was used as a basis for maximum likelihood estimation, the averages of the 100 parameter estimates are always very close to the true parameter values used to generate the samples. Furthermore, the estimated asymptotic variances associated with the averages of the parameter estimates seem to accurately reflect the degree of precision with which such averages approximate the true underlying parameter values. When equation (3) was used to specify the likelihood function, however, λ is consistently underestimated. The averages of the estimates for μ and σ^2 are not any better. Furthermore, the estimated asymptotic variances are extremely misleading as to the precision with which the estimates approximate the true underlying parameter values.

Conclusions

This study establishes that the function currently used as a basis for implementing the Box-Cox methodology is not a bona fide probability density function. It derives the proper density function associated with the random variable underlying the Box-Cox methodology. Furthermore, it shows that using the function proposed by Box and Cox may result on considerable bias and inconsistency of the parameter estimates, and very misleading estimates of the asymptotic variances associated with them.

Finally, although space limitations do not allow further discussion regarding the proper probability density function (equation (6)), it is important to mention that maximum likelihood estimation of the transformed linear model based on equation (6) is as simple as maximum likelihood estimation based on the function proposed by Box and Cox since equation (6) can be concentrated in a similar fashion than equation (3).

Appendix 1

First of all, notice that even if we ignore all the arguments presented in the paper regarding the transformation $Y = (\lambda G+1)^{1/\lambda}$, the domain of $f_Y^B(y)$ has to be limited between zero and infinity because neither $Y^{\lambda-1}$ nor Y^{λ} $0<\lambda<1$ is defined for y less than zero. Using a little algebra on the exponent:

$$\int_0^\infty f_Y^B(y) = \int_0^\infty (y^{\lambda-1})(2\pi\sigma^2)^{-1/2} \exp(-.5[(y^\lambda-[\lambda\mu+1])/\lambda\sigma]^2)\, dy$$

let $X = Y^\lambda$, $0<Y<\infty$ $0<\lambda<1$. Thus, $0<X<\infty$.

$Y = X^{1/\lambda}$, a one-to-one transformation since $X>0$.

$dY = 1/\lambda\ X^{(1/\lambda)-1}\, dX$.

Therefore,

$$\int_0^\infty f_Y^B(y) = \int_0^\infty (x^{1/\lambda})^{\lambda-1}\ (2\pi\sigma^2)^{-1/2} \exp^{-.5[(x-[\lambda\mu+1])/\lambda\sigma]^2}\ (1/\lambda)\ x^{(1/\lambda)-1}\, dx$$

$$= \int_0^\infty (2\pi\lambda^2\sigma^2)^{-1/2} \exp^{-.5[(x-[\lambda\mu+1])/\lambda\sigma]^2}\, dx \quad < 1$$

Which is the integral from zero to infinity of a normal density function and it integrates out to less than one since the integral from minus infinity to infinity of a normal density function equals one.

On the other hand; also using a little algebra on the exponents:

$$\int_0^\infty f_Y^O(y) = \int_0^\infty (y^{\lambda-1})(2\pi\sigma^2)^{-1/2} (\exp(-.5[(y^\lambda-[\lambda\mu+1])/\lambda\sigma]^2)$$

$$+\exp(-.5[(-[y^\lambda]-[\lambda\mu+1])/\lambda\sigma]^2))\, dy$$

$$= \int_0^\infty f_Y^B(y) + \int_0^\infty (y^{\lambda-1})(2\pi\sigma^2)^{-1/2} \exp(-.5[(-[y^\lambda]-[\lambda\mu+1])/\lambda\sigma]^2)\, dy$$

Applying the same transformation outlined above, the second term becomes:

$$= \int_0^\infty (2\pi\lambda^2\sigma^2)^{-1/2} \exp^{-.5[(-x-[\lambda\mu+1])/\lambda\sigma]^2}\, dx$$

