

## Testing the Semiparametric Box-Cox Model with the Bootstrap

N.E. Savin, Henry B. Tippie College of Business, Department of Economics,  
University of Iowa, 108 John Pappajohn Business Bldg., Iowa City, IA 52242-1000. Tel:  
(319) 335-0855, FAX (319) 335-1956, [gene-savin@uiowa.edu](mailto:gene-savin@uiowa.edu).

Allan H. Würtz, Department of Economics, University of Aarhus, Building 322,  
DK 8000 Aarhus C, Denmark and Center for Applied Microeconomics. Tel: 45-8942-  
1507; FAX:45-8613-6334 [awurtz@econ.dk](mailto:awurtz@econ.dk).

September 1, 2002

### Abstract

This paper considers tests of the transformation parameter of the Box-Cox model when the distribution of the error is unknown. Monte Carlo experiments are carried out to investigate the rejection probabilities of the GMM-based Wald and Lagrange Multiplier (LM) tests when the null hypothesis is true. The results show that the differences between empirical and nominal levels can be large when asymptotic critical values are used. In most cases, the bootstrap reduces the differences between the empirical and nominal levels, and, in many cases, essentially removes the distortions in levels that occur with asymptotic critical values. Experiments are also carried out to investigate the ability of the bootstrap to provide improved finite-sample critical values with Wald tests based on the semiparametric estimation procedure recently developed by Foster, Tian and Wei (2001).

*JEL Classification:* C13, C14

Field Designations: 17

## 1. INTRODUCTION

The Box-Cox (1964) regression model is a transformation model of the form

$$(1.1) \quad T(Y, \alpha) = X' \beta + U,$$

where  $T$  is a strictly increasing function,  $Y$  is an observed positive dependent variable,  $X$  is an observed  $K$  dimensional random column vector,  $\beta$  is a vector of constant parameters that is conformable with  $X$ , and  $U$  is an unobserved random variable that is independent of  $X$ . Let the cumulative distribution function of  $U$  be denoted by  $F$ . It is assumed that  $E(U) = 0$ ,  $V(U) < \infty$  for all  $x$  in the support of  $X$  and that  $F$  is unknown.

The Box-Cox transformation is

$$(1.2) \quad T(y, \alpha) = \begin{cases} \frac{y^\alpha - 1}{\alpha}, & \text{if } \alpha \neq 0, y \geq 0, \\ \log y, & \text{if } \alpha = 0, y \geq 0. \end{cases}$$

The transformation provides a flexible parameterization of the relation between  $Y$  and  $X$ . In particular, the model is a linear model if  $\alpha = 1$ , a power transformation model if  $\alpha \neq 0$  or 1, and a log linear model if  $\alpha = 0$ .

If  $F$  is known or known up to finite dimensional parameters, then  $\alpha$  and  $\beta$  and any parameters of  $F$  can be estimated by maximum likelihood. A widely used procedure, which was suggested by Box and Cox (1964), is to estimate  $\alpha$  and  $\beta$  by maximum likelihood (ML), assuming that  $U$  is normally distributed. The resulting estimator of  $\alpha$  and  $\beta$  is referred to as the Box-Cox ML estimator. The Box-Cox ML estimator is discussed in many econometric textbooks, for example, Amemiya (1985), Greene (2000), Mittelhammer, Judge and Miller (2000) and Ruud (2000).

The assumption of normality cannot be strictly true, however. The Box-Cox transformation  $T(y, \alpha)$  is bounded from below (above) if  $\alpha > 0$  ( $\alpha < 0$ ) unless  $\alpha$  is an odd integer or 0. Thus, the Box-Cox transformation cannot be applied to models in which the dependent variable can be negative or the distribution of  $U$  has unbounded support, and, hence, this rules out the case where  $U$  is normally distributed.

In practice, however,  $F$  is often unknown. Thus, an empirically relevant statistical problem is to obtain consistent estimators of  $\alpha$  and  $\beta$  when  $F$  is unknown. A solution proposed by Amemiya and Powell (1981) is to use the nonlinear two-stage least squares (NL2SLS) estimator of  $\alpha$  and  $\beta$ . The NL2SLS estimator is a generalized method of moments (GMM) estimator, and it is the efficient GMM estimator for the choice of instruments used by Amemiya and Powell if  $U$  is independent of  $X$ . Horowitz (1998) discusses GMM estimation of  $\alpha$  and  $\beta$ .

Khazzoom (1989) pointed out that the NL2SLS estimates for this model are ill-defined for data sets in which the dependent variable always exceeds (or is exceeded by) one. The non-negative GMM objective function has a global minimum of zero as  $\alpha$  tends to minus infinity when  $y > 1$  and infinity when  $y < 1$ . Powell (1996) has proposed a simple rescaling of the GMM objective function that helps ensure the estimates are interior points of the parameter space.

The focus of this paper is on testing the transformation parameter  $\alpha$  in the Box-Cox model when  $F$  is unknown. This null is tested using the GMM-based Wald and Lagrange Multiplier (LM) tests proposed by Newey and West (1987). The test of the null is based on an estimator of the *Type I critical value*. Horowitz and Savin (2000) define this critical value as one that would be obtained if the exact finite sample distribution of the

test statistic under the true data generation process is known. In our setting, the true Type I critical value is unknown because the null hypothesis is composite: the exact finite-sample distribution of the test statistic depends on  $\beta$  and  $F$ , population parameters not specified by the null. Thus, an approximation to the Type I critical value is required to implement the test.

An approximation to the Type I critical value can be obtained by using the first-order asymptotic distribution of the test statistic to approximate its finite-sample distribution. The approximation is useful because most test statistics in econometrics are asymptotically pivotal: their asymptotic distributions do not depend on unknown population parameters when the null hypothesis being tested is true. In particular, this is true for the GMM-based Wald and LM statistics employed to test null hypotheses about the transformation parameter. Hence, an approximate Type I critical value can be obtained from first-order asymptotic distribution theory without knowledge of the true data generation process.

However, the Monte Carlo experiments carried out in this paper show that the first-order asymptotic distribution is often a poor approximation to the true, finite-sample distributions for the sample sizes available in applications. In other contexts, many investigators have found that the asymptotic approximation for GMM based tests is poor; for example, see the 1996 special issue of the Journal of Business and Economics.

The bootstrap often provides a practical method for improving upon first -order asymptotic approximations. It is a method for estimating the distribution of a statistic or a feature of the distribution, such as a moment or a quantile. The bootstrap can be implemented for model (1.1) - (1.2) by randomly resampling  $(Y, X)$  pairs with

replacement or by randomly resampling GMM residuals with replacement, provided the bootstrap takes into account that  $Y$  can only take positive values. This paper reports the numerical performance of the bootstrap for the two resampling schemes. The Monte Carlo experiments show that when bootstrap critical values obtained by resampling residuals are used, the differences between the empirical and nominal levels of the tests are often very small.

In the context of the Box-Cox model, the linear model can be tested against models that are indexed by the transformation parameter. For example, the linear model can be tested against the log-linear model by testing the null hypothesis that  $\alpha = 1$  against the alternative  $\alpha = 0$ . The tests have to be able to discriminate between alternative models in order to be useful. This paper also carries out a Monte Carlo investigation of the powers of the tests with bootstrap critical values.

Recently, Foster, Tian and Wei (2001) have proposed an alternative to GMM estimation when  $U$  and  $X$  are independent and the distribution of  $U$  is unknown. Monte Carlo experiments are conducted to investigate the performance of Wald and LM tests when the model is estimated using the Foster, Tian and Wei (FTW) estimator. In the experiments reported here, the tests are carried out with asymptotic critical values and bootstrap critical values.

The organization of the paper is the following. Section 2 reviews the GMM estimation of the Box-Cox model, Section 3 introduces the GMM-based Wald and LM tests, Section 4 describes the calculation of the bootstrap critical values, and Section 5 presents the design of the experiments and the methods used to calculate the empirical rejection probabilities. Section 6 reports the Monte Carlo evidence on the numerical

performance of the GMM-based tests with asymptotic and bootstrap critical values, and Section 7 presents the results of the experiments using the FTW estimator. The concluding comments are contained in Section 8.

