# RESCALED METHOD-OF-MOMENTS ESTIMATION FOR THE BOX-COX REGRESSION MODEL

**James L. Powell**
**Princeton University**

**Econometric Research Program**
**Research Memorandum No. 370**

**June 1994**

**Abstract**

This note addresses a problem in the routine application of nonlinear two-stage least squares or generalized method-of-moment estimation methods to the Box-Cox regression model - namely, existence of an inconsistent minimizer at infinity when the dependent variable always exceeds (or is exceeded by) one. The proposed solution is to rescale the minimand for the estimation criterion by a power of the geometric mean of the dependent variable, which corresponds to rescaling the dependent variable by its geometric mean in a reparametrization of the model. This rescaling of the estimation criterion eliminates the root at infinity except for pathological configurations of the data, but does not affect the asymptotic distribution of a consistent root of the minimization problem.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, U.S.A.

This note suggests a simple modification to the GMM criterion to help ensure that the minimizers are interior points of the parameter space. The modification is similar in spirit to the rescaling of the Box-Cox transformation by a Jacobian term, as proposed by Box and Cox (1964) and Hinkley and Runger (1984). The suggested rescaling of the GMM criterion typically eliminates the pathological behavior of the minimization problem, but this is purely a global effect; the local properties of the consistent minimizer of the rescaled criterion are the same as for the (unscaled) GMM estimator.

## 2. The Model and Proposed Estimator

The Box-Cox regression model analyzed here is the same as was studied in Amemiya and Powell (1981); given the (p-dimensional) regression vector $x_i$ and scalar error term $\varepsilon_i$, the dependent variable $y_i$ satisfies the relation

$$z(y_i, \lambda_o) = x_i'\beta_o + \varepsilon_i, \qquad i = 1, \dots, n, \tag{2.1}$$

where $\beta_o$ and $\lambda_o$ are unknown parameters and $z(u, \lambda)$ is the Box-Cox transformation (Box and Cox (1964)), defined as

$$z(y, \lambda) = 1(\lambda \neq 0)\cdot\lambda^{-1}(y^{\lambda} - 1) + 1(\lambda = 0)\cdot\log(y) . \tag{2.2}$$

[The symbol "1(A)" denotes the indicator function of the statement "A".] Thus, the dependent variable is generated as

$$y_i = h(x_i'\beta_o + \varepsilon_i, \lambda_o) , \qquad i = 1, \dots, n, \tag{2.3}$$

where $h(\cdot)$ is the inverse transform

$$h(u, \lambda) = 1(\lambda \neq 0)\cdot(1 + \lambda u)^{1/\lambda} + 1(\lambda = 0)\cdot\exp(u) . \tag{2.4}$$

Estimation of the unknown parameters $\beta_o$ and $\lambda_o$ for this model traditionally proceeds by assuming the error terms $\varepsilon_i$ are i.i.d. and Gaussian; the conditional likelihood for the $\{y_i\}$ can then be obtained from (2.3). However, assumption of

$$m_n(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (z(y_i, \lambda) - x_i'\beta)\cdot w_i, \tag{2.6}$$

a generalized method-of-moments (GMM) estimator of $\beta_o$ and $\lambda_o$ can be defined to minimize the quadratic form

$$S_n(\beta, \lambda) \equiv [m_n(\beta, \lambda)]' A_n [m_n(\beta, \lambda)]. \tag{2.7}$$

Under suitable regularity conditions (discussed below), this estimator will be consistent if $A_n$ converges in probability to a positive definite matrix. Amemiya and Powell (1981) considered the special case $A_n = n^{-1} \sum_i w_i w_i'$, which yields the nonlinear two-stage least squares (NL2S) estimator proposed by Amemiya (1974). This choice would be appropriate if the error terms happened to be homoskedastic, but as Hansen (1982) has noted, a more efficient estimator is obtained if $A_n$ converges in probability to the inverse of the covariance matrix of $\varepsilon_i \cdot w_i$, which is not proportional to $n^{-1} \sum_i w_i w_i'$ in general.

Consistency of the estimator minimizing (2.7) is established by verification of three conditions: compactness of the parameter space; convergence in probability of the minimand $S_n$ to its expected value, uniformly in $\beta$ and $\lambda$; and uniqueness of the solutions $\beta_o$ and $\lambda_o$ satisfying the moment condition (2.5). While the uniform convergence condition can be established with relatively weak regularity conditions, the compactness and identification requirements turn out to be much more important in this case, due to a peculiarity of the transformation function $z(y, \lambda)$. As pointed out by Khazzoom (1989), if $y > 1$, $z(y, \lambda) \to 0$ as $\lambda \to -\infty$ (similarly, for $y < 1$, $z(y, \lambda) \to 0$ as $\lambda \to \infty$). This implies that compactness of the parameter space plays a crucial role in uniqueness of the solution of (2.5), since

$$\Pr\{y_i > 1\} = 1 \quad \Longrightarrow \quad \lim_{\lambda \to -\infty,\ \beta \to 0} E[z(y_i, \lambda) - x_i'\beta] = 0, \tag{2.8}$$

with an analogous result if $\Pr\{y_i < 1\} = 1$. Put differently, each residual $\varepsilon_i = z(y_i, \lambda) - x_i'\beta$ can be set to zero by setting $\lambda = -\infty$ and $\beta = 0$ if each $y_i > 1$.

