Estimation with inequality constraints on the parameters: dealing with truncation of the sampling distribution.

William A. Barnett and Ousmane Seck

University of Kansas

28. August 2008

Estimation with Inequality Constraints on the Parameters: Dealing with Truncation of the Sampling Distribution.

William A. Barnett, University of Kansas
Ousmane Seck, California State University at Fullerton

August 28, 2008

1 Introduction

Theoretical constraints on economic-model parameters often are in the form of inequality restrictions. For example, many theoretical results are in the form of monotonicity or nonnegativity restrictions. Inequality constraints can truncate sampling distributions of parameter estimators, so that asymptotic normality no longer is possible. Sampling theoretic asymptotic inference is thereby greatly complicated or compromised. We use numerical methods to investigate the resulting sampling properties of inequality-constrained estimators produced by popular methods of imposing inequality constraints. In particular, we investigate the possible bias in the asymptotic standard errors of estimators of inequality constrained estimators, when the constraint is imposed by the popular method of squaring. That approach is known to violate a regularity condition in the available asymptotic proofs regarding the unconstrained estimator, since the sign of the unconstrained estimator, prior to squaring, is nonidentified.

2 Example

As an illustration, consider this simple classical linear regression model $y_t = \beta x_t + \epsilon_t$ for $t = 1, ..., n$, where the disturbance ϵ_t is assumed to be normally distributed with mean zero at every observation. Let $\mathbf{y} = (y_1, ..., y_n)^T$, $\mathbf{x} = (x_1, ..., x_n)^T$, and $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^T$, so that the regression model can be written as $\mathbf{y} = \beta\mathbf{x} + \boldsymbol{\epsilon}$, and let the covariance matrix of $\boldsymbol{\epsilon}$ be $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix. Suppose that the unconstrained least squares estimate of the model's one parameter is $\hat{\beta} = 1$ with standard error of 2.

Suppose that prior information about the parameter is available in the form of a nonnegativity constraint. When nonnegativity is imposed, the constrained estimator would impute zero probability to the area to the left of the origin. The region not satisfying the constraint in figure 1 would be replaced by a probability mass function concentrated at zero with height 0.3015 in our example. The result is a mixed discrete-continuous distribution in the form of a truncated normal distribution. Inferences based on the standard error of the unconstrained estimator or on asymptotic normality of the constrained estimator would be compromised. The sampling distribution of the estimator, with and without inequality constraint, is displayed in figure 1. To address problems stemming from truncation of sampling distributions, different techniques have been proposed in the literature, some using the sampling theoretic approach and same using the Bayesian approach. In this paper we focus on the sampling theoretic approach and its asymptotic properties.

Figure 1: Sampling distribution of the estimator, with and without inequality constraint.

3 Sampling Theoretic Approaches

We consider the following transformation approach, widely used to impose inequality constraints in econometrics. If g is a continuous function of θ, and β is the constrained parameter, each approach acquires point estimates of β from the transformation $\beta = g(\theta)$, where g is chosen such that $g(\theta)$ satisfies the relevant inequality constraint for all unconstrained values of θ. The constrained parameter β is replaced within the regression by $\beta = g(\theta)$, and the parameter θ is estimated without constraints. The unconstrained parameter can be estimated by maximum likelihood, and the constrained parameter estimate can be recovered from the invariance property of maximum likelihood estimator[1]. No compromise in the approach to point estimation is implied by truncation of the sampling distribution, but computation of the standard error of the constrained estimator presents problems.

The " method of squaring " and the exponential functional form are two commonly used transformations, g. For example, to constrain the parameter β to be nonnegative, the "method of squaring" transformation, $\beta = \theta^2$, could be used. Then substitute θ^2 for β in the equation to be estimated and estimate θ without constraints. Alternatively an exponential transformation could be employed by defining $\beta = \exp(\theta)$ and then proceeding as in the method of squaring. This exponential transformation can be used, if β must be strictly positive. But that approach has an obvious problem when the constraint is binding, so is much less widely used than the method of squaring.

In the next three subsections, we present competing techniques for determining the standard errors of the estimates.

3.1 The Delta Method

The delta method exploits the asymptotic properties of the estimators. Under certain additional assumptions, if $\mathbf{g}(\boldsymbol{\theta})$ is a vector of continuous functions of the vector of parameters, $\boldsymbol{\theta}$, such that $\Gamma = \partial \mathbf{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T$ and if $\hat{\boldsymbol{\theta}}$ has asymptotic distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{V}$, then $\hat{\boldsymbol{\beta}} = \boldsymbol{g}(\hat{\boldsymbol{\theta}})$ has a limiting distribution, with mean $\mathbf{g}(\boldsymbol{\theta})$ and covariance matrix $\Gamma \mathbf{V} \Gamma^T$ [2].

Two problems arise when using this approach. The first is that the sample size in economic applications often is small. To avoid having our results compromised, we will increase our sample sizes sequentially to assure that small sample size is not a source of efficiency loss.

The second problem, on which we focus, is related to the choice of the functional form used for the transformation of parameters. If the function $\mathbf{g}$ is continuous but not bijective, the signs of the unconstrained parameters, $\boldsymbol{\theta}$, may be nonidentified. For example, when using the method of squaring to impose nonnegativity on $\beta_i = g_i(\theta_i)$, the estimation of $g_i(\hat{\theta}_i)$ cannot distinguish between $-\hat{\theta}_i$ and $+\hat{\theta}_i$. Hence, one of the regularity conditions is violated in the asymptotic proof with the delta method. We investigate the extent of the damage by using the delta method, when the sign of θ_i is nonidentified.

