TRUNCATED REGRESSION IN EMPIRICAL ESTIMATION

Thomas L. Marsh[a] and Ron C. Mittelhammer[b]

Paper presented at the Western Agricultural Economics Association Annual Meetings,
Vancouver, British Columbia,
June 29-July 1, 2000

[a]Assistant Professor, Department of Agricultural Economics, Kansas State University, Manhattan, KS, 66506-4011, 785-532-4913, tlmarsh@agecon.ksu.edu;

[b]Professor, Department of Agricultural Economics and Adjunct Professor, Program in Statistics, Washington State University, Pullman, WA, 99164-0621

Truncated Regression in Empirical Estimation

Abstract: In this paper we illustrate the use of alternative truncated regression estimators for the general linear model. These include variations of maximum likelihood, Bayesian, and maximum entropy estimators in which the error distributions are doubly truncated. To evaluate the performance of the estimators (e.g., efficiency) for a range of sample sizes, Monte Carlo sampling experiments are performed. We then apply each estimator to a factor demand equation for wheat-by-class.

Key Words: doubly truncated samples, Bayesian regression, maximum entropy, wheat-by-class

Introduction

In empirical applications, economists are increasingly estimating regression models that are truncated in nature.[1] Most commonly, truncated regression has been associated with singly truncated distributions (for example, see Maddala). However, doubly truncated distributions also arise in practice and are receiving attention in the economic literature (Bhattacharya, Chaturvedi, and Singh; Cohen; Maddala; Nakamura and Nakamura; Schneider). Accounting for truncated random variables in regression analysis is important because ordinary least squares estimators can be inefficient, biased, or inconsistent (Maddala) otherwise. Moreover, in many empirical modeling situations, it is imminently reasonable to assume that supports of the dependent variables and/or error terms are not unbounded in R^m space, but rather are contained in a finitely-bounded subset of real space. For example, modeling demand-share equations, as is done later in the empirical application of this paper, is a clear illustration of such a situation.

Alternative estimators have been introduced into the economic/econometric literature that allow a priori information, in various forms, to be introduced in estimation.[2] These include variations of constrained maximum likelihood, Bayesian, and maximum entropy estimators. In the Bayesian framework, parameters are treated as random variables in the sense that parameters have associated probability distributions that describes the state of knowledge about the parameters [Mittelhammer,

[1]A truncated sample is one which the values of the explanatory variables are observed only if the value of the dependent variable is observed (Judge et al.).

[2]In practice, problems arise for which prior beliefs are held about the signs, magnitudes, distributions, or ranges of plausible values for each of the unknowns. For instance, uncertainty about the value of an unknown parameter can be expressed in terms of a probability distribution using Bayesian analysis (Judge et al; Mittelhammer, Judge, and Miller).

Judge, and Miller (2000, chapters 22-24)]. In contrast, generalized maximum entropy estimators, introduced by Golan, Judge, and Miller (1996) with asymptotic property extensions by Mittelhammer and Cardell (1998), not only assume compactness of the parameter space but also finite support of the error distribution. For empirical problems with inherent uncertainties, restrictions, or truncations, the appeal of estimators such as constrained maximum likelihood, Bayesian, and maximum entropy estimators is that they offer systematic - as opposed to ad hoc - frameworks for incorporating a priori information into an econometric model.

In this paper we focus on alternative truncated regression estimators for the general linear model (GLM). Our objective is to illustrate the implementation and finite sampling properties of a range of maximum likelihood, Bayesian, and maximum entropy estimators for the case of doubly truncated error distributions. To evaluate the performance of the estimators for a range of sample sizes, Monte Carlo sampling experiments are performed. We focus on small- to- medium sized sample performance of the various estimators, and their performance relative to ordinary least squares (OLS). The performance metric used is mean square error between the empirical and true parameter values. In performing comparisons, we examine explicitly the potential benefits of truncating error distributions and imposing related parameter restrictions (i.e., gains in econometric efficiency) and the econometric costs of imposing such constraints (i.e., biasedness).

To complete the paper we apply each estimator to a factor demand equation for wheat-by-class expressed in share form. We illustrate issues that arise in empirical estimation contexts, including setting parameter restrictions and the flexibility of doing so across models. Finally, we note practical problems in modeling truncating error distributions, and make recommendations for additional research.