This identification issue did not arise in the theoretical calculations in Amemiya

The rescaled GMM criterion function $Q_n(\cdot)$ is clearly less likely than $S_n(\cdot)$ be minimized by values on the boundary of the parameter space. If, for example, $y_i > 1$ for all $i$, the value of $S_n(\beta, \lambda)$ can be made arbitrarily small by letting $\lambda$ tend to $-\infty$; in this case, though, $\dot{y}$ also exceeds one, so the denominator of $Q_n(\cdot)$ also tends to zero as $\lambda$ tends to $-\infty$. Since $|z(y_i, \lambda)|/\dot{y}^{\lambda} \to \infty$ if either $\lambda \to \infty$ and $|y_i| > \dot{y}$ or if $\lambda \to -\infty$ and $|y_i| < \dot{y}$, it follows that $\|\dot{m}(\beta, \lambda)\| \to \infty$, and thus $Q_n(\beta, \lambda) \to \infty$, as $|\lambda| \to \infty$, as long as the regressors $x_i$ and instruments $w_i$ are sufficiently variable and the fraction of observations with $|y_i| > \dot{y}$ is not too close to either zero or one.

Unfortunately, the rescaling of the original GMM function $S_n(\beta, \lambda)$ by $\dot{y}^{-2\lambda}$ cannot guarantee that a unique and finite minimizing value $\lambda$ will exist. Consider the special case when there are no regressors (i.e., $\beta_o = 0$ is known) and (2.5) is satisfied for some scalar sequence $w_i$; that is, for some finite value of $\lambda_o$, $E[z(y_i, \lambda_o)\cdot w_i] = 0$. (For example, $y_i$ may be uniformly distributed on $(0, 2)$ and independent of $w_i$, so this moment condition will hold uniquely for $\lambda_o = 1$ if $E[w_i] \neq 0$.) In this case, the rescaled function $Q_n(0, \lambda)$ will be minimized by any $\lambda$ which solves

$$\frac{1}{n}\sum_{i=1}^{n} \frac{z(y_i, \lambda)}{\dot{y}^{\lambda}}\cdot w_i = 0. \tag{2.13}$$

However, suppose it happens that $w_i = 0$ for all observations for which $|y_i| < \dot{y}$ (which could occur, with positive probability, if $w_i$ were Bernoulli and independent of $y_i$). In this case, since $z(y_i, \lambda)/\dot{y}^{\lambda} \to 0$ as $\lambda \to \infty$ if $|y_i| < \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to \infty$; similarly, if $w_i = 0$ whenever $|y_i| > \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to -\infty$.

Though such aberrant behavior of the criterion $Q_n(\beta, \lambda)$ is possible, it only occurs for pathological configurations of the instruments $w_i$ (and, in general, of the regressors $x_i$). In the foregoing example, if the $\{y_i, i = 1, \ldots, n\}$ are distinct, which occurs with probability one if they are continuously distributed, then $Q_n(0, \lambda) \to \infty$ as $\lambda \to \infty$ $(\lambda \to -\infty)$ unless $w_i = 0$ whenever $y_i < \dot{y}$ $(y_i > \dot{y})$. While it is difficult to give more general conditions to ensure that $Q_n(\beta, \lambda) \to \infty$ as $|\lambda| \to \infty$, it seems evident that this would be virtually assured in practice.

the minimizing values $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ can be obtained by the usual GMM formula,

$$\hat{\beta} = (D_n' A_n D_n)^{-1} D_n A_n Z_n(\hat{\lambda}), \qquad \text{for} \quad Z_n(\lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} w_i z_i(y_i, \lambda). \tag{3.6}$$

Given the close relation between the original and rescaled GMM minimization problems, it is not surprising that the rescaling of the criterion does not affect the first-order asymptotic behavior of the estimators of $\lambda_o$ and $\beta_o$. Because $\partial S_n(\beta, \lambda)/\partial\beta$ is proportional to $\partial Q_n(\beta, \lambda)/\partial\beta$, the only difference in the first-order conditions for the two minimization problems appears in the condition for the transformation parameter $\lambda$, with

$$\frac{\partial Q_n(\beta, \lambda)}{\partial\lambda} = \left[ \frac{\partial S_n(\beta, \lambda)}{\partial\lambda} - 2 \ln(\dot{y}) \cdot S_n(\beta, \lambda) \right] \cdot (\dot{y})^{-2\lambda}. \tag{3.7}$$

But if $\hat{\beta}$ and $\hat{\lambda}$ are root-n-consistent estimators (which follows from imposition of the regularity conditions given in, say, Amemiya, 1974), then $S_n(\hat{\beta}, \hat{\lambda}) = O_p(n^{-1})$, since it is a quadratic form in sample moment functions (evaluated at consistent estimators) which are converging to zero at a root-n rate. Hence, when evaluated at the consistent roots,

$$\frac{\partial Q_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} = \frac{\partial S_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} + O_p(n^{-1}), \tag{3.8}$$

which implies that the (consistent) minimizers of $S_n(\cdot)$ and $Q_n(\cdot)$ have the same asymptotic distribution by the usual Taylor's series expansions. This means that the standard formulae for the asymptotic distribution and asymptotic covariance matrix estimators for GMM estimators apply directly to the minimizers of the rescaled criterion $Q_n(\beta, \lambda)$, and that any large-sample distributional formulae for unscaled GMM estimators of the Box-Cox regression model (such as those given in Amemiya and Powell, 1981) are still valid even if the rescaled criterion is used to obtain estimators which are not on the boundary of the parameter space.