## 2. GMM ESTIMATORS

This section introduces the Box-Cox model employed in the Monte Carlo study, reviews GMM estimation of the parameters and presents the rescaling procedure proposed by Powell (1996) to address the problem noted by Khazzoom (1989).

The model simulated in the Monte Carlo experiments is

$$(2.1) \quad T(Y, \alpha) = \beta_0 + \beta_1 X + U$$

where  $X$  is a scalar random variable and  $X$  and  $U$  are independent. The instruments used are those employed by Amemiya and Powell (1981), namely,  $1, X$  and  $X^2$ . With this set of instruments, the number of moment conditions is equal to the number of the parameters, and hence parameters are exactly identified. In the exactly identified case, NL2SLS is (trivially) the efficient GMM estimator.

Denote the estimation data by  $\{Y_i, X_i: i = 1, \dots, n\}$  and assume that they are a random sample from the joint distribution of  $(Y, X)$ . Let  $\theta = (\alpha, \beta)'$ ,

$$U_i(\theta) = T(Y_i, \alpha) - X_i' \beta \text{ and } U(\theta) = (U_1(\theta), \dots, U_n(\theta))' . \text{ Also let } W = [W_1, \dots, W_n]' \text{ denote the}$$

matrix of instruments where  $W_i$  is a vector of functions of  $X_i$ . Finally, let  $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)'$

where  $\hat{\alpha}_n$  and  $\hat{\beta}_n$  denote the unconstrained GMM estimators of  $\alpha$  and  $\beta$ , respectively.

The unconstrained GMM estimator solves

$$(2.2) \quad \underset{\theta}{\text{minimize}} : S_n(\theta) = U(\theta)' W \Omega_n W' U(\theta)$$

where the weight matrix  $\Omega_n$  is a positive definite, possibly stochastic matrix. One possible choice of the weight matrix is  $\Omega_n = [W'W]^{-1}$ , in which case (2.2) gives the NL2SLS estimator of Amemiya (1974, 1985). This choice is asymptotically efficient if the errors  $U_i$  are homoskedastic. Amemiya and Powell (1981) and Amemiya (1985) discuss the use of NL2SLS for estimation of the Box-Cox model. The weight matrix does not matter in the exactly identified case, provided the sample moment conditions are solved by the unconstrained GMM estimator.

The change in the NL2SLS estimate of  $\beta$  due to a rescaling of  $X$  is the same as the change in the ordinary least squares (OLS) estimate in the linear regression model. By contrast, the effect of rescaling  $Y$  depends on whether the parameters are exactly or overidentified. In the exactly identified case, rescaling  $Y$  has no effect on the NL2SLS estimate of  $\alpha$ ; only  $\beta$  is affected. In the overidentified case, rescaling  $Y$  changes the estimates of both  $\alpha$  and  $\beta$ .

The consistency of the estimator minimizing (2.2) is established by verification of three conditions: compactness of the parameter space; convergence in probability of the objective function  $S_n$  to its expected value, uniformly in  $\alpha$  and  $\beta$ ; and uniqueness of the solutions satisfying the population moment condition  $E\{W\{T(Y, \alpha) - X'\beta\}\} = 0$ . The compactness and identification conditions turn out to be demanding due to the nature of the transformation function,  $T(Y, \alpha)$ .

As Khazzoom (1989) notes, if  $y > 1$ , then  $T(y, \alpha) \rightarrow 0$  as  $\alpha \rightarrow -\infty$ , and, similarly, if  $y < 1$ ,  $T(y, \alpha) \rightarrow 0$  as  $\alpha \rightarrow \infty$ . This implies that compactness of the parameter space plays a crucial role in the uniqueness of the solution of the population moment condition.

In particular, each residual  $U_i(\theta) = T(y_i, \alpha) - x_i'\beta$  can be set equal to 0 by setting  $\alpha = -\infty$  and  $\beta = 0$  if each  $y_i > 1$ . The resulting pathology of the objective function is important in practice since in many data sets all values of the dependent variable exceed one.

To avoid the problem associated with the scaling of the dependent variable, Powell (1996) suggested the following rescaling of the GMM objective function:

$$(2.3) \quad Q_n(\theta) = S_n(\theta) \cdot (\bar{y})^{-2\alpha},$$

where the GMM objective function  $S_n$  is given in (2.2) and  $\bar{y}$  is the geometric mean of the absolute values of the dependent variable:

$$(2.4) \quad \bar{y} \equiv \exp \left\{ \frac{1}{n} \sum_{i=1}^n \log(|y_i|) \right\}.$$

The rescaled GMM objective function  $Q_n$  is less likely than  $S_n$  to be minimized by values on the boundary of the parameter space. However, as Powell (1996) notes, rescaling the original GMM function by  $\bar{y}^{-2\alpha}$  cannot guarantee that a unique and finite minimizing value of  $\alpha$  will exist. Following Powell, the estimator based on the rescaled GMM objective function is denoted RNL2SLS.

The estimation procedure for rescaled GMM simplifies to a one-dimensional grid search and similarly for the original GMM. The objective function

$$(2.5) \quad Q_n(\theta) = [U(\theta)'W / \bar{y}^\alpha]' \Omega [W'U(\theta) / \bar{y}^\alpha]$$

can be concentrated as a function of  $\alpha$  only. The reason is that for given  $\alpha$  the optimal  $\beta$  in (2.5) is

$$(2.6) \quad \beta(\alpha) = \left[ \left( \sum_{i=1}^n W_i X_i' \right)' \Omega \left( \sum_{i=1}^n W_i X_i' \right) \right]^{-1} \left( \sum_{i=1}^n W_i X_i' \right)' \Omega \sum_{i=1}^n W_i (T(Y_i, \alpha))$$

since  $\dot{y}^{-\alpha}$  cancels. The concentrated objective function is obtained by substituting (2.6) into (2.5), which gives

$$(2.7) \quad Q_n(\alpha) = Q_n(\alpha, \beta(\alpha)) = S_n(\alpha, \beta(\alpha)) / \dot{y}^{2\alpha},$$

Note that if NL2SLS and RNL2SLS give the same estimate of  $\alpha$ , then they both give the same estimate of  $\beta$ .

Powell (1996) argues that the original and rescaled GMM estimators have the same asymptotic distribution. Hence, the standard formulae for the first-order asymptotic distribution and asymptotic covariance matrix estimators for GMM estimators apply directly to the rescaled estimators.

### 3. GMM-BASED TESTS

This section introduces the null hypotheses tested in the experiments and presents the GMM-based Wald and LM test statistics proposed by Newey and West (1987).

The null hypotheses specify the value of the transformation parameter:  $H_0: \alpha = \alpha_0$ . Two values of  $\alpha$  are considered:  $\alpha_0 = 0$  and 1. The first value specifies a log-linear model and the second a linear model.

The GMM-based Wald test statistic is derived from the asymptotic distributional properties of the unconstrained GMM estimator. Hansen (1982) showed under mild regularity conditions that  $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n')'$  is a consistent estimator of  $\theta$  and that  $\hat{\theta}_n$  is asymptotically normally distributed:

$$(3.1) \quad n^{1/2}(\hat{\theta}_n - \theta) \rightarrow^d N(0, V)$$

where

$$(3.2) \quad V = (D' \Omega D)^{-1},$$

with  $D = E \frac{\partial}{\partial \theta} W[T(Y, \alpha) - X\beta]$  and  $\Omega = p \lim_{n \rightarrow \infty} \Omega_n$ . Letting  $U_\theta = \partial U(\theta) / \partial \theta$  and,

$\hat{U}_\theta = \partial U(\hat{\theta}_n) / \partial \theta$ ,  $V$  can be estimated by replacing  $D$  in (3.2) by  $W'\hat{U}_\theta$  and  $\Omega$  by  $\Omega_n$ .

Thus, (3.1) and (3.2) with  $V$  replaced by

$$(3.3) \quad \hat{V}_n = \hat{U}_\theta' W \Omega_n W' \hat{U}_\theta$$

makes it possible to carry out inference in sufficiently large samples.