[1] The maximum likelihood estimator of $\beta = g(\theta)$ is $g(\hat{\theta}_{ML})$

[2] We use the superscript T to designate transpose of a matrix. In the case of linear least square estimation, the covariance matrix $\mathbf{V}$ is $\frac{\sigma^2}{n}\mathbf{Q}^{-1}$, where $\mathbf{Q}$ is the limit of $(\mathbf{X}^T\mathbf{X})/n$ as n goes to infinity. In nonlinear least square estimation of the model $\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{\beta}, \boldsymbol{X}) + \boldsymbol{\epsilon}$, the covariance matrix $\mathbf{V}$ is found by replacing $\mathbf{Q}$ by $\mathbf{Q}^0 = plim(\mathbf{X}^0)^T\mathbf{X}^0 = plim(\frac{1}{n}\sum(\partial \mathbf{h}(\mathbf{x}_i,\boldsymbol{\beta})/\partial\boldsymbol{\beta})(\partial \mathbf{h}(\mathbf{x}_i,\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T))$, where $\mathbf{X}$ is the matrix having as its rows the vectors $\mathbf{x}_i^T : i = 1, ..., n$

It should be observed that the delta method usually is used, with $\hat{\boldsymbol{\theta}}$ assumed to be asymptotically normal and the stronger conclusion than we use is that $\hat{\boldsymbol{\beta}} = \boldsymbol{g}(\hat{\boldsymbol{\theta}})$ is asymptotically normal. But since we are exploring the implications of truncation of the distribution of $\hat{\boldsymbol{\beta}} = \boldsymbol{g}(\hat{\boldsymbol{\theta}})$, asymptotic normality is not possible. Our concern is only with the first two moments of the limiting distribution [3].

3.2 The Jackknife

The jackknife is a resampling technique that consists in creating n samples from an observed sample of size n, by deleting one observation each time. The resulting n samples are of size $n-1$. The statistic of interest is estimated using each sample, and the n estimates are combined to obtain the mean and the standard errors. Wu (1986) refers to this approach as the delete-one jackknife. In large samples, Miller (1974) proves that this technique produces consistent results for bias and variance estimation [4].

Another jackknifing technique known as the delete-k jackknife consists in deleting an arbitrary number,k, of observations. Some method must be selected for choosing k. Wu (1985, 1986) shows that in practice, if one chooses $n-k = .72n$, where n is sample size, the delete-k jackknife possesses "nice asymptotic properties."

3.3 The Bootstrap

The bootstrap is a computer-based resampling method for assigning a measure of accuracy to a statistical estimate (Efron 1979). In regression analysis, the bootstrap method often is used to estimate finite-sample standard errors, when asymptotic standard errors are unreliable. Consider the regression equation, $\mathbf{y} = h(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is a vector of k regressors and $\boldsymbol{\theta}$ is a vector of parameters. Two frequently used methods are bootstrapping the fitted residuals or bootstrapping the pairs, $(\mathbf{X}, \mathbf{y})$, where $\mathbf{X}$ is the $n \times k$ matrix of k regressors and $\mathbf{y}$ is the n observations on the dependent variable.

Bootstrapping the residuals consists in creating J bootstrap samples,

$X_j^* = \{(\mathbf{x}_1, h(\mathbf{x}_1, \hat{\boldsymbol{\theta}} + \epsilon_{j1}^*), (\mathbf{x}_2, h(\mathbf{x}_2, \hat{\boldsymbol{\theta}}) + \epsilon_{j2}^*), ..., (\mathbf{x}_n, h(\mathbf{x}_n, \hat{\boldsymbol{\theta}}) + \epsilon_{jn}^*)\}$, for j=1,2,..., J,

where $\mathbf{x}_i$ is the i^{th} row of matrix $\mathbf{X}$ and $(\epsilon_{j1}^*, \epsilon_{j2}^*, ..., \epsilon_{jn}^*)$ are the errors drawn with replacement from the residuals during the j'th bootstrap, when estimating $\mathbf{y} = h(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$ [5].

[3] As we discuss below, problems with higher order moments are unavoidable.

[4] Wu (1986) warns about the theoretical difficulties in generating confidence intervals and in estimating variances, when the functional form is non-smooth. But all of the transformations we use in reparameterizing are smooth.

[5] This resampling method assumes that the errors are independently and identically distributed. That assumption is violated in the presence of heteroskedastic or autocorrelated errors. Extensions that correct for those problems exist. See, among others, Efron and Tibshirani (1986).