The Econometric Estimators

Consider the (GLM) with N observations $y_i = x_{i\cdot} b + u_i$, $i = 1,..., N$. In this equation y_i is the i^{th} observation on the dependent variable, $x_{i\cdot}$ is a ($1\times K$) row vector of explanatory values for each observation i, b is a ($K\times 1$) column vector of parameters, and u_i is the model residual associated with observation i.

Three estimators for the GLM that account explicitly for the truncated nature of the dependent variable are considered and compared to ordinary least squares (OLS). The first estimator considered is the standard regression model where the error distribution is doubly truncated, referred to as the truncated regression model (TRM). The TRM is estimated in a standard manner using constrained maximum likelihood based on a truncated likelihood function. The second estimator considered is a variation of the standard Bayesian estimator. It assumes not only prior information on the parameters in the usual Bayesian way, but also assumes that the error distribution is doubly truncated via the likelihood function of the standard regression model. We emphasize that the parameter space includes two truncation parameters indicating the points at which the error distribution is truncated from above and below. Since it would be expected that there would be uncertainty regarding the values of the truncation parameters in any given empirical application, a prior probability distribution is introduced on these parameters, as well as the other unknown parameters in the model. The Bayesian doubly truncated estimator, or BTR, utilizes Monte Carlo simulation in the estimation of the parameters of the model, which is justified on the basis of laws of large numbers and central limit theorems.

The third estimator is the data-constrained maximum entropy estimator for the general linear model (Golan, Judge, and Miller 1996; Mittelhammer and Cardell 1998), or GME. Here finite support of the parameter space is assumed in addition to a truncated *symmetric* error distribution. The parameters for the GME model are estimated via nonlinear optimization by maximizing an entropy function subject to data and other constraints. To motivate the functional specification of the estimators more fully, we first review the principal characteristics of the doubly truncated regression model.

Maximum Likelihood in the Truncated Case

Under standard assumptions of normality, the likelihood function for the GLM is given by

(1) $$L(b, s \mid y, x) = (2ps^{2})^{-n/2} \exp(-(y - xb)'(y - xb)/2s^{2})$$

Maximizing (1) yields an estimator equivalent to the standard OLS estimator, $\hat{b}^{OLS}$. Under general regularity conditions the maximum likelihood estimators are consistent and asymptotically normally distributed (Mittelhammer, Judge, and Miller(2000), chapter 3).

Doubly truncating the residuals, as $c < u_i < d, i = 1,..., N$, yields a truncated regression model (TRM). The TRM has the log-likelihood function[3]

(2) $$L(b, s \mid y, x) = (2ps^{2})^{-n/2} \frac{\exp(-(y - xb)'(y - xb)/2s^{2})}{\left[F\left(\frac{d}{s}\right) - F\left(\frac{c}{s}\right)\right]^{n}} \Pi_{i=1}^{n} I_{(c,d)}(y_i - x_{i\cdot} b)$$

where F is the standard normal cdf, c is the lower bound and d is the upper bound for error truncation, and $I_{(c,d)}(u_i)$ is the standard indicator function taking the value 1 when $u_i \in (c, d)$ and 0 otherwise.

[3]The doubly truncated regression model in (2) is based on truncating the residuals of the GLM. See Cohen, Maddala, or Schneider for an alternative formulation that doubly truncates the dependent variable.

Maximizing (2) yields a truncated regression estimator, or $\hat{b}^{TRM}$.

Bayesian Doubly Truncated Regression

Similar to maximum likelihood methods, the Bayesian approach uses the likelihood function to link the information contained in a sample of data and the value of the unknown parameters of an econometric model. In addition, the Bayesian method factors additional information into problem solutions via the specification of a *prior* pdf, B(C), on the parameter vector of the statistical model. In general the marginal posterior distribution for b is defined as

$$p(b \mid \mathbf{y}, \mathbf{x}) = \frac{\int_0^\infty L(b, \sigma \mid \mathbf{y}, \mathbf{x})\, \pi(b, \sigma) d\sigma}{\int_{R^k} \left[\int_0^\infty L(b, \sigma \mid \mathbf{y}, \mathbf{x})\, \pi(b, \sigma) d\sigma \right] db}$$

for a likelihood function $L(\beta, \sigma \mid \mathbf{y}, \mathbf{x})$ and prior pdf $p(b, s)$. Furthermore, if we assume the prior on the b values is independent of the prior on F values, so that $\pi(\beta, \sigma) \equiv \pi(\beta)\pi(\sigma)$, then the marginal posterior on b can be simplified to

$$p(\beta \mid \mathbf{y}, \mathbf{x}) = \frac{L_*(\beta \mid \mathbf{y}, \mathbf{x})\, \pi(\beta)}{\int_{R^k} [L_*(\beta \mid \mathbf{y}, \mathbf{x})\, \pi(\beta)] db}$$

where

$$L_*(\beta \mid \mathbf{y}, \mathbf{x}) = \int_0^\infty L(\beta, \sigma \mid \mathbf{y}, \mathbf{x})\, \pi(\sigma) d\sigma$$

is effectively the marginal likelihood function for b after F has been integrated out of the likelihood function.