# RESCALED METHOD-OF-MOMENTS ESTIMATION FOR THE BOX-COX REGRESSION MODEL

James L. Powell
Princeton University

Econometric Research Program
Research Memorandum No. 370

June 1994

**Abstract**

This note addresses a problem in the routine application of nonlinear two-stage least squares or generalized method-of-moment estimation methods to the Box-Cox regression model - namely, existence of an inconsistent minimizer at infinity when the dependent variable always exceeds (or is exceeded by) one. The proposed solution is to rescale the minimand for the estimation criterion by a power of the geometric mean of the dependent variable, which corresponds to rescaling the dependent variable by its geometric mean in a reparametrization of the model. This rescaling of the estimation criterion eliminates the root at infinity except for pathological configurations of the data, but does not affect the asymptotic distribution of a consistent root of the minimization problem.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, U.S.A.

This note suggests a simple modification to the GMM criterion to help ensure that the minimizers are interior points of the parameter space. The modification is similar in spirit to the rescaling of the Box-Cox transformation by a Jacobian term, as proposed by Box and Cox (1964) and Hinkley and Runger (1984). The suggested rescaling of the GMM criterion typically eliminates the pathological behavior of the minimization problem, but this is purely a global effect; the local properties of the consistent minimizer of the rescaled criterion are the same as for the (unscaled) GMM estimator.

## 2. The Model and Proposed Estimator

The Box-Cox regression model analyzed here is the same as was studied in Amemiya and Powell (1981); given the (p-dimensional) regression vector $x_i$ and scalar error term $\varepsilon_i$, the dependent variable $y_i$ satisfies the relation

$$z(y_i, \lambda_o) = x_i'\beta_o + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\beta_o$ and $\lambda_o$ are unknown parameters and $z(u, \lambda)$ is the Box-Cox transformation (Box and Cox (1964)), defined as

$$z(y, \lambda) = 1(\lambda \neq 0)\cdot\lambda^{-1}(y^{\lambda} - 1) + 1(\lambda = 0)\cdot\log(y) . \tag{2.2}$$

[The symbol "1(A)" denotes the indicator function of the statement "A".] Thus, the dependent variable is generated as

$$y_i = h(x_i'\beta_o + \varepsilon_i, \lambda_o) , \qquad i = 1, \ldots, n, \tag{2.3}$$

where $h(\cdot)$ is the inverse transform

$$h(u, \lambda) = 1(\lambda \neq 0)\cdot(1 + \lambda u)^{1/\lambda} + 1(\lambda = 0)\cdot\exp(u) . \tag{2.4}$$

Estimation of the unknown parameters $\beta_o$ and $\lambda_o$ for this model traditionally proceeds by assuming the error terms $\varepsilon_i$ are i.i.d. and Gaussian; the conditional likelihood for the $\{y_i\}$ can then be obtained from (2.3). However, assumption of

$$m_n(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (z(y_i, \lambda) - x_i'\beta) \cdot w_i, \tag{2.6}$$

a generalized method-of-moments (GMM) estimator of $\beta_o$ and $\lambda_o$ can be defined to minimize the quadratic form

$$S_n(\beta, \lambda) \equiv [m_n(\beta, \lambda)]' A_n [m_n(\beta, \lambda)]. \tag{2.7}$$

Under suitable regularity conditions (discussed below), this estimator will be consistent if $A_n$ converges in probability to a positive definite matrix. Amemiya and Powell (1981) considered the special case $A_n = n^{-1} \sum_i w_i w_i'$, which yields the nonlinear two-stage least squares (NL2S) estimator proposed by Amemiya (1974). This choice would be appropriate if the error terms happened to be homoskedastic, but as Hansen (1982) has noted, a more efficient estimator is obtained if $A_n$ converges in probability to the inverse of the covariance matrix of $\varepsilon_i \cdot w_i$, which is not proportional to $n^{-1} \sum_i w_i w_i'$ in general.

Consistency of the estimator minimizing (2.7) is established by verification of three conditions: compactness of the parameter space; convergence in probability of the minimand $S_n$ to its expected value, uniformly in $\beta$ and $\lambda$; and uniqueness of the solutions $\beta_o$ and $\lambda_o$ satisfying the moment condition (2.5). While the uniform convergence condition can be established with relatively weak regularity conditions, the compactness and identification requirements turn out to be much more important in this case, due to a peculiarity of the transformation function $z(y, \lambda)$. As pointed out by Khazzoom (1989), if $y > 1$, $z(y, \lambda) \to 0$ as $\lambda \to -\infty$ (similarly, for $y < 1$, $z(y, \lambda) \to 0$ as $\lambda \to \infty$). This implies that compactness of the parameter space plays a crucial role in uniqueness of the solution of (2.5), since

$$\Pr\{y_i > 1\} = 1 \quad ==> \quad \lim_{\lambda \to -\infty,\ \beta \to 0} E[z(y_i, \lambda) - x_i'\beta] = 0, \tag{2.8}$$

with an analogous result if $\Pr\{y_i < 1\} = 1$. Put differently, each residual $\varepsilon_i = z(y_i, \lambda) - x_i'\beta$ can be set to zero by setting $\lambda = -\infty$ and $\beta = 0$ if each $y_i > 1$.