Alternatively, bootstrapping $(\mathbf{X}, \mathbf{y})$ proceeds as follows. The matrix $\mathbf{X}$ of n observations on the k exogenous variables, $\mathbf{x}$, and the vector $\mathbf{y}$ of n observations on the one endogenous variable, $\mathbf{y}$, are bootstrapped J times, creating $X_j^* = \{(y_{j1}, \mathbf{x}_{j1}), (y_{j2}, \mathbf{x}_{j2}), ..., (y_{jn}, \mathbf{x}_{jn})\}$ for j = 1, 2,..., J, where $(y_{ji}, \mathbf{x}_{ji})$ is the i^{th} draw with replacement from the original sample during the j'th bootstrap. After estimating the model on the J bootstrap samples, we obtain the bootstrap sample estimates of the parameters, $\hat{\boldsymbol{\theta}}_{\mathbf{1}}, \hat{\boldsymbol{\theta}}_{\mathbf{2}}, ..., \hat{\boldsymbol{\theta}}_{\boldsymbol{J}}$. Assuming $\hat{\boldsymbol{\theta}}_{\boldsymbol{j}} = \{\hat{\theta}_j^1, ..., \hat{\theta}_j^k\}$, then the J bootstrap replications of $\hat{\theta}^r (r \in \{1, 2, ..., k\})$ can be used to compute the estimate of the standard error, $\hat{\sigma}(\hat{\theta}^r)$ of $\hat{\theta}^r$, as follows [6]:

$$\hat{\sigma}(\hat{\theta}^r) = \sqrt{\frac{\sum_{i=1}^{J} [\hat{\theta}_i^r - \hat{\theta}^{r*}]^2}{J - 1}} \tag{1}$$

where

$$\hat{\theta}^{r*} = \frac{\sum_{i=1}^{J} \hat{\theta}_i^r}{J} \tag{2}$$

4 A Nonlinear Money Demand Function Illustration

In this section we describe a typical model having the ability to estimate the elasticity of substitution between two goods. That model will be used in the remainder of this paper to provide parameter values used as a "norm" for illustration in the figures. To conserve on journal space, we are presenting plots of results only with parameter estimates acquired from that illustration. But results with only one vector of parameter values are of limited value, without confirmation that the results are robust to the parameter value choices. In fact, we ran our Monte Carlo simulations with different values of the parameters. Since we found our results to be robust to different parameter settings, we are providing the plots only for our one (admittedly arbitrary, but currently interesting) calibrated "norm" settings of model parameters [7].

4.1 Problem Description

In producing our parameter setting norm, we decided to look at the relationship between two components of financial transactions balances. The degree of substitution among monetary assets is an important issue that has macroeconomic consequences and has been the subject of many published papers and books. The commonly published statistics on monetary aggregates use simple sum aggregation. Such summation aggregation implies that the assets are regarded by consumers as perfect substitutes. When different goods are perfect one-to-one substitutes, utility maximizers will hold the asset with the lowest opportunity cost. But investors' portfolios of monetary assets usually include a variety of assets with different opportunity costs. Hence, monetary assets are revealed not to be regarded as perfect substitutes.

[6] See Efron and Tibshirani (1993).

[7] The SAS code and outputs with other parameter settings are available upon request.

Knowledge of the elasticities of substitution among monetary assets is highly relevant to determining bias, when assets are aggregates using simple sum aggregation.

In the two-good case, the constant elasticity of substitution (CES) utility function is both flexible and globally regular. Hence, the CES is a suitable choice for our illustration.

4.2 Data Description and Model Design

Monetary Services Index (MSI) data are supplied for the United States by the Federal Reserve Bank of St. Louis on a regular basis. MSI data accurately measure the flow of monetary services received by households from monetary assets [8]. These data are based upon Divisia aggregation over highly disaggregated component data. We extract from these input data two elements between January 1992 and August 2005: the seasonally adjusted savings deposits at commercial banks net of money market deposit accounts ($q^{(1)}$) and the seasonally adjusted savings deposits at thrift institutions net of money market deposit accounts ($q^{(2)}$).

We estimate a 2-good demand function system derived from a C.E.S. utility function of the form:

$$U(q^{(1)}, q^{(2)}) = A[\alpha_1(q^{(1)})^{\rho} + \alpha_2(q^{(2)})^{\rho}]^{1/\rho} \tag{3}$$

where $\alpha_1 + \alpha_2 = 1$, $\rho < 1$, and A is a positive scalar, which can be normalized to 1. When a representative consumer is maximizing utility subject to the budget constraint, the demand function for commodity 1 can be written in budget share form as follows:

$$w_t^{(1)} = \frac{\alpha_1^{\sigma}(\pi_t^{(1)})^{1-\sigma}}{\alpha_1^{\sigma}(\pi_t^{(1)})^{1-\sigma} + \alpha_2^{\sigma}(\pi_t^{(2)})^{1-\sigma}} \tag{4}$$

where the elasticity of substitution between the two goods is σ , with $\sigma = 1/(1-\sigma)$. The constraint $\rho < 1$ implies $\sigma > 0$. The subscript t represents time, $w^{(1)}$ is the share of savings at commercial banks, and $\pi_t^{(1)}$ and $\pi_t^{(2)}$ are the user costs of savings deposits at commercial banks and at thrift institutions respectively. The formula for monetary services user costs was derived in Barnett (1978,1980). With the parameter α_2^{σ} normalized to be 1, we change the notation for α_1^{σ} to γ, leaving two parameters to be estimated: γ and σ.