The Wald test statistic for testing  $H_0: \alpha = \alpha_0$  is

$$(3.4) \quad Wald = \frac{n(\hat{\alpha}_n - \alpha_0)^2}{\hat{s}_n^2},$$

where  $\hat{s}_n^2$  is the first diagonal element in  $\hat{V}_n$ . The Wald statistic (3.4) is distributed asymptotically as chi-square variables with one degree of freedom when the null hypothesis is true. The GMM estimators that can be used in computing (3.4) include as special cases the NL2SLS and RNL2SLS estimators. The principle disadvantage of the GMM-based Wald statistic is that it is not invariant to reparametrization of the null hypothesis or rescaling of the dependent variable. Spitzer (1984) has shown a similar lack of invariance for the Wald statistic based on the Box-Cox ML estimator; see also Drucker (2000).

Newey and West (1987) have developed an LM test based on the constrained GMM estimator. This LM test is presented in Greene (2000). Suppose the constrained estimator, denoted by  $\tilde{\theta}_n = (\tilde{a}_n, \tilde{b}_n')'$ , solves (2.2) subject to a constraint of the form  $H_0: h(\theta) = 0$ . Then the GMM-based LM statistic is

$$(3.5) \quad LM = \frac{\partial S_n(\tilde{\theta}_n)}{\partial \theta'} \left[ Var\left(\frac{\partial S_n(\tilde{\theta}_n)}{\partial \theta}\right) \right]^{-1} \frac{\partial S_n(\tilde{\theta}_n)}{\partial \theta}.$$

The LM statistic can also be written as

$$(3.6) \quad LM = n \cdot \tilde{U}' P_W \tilde{U}_\theta [\tilde{U}_\theta' P_W \tilde{U}_\theta]^{-1} \tilde{U}_\theta' P_W \tilde{U} / \tilde{U}' \tilde{U} ,$$

where  $P_W = W(W'W)^{-1}W'$ ,  $\tilde{U} = U(\tilde{\theta}_n)$  and  $\tilde{U}_\theta = \partial U(\tilde{\theta}_n) / \partial \theta$ . The (3.6) version of the LM statistic is  $n \cdot R^2$  from a regression of  $\tilde{U}$  on  $P_W \tilde{U}_\theta$ . That is, the LM statistic can be obtained from regressing  $\tilde{U}_\theta$  on  $W$ , calculating the predicted value, and then calculating  $n \cdot R^2$  from a regression of the restricted residual on these predicted values. The constrained NL2SLS and RNL2SLS estimates of  $\alpha$  are the same, and, hence, the constrained NL2SLS and RNL2SLS estimates of  $\beta$  are the same. As a result, the values of the LM statistic for NL2SLS and RNL2SLS are also the same.

Newey (2001) shows that the calculation of the GMM-based LM test statistic simplifies when the constraint imposed by the null hypothesis is  $H_0: \alpha = \alpha_0$  and  $X$  is included among the instruments. Note first that by having  $X_i$  included in  $W_i$ , the constrained estimator is  $\tilde{\theta}_n = (\alpha_0, \tilde{b}_n')'$ , where  $\tilde{\beta}$  is the OLS estimator obtained by regressing  $T(Y_i, \alpha_0)$  on  $X_i$ . Therefore the constrained residual vector  $\tilde{U}$  is just the residual vector from the OLS regression of  $T(Y_i, \alpha_0)$  on  $X_i$ . Also  $\tilde{U}_\theta = \partial U(\tilde{\theta}_n) / \partial \theta = [\tilde{T}_\alpha, -X]$ , where  $\tilde{T}_\alpha = (\partial T(y_1, \alpha_0) / \partial \alpha, \dots, \partial T(y_n, \alpha_0) / \partial \alpha)'$  and  $X = [X_1, \dots, X_n]'$ . Furthermore, if  $X_i$  is included in  $W_i$ , then  $P_W \tilde{U}_\theta = [P_X \tilde{T}_\alpha, -X]$ . Thus, the LM statistic for testing  $\alpha$  can be obtained in three steps as follows:

1. Obtain the OLS residuals from regressing  $T(Y_i, \alpha_0)$  on  $X_i$ .
2. Obtain the predicted values from regressing  $T(Y_i, \alpha_0)$  on  $W_i$ .

3. Calculate the test statistic as  $n \cdot R^2$  from regressing the residuals from step 1 on the predicted values from step 2 and the  $X_i$ 's .

The GMM-based LM statistic is invariant to reparametrization of the null hypothesis, but not always to the rescaling of the dependent variable. Invariance to rescaling depends on whether the parameters are exactly identified. The LM statistic is invariant to rescaling of the dependent variable in the exactly identified case, but not in the overidentified case.

#### 4. BOOTSTRAP CRITICAL VALUES

This section describes the Monte Carlo procedure for computing the bootstrap critical values. The description is given for two resampling schemes, a nonparametric and a parametric scheme. In the nonparametric scheme,  $(Y, X)$  pairs are randomly sampled with replacement, and in the parametric scheme, residuals are randomly sampled with replacement. Resampling the  $(Y, X)$  pairs is used only for the GMM-based Wald test.

When resampling  $(Y, X)$  pairs, the Monte Carlo procedure for computing the bootstrap critical value for the Wald test is the following:

NP1. Generate a bootstrap sample of size  $n$  by random sampling  $(Y, X)$  pairs from the estimation data with replacement.

NP2. Compute the unconstrained GMM estimators of  $\theta$  and  $V$  from the bootstrap sample. Call the results  $\hat{\theta}_n^* = (\hat{a}_n^*, \hat{b}_n^{*'})'$  and  $\hat{V}_n^*$ .

NP3. Compute the bootstrap version of the Wald statistic,

$$(4.1) \quad Wald^* = \frac{n(\hat{a}_n^* - \hat{a}_n)^2}{\hat{s}_n^{2*}}$$

where  $\hat{s}_n^{2*}$  is the first element of  $\hat{V}_n^*$ . Note that  $Wald^*$  is centered by replacing  $\alpha_0$  by  $\hat{a}_n$ .

NP4. Obtain the empirical distribution function (EDF) of the test statistic  $Wald^*$  by repeating steps NP2 and NP3 many times. The bootstrap critical value is obtained from the EDF. For example, the 0.01 Type I critical value is estimated by the 0.99 quantile of the EDF of  $Wald^*$ . Let  $z_{n,0.01}^*$  denote the bootstrap critical value for the nominal 0.01 level test. The 0.05 and 0.10 bootstrap critical values are obtained similarly.

The Monte Carlo procedure using resampled residuals is based on the functional form of the model (1.1) - (1.2). The bootstrap critical values are calculated as follows:

P1. Estimate  $\theta_0$  by constrained GMM using the estimation sample  $\{Y_i, X_i: i = 1, \dots, n\}$  and compute the constrained GMM residuals  $\tilde{U}_1, \dots, \tilde{U}_n$ .

P2. Generate the bootstrap sample by setting  $Y_i^* = [\alpha_0(\tilde{b}_{0n} + X_i^* \tilde{b}_{1n} + U_i^*) + 1]^{1/a_0}$  where  $U_i^*$  is sampled randomly with replacement from the  $\tilde{U}_i$ . The  $X_i$  are fixed in repeated samples.

P3. Estimate  $\theta_0$  by unconstrained GMM using the bootstrap sample and compute the Wald statistic

$$(4.2) \quad Wald^{**} = \frac{n(\hat{a}_n^* - \alpha_0)^2}{\hat{S}_n^{2*}}.$$

P4. Obtain the EDF of  $Wald^{**}$  by repeating steps P2 and P3 many times. The bootstrap critical values are obtained from the EDFs. The 0.01 Type I critical value is estimated by the 0.99 quantile of the EDF. The 0.05 and 0.10 bootstrap critical values are obtained similarly.

The advantage of resampling residuals compared to resampling  $(Y, X)$  pairs is numerical accuracy. Monte Carlo evidence (Horowitz (1997)) indicates that the numerical accuracy of the bootstrap tends to be much higher when residuals are

resampled than when  $(Y, X)$  pairs are resampled. This because the functional form of the model is exploited and the null hypothesis is imposed when obtaining the parameter estimate.

In this paper, there is no need to recenter the GMM moment conditions due to over-identifying restrictions. In general, when there are over-identifying restrictions, the GMM moment conditions have to be recentered for the bootstrap to improve upon first-order approximations. This is because the sample analog of  $E(WU) = 0$  is not satisfied in the estimation sample, and, hence, the bootstrap implements a moment condition that does not hold in the population that the bootstrap samples.