This identification issue did not arise in the theoretical calculations in Amemiya

The rescaled GMM criterion function $Q_n(\cdot)$ is clearly less likely than $S_n(\cdot)$ be minimized by values on the boundary of the parameter space. If, for example, $y_i > 1$ for all $i$, the value of $S_n(\beta, \lambda)$ can be made arbitrarily small by letting $\lambda$ tend to $-\infty$; in this case, though, $\dot{y}$ also exceeds one, so the denominator of $Q_n(\cdot)$ also tends to zero as $\lambda$ tends to $-\infty$. Since $|z(y_i, \lambda)|/\dot{y}^{\lambda} \to \infty$ if either $\lambda \to \infty$ and $|y_i| > \dot{y}$ or if $\lambda \to -\infty$ and $|y_i| < \dot{y}$, it follows that $\|\dot{m}(\beta, \lambda)\| \to \infty$, and thus $Q_n(\beta, \lambda) \to \infty$, as $|\lambda| \to \infty$, as long as the regressors $x_i$ and instruments $w_i$ are sufficiently variable and the fraction of observations with $|y_i| > \dot{y}$ is not too close to either zero or one.

Unfortunately, the rescaling of the original GMM function $S_n(\beta, \lambda)$ by $\dot{y}^{-2\lambda}$ cannot guarantee that a unique and finite minimizing value $\lambda$ will exist. Consider the special case when there are no regressors (i.e., $\beta_o = 0$ is known) and (2.5) is satisfied for some scalar sequence $w_i$; that is, for some finite value of $\lambda_o$, $E[z(y_i, \lambda_o)\cdot w_i] = 0$. (For example, $y_i$ may be uniformly distributed on $(0, 2)$ and independent of $w_i$, so this moment condition will hold uniquely for $\lambda_o = 1$ if $E[w_i] \neq 0$.) In this case, the rescaled function $Q_n(0, \lambda)$ will be minimized by any $\lambda$ which solves

$$\frac{1}{n}\sum_{i=1}^{n} \frac{z(y_i, \lambda)}{\dot{y}^{\lambda}}\cdot w_i = 0. \tag{2.13}$$

However, suppose it happens that $w_i = 0$ for all observations for which $|y_i| < \dot{y}$ (which could occur, with positive probability, if $w_i$ were Bernoulli and independent of $y_i$). In this case, since $z(y_i, \lambda)/\dot{y}^{\lambda} \to 0$ as $\lambda \to \infty$ if $|y_i| < \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to \infty$; similarly, if $w_i = 0$ whenever $|y_i| > \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to -\infty$.

Though such aberrant behavior of the criterion $Q_n(\beta, \lambda)$ is possible, it only occurs for pathological configurations of the instruments $w_i$ (and, in general, of the regressors $x_i$). In the foregoing example, if the $\{y_i, i = 1, \ldots, n\}$ are distinct, which occurs with probability one if they are continuously distributed, then $Q_n(0, \lambda) \to \infty$ as $\lambda \to \infty$ $(\lambda \to -\infty)$ unless $w_i = 0$ whenever $y_i < \dot{y}$ $(y_i > \dot{y})$. While it is difficult to give more general conditions to ensure that $Q_n(\beta, \lambda) \to \infty$ as $|\lambda| \to \infty$, it seems evident that this would be virtually assured in practice.

the minimizing values $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ can be obtained by the usual GMM formula,

$$\hat{\beta} = (D_n' A_n D_n)^{-1} D_n A_n Z_n(\hat{\lambda}), \qquad \text{for} \quad Z_n(\lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} w_i z_i(y_i, \lambda). \tag{3.6}$$

Given the close relation between the original and rescaled GMM minimization problems, it is not surprising that the rescaling of the criterion does not affect the first-order asymptotic behavior of the estimators of $\lambda_o$ and $\beta_o$. Because $\partial S_n(\beta, \lambda)/\partial\beta$ is proportional to $\partial Q_n(\beta, \lambda)/\partial\beta$, the only difference in the first-order conditions for the two minimization problems appears in the condition for the transformation parameter $\lambda$, with

$$\frac{\partial Q_n(\beta, \lambda)}{\partial\lambda} = \left[ \frac{\partial S_n(\beta, \lambda)}{\partial\lambda} - 2 \ln(\dot{y}) \cdot S_n(\beta, \lambda) \right] \cdot (\dot{y})^{-2\lambda}. \tag{3.7}$$

But if $\hat{\beta}$ and $\hat{\lambda}$ are root-n-consistent estimators (which follows from imposition of the regularity conditions given in, say, Amemiya, 1974), then $S_n(\hat{\beta}, \hat{\lambda}) = O_p(n^{-1})$, since it is a quadratic form in sample moment functions (evaluated at consistent estimators) which are converging to zero at a root-n rate. Hence, when evaluated at the consistent roots,

$$\frac{\partial Q_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} = \frac{\partial S_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} + O_p(n^{-1}), \tag{3.8}$$

which implies that the (consistent) minimizers of $S_n(\cdot)$ and $Q_n(\cdot)$ have the same asymptotic distribution by the usual Taylor's series expansions. This means that the standard formulae for the asymptotic distribution and asymptotic covariance matrix estimators for GMM estimators apply directly to the minimizers of the rescaled criterion $Q_n(\beta, \lambda)$, and that any large-sample distributional formulae for unscaled GMM estimators of the Box-Cox regression model (such as those given in Amemiya and Powell, 1981) are still valid even if the rescaled criterion is used to obtain estimators which are not on the boundary of the parameter space.