4.3 Econometric Results

We employ maximum likelihood estimation of the model represented by equation 4. Since the two expenditure shares sum to one, the second equation will be omitted from the estimation and can be

[8] For details on the theory and methodology relevant to these indexes, see Barnett (1977, 1978, 1980) and Anderson, Jones and Nesmith, 1997.

recovered from equation 4. The model is estimated with an additive AR(1) error term. The parameter estimates of equation 4 with an additive autoregressive error structure are shown in table 1. Note the finding that substitution among the two assets, savings deposits at commercial banks and savings deposits at thrift institutions, is very low ($\sigma = .21$). Even though both are savings deposits, simple sum aggregation over them would not be justifiable, since the services produced by the two types of savings deposits are far from perfect substitutes. We were surprised by just how low that elasticity of substitution was for savings deposits at different institution types. In addition, since this minor step in our procedure is only to produce a calibration norm for illustration figures from our Monte Carlo experiments, we felt that such a low elasticity of substitution cannot be viewed as adequately typical. So in generating simulated data for our initial Monte Carlo experiments, we adjusted the elasticity of substitution upwards to 0.37. We round γ only slightly upwards to 2.8. The figures in this paper are conditional upon those initial calibrated settings for parameters, but the figures produced the same conclusions with other parameter settings.

σ	γ	ρ
0.21	2.728	1.004
(0.42)	(0.15)	(.002)

Table 1: Parameter estimates (Standard errors in parenthesis)

5 Monte Carlo Experiment

The two goods we simulate are assumed to be substitutable to some degree, so that the two goods (perhaps monetary assets, but only used as an illustration in the one calibrated case) are subject to the inequality constraint $\sigma > 0$. With the simulated data described below, we estimate the demand model with the simulated data subject to that inequality constraint, using the method of squaring by applying the reparameterization, $\sigma = 10^{-20} + 0.01\theta^2$, while alternatively the exponential transformation approach is implemented by applying the reparameterization, $\sigma = 0.00001 \exp \theta$. The next sections describe the data generation process and the estimation method, followed by the results. There are two objectives of our Monte Carlo experiment: (1) assess the potential damage to the asymptotic properties of $\sigma = g(\theta)$ resulting from the indeterminacy of the sign of the squared parameter θ in the method of squaring [9] and (2) determine the asymptotic properties of the constrained parameter when the jackknife and the bootstrap are used to calculate the finite sample standard errors, with sample sizes permitted to increase to large values.

[9] In this context, $g(\theta) = 10^{-20} + 0.01\theta^2$

The parameters (σ, γ) are set at various values, but since our results were robust to the setting of those parameters, we provide illustrative figures only for the case calibrated to have $(\sigma, \gamma) = (0.37, 2.8)$.

5.1 Data Generation Process

The data generation process proceeds in six steps, following the setting of the values of the parameters.

Step 1: Generate three series of 100,000 random numbers that will be the seeds for generating two user costs series and the white noise errors.

Step 2: Generate two stationary series containing S observations and representing the unit costs of two categories of assets $[\pi_t^{(1)}$ and $\pi_t^{(2)}$, $t = 1, 2, 3, ..., S]$. We generated that data from the following simple stationary specifications: $\pi_t^{(1)} = 2 + 6\nu_1$ and $\pi_t^{(2)} = 1 + 5\nu_2$, where ν_1 and ν_2 are uniformly distributed between 0 and 1 [10].

Step 3: Use equation 4 to generate a series of expenditure shares of asset 1, $w_t^{(1)}$, with the true values of the parameters set at $\sigma = 0.37$, $\gamma = 2.8$. The expenditure share of monetary asset 2 are then derived from $w_t^{(1)} + w_t^{(2)} = 1$.

Step 4: Generate a white noise error term series with mean zero and standard deviation equal to 0.04.

Step 5: Add the errors created in step 4 to the series of expenditure shares of asset 1 from step 3. The resulting realized stochastic shares are designated by *fw1*.

Step 6: The set of increasing sample sizes are chosen to be:

$\mathbf{S} \in \{30, 45, 60, 100, 200, 400, 800, 1000, 2000, 3000, 4000, ..., 100000)\}$.

5.2 Estimation Techniques

With the simulated data, maximum likelihood is used to estimate equation 4 with replaced by fw1. The positivity constraint on σ is imposed using the method of squaring with $\sigma = 10^{-20} + 0.01\theta^2$ and alternatively by using the exponential transformation, $\sigma = 0.00001 \exp \theta$. Our primary objective is to determine whether $Y = \sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ has a limiting distribution providing accurate measures of its standard deviation. Other properties of the limiting distribution are not relevant to this study, and figure 1 demonstrates that limiting normality is impossible for Y with the distribution of $g(\hat{\theta})$ being truncated at the origin.

[10] We considered using simulated autogressive price data, but the nature of those stochastic processes seems unrelated to the truncation and sign-identification issues that are our focus.

Nevertheless, it is possible that enough properties of the limiting distribution may be undamaged so that limiting normality of Y cannot be rejected empirically. Since we are only concerned with the first two moments, the unavoidable errors in the higher order moments (that do not exist with the normal distribution) need not concern us. In fact our objective is focused solely on convergence of the standard deviation, which remains possible, even if the distribution cannot converge to a limiting normal.

For every generated sample of size S, we estimate the model using the method of squaring first and then by using the exponential transformation. If the parameter estimation converges as S increases with the method of squaring, we consider the trial to be successful. This procedure is repeated 1000 times and the parameter estimates from the first 220 successful experiments are collected to compute $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$, with N being the sample size, set at the increasing values of S [11].

We first look at the evolution of the finite-sample estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$, as N diverges to infinity, since those standard deviations are the focus of this paper. If a limiting distribution exists, the variance should be stationary as the sample size increases. Although limiting normality is impossible with truncated distribution, we also compare with the known quantiles for the normal distribution. Finally, we use three normality tests: the Kolmogorov-Smirnov, the Cramer-von Mises, and the Anderson-Darling tests.