When resampling the  $\tilde{U}_i$ , it may happen that the value of  $Y_i^*$  cannot be calculated (the implied value is complex) due to estimation error in the  $\tilde{U}_i$ . In such cases,  $Y_i^*$  is chosen to be  $Y_i^* = 0.0001$ . This procedure is valid given that the estimation error vanishes rapidly as the sample size increases.

The Monte Carlo procedure for computing the bootstrap critical value for the LM test when resampling residuals is similar to the one used for the Wald test. The main difference occurs in step P3. For the LM test, estimate  $\theta_\theta$  in the bootstrap sample by constrained GMM and compute the LM statistic

$$(4.3) \quad LM = \frac{\partial S_n^*(\tilde{\theta}_n^*)}{\partial \theta'} \left[ Var\left(\frac{\partial S_n^*(\tilde{\theta}_n^*)}{\partial \theta}\right) \right]^{-1} \frac{\partial S_n^*(\tilde{\theta}_n^*)}{\partial \theta},$$

where  $S_n^*$  is the objective function (4.4), and  $\tilde{\theta}_n^*$  is the constrained GMM estimator of  $\theta$ .

## 5. DESIGN OF EXPERIMENTS AND COMPUTATIONS

This section presents the design of the Monte Carlo experiments used to investigate the ability of the bootstrap to reduce the distortions in the level of the Wald and LM tests that occur when asymptotic critical values are used.

Two different specifications for the distribution function of  $U$  are considered for the Box-Cox model. The first specification is a truncated normal distribution suggested by Poirier (1978). Let  $U$  be  $N(0, (0.5)^2)$  with left truncation point  $-1$ . The second is an exponential distribution for  $U$  with parameter  $\lambda = 4$ . In both specifications, the distribution of  $U$  is corrected to have mean 0; for example, in the exponential case,  $U-1/4$  is used instead of  $U$ . Foster, Tian and Wei (2001) use the exponential distribution in their Monte Carlo experiments. The values of  $X$  are obtained by random sampling the following marginal distributions of  $X$ : uniform  $[-0.5, 0.5]$ , lognormal based on  $N(0, 1)$  and exponential with  $\lambda = 1$ .

The above specifications of the distribution function of  $U$  and of  $X$  are combined to produce four basic specifications of the model. The specifications are the following:

Model 1:  $\beta_0 = 1, \beta_1 = 1, F$  truncated normal,  $\sigma = 0.5, X$  uniform  $[-0.5, 0.5]$ .

Model 2:  $\beta_0 = 0.1, \beta_1 = 1, F$  truncated normal,  $\sigma = 0.5, X$  lognormal. ( $\beta_0 = 0.1$  instead of  $\beta_0 = 0$  to avoid negative values of  $Y$ .)

Model 3:  $\beta_0 = 0, \beta_1 = 1, F$  exponential,  $X$  uniform  $[-0.5, 0.5]$ .

Model 4:  $\beta_0 = 0, \beta_1 = 1, F$  exponential,  $X$  exponential.

Using these models, tests of the null hypothesis  $H_0: \alpha = \alpha_0, \alpha_0 = 0$  and  $1$ , are conducted at three nominal levels:  $0.01, 0.05$  and  $0.10$ . The sample sizes investigated are  $n = 25$  and  $50$ .

The rejection probabilities of the tests when the null is true were estimated by conducting Monte Carlo experiments. The number of Monte Carlo replications in each experiment with the GMM estimators is 10,000. Each replication consists of the following steps:

MC1. Generate an estimation dataset of size  $n$  by random sampling from the model (1.1) - (1.2) with the null hypothesis  $H_0: \alpha = \alpha_0$  imposed. Compute the value of the Wald statistic  $Wald$  and the value of the LM statistic  $LM$  for testing  $H_0$ .

MC2. Generate a bootstrap sample of size  $n$  for the bootstrap-based test. Compute the bootstrap critical value for the Wald test using NP1- NP4 when the bootstrap samples are generated by resampling  $(Y, X)$  pairs and denote the estimated 0.01 critical value by  $z_{n,0.01}^*$ . Compute the bootstrap critical value for the Wald test using P1- P4 when the bootstrap samples are generated by resampling residuals and denote the estimated 0.01 critical value by  $z_{n,0.01}^{**}$ . In MC2 the EDF is obtained from 999 bootstrap replications.

MC3. Reject  $H_0$  at the nominal 0.01 level with the asymptotic critical value if  $Wald > 6.66$ , with the bootstrap critical value based on resampling  $(Y, X)$  pairs if  $Wald^* > z_{n,0.01}^*$  and with the bootstrap critical value based on resampling residuals if  $Wald^{**} > z_{n,0.01}^{**}$ . The rules are similar for the nominal 0.05 and 0.10 Wald tests and for the nominal 0.01,0.05 and 0.10 LM tests.

The powers of the Wald and LM tests with asymptotic and bootstrap critical values are also estimated by conducting Monte Carlo experiments. In each replication of the power experiments, the first step consists of generating the estimation dataset under the alternative hypothesis instead of under the null. The remaining steps in each replication are the same as those above.

In the experiments, the unconstrained GMM estimate is calculated by minimizing the objective function over a grid of values of the transformation parameter  $\alpha$ . To speed up the calculations the grid search is implemented in two steps. The first step is to use a coarse grid with width  $\delta$  and precision  $p$ . The precision is the distance between two neighboring points in the grid. The coarse grid is located with the true value in the middle of the grid. Suppose the minimum of the objective function is located at a point inside the coarse grid. Denoting this point by  $P$ . The second step is to construct a fine grid with a width  $\delta'$  that is two times the precision of the coarse grid,  $\delta' = 2p$  and a precision  $p'$ . The fine grid is located with the point  $P$  in the middle. Thus, the fine grid evaluates the objective function between two points in the coarse grid that are neighbors to the point  $P$ .

In some cases, the solution using the coarse grid is a point on the boundary. If this occurs, then the coarse grid is shifted to cover the interval in the neighborhood of the boundary solution. Then a new minimum is located, and the algorithm proceeds to the fine grid. The sample is discarded if the new solution is also a boundary solution. Therefore, the estimate of the rejection probability under  $H_0$  is computed as  $R/G$  where  $R$  is the number of rejections of  $H_0$  in  $G$  non-deleted estimation samples.

The total precision of this algorithm is a grid with a total width of  $3\delta$  and precision  $p'$ . This algorithm is faster than simply using a grid with a width  $3\delta$  and a precision  $p'$  because it is usually not necessary to shift the coarse grid to find the minimum. Also the fine grid is only evaluated around the minimum located with the coarse grid. This algorithm worked well for the objective function of the Box-Cox model.

## 6. EMPIRICAL REJECTION PROBABILITIES

This section reports the empirical rejection probabilities of the Wald and LM tests based on the GMM estimators. The results illustrate the numerical performance of the tests with asymptotic critical values and those with bootstrap critical values. The empirical rejection probabilities under  $H_0$  are reported for Models 1-4 in Tables I-IV, respectively.

A summary of the main features of the results is the following:

*Asymptotic Critical Values.* The results are first summarized for the Wald tests with asymptotic critical values. The tables show that the model influences the differences between the empirical and nominal levels. In Table I, the empirical levels are much smaller than the nominal levels for Model 1. By contrast, Table IV shows that the differences between the empirical and nominal levels are usually very small for Model 4.

Table II shows that the empirical levels are sensitive to the estimation method and the null hypothesis. Using NL2SLS estimation, the empirical levels are much smaller than the nominal levels when testing  $H_0: \alpha = 0$  and larger than the nominal levels when testing  $H_0: \alpha = 1$ . On the other hand, when the model is estimated by RNL2SLS, empirical levels are larger than the nominal levels.

The tables also show that the experimental evidence is mixed when comparing RNL2SLS and NL2SLS. The differences between the empirical and nominal levels of the RNL2SLS Wald tests are often, but not always, smaller than those of the NL2SLS Wald tests. Note that in Table IV the empirical rejection probabilities for the NL2SLS and RNL2SLS Wald tests with asymptotic critical values are identical when testing  $H_0: \alpha = 1$ . The explanation here is that NL2SLS and RNL2SLS give the same estimate of  $\alpha$ . When this occurs, both estimation methods produce the same estimate of  $\beta$ .