To evaluate that integral, let Z = -X (a one-to-one transformation) so that X = -Z, $0>Z>-\infty$, dX = -dZ. Applying the transformation technique:

$$= \int_{0}^{-\infty} -(2\pi\lambda^2\sigma^2)^{-1/2} \exp^{-.5[(z-[\lambda\mu+1])/\lambda\sigma]^2} dz$$

$$= \int_{-\infty}^{0} (2\pi\lambda^2\sigma^2)^{-1/2} \exp^{-.5[(z-[\lambda\mu+1])/\lambda\sigma]^2} dz.$$

It follows that:

$$\int_{0}^{\infty} f_Y^0(y) = \int_{0}^{\infty} f_Z(z) + \int_{-\infty}^{0} f_Z(z) = \int_{-\infty}^{\infty} f_Z(z) = 1$$

since $f_Z(z)$ is a normal probability density function.

Appendix 2

To deal with transformations that are not one-to-one, a generalization of the standard transformation technique has to be used (Mood et al. pg. 201).

Let $Y = (\lambda G+1)^{1/\lambda} = (N)^{1/\lambda}$ or $Y = (|\lambda G+1|)^{1/\lambda} = (|N|)^{1/\lambda}$ if $1/\lambda$ is such that the first specification of the transformation is not defined for negative values of N. In both cases:

$N = -(Y^\lambda)$ if $N = \{ n:n \in N,\ n < 0 \}$.

$N = (Y^\lambda)$ if $N = \{ n:n \in N,\ n > 0 \}$.

Then, following Mood

$$f_Y^0(y) = |\lambda y^{\lambda-1}|\ f_N(y^\lambda) + |-\lambda y^{\lambda-1}|\ f_N(-(y^\lambda))$$

where $f_N(.)$ is the probability density function associated with the random variable N which is normal with mean $\lambda\mu+1$ and variance $(\lambda\sigma)^2$. It then follows that:

$$f_Y^0(y)=(y^{\lambda-1})(2\pi\sigma^2)^{-1/2}\ (\exp(-.5[(y^\lambda-[\lambda\mu+1])/\lambda\sigma]^2)$$

$$+\exp(-.5[(-[y^\lambda]-[\lambda\mu+1])/\lambda\sigma]^2))\ dy \qquad 0<y<\infty.$$

Which, after using a little algebra on the exponents becomes equation (6).

References

Bickel, P. J. and K. A. Doksum. "An Analysis of Transformations Revisited," Journal of the American Statistical Association. 76, 296-311.

Box, George E. P. and David R. Cox. "An Analysis of Transformations." Journal of the Royal Statistical Association. Ser. B, 26 (1964):211-52. 1964.

Burbidge, John B., Lonnie Magee and A. Leslie Robb. "Alternative Transformations to Handle Extreme Values of the Dependent Variable," Journal of the American Statistical Association. Vol. 83, No. 401, Theory and Methods. March 1988.

Davidson, R. and J. G. MacKinnon. "Model Specification Tests Based on Artificial Linear Regressions," International Economic Review. 25, 485-502. 1984.

Draper, Norman R. and David R. Cox. "On Distributions and Their Transformation to Normality," Journal of the Royal Statistical Society. Ser. B, 31, 472-476. 1969.

MacKinnon, J. G. and L. J. Magee. "Transforming the Dependent Variable in Regression Models." Working Paper 87-08, McMaster University, Department of Economics. 1987.

Magee, L. "The Behavior of a Modified Box-Cox Regression Model When Some Values of the Dependent Variable Are Close to Zero." Unpublished manuscript, McMaster University, Department of Economics (submitted to the Review of Economics and Statistics. 1987.

Mood, A. M., F. A. Graybill and D. C. Boes. Introduction to the Theory of Statistics. 3rd ed. New York: McGraw-Hill. 1974.

Poirier, Dale J. "The Use of the Box-Cox Transformation in Limited Dependent Variable Models," Journal of the American Statistical Association. Vol., 73, No. 362, Applications Section. June 1978.

Spitzer, J. J. "A Primer on Box-Cox Estimation," Review of Economics and Statistics. 64, 307-313. 1982.

Spitzer, John J. "A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms," Journal of Econometrics. 5, 117-128. 1977.