These tests are based on the empirical distribution function (EDF). The Kolmogorov-Smirnov test statistic D is based on the largest vertical difference (sup norm) between the EDF, $(F_n(x))$, and the theoretical distribution function $F(x)$ so that $D = Sup_x|F_n(x) - F(x)|$. The Anderson-Darling and the Cramer-von Mises tests use the weighted square difference as the norm. The Cramer-von Mises test weights are constant and equal to 1, while the Anderson-Darling weights are given by $F(x)(1-F(x))$ The tails are weighted more in the Anderson-Darling test than in the Kolmogorov-Smirnov or the Cramer-von Mises tests. With each of the three tests, the smaller the test statistic, the closer the empirical distribution is to the normal distribution. We cannot take seriously limiting normality with truncation, since the normal distribution has no moments higher than the second moment, while a truncated distribution does. Nevertheless, empirical inability to reject limiting normality could strengthen our ability to use the first two moments from the limiting distribution in producing asymptotic inferences, since the first two moments have particularly heavy influence on normality tests.

5.3 Estimation Results

The results about the asymptotic properties of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ are summarized in tables 2a,b and in figure 2 - 5. The method of squaring was implemented by defining $g(\theta) = 10^{-20} + 0.01\theta^2$ and the

[11] This number of replications, 1000, is arbitrary but its only importance is to guarantee that each sample of parameter estimates will have 220 observations.

exponential transformation by defining $\sigma = 0.00001e^{\theta}$ [12]. We have not attempted to weaken the existing asymptotic proofs for the delta method to permit the nonidentified sign of the unconstrained parameter estimates. But our Monte Carlo results demonstrate that the nonidentified sign does not compromise the asymptotic standard errors. It should be emphasized that the regularity assumptions in the existing proofs are sufficient but not necessary for the results on the variance of the limiting distribution.

Figure 2 exhibits the estimated standard deviation of the limiting distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ with the two reparameterizations (method of squaring and exponential transformation). These results were acquired from the delta method's asymptotic distribution theory, but with increasing simulated sample sizes. The results are almost identical, which demonstrates that the estimated asymptotic standard errors do not depend on the transformation used to impose the inequality constraint, or the nonidentification of the sign of the unconstrained parameter with the method of squaring. The exponential transformation and the method of squaring perform equally well. As the sample size increases, the estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ converges to approximately 0.42 in both cases. This convergence tends to support the use of the asymptotic theory.

The results in figure 2 are consistent with those in the first plot ($Std1$) of figure 3, which shows the directly computed finite sample estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ from the Monte Carlo simulation results. The standard error again converges to approximately 0.42 as the sample size increases. We view 0.42 thereby as being the correct limiting standard deviation against which all other computations should be compared[13].

The second and third plots (Std2 and Std3) in figure 3 show the evolution of the finite sample estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ for increasing sample size, when the bootstrap and the jackknife are utilized. The jackknifed standard deviation appears to be stationary around 0.22, which is almost half the table 1 standard deviation of the constrained estimator.

The bootstrap performs better than the jackknife, since the bootstrapped standard deviation does converge to the Std1 estimated standard deviation of the limiting distribution of Y, as the sample size increases, while the jackknifed standard deviations are consistently lower than the bootstrapped standard deviation. Figure 4 shows that this result is a consequence of the relatively small proportion, k, of jackknife observations deleted. After almost 90 percent of the sample is deleted, the jackknifed finite-sample standard deviation of Y does converge to the estimated standard deviation of the limiting distribution of

[12] As mentioned in a prior footnote above, we also ran our model with different values of the constrained parameter (elasticities of substitution), and those results are available upon request.

[13] This Delta method standard deviation converges to the table 1 standard errors of the constrained parameter, regardless of the distribution of the unconstrained parameter and regardless of whether or not the sign of the unconstrained parameter is identified. But we view this as being a coincidence. In Table 1, we are using real monetary asset user cost data, while in Figure 3, we are using simulated user cost data. Also in Figure 3, we are plotting the standard deviation of the limiting distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$, while in table 1, we provide the standard error of the estimate of $g(\theta)$.

Y. These results strongly argue against the jackknife, in such applications as consumer demand modeling, where very large sample size is the exception rather than the rule.

The bootstrap standard deviation of Y performs very similarly to the estimated standard deviation from the theoretical limiting distribution, as figure 5 shows. Not only are the two very similar to each other at all sample sizes, but converge to each other as sample size grows.

As the sample size increases, the normality of the limiting distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ from both the bootstrap and the jackknife cannot be rejected. This is despite the fact that normality is impossible, as a result of the truncation displayed in figure 1. As displayed in table 2b, we cannot reject the null hypothesis of normality at the 15 percent level with the Kolmogorov-Smirnov test and at 25 percent with both the Cramer-von Mises and the Anderson-Darling tests. In addition, as displayed in table 2a, the estimated quantiles of the normal distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ converge to the observed quantiles, as the sample size diverges to infinity. While we know that limiting normality is impossible for a truncated distribution, we are only concerned in this paper about whether or not the asymptotic theory is adequate for certain properties — in particular standard errors. Our numerical experiments demonstrate that the asymptotic theory, using the delta method, is undamaged by the truncation. Our results with tests of limiting normality suggest that there are properties of the limiting distribution that also are undamaged, at least approximately, but we do not pursue the implications for other properties of the limiting distribution. Clearly higher order limiting moments cannot be used, since the normal distribution has no moments higher than the second moment, while the truncated distribution in table 1 displays existence of higher order moments, such as skewness towards the right.