Turning to the LM tests, the differences between the empirical and nominal levels are much smaller for the LM tests than the Wald tests, both for  $n = 25$  and  $n = 50$ . Indeed, inspection of Tables I-IV shows that when  $n = 50$  the differences between the empirical and nominal levels are, almost without exception, essentially zero for all models.

A striking feature of Tables I-IV is that the empirical levels of the LM tests do not depend on the hypothesized value of  $\alpha$ . In particular, the levels for the test of  $H_0: \alpha = 0$  are identical to the levels for the test of  $H_0: \alpha = 1$ . The key to the explanation is provided by the Newey (2001) procedure for calculating the LM statistic, which is presented in Section 3. In this procedure, the LM statistic is obtained from regressing  $T(Y_i, \alpha_0)$  on  $X_i$  and  $W_i$ . In the Monte Carlo experiments, when  $H_0: \alpha = \alpha_0$  is true, the value of  $T(Y_i, \alpha_0)$  is the same no matter what the value of  $\alpha_0$ . This is because  $T(Y_i, \alpha_0) = \beta_0 + \beta_1 X_i + U_i$  and  $\beta_0 + \beta_1 X_i + U_i$  is the same independently of  $\alpha_0$ ; that is,  $\beta_0, \beta_1, X_i$ , and  $U_i$  are determined independently of  $\alpha_0$ . Even though  $Y_i$  is different for different values of  $\alpha_0$ ,  $Y_i$  only enters in the calculation of the LM statistic through  $T(Y_i, \alpha_0)$ . Hence, the results of the steps 1 and 2 of the Newey procedure do not depend on the value of  $\alpha_0$ . The same argument applies to the LM tests based on bootstrap critical values.

*Bootstrap Critical Values: (Y, X) Pairs.* The experiments investigated the ability of the bootstrap critical values obtained by resampling  $(Y, X)$  pairs to reduce the distortions in the levels of the Wald tests that occur when asymptotic critical values are used. The results are generally negative: the bootstrap based on resampling  $(Y, X)$  pairs does not reduce the distortions in the levels for most of the cases considered. Table I and Table III show that the distortions are often larger in Models 2 and 3, not smaller, when

bootstrap critical values obtained by resampling  $(Y, X)$  pairs are used. The poor numerical performance of these bootstrap critical values is disappointing, but not surprising because resampling  $(Y, X)$  pairs does not impose the null hypothesis in population that the bootstrap samples.

Monte Carlo experiments were also carried out to investigate the rejection probabilities of the NL2SLS Wald test with bootstrap critical values obtained from resampling  $(Y, X)$  pairs when  $n = 100$ . At this sample size, the distortions in the levels of the tests tended to be much reduced.

*Bootstrap Critical Values: Residuals.* The results for the Wald tests show that bootstrap critical values obtained by resampling residuals reduce, in most cases, the differences between the empirical and nominal levels that occur when asymptotic critical values are used. In some cases, however, the bootstrap does not remove the distortions. This is shown in Table I, especially for  $n = 25$ . In other cases, the bootstrap essentially eliminates the level distortions present with asymptotic critical values. For example, this is illustrated for Model 3 by the results for the NL2SLS Wald test of  $H_0: \alpha = 0$  with  $n = 50$  in Table III. Again, the experimental evidence is mixed when comparing RNL2SLS and NL2SLS.

As noted above, distortions in the levels of LM tests with asymptotic critical values, only occur when  $n = 25$ . The tabled results show that these distortions are essentially removed when the LM tests use bootstrap critical values obtained by resampling residuals.

Finally, Monte Carlo power experiments were performed to investigate the ability of the Wald tests to discriminate between alternative values of the transformation

parameter. The powers were computed for NL2SLS Wald tests with bootstrap critical values obtained by resampling residuals. The empirical powers were calculated for a 0.05 level tests of  $H_0: \alpha = 1$  when the values of  $\alpha$  are equal to 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95,  $n = 25$ . Figure 1 illustrates the empirical powers in the case of Model 2:  $\beta_0 = 0.1$ ,  $\beta_1 = 1$ ,  $F$  truncated normal,  $\sigma = .5$ ,  $X$  lognormal, and Model 3:  $\beta_0 = 0$ ,  $\beta_1 = 1$ ,  $F$  exponential,  $X$  uniform  $[-0.5, 0.5]$ .

In Figure 1, the solid line shows the empirical powers for the tests with asymptotic critical values, and the dashed line shows the empirical powers for the tests with bootstrap critical values. For Model 2, the test using asymptotic critical values appears to have a bit higher power, which is partly because the test over-rejects under the null when asymptotic critical values are used. The empirical powers for Model 3 are dramatically lower than those for Model 2. This illustrates that the experimental design can make a substantial difference as to whether the test can discriminate among alternatives. Here, the powers are lower using asymptotic critical values, in part because the test under-rejects when asymptotic critical values are used.

We conducted additional power experiments. These show that the Wald tests based on the NL2SLS and the RNL2SLS estimators have about the same power when they use bootstrap critical values. The LM tests appears to have higher power than the NL2SLS Wald test for Model 2 and lower power for Model 3, again when the tests are based on bootstrap critical values.

## 7. FTW ESTIMATOR

This section introduces the semiparametric estimator of the parameters of the Box-Cox model recently proposed by Foster, Tian and Wei (2001) (henceforth FTW))

and reports the results of a Monte Carlo investigation of the Wald test based on the FTW estimator.

FTW (2001) motivate their estimation procedure by considering the case where  $\alpha$  is known. Again let  $\{Y_i, X_i: i = 1, \dots, n\}$  be a random sample from  $(Y, X)$ . Then the least squares estimator of  $\beta$  is

$$(7.1) \quad b_n(a) = [\sum_i X_i X_i']^{-1} \sum_i X_i' T(Y_i, \alpha).$$

To estimate  $\alpha$ , consider the process  $\{I(Y_i < t), t \geq 0\}$ , where  $I(\cdot)$  is the indicator function.

The expected value of  $I(Y_i < t)$  is

$$(7.2) \quad E[I(Y_i < t)] = P(Y_i < t) = P[T(Y_i, \alpha) - X_i' \beta < (T(t, \alpha) - X_i' \beta)] = F(T(t, \alpha) - X_i' \beta).$$

A consistent estimate of the distribution function of  $U$  can be obtained using the empirical distribution of the residuals:

$$(7.3) \quad F_n(t) = \frac{1}{n} \sum_{i=1}^n [I(T(Y_i, \alpha) - X_i' b_n(\alpha)) < t].$$

The FTW estimator of  $\alpha$  solves:

$$(7.4) \quad \underset{\alpha}{\text{minimize}} S_n(\alpha) = \frac{1}{n} \sum_{i=1}^n \int_0^\infty [I(Y_i < t) - F_n(T(t, \alpha) - X_i' b_n(\alpha))]^2 dh(t),$$

where  $h(\cdot)$  is a strictly increasing deterministic weight function. The resulting estimate of  $\alpha$  is used to obtain the estimate of  $\beta$  via (7.1). A limitation of this procedure is that  $X_i' \beta$  cannot be degenerate; that is,  $\beta$  cannot be equal to 0.

FTW use the theory of U processes to show for large  $n$  that the estimator of  $\alpha$  is the unique global minimizer of  $S_n(\alpha)$  and is strongly consistent for  $\alpha_0$  and that the estimator of  $\beta$  is strongly consistent for  $\beta_0$ . The authors also show that the joint

distribution of the estimators is asymptotically normal with a finite covariance matrix that involves the unknown density function of  $F$ , which may not be well estimated by directly using nonparametric density estimation. They propose a new resampling method that they claim provides reliable estimates of the covariance matrix.

In the Monte Carlo experiments reported here, the empirical rejection probabilities of the FTW Wald tests are investigated using the four models considered previously. The tests are carried out with asymptotic critical values and with bootstrap critical values obtained by resampling residuals. The sample size in the experiments is restricted to  $n = 25$  because the simulations are very time consuming, which is a consequence of the method used by FTW to compute an estimate of the covariance matrix. For the same reason, only 500 Monte Carlo replications are used in each experiment. The bootstrap critical values are computed using 199 bootstrap replications

The Monte Carlo results are presented in Table V. The empirical levels of the FTW Wald tests based on the asymptotic critical value are larger than the nominal levels, except for Model 1. In Model 2, 3 and 4, the differences between the empirical and nominal levels are especially large when testing  $H_0: \alpha = 0$ . The evidence from these experiments suggests that there may be no advantage in using the FTW Wald test instead of the GMM Wald test when asymptotic critical values are employed. In most cases, the bootstrap reduces the distortions in the levels of the Wald tests that occur with asymptotic critical values. In the case of Model 1, 2 and 4, the bootstrap essentially removes the distortions in the levels that occur with asymptotic critical values.