# RESCALED METHOD-OF-MOMENTS ESTIMATION FOR THE BOX-COX REGRESSION MODEL

James L. Powell
Princeton University

Econometric Research Program
Research Memorandum No. 370

June 1994

## Abstract

This note addresses a problem in the routine application of nonlinear two-stage least squares or generalized method-of-moment estimation methods to the Box-Cox regression model - namely, existence of an inconsistent minimizer at infinity when the dependent variable always exceeds (or is exceeded by) one. The proposed solution is to rescale the minimand for the estimation criterion by a power of the geometric mean of the dependent variable, which corresponds to rescaling the dependent variable by its geometric mean in a reparametrization of the model. This rescaling of the estimation criterion eliminates the root at infinity except for pathological configurations of the data, but does not affect the asymptotic distribution of a consistent root of the minimization problem.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, U.S.A.

This note suggests a simple modification to the GMM criterion to help ensure that the minimizers are interior points of the parameter space. The modification is similar in spirit to the rescaling of the Box-Cox transformation by a Jacobian term, as proposed by Box and Cox (1964) and Hinkley and Runger (1984). The suggested rescaling of the GMM criterion typically eliminates the pathological behavior of the minimization problem, but this is purely a global effect; the local properties of the consistent minimizer of the rescaled criterion are the same as for the (unscaled) GMM estimator.

## 2. The Model and Proposed Estimator

The Box-Cox regression model analyzed here is the same as was studied in Amemiya and Powell (1981); given the (p-dimensional) regression vector $x_i$ and scalar error term $\varepsilon_i$, the dependent variable $y_i$ satisfies the relation

$$z(y_i, \lambda_o) = x_i'\beta_o + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\beta_o$ and $\lambda_o$ are unknown parameters and $z(u, \lambda)$ is the Box-Cox transformation (Box and Cox (1964)), defined as

$$z(y, \lambda) = 1(\lambda \neq 0)\cdot\lambda^{-1}(y^{\lambda} - 1) + 1(\lambda = 0)\cdot\log(y) . \tag{2.2}$$

[The symbol "1(A)" denotes the indicator function of the statement "A".] Thus, the dependent variable is generated as

$$y_i = h(x_i'\beta_o + \varepsilon_i, \lambda_o) , \qquad i = 1, \ldots, n, \tag{2.3}$$

where $h(\cdot)$ is the inverse transform

$$h(u, \lambda) = 1(\lambda \neq 0)\cdot(1 + \lambda u)^{1/\lambda} + 1(\lambda = 0)\cdot\exp(u) . \tag{2.4}$$

Estimation of the unknown parameters $\beta_o$ and $\lambda_o$ for this model traditionally proceeds by assuming the error terms $\varepsilon_i$ are i.i.d. and Gaussian; the conditional likelihood for the $\{y_i\}$ can then be obtained from (2.3). However, assumption of

$$m_n(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (z(y_i, \lambda) - x_i'\beta) \cdot w_i, \qquad (2.6)$$

a generalized method-of-moments (GMM) estimator of $\beta_o$ and $\lambda_o$ can be defined to minimize the quadratic form

$$S_n(\beta, \lambda) \equiv [m_n(\beta, \lambda)]' A_n [m_n(\beta, \lambda)]. \qquad (2.7)$$

Under suitable regularity conditions (discussed below), this estimator will be consistent if $A_n$ converges in probability to a positive definite matrix. Amemiya and Powell (1981) considered the special case $A_n = n^{-1} \sum_i w_i w_i'$, which yields the nonlinear two-stage least squares (NL2S) estimator proposed by Amemiya (1974). This choice would be appropriate if the error terms happened to be homoskedastic, but as Hansen (1982) has noted, a more efficient estimator is obtained if $A_n$ converges in probability to the inverse of the covariance matrix of $\varepsilon_i \cdot w_i$, which is not proportional to $n^{-1} \sum_i w_i w_i'$ in general.

Consistency of the estimator minimizing (2.7) is established by verification of three conditions: compactness of the parameter space; convergence in probability of the minimand $S_n$ to its expected value, uniformly in $\beta$ and $\lambda$; and uniqueness of the solutions $\beta_o$ and $\lambda_o$ satisfying the moment condition (2.5). While the uniform convergence condition can be established with relatively weak regularity conditions, the compactness and identification requirements turn out to be much more important in this case, due to a peculiarity of the transformation function $z(y, \lambda)$. As pointed out by Khazzoom (1989), if $y > 1$, $z(y, \lambda) \to 0$ as $\lambda \to -\infty$ (similarly, for $y < 1$, $z(y, \lambda) \to 0$ as $\lambda \to \infty$). This implies that compactness of the parameter space plays a crucial role in uniqueness of the solution of (2.5), since

$$\Pr\{y_i > 1\} = 1 \quad \Longrightarrow \quad \lim_{\lambda \to -\infty,\ \beta \to 0} E[z(y_i, \lambda) - x_i'\beta] = 0, \qquad (2.8)$$

with an analogous result if $\Pr\{y_i < 1\} = 1$. Put differently, each residual $\varepsilon_i = z(y_i, \lambda) - x_i'\beta$ can be set to zero by setting $\lambda = -\infty$ and $\beta = 0$ if each $y_i > 1$.