6 Conclusion

In this paper, our goal is to investigate the empirical implication of inequality constraints imposed on the parameters of a regression. In particular, we are interested in knowing the asymptotic implications of the nonidentified sign of the unconstrained parameter in the method of squaring. While that nonidentified sign violates the regularity conditions of the currently available asymptotic proofs with the delta method, we cannot rule out the possibility that the usual asymptotic properties of the constrained parameter still apply, despite the unavailability of a theoretical proof. As a result, we explore that issue using numerical Monte Carlo methods. Results with the popular method of squaring were compared to results with the exponential transformation, which violates different regularity conditions of available theoretical asymptotic proofs[14]. We find that the theoretical regularity conditions violations do not affect the usefulness of existing asymptotic theory in determining standard errors of the constrained parameter

[14] Any transformation that produces truncated sampling distribution for the transformed parameters inherently must violate the existing proofs, which produce the excessively strong result of asymptotic normality.

estimates by the delta method. In addition, the results were not sensitive to the functional form used to impose the inequality constraint. Our second result compares the estimated standard errors from the jackknife and the bootstrap. We find that the finite sample bootstrapped standard errors and the estimated standard errors from the limiting distribution of the constrained parameter estimate converge to each other. However, the finite sample jackknifed standard errors is an increasing function of the proportion of the sample deleted within that procedure. For that reason, the bootstrap dominates the jackknife, even though the finite sample jackknifed standard errors are lower than the finite sample bootstrapped standard errors.

7 References

Barnett, WA. 1977. Recursive subaggregation and a generalized hypocycloidal demand model, **Econometrica** vol. 45, issue 5, pages 1117-36

Barnett, WA. 1978. The User Cost of Money. **Economics Letters**, Vol. 1, No. 2.

Barnett, WA. 1980. Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory. **Journal of Econometrics**, September.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. **The Annals of Statistics**, 7, 1-26.

Efron, B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. **Biometrika**, 68, 589-599.

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. Society of Industrial and Applied Mathematics CBMS-NSF Monographs, 38.

Efron, B. 1985. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. **Biometrika**, 68, 589-599.

Efron, B., Tibshirani, R J. (1993). An introduction to the bootstrap. New York: Chapman and Hall.

Miller, R. G. 1974. The jackknife–A review. **Biometrika** 61:115.

Theil, H. 1971. Principles of Econometrics. North Holland, Amsterdam.

Wu, CFJ. 1985. "Statistical methods based on data resampling". Special invited paper presented at IMS meeting in Stony Brook.

Wu, CFJ. 1986. "Jackknife, bootstrap and other resampling plans in regression analysis", **Ann.Statist** 14. pp.1261-1350.

	BOOTSTRAP 100		JACKKNIFE N=100	
	-----*Quantiles*------		-----*Quantiles*------	
Percent	Observed	Observed	Estimated	Estimated
1.0	-1.65570	-0.49016	-0.554355	-1.667597
5.0	-1.25156	-0.40389	-0.391958	-1.179081
10.0	-1.02117	-0.32782	-0.305385	-0.918654
25.0	-0.43871	-0.16927	-0.160725	-0.483493
50.0	0.03750	0.00314	0.000003	0.000003
75.0	0.47197	0.18360	0.160731	0.483498
90.0	0.92255	0.29008	0.305390	0.918660
95.0	1.14766	0.33082	0.391964	1.179086
99.0	1.46774	0.54913	0.554361	1.667603

	BOOTSTRAP 30,000		JACKKNIFE N=30,000	
	-----*Quantiles*------		-----*Quantiles*------	
Percent	Observed	Observed	Estimated	Estimated
1.0	-0.08953	-0.51335	-0.51335	-0.094693
5.0	-0.06416	-0.40140	-0.40140	-0.066954
10.0	-0.05586	-0.30727	-0.30727	-0.052167
25.0	-0.03173	-0.14700	-0.14700	-0.027458
50.0	0.00216	0.01539	0.01539	-0.000004
75.0	0.02984	0.14528	0.14528	0.027450
90.0	0.05063	0.30632	0.30632	0.052159
95.0	0.06246	0.41219	0.41219	0.066947
99.0	0.09395	0.48937	0.48937	0.094686

***Table 2a*:** Normality tests for $\sqrt{N}\left[g(\hat{\theta})-Eg(\hat{\theta})\right]$, where $g(\theta)=\sigma$. Quantiles for limiting normal distribution of Y.