The bootstrap is very time consuming because of the fact that the estimation of the covariance matrix also involves simulation. Moreover, several tuning parameters have

to be chosen to implement the simulation approach. An alternative approach is to use a non-Studentized test statistic to test the hypothesis. The non-Studentized test statistic proposed here is  $\sqrt{n}(\bar{a}_n - \alpha_0)$  where  $\bar{a}_n$  is the FTW estimator of  $\alpha$ . This statistic has the advantage that it does not require calculation of the covariance matrix. This approach is motivated by the fact that higher-order approximations to the distributions of statistics that are not asymptotically pivotal can be obtained through the use of bootstrap iteration (Beran (1988)). The idea is to obtain the critical values of the non-Studentized test using the double bootstrap. Although the double bootstrap itself is computationally intensive, it is, nevertheless, less so than bootstrapping the FTW Wald statistic.

Table V also reports the results of Monte Carlo experiments using the non-Studentized tests with single and double bootstrap critical values obtained by resampling residuals. Again, 500 Monte Carlo replications are used in each experiment. Both the single bootstrap and double bootstrap results are based on 199 bootstrap replications.

Table V shows that the differences between the empirical and nominal levels are small when critical values based on the single bootstrap are used. There are essentially no distortions in the levels of the non-Studentized tests. The exception occurs when testing  $H_0: \alpha = 0$  in Design 3. The empirical levels when the critical values are based on the double bootstrap tend to be similar to the empirical levels based on the single bootstrap. This is surprising because the single bootstrap applied to statistics that are not asymptotically pivotal does not provide higher-order approximations to their distributions. This does not imply, of course, that the single bootstrap cannot be better than the asymptotically FTW Wald test. The comparison of the numerical performance of the asymptotically FTW Wald test and the bootstrap for the non-Studentized tests

suggests that applying the bootstrap to the non-Studentized test is competitive with calculating the variance of the FTW estimator and using the asymptotically FTW Wald test.

Finally, in a small power experiment, the FTW Wald test appears to have a bit higher power than the GMM Wald tests when both tests use bootstrap critical values. This result also appears to carry over to the case of the non-Studentized test based on the FTW estimator.

## 8. CONCLUDING COMMENTS

In this section, the results are briefly reviewed and three topics are recommended for further research. The first involves the NL2SLS and RML2SLS estimators, the second the wild bootstrap with GMM-based tests, and the third is the double bootstrap with tests based on the new estimation method proposed by FTW (2001).

This study has focused on testing the transformation parameter in a Box-Cox model where  $U$  is independent of  $X$ . In a setting where  $U$  and  $X$  are independent, bootstrap critical values can be obtained by randomly resampling  $(Y, X)$  pairs with replacement or by randomly resampling residuals with replacement. The Monte Carlo experiments show that the bootstrap often essentially eliminates the level distortions that occur with asymptotic critical values when the bootstrap critical values are obtained by resampling residuals.

Two versions of the GMM estimator developed for the Box-Cox model are NL2SLS and RNL2SLS where the latter was designed to address certain shortcomings of the former. The experiments compared Wald tests based on the NL2SLS estimator with Wald tests based on the RNL2SLS estimator. The results are somewhat mixed. The

differences between the empirical and nominal levels for the RNL2SLS Wald tests are often, but not always, smaller than those for the NL2SLS Wald tests. This is true when asymptotic or bootstrap critical values are used. As a consequence, more evidence is needed to determine which estimator among the two is preferable for testing purposes.

In applications, the  $U$  may have heteroskedasticity of unknown form. In particular, the variance of  $U$  may depend on the value of  $X$ . In this situation, the bootstrap can be implemented by resampling  $(Y, X)$  pairs. However, in our experiments, the Wald and LM tests using bootstrap critical values obtained from resampling  $(Y, X)$  pairs often do not provide satisfactory control over the Type I error. An alternative to resampling  $(Y, X)$  pairs is to use the wild bootstrap. Liu (1988) introduced the wild bootstrap following a suggestion by Wu (1986). Horowitz (1997) reports the performance of the wild bootstrap in experiments using a linear regression model with heteroskedasticity of unknown form. The results show that using critical values obtained from the wild bootstrap substantially reduces the error in the rejection probability under the null hypothesis. In the case of the Box-Cox model, the wild bootstrap has the drawback that it does not constrain the value of  $Y$  to be positive. Adapting the wild bootstrap to a Box-Cox model is a topic that merit further research.

FTW (2001) have recently proposed a semiparametric estimation procedure for the Box-Cox model. A small Monte Carlo experiment was carried out to investigate the Wald test based on the FTW estimator. The results show that the differences between the empirical and the nominal levels can be large when the test uses asymptotic critical values. The bootstrap reduces, and often eliminates, the distortions that occur with asymptotic critical values. But there is a complication. The bootstrap is very time

consuming because of the fact that the estimation of the covariance matrix also involves simulation. The alternative explored here is to avoid the computation of a covariance matrix estimate by using a non-Studentized test and to obtain the critical values of the non-Studentized test by using the double bootstrap. The double bootstrap approach appears to be promising and one that merits further investigation. Indeed, what is surprising is that the empirical levels are often close to the nominal levels when the non-Studentized test uses critical values based on the single bootstrap.

### **Acknowledgements**

The authors thank Don Andrews, editor, the referees, and Joel Horowitz, George Judge, Whitney Newey and as well as the participants of the NSF Summer Symposium '01 on Identification and Inference for Econometric Models, UC Berkeley, for useful comments and suggestions. Allan Würtz acknowledges the support from the Center for Applied Microeconometrics (CAM). The activities of CAM are financed from a grant from the Danish National Research Foundation.

### **REFERENCES**

- Amemiya, T. (1974): "The Nonlinear Two-stage Least-squares Estimator," *Journal of Econometrics*, 2, 105-110.
- Amemiya, T. (1985): *Advanced Econometrics*, Cambridge: Harvard University Press.
- Amemiya, T. and J. L. Powell (1981): "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-linear Two-stage Least-squares Estimator," *Journal of Econometrics*, 17, 351-381.

Beran, R. (1998): Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements,” *Journal of the American Statistical Association*, Vol. 83, No. 403, 687-697.

Box, G.E.P. and D.R. Cox (1964): “An Analysis of Transformations,” *Journal of the Royal Statistical Society, Series B* 26, 296-311.

Drucker, D. M. (2000): “On the Manipulability of Wald tests in Box-Cox Regression Models (sg1301),” *Stata Technical Bulletin*, 54, 36-42.

Foster, A.M., L. Tian, L. and J. L. Wei (2001): “Estimation of the Box-Cox Transformation Model Without Assuming Parametric Error Distribution,” *Journal of the American Statistical Association*, 96, 1097-1101.

Greene, W. H. (2000): *Econometric Analysis, second edition*. New York: Macmillian Publishing Company.

Hall, P and J.L. Horowitz (1996): “Bootstrap Critical Values for Test Based on Generalized-Method-of-Moments Estimators,” *Econometrica*, 64,891-916.

Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029-1054.

Horowitz, J.L. (1997): “Bootstrap Methods in Econometrics,” in *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress*, ed. by D.M. Kreps and K.R.Wallis. Cambridge: Cambridge University Press, pp. 188-222.

Horowitz, J. L. (1998): *Semiparametric Methods in Econometrics*. New York: Springer-Verlag.

Horowitz, J.L. (1999): "The Bootstrap in Econometrics," in *Handbook of Econometrics, Vol 5*, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, Forthcoming.

Horowitz, J.L. and N.E. Savin (2000): "Empirically relevant critical values for hypothesis tests; a bootstrap approach," *Journal of Econometrics*, 95, 375-389.