This identification issue did not arise in the theoretical calculations in Amemiya

The rescaled GMM criterion function $Q_n(\cdot)$ is clearly less likely than $S_n(\cdot)$ be minimized by values on the boundary of the parameter space. If, for example, $y_i > 1$ for all $i$, the value of $S_n(\beta, \lambda)$ can be made arbitrarily small by letting $\lambda$ tend to $-\infty$; in this case, though, $\dot{y}$ also exceeds one, so the denominator of $Q_n(\cdot)$ also tends to zero as $\lambda$ tends to $-\infty$. Since $|z(y_i, \lambda)|/\dot{y}^{\lambda} \to \infty$ if either $\lambda \to \infty$ and $|y_i| > \dot{y}$ or if $\lambda \to -\infty$ and $|y_i| < \dot{y}$, it follows that $\|\dot{m}(\beta, \lambda)\| \to \infty$, and thus $Q_n(\beta, \lambda) \to \infty$, as $|\lambda| \to \infty$, as long as the regressors $x_i$ and instruments $w_i$ are sufficiently variable and the fraction of observations with $|y_i| > \dot{y}$ is not too close to either zero or one.

Unfortunately, the rescaling of the original GMM function $S_n(\beta, \lambda)$ by $\dot{y}^{-2\lambda}$ cannot guarantee that a unique and finite minimizing value $\lambda$ will exist. Consider the special case when there are no regressors (i.e., $\beta_o = 0$ is known) and (2.5) is satisfied for some scalar sequence $w_i$; that is, for some finite value of $\lambda_o$, $E[z(y_i, \lambda_o)\cdot w_i] = 0$. (For example, $y_i$ may be uniformly distributed on $(0, 2)$ and independent of $w_i$, so this moment condition will hold uniquely for $\lambda_o = 1$ if $E[w_i] \neq 0$.) In this case, the rescaled function $Q_n(0, \lambda)$ will be minimized by any $\lambda$ which solves

$$\frac{1}{n}\sum_{i=1}^{n} \frac{z(y_i, \lambda)}{\dot{y}^{\lambda}}\cdot w_i = 0. \tag{2.13}$$

However, suppose it happens that $w_i = 0$ for all observations for which $|y_i| < \dot{y}$ (which could occur, with positive probability, if $w_i$ were Bernoulli and independent of $y_i$). In this case, since $z(y_i, \lambda)/\dot{y}^{\lambda} \to 0$ as $\lambda \to \infty$ if $|y_i| < \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to \infty$; similarly, if $w_i = 0$ whenever $|y_i| > \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to -\infty$.

Though such aberrant behavior of the criterion $Q_n(\beta, \lambda)$ is possible, it only occurs for pathological configurations of the instruments $w_i$ (and, in general, of the regressors $x_i$). In the foregoing example, if the $\{y_i, i = 1, \ldots, n\}$ are distinct, which occurs with probability one if they are continuously distributed, then $Q_n(0, \lambda) \to \infty$ as $\lambda \to \infty$ $(\lambda \to -\infty)$ unless $w_i = 0$ whenever $y_i < \dot{y}$ $(y_i > \dot{y})$. While it is difficult to give more general conditions to ensure that $Q_n(\beta, \lambda) \to \infty$ as $|\lambda| \to \infty$, it seems evident that this would be virtually assured in practice.

the minimizing values $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ can be obtained by the usual GMM formula,

$$\hat{\beta} = (D_n' A_n D_n)^{-1} D_n A_n Z_n(\hat{\lambda}), \qquad \text{for} \quad Z_n(\lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} w_i z_i(y_i, \lambda). \tag{3.6}$$

Given the close relation between the original and rescaled GMM minimization problems, it is not surprising that the rescaling of the criterion does not affect the first-order asymptotic behavior of the estimators of $\lambda_o$ and $\beta_o$. Because $\partial S_n(\beta, \lambda)/\partial\beta$ is proportional to $\partial Q_n(\beta, \lambda)/\partial\beta$, the only difference in the first-order conditions for the two minimization problems appears in the condition for the transformation parameter $\lambda$, with

$$\frac{\partial Q_n(\beta, \lambda)}{\partial\lambda} = \left[ \frac{\partial S_n(\beta, \lambda)}{\partial\lambda} - 2 \ln(\dot{y}) \cdot S_n(\beta, \lambda) \right] \cdot (\dot{y})^{-2\lambda}. \tag{3.7}$$

But if $\hat{\beta}$ and $\hat{\lambda}$ are root-n-consistent estimators (which follows from imposition of the regularity conditions given in, say, Amemiya, 1974), then $S_n(\hat{\beta}, \hat{\lambda}) = o_p(n^{-1})$, since it is a quadratic form in sample moment functions (evaluated at consistent estimators) which are converging to zero at a root-n rate. Hence, when evaluated at the consistent roots,

$$\frac{\partial Q_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} = \frac{\partial S_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} + o_p(n^{-1}), \tag{3.8}$$

which implies that the (consistent) minimizers of $S_n(\cdot)$ and $Q_n(\cdot)$ have the same asymptotic distribution by the usual Taylor's series expansions. This means that the standard formulae for the asymptotic distribution and asymptotic covariance matrix estimators for GMM estimators apply directly to the minimizers of the rescaled criterion $Q_n(\beta, \lambda)$, and that any large-sample distributional formulae for unscaled GMM estimators of the Box-Cox regression model (such as those given in Amemiya and Powell, 1981) are still valid even if the rescaled criterion is used to obtain estimators which are not on the boundary of the parameter space.