Sample size=100

TESTS	BOOTSTRAP		JACKKNIFE	
	--Statistic--	--p Value--	--Statistic--	--p Value--
Kolmogorov-Smirnov	D 0.0467	Pr > D > 0.15	D 0.057	Pr > D < 0.010
Cramer-von Mises	W^2 0.0704	Pr > W^2 > 0.25	W^2 0.201	Pr > W^2 < 0.005
Anderson-Darling	A^2 0.4567	Pr > A^2 > 0.25	A^2 1.28	Pr > A^2 < 0.005

Sample size=30,000

TESTS	BOOTSTRAP		JACKKNIFE	
	--Statistic--	--p Value--	--Statistic--	--p Value--
Kolmogorov-Smirnov	D 0.0368	Pr > D > 0.150	D 0.035	Pr > D > 0.150
Cramer-von Mises	W^2 0.0585	Pr > W^2 > 0.250	W^2 0.049	Pr > W^2 > 0.250
Anderson-Darling	A^2 0.3831	Pr > A^2 > 0.250	A^2 0.271	Pr > A^2 > 0.250

***Table 2b*:** Normality tests for $\sqrt{N}\left[g(\hat{\theta})-Eg(\hat{\theta})\right]$, where $g(\theta)=\sigma$. Goodness of fit tests for limiting normal distribution of Y.

Figure 2: Estimated standard deviation of the theoretical limiting distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ as the sample size, N, increases.

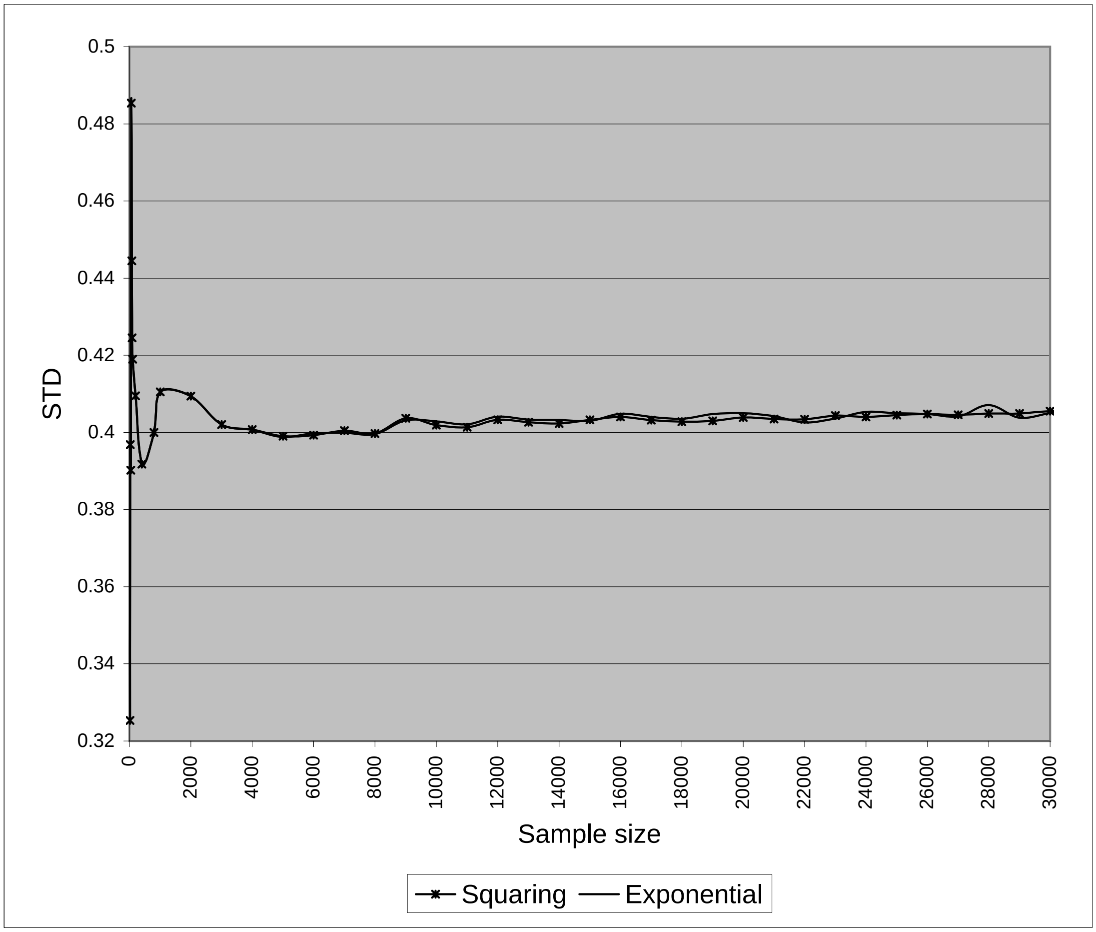

Figure 3: Finite sample estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ as the sample size, N, increases.