Khazzoom, D.J. (1989): A Note on the Application of the Non-linear Two-Stage Least-squares Estimator to a Box-Cox Transformed Model," *Journal of Econometrics*, 42, 377- 379.

Liu, R. Y. (1988): "Bootstrap Procedures Under Some Non-i.i.d.Models," *Annals of Statistics*, 16, 1696-1708.

Mittlehammer, R. C., G. C. Judge, and D. J. Miller (2000): *Econometric Foundations*. Cambridge: Cambridge University Press.

Newey, W.K. and K.D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777-787.

Newey, W. K. (2001): Personal communication.

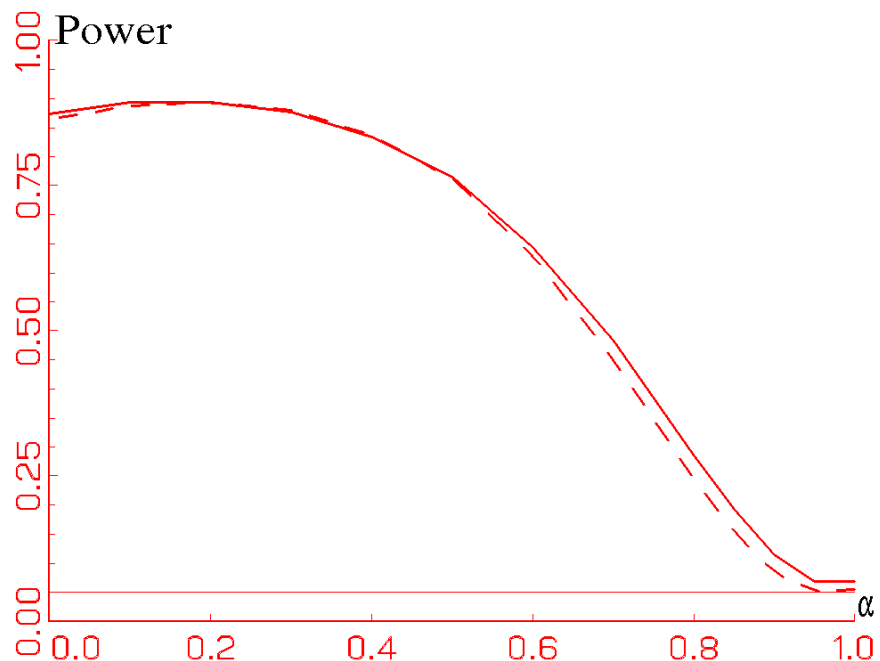
Poirier, D. J. (1978): "The Use of the Box-Cox Transformation in Limited Dependent Variable Models," *Journal of American Statistical Association*, 73, 284-287.

Powell, J. L. (1996): "Rescaling Methods-of -Moments Estimation for the Box-Cox Regression Model," *Economic Letters*, 51, 259-265.

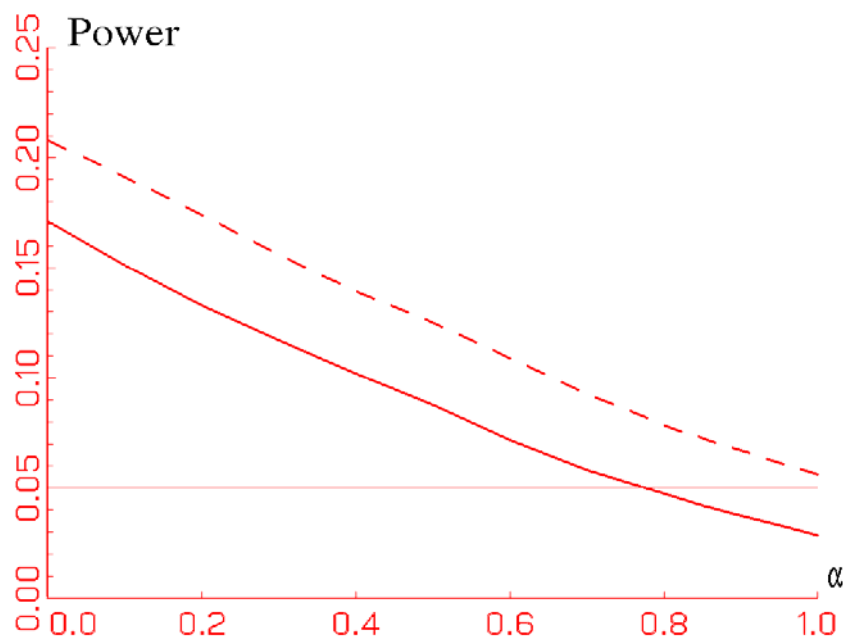
Ruud, P (2000): *Classical Econometric Theory*. New York: Oxford University Press.

Spitzer, J.J. (1984): "Variance Estimates in Models of the Box-Cox Transformation: Implications for Estimation and Hypothesis Testing," *The Review of Economics and Statistics*, 66, 645-652.

Wu, C. F. J. (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14, 1261-1295.



Model 2:  $F$  truncated normal,  $\beta_0 = 0.1$ ,  $\beta_1 = 1$ ,  $\sigma = .5$ ,  $X$  lognormal,  $n = 25$



Model 3:  $F$  exponential,  $\beta_0 = 0$ ,  $\beta_1 = 1$ ,  $X$  uniform  $[-0.5, 0.5]$ ,  $n = 25$

Figure 1.— Empirical powers of nominal 0.05 level NL2SLS-based Wald tests of  $H_0: \alpha = 1$  are given for  $\alpha$  equal to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 1.0 using asymptotic critical values (solid line) and bootstrap critical values (dashed line) obtained by randomly resampling residuals with replacement.

**TABLE I**

Empirical Rejection Probabilities (Percent) of Wald and LM Tests for Model 1:  $\beta_0 = 1$ ,  $\beta_1 = 1$ ,  $F$  truncated normal,  $\sigma = 0.5$ ,  $X$  uniform  $[-0.5, 0.5]$

Critical Values	Hypothesis	Nominal Rejection Probabilities								
		1	5	10	1	5	10	1	5	10
		Wald						LM		
		NL2SLS			RNL2SLS			NL2SLS		
n = 25										
Asymptotic										
	$\alpha = 0$	0.00	0.05	0.29	0.44	1.34	2.64	1.04	5.83	11.4
	$\alpha = 1$	0.00	0.17	0.78	1.21	3.26	5.62	1.04	5.83	11.4
Bootstrap (Y, X) Pair										
	$\alpha = 0$	0.03	0.17	0.49	0.00	0.01	0.02			
	$\alpha = 1$	0.03	0.03	1.89	0.00	0.00	0.01			
Bootstrap Residuals										
	$\alpha = 0$	0.69	3.23	7.10	0.03	3.05	6.98	1.10	5.17	9.89
	$\alpha = 1$	0.85	3.74	7.80	0.44	3.61	8.04	1.10	5.17	9.89
n = 50										
Asymptotic	$\alpha = 0$	0.00	0.13	0.70	0.16	0.90	2.16	1.02	5.23	10.4
	$\alpha = 1$	0.03	0.43	1.66	0.72	2.15	4.45	1.02	5.23	10.4
Bootstrap (Y, X) Pair										
	$\alpha = 0$	0.00	0.17	0.78	0.00	0.00	0.05			
	$\alpha = 1$	0.06	0.75	3.84	0.00	00.0	0.01			
Bootstrap Residuals										
	$\alpha = 0$	0.83	3.84	8.16	0.11	3.31	7.86	1.06	5.10	9.87
	$\alpha = 1$	0.90	4.56	9.17	0.41	3.43	8.41	1.06	5.10	9.87

The empirical rejection probabilities are computed using 10,000 Monte Carlo replications and 999 bootstrap replications. The 95 percent confidence intervals for the 0.01, 0.05 and 0.10 levels are (0.80, 1.12), (4.57, 5.43) and (9.41, 10.59), respectively; the 99 percent confidence intervals are (0.744, 1.26), (4.44, 5.56) and (9.23, 10.8), respectively.