# RESCALED METHOD-OF-MOMENTS ESTIMATION FOR THE BOX-COX REGRESSION MODEL

James L. Powell
Princeton University

Econometric Research Program
Research Memorandum No. 370

June 1994

**Abstract**

This note addresses a problem in the routine application of nonlinear two-stage least squares or generalized method-of-moment estimation methods to the Box-Cox regression model - namely, existence of an inconsistent minimizer at infinity when the dependent variable always exceeds (or is exceeded by) one. The proposed solution is to rescale the minimand for the estimation criterion by a power of the geometric mean of the dependent variable, which corresponds to rescaling the dependent variable by its geometric mean in a reparametrization of the model. This rescaling of the estimation criterion eliminates the root at infinity except for pathological configurations of the data, but does not affect the asymptotic distribution of a consistent root of the minimization problem.

Econometric Research Program
Princeton University
203 Fisher Hall
Princeton, NJ 08544-1021, U.S.A.

This note suggests a simple modification to the GMM criterion to help ensure that the minimizers are interior points of the parameter space. The modification is similar in spirit to the rescaling of the Box-Cox transformation by a Jacobian term, as proposed by Box and Cox (1964) and Hinkley and Runger (1984). The suggested rescaling of the GMM criterion typically eliminates the pathological behavior of the minimization problem, but this is purely a global effect; the local properties of the consistent minimizer of the rescaled criterion are the same as for the (unscaled) GMM estimator.

## 2. The Model and Proposed Estimator

The Box-Cox regression model analyzed here is the same as was studied in Amemiya and Powell (1981); given the (p-dimensional) regression vector $x_i$ and scalar error term $\varepsilon_i$, the dependent variable $y_i$ satisfies the relation

$$z(y_i, \lambda_o) = x_i'\beta_o + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\beta_o$ and $\lambda_o$ are unknown parameters and $z(u, \lambda)$ is the Box-Cox transformation (Box and Cox (1964)), defined as

$$z(y, \lambda) = 1(\lambda \neq 0)\cdot\lambda^{-1}(y^{\lambda} - 1) + 1(\lambda = 0)\cdot\log(y) . \tag{2.2}$$

[The symbol "1(A)" denotes the indicator function of the statement "A".] Thus, the dependent variable is generated as

$$y_i = h(x_i'\beta_o + \varepsilon_i, \lambda_o) , \qquad i = 1, \ldots, n, \tag{2.3}$$

where $h(\cdot)$ is the inverse transform

$$h(u, \lambda) = 1(\lambda \neq 0)\cdot(1 + \lambda u)^{1/\lambda} + 1(\lambda = 0)\cdot\exp(u) . \tag{2.4}$$

Estimation of the unknown parameters $\beta_o$ and $\lambda_o$ for this model traditionally proceeds by assuming the error terms $\varepsilon_i$ are i.i.d. and Gaussian; the conditional likelihood for the $\{y_i\}$ can then be obtained from (2.3). However, assumption of

$$m_n(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} (z(y_i, \lambda) - x_i'\beta) \cdot w_i, \qquad (2.6)$$

a generalized method-of-moments (GMM) estimator of $\beta_o$ and $\lambda_o$ can be defined to minimize the quadratic form

$$S_n(\beta, \lambda) \equiv [m_n(\beta, \lambda)]' A_n [m_n(\beta, \lambda)]. \qquad (2.7)$$

Under suitable regularity conditions (discussed below), this estimator will be consistent if $A_n$ converges in probability to a positive definite matrix. Amemiya and Powell (1981) considered the special case $A_n = n^{-1} \sum_i w_i w_i'$, which yields the nonlinear two-stage least squares (NL2S) estimator proposed by Amemiya (1974). This choice would be appropriate if the error terms happened to be homoskedastic, but as Hansen (1982) has noted, a more efficient estimator is obtained if $A_n$ converges in probability to the inverse of the covariance matrix of $\varepsilon_i \cdot w_i$, which is not proportional to $n^{-1} \sum_i w_i w_i'$ in general.

Consistency of the estimator minimizing (2.7) is established by verification of three conditions: compactness of the parameter space; convergence in probability of the minimand $S_n$ to its expected value, uniformly in $\beta$ and $\lambda$; and uniqueness of the solutions $\beta_o$ and $\lambda_o$ satisfying the moment condition (2.5). While the uniform convergence condition can be established with relatively weak regularity conditions, the compactness and identification requirements turn out to be much more important in this case, due to a peculiarity of the transformation function $z(y, \lambda)$. As pointed out by Khazzoom (1989), if $y > 1$, $z(y, \lambda) \to 0$ as $\lambda \to -\infty$ (similarly, for $y < 1$, $z(y, \lambda) \to 0$ as $\lambda \to \infty$). This implies that compactness of the parameter space plays a crucial role in uniqueness of the solution of (2.5), since

$$\Pr\{y_i > 1\} = 1 \quad ==> \quad \lim_{\lambda \to -\infty,\ \beta \to 0} E[z(y_i, \lambda) - x_i'\beta] = 0, \qquad (2.8)$$

with an analogous result if $\Pr\{y_i < 1\} = 1$. Put differently, each residual $\varepsilon_i = z(y_i, \lambda) - x_i'\beta$ can be set to zero by setting $\lambda = -\infty$ and $\beta = 0$ if each $y_i > 1$.