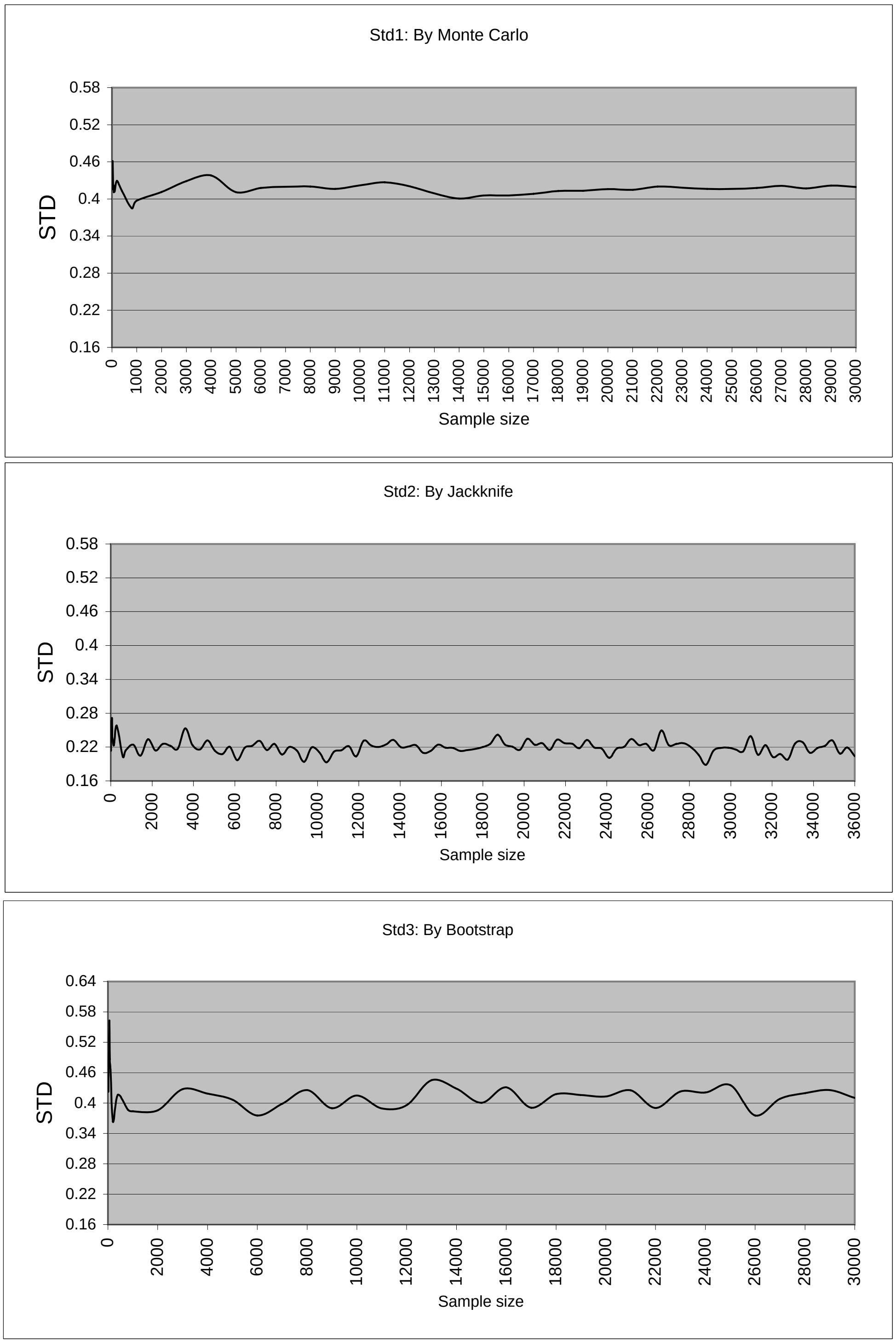

Figure 4: Finite sample estimated standard deviation of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ where $N = 800$, as the number of Jackknife replications, k, increases.

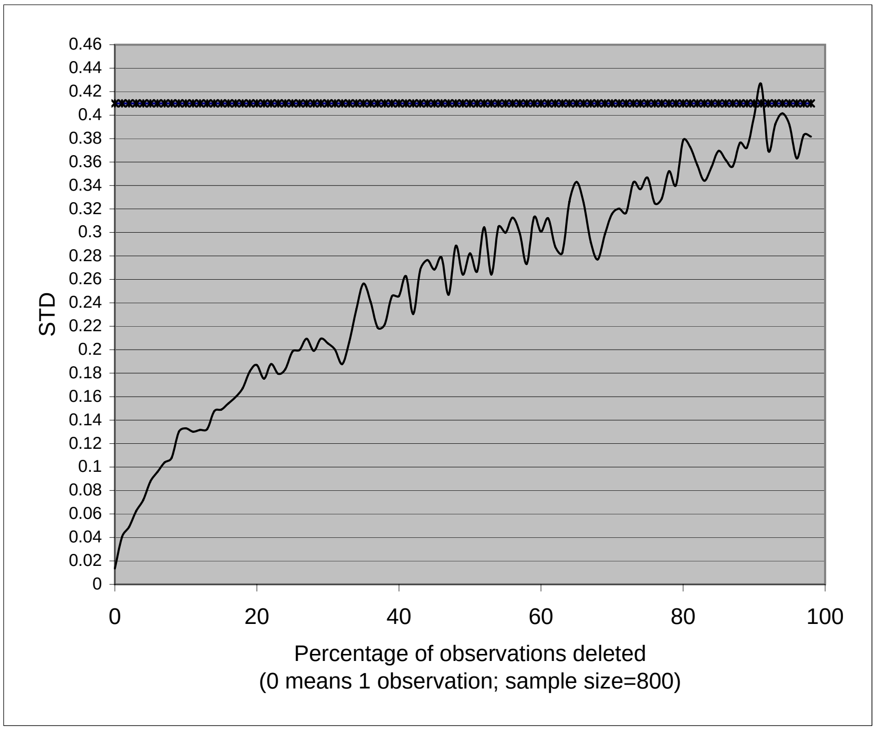

Figure 5: Bootstrapped versus asymptotic standard deviation of the limiting distribution of $\sqrt{N}[g(\hat{\theta}) - Eg(\hat{\theta})]$ as N increases to 2000.