**TABLE II**

Empirical Rejection Probabilities (Percent) of Wald and LM Tests for Model 2:  $\beta_0 = 0.1$ ,  
 $\beta_1 = 1$ ,  $F$  truncated normal,  $\sigma = 0.5$ ,  $X$  lognormal

Critical Values	Hypothesis	Nominal Rejection Probabilities								
		1	5	10	1	5	10	1	5	10
		Wald						LM		
		NL2SLS			RNL2SLS			NL2SLS		
n = 25										
Asymptotic										
	$\alpha = 0$	1.05	2.89	5.39	2.78	7.44	12.4	1.09	6.20	12.2
	$\alpha = 1$	2.38	6.87	11.6	2.43	7.01	11.8	1.09	6.20	12.2
Bootstrap ( $Y, X$ ) Pair										
	$\alpha = 0$	0.69	3.05	6.49	0.35	1.94	4.45			
	$\alpha = 1$	1.20	5.51	11.7	0.97	4.72	10.5			
Bootstrap Residuals										
	$\alpha = 0$	1.06	4.27	8.70	1.19	5.33	10.2	1.10	5.44	10.8
	$\alpha = 1$	1.19	5.45	10.2	1.16	5.41	10.2	1.10	5.44	10.8
n = 50										
Asymptotic	$\alpha = 0$	0.38	1.63	3.17	1.91	6.71	11.7	1.06	5.35	10.8
	$\alpha = 1$	1.46	5.66	10.9	1.50	5.78	11.1	1.06	5.35	10.8
Bootstrap ( $Y, X$ ) Pair										
	$\alpha = 0$	0.43	1.98	4.44	0.32	1.31	2.67			
	$\alpha = 1$	1.26	5.64	11.4	1.20	5.55	11.2			
Bootstrap Residuals										
	$\alpha = 0$	0.81	3.97	8.53	1.02	4.72	9.33	1.10	5.25	10.1
	$\alpha = 1$	1.05	4.97	10.1	1.04	4.97	10.0	1.10	5.25	10.1

See notes to Table I.

**TABLE III**  
Empirical Rejection Probabilities (Percent) of GMM Wald and LM Tests for Model 3:  
 $\beta_0 = 0, \beta_1 = 1, F$  exponential,  $X$  uniform  $[-0.5, 0.5]$

Critical Values	Hypothesis	Nominal Rejection Probabilities								
		1	5	10	1	5	10	1	5	10
		Wald						LM		
		NL2SLS			RNL2SLS			NL2SLS		
n = 25										
Asymptotic										
	$\alpha = 0$	0.31	1.86	4.32	0.27	1.63	3.95	1.15	5.31	10.6
	$\alpha = 1$	0.64	2.85	6.17	0.58	2.63	5.72	1.15	5.31	10.6
Bootstrap (Y, X) Pair*										
	$\alpha = 0$	0.32	1.67	4.44	0.03	0.40	1.91			
	$\alpha = 1$	0.44	2.51	6.51	0.01	0.61	3.28			
Bootstrap Residuals*										
	$\alpha = 0$	1.25	5.24	9.83	0.76	4.49	9.00	1.22	4.79	9.44
	$\alpha = 1$	1.41	5.62	10.3	1.26	5.14	9.75	1.22	4.79	9.44
n = 50										
Asymptotic	$\alpha = 0$	0.34	2.91	6.39	0.31	2.67	5.95	0.73	5.04	10.6
	$\alpha = 1$	0.96	4.45	8.13	0.91	4.20	7.77	0.73	5.04	10.6
Bootstrap (Y, X) Pair*										
	$\alpha = 0$	0.35	2.52	7.35	0.00	0.57	3.27			
	$\alpha = 1$	0.71	4.62	9.22	0.02	0.98	5.81			
Bootstrap* Residuals										
	$\alpha = 0$	1.09	5.40	10.4	0.77	4.85	9.82	0.84	4.84	9.95
	$\alpha = 1$	1.11	5.57	10.4	1.02	5.38	10.2	0.84	4.84	9.95

See notes to Table I.

**TABLE IV**

Empirical Rejection Probabilities (Percent) of Wald and LM Tests for Model 4:  $\beta_0 = 0$ ,  $\beta_1 = 1$ ,  $F$  exponential,  $X$  exponential

Critical Values	Hypothesis	Nominal Rejection Probabilities								
		1	5	10	1	5	10	1	5	10
		Wald						LM		
		NL2SLS			RNL2SLS			NL2SLS		
n = 25										
Asymptotic										
	$\alpha = 0$	1.89	6.41	11.3	1.95	6.65	11.7	1.58	5.89	11.2
	$\alpha = 1$	1.82	6.23	11.2	1.82	6.22	11.3	1.58	5.89	11.2
Bootstrap ( $Y, X$ ) Pair										
	$\alpha = 0$	0.75	3.61	8.39	0.73	3.53	8.24			
	$\alpha = 1$	0.55	3.50	8.12	0.48	3.21	7.77			
Bootstrap Residuals										
	$\alpha = 0$	1.26	5.03	9.72	1.26	5.01	9.74	1.47	5.23	9.85
	$\alpha = 1$	1.32	4.91	9.70	1.30	4.90	9.68	1.47	5.23	9.85
n = 50										
Asymptotic	$\alpha = 0$	1.33	5.47	10.2	1.46	5.66	10.5	1.42	5.32	10.3
	$\alpha = 1$	1.39	5.48	10.2	1.39	5.48	10.2	1.42	5.32	10.3
Bootstrap ( $Y, X$ ) Pair										
	$\alpha = 0$	0.83	4.97	10.5	0.81	5.01	10.5			
	$\alpha = 1$	0.71	4.92	10.3	0.70	4.89	10.3			
Bootstrap Residuals										
	$\alpha = 0$	1.19	4.87	9.83	1.27	4.92	9.57	1.27	4.95	10.0
	$\alpha = 1$	1.24	4.99	9.74	1.24	4.99	9.74	1.27	4.95	10.0

See notes to Table I.

**TABLE V**  
Empirical Rejection Probabilities (Percent) of FTW Tests

Form of Test and Critical Value	Nominal Rejection Probabilities					
	1	5	10	1	5	10
	Hypotheses					
	$\alpha = 0$			$\alpha = 1$		
Model 1: $\beta_0 = 1, \beta_1 = 1, F$ truncated normal, $\sigma = 0.5, X$ uniform $[-0.5, 0.5]$						
Wald						
Asymptotic	1.00	5.20	12.4	1.80	5.20	10.2
Bootstrap	1.40	5.40	10.8	1.20	5.20	10.6
Non-Studentized						
Single Bootstrap	0.80	3.40	8.20	0.40	5.62	11.4
Double Bootstrap	0.40	4.00	8.40	0.80	5.20	10.6
Model 2: $\beta_0 = 0.1, \beta_1 = 1, F$ truncated normal, $\sigma = 0.5, X$ lognormal						
Wald						
Asymptotic	9.80	17.8	23.4	3.20	7.80	15.6
Bootstrap	1.20	4.20	8.80	0.80	4.80	11.6
Non-Studentized						
Single Bootstrap	1.40	3.80	9.60	0.80	4.80	9.20
Double Bootstrap	0.80	5.20	10.6	0.80	4.20	6.80
Model 3: $\beta_0 = 0, \beta_1 = 1, F$ exponential, $X$ uniform $[-0.5, 0.5]$						
Wald						
Asymptotic	3.40	10.8	17.4	2.80	8.40	15.2
Bootstrap	2.60	7.40	12.8	2.40	7.60	12.2
Non-Studentized						
Single Bootstrap	1.20	5.80	10.6	0.40	2.80	7.00
Double Bootstrap	0.60	4.80	8.80	0.20	3.20	6.60
Model 4: $\beta_0 = 0, \beta_1 = 1, F$ exponential, $X$ exponential						
Wald						
Asymptotic	7.20	13.6	17.6	1.80	8.0	12.4
Bootstrap	0.60	3.00	9.20	1.00	5.80	10.2
Non-Studentized						
Single Bootstrap	1.40	5.00	8.40	0.80	3.60	8.80
Double Bootstrap	0.20	4.00	8.60	0.40	4.20	9.20

The empirical rejection probabilities are computed using 500 Monte Carlo replications and 199 bootstrap replications in both the single and double bootstrap. The 95 percent confidence intervals for the 0.01, 0.05 and 0.10 levels are (.013, 1.87), (3.09, 6.91) and (7.37, 12.63), respectively; the 99 percent confidence intervals are (-0.15, 2.15), (2.49, 7.51) and (6.54, 13.5), respectively.