This identification issue did not arise in the theoretical calculations in Amemiya

The rescaled GMM criterion function $Q_n(\cdot)$ is clearly less likely than $S_n(\cdot)$ be minimized by values on the boundary of the parameter space. If, for example, $y_i > 1$ for all i, the value of $S_n(\beta, \lambda)$ can be made arbitrarily small by letting $\lambda$ tend to $-\infty$; in this case, though, $\dot{y}$ also exceeds one, so the denominator of $Q_n(\cdot)$ also tends to zero as $\lambda$ tends to $-\infty$. Since $|z(y_i, \lambda)|/\dot{y}^{\lambda} \to \infty$ if either $\lambda \to \infty$ and $|y_i| > \dot{y}$ or if $\lambda \to -\infty$ and $|y_i| < \dot{y}$, it follows that $\|\dot{m}(\beta, \lambda)\| \to \infty$, and thus $Q_n(\beta, \lambda) \to \infty$, as $|\lambda| \to \infty$, as long as the regressors $x_i$ and instruments $w_i$ are sufficiently variable and the fraction of observations with $|y_i| > \dot{y}$ is not too close to either zero or one.

Unfortunately, the rescaling of the original GMM function $S_n(\beta, \lambda)$ by $\dot{y}^{-2\lambda}$ cannot guarantee that a unique and finite minimizing value $\lambda$ will exist. Consider the special case when there are no regressors (i.e., $\beta_o = 0$ is known) and (2.5) is satisfied for some scalar sequence $w_i$; that is, for some finite value of $\lambda_o$, $E[z(y_i, \lambda_o)\cdot w_i] = 0$. (For example, $y_i$ may be uniformly distributed on (0, 2) and independent of $w_i$, so this moment condition will hold uniquely for $\lambda_o = 1$ if $E[w_i] \neq 0$.) In this case, the rescaled function $Q_n(0, \lambda)$ will be minimized by any $\lambda$ which solves

$$\frac{1}{n}\sum_{i=1}^{n} \frac{z(y_i, \lambda)}{\dot{y}^{\lambda}}\cdot w_i = 0. \tag{2.13}$$

However, suppose it happens that $w_i = 0$ for all observations for which $|y_i| < \dot{y}$ (which could occur, with positive probability, if $w_i$ were Bernoulli and independent of $y_i$). In this case, since $z(y_i, \lambda)/\dot{y}^{\lambda} \to 0$ as $\lambda \to \infty$ if $|y_i| < \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to \infty$; similarly, if $w_i = 0$ whenever $|y_i| > \dot{y}$, $Q_n(0, \lambda) \to 0$ as $\lambda \to -\infty$.

Though such aberrant behavior of the criterion $Q_n(\beta, \lambda)$ is possible, it only occurs for pathological configurations of the instruments $w_i$ (and, in general, of the regressors $x_i$). In the foregoing example, if the $\{y_i, i = 1, \ldots, n\}$ are distinct, which occurs with probability one if they are continuously distributed, then $Q_n(0, \lambda) \to \infty$ as $\lambda \to \infty$ $(\lambda \to -\infty)$ unless $w_i = 0$ whenever $y_i < \dot{y}$ $(y_i > \dot{y})$. While it is difficult to give more general conditions to ensure that $Q_n(\beta, \lambda) \to \infty$ as $|\lambda| \to \infty$, it seems evident that this would be virtually assured in practice.

the minimizing values $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ can be obtained by the usual GMM formula,

$$\hat{\beta} = (D_n' A_n D_n)^{-1} D_n A_n Z_n(\hat{\lambda}), \qquad \text{for} \quad Z_n(\lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} w_i z_i(y_i, \lambda). \tag{3.6}$$

Given the close relation between the original and rescaled GMM minimization problems, it is not surprising that the rescaling of the criterion does not affect the first-order asymptotic behavior of the estimators of $\lambda_o$ and $\beta_o$. Because $\partial S_n(\beta, \lambda)/\partial\beta$ is proportional to $\partial Q_n(\beta, \lambda)/\partial\beta$, the only difference in the first-order conditions for the two minimization problems appears in the condition for the transformation parameter $\lambda$, with

$$\frac{\partial Q_n(\beta, \lambda)}{\partial\lambda} = \left[ \frac{\partial S_n(\beta, \lambda)}{\partial\lambda} - 2 \ln(\dot{y}) \cdot S_n(\beta, \lambda) \right] \cdot (\dot{y})^{-2\lambda}. \tag{3.7}$$

But if $\hat{\beta}$ and $\hat{\lambda}$ are root-n-consistent estimators (which follows from imposition of the regularity conditions given in, say, Amemiya, 1974), then $S_n(\hat{\beta}, \hat{\lambda}) = O_p(n^{-1})$, since it is a quadratic form in sample moment functions (evaluated at consistent estimators) which are converging to zero at a root-n rate. Hence, when evaluated at the consistent roots,

$$\frac{\partial Q_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} = \frac{\partial S_n(\hat{\beta}, \hat{\lambda})}{\partial\lambda} + O_p(n^{-1}), \tag{3.8}$$

which implies that the (consistent) minimizers of $S_n(\cdot)$ and $Q_n(\cdot)$ have the same asymptotic distribution by the usual Taylor's series expansions. This means that the standard formulae for the asymptotic distribution and asymptotic covariance matrix estimators for GMM estimators apply directly to the minimizers of the rescaled criterion $Q_n(\beta, \lambda)$, and that any large-sample distributional formulae for unscaled GMM estimators of the Box-Cox regression model (such as those given in Amemiya and Powell, 1981) are still valid even if the rescaled criterion is used to obtain estimators which are not on the boundary of the parameter space.