To specify the posterior pdf for the GLM with the doubly truncated error distribution in more functionally explicit form, we make several basic assumptions. First of all, it is assumed that the likelihood function $L(\beta, \sigma|\mathbf{y}, \mathbf{x})$ is defined to be the truncated likelihood function in (2), expanded to include *c* and *d* as unknown parameters, i.e., the likelihood function is algebraically identical to (2) with expanded parameter arguments as $L(\beta, \sigma, c, d\,|\mathbf{y}, \mathbf{x})$. We also let the prior pdf on F be the standard parameter transformation-invariant ignorance prior B(F)%F $^{-1}$, which implies $\pi(\beta, c, d, \sigma) = \pi(\beta, c, d)\sigma^{-1}$. Finally, we assume the prior pdf on $\xi \equiv [\beta', c, d]'$ is uniform, as

$$\pi(\beta, c, d) = \pi(\xi) = \left(\prod_{i=1}^{k+2}[\xi_{\ell_i} - \xi_{h_i}]^{-1}\right) I_A(\xi) \text{ where } A = \times_{i=1}^{k+2}[\xi_{\ell_i}, \xi_{h_i}].$$

Here, again, $I_A(\xi)$ is the standard indicator function taking the value 1 when $\xi \in A$ and 0 otherwise. The uniform prior distribution is a particularly useful choice of prior because it effectively allows prior inequality restrictions to be imposed on *b*. Based on these assumptions, the joint posterior pdf is proportional to

$$p(\xi, \sigma|\, \mathbf{y}, \mathbf{x}) \propto \left(\frac{\left(\dfrac{\sigma^{-n-1}}{(2\pi)^{n/2}}\right) e^{\left[-(\mathbf{y}-\mathbf{x}b)'(\mathbf{y}-\mathbf{x}b)/(2\sigma^2)\right]}}{\left[\left(F\left(\dfrac{d}{\sigma}\right) - F\left(\dfrac{c}{\sigma}\right)\right)\right]^n} \right) \left(\prod_{i=1}^{n} I_{(c,d)}(y_i - \mathbf{x}_{i\bullet}\beta)\right) \times \left(\prod_{i=1}^{k+2}\left[\xi_{\ell_i} - \xi_{h_i}\right]^{-1} I_A(\xi)\right)$$

which is a product of 1/F, the likelihood function in (2), and the uniform prior $\pi(\xi)$. Note that integrating out F from the marginal posterior pdf on *x* is not readily analytically tractable.

To account for intractable analytical integration, Monte Carlo integration can be used to estimate moments of the posterior distribution. Let $\theta = [\beta', c, d, \sigma]'$ and suppose $p(q)$ is a posterior pdf from which iid random outcomes $\theta^*_{(i)}$, i = 1, . . . , n*, can be generated. Then we can rely on laws of large numbers to estimate the expectation of $g(\Theta)$ as simply a sample mean of Monte Carlo repetitions,

$$\hat{E}[g(\Theta)] = \frac{1}{n^*}\sum_{i=1}^{n^*} g(\theta^*_i) \overset{as}{\rightarrow} E[g(\Theta)]$$

More sophisticated Monte Carlo methods are required when it is not possible to directly sample from $p(q)$, as is the case in the empirical work contained in this paper. One approach to Monte Carlo integration is the use of importance sampling to estimate moments based on the kernel of the posterior distribution (Mittelhammer, Judge, and Miller; Van Dijk, Hop and Louter).[4] Let $h(q)$ be a density function from which samples are easily obtained.[5] Historically $h(q)$ is called the importance function. Let $w(q)$ denote the importance weight, which is defined as $w(q) = p(q)/h(q)$ on the region where $h(q) > 0$. Then $E[g(\Theta)]$ is approximated by the sampling importance estimator

$$\hat{E}[g(\Theta)] = \left[\frac{1}{n^*}\sum_{i=1}^{n^*} g(\theta^*_i)w(\theta^*_i)\right] \Bigg/ \frac{1}{n^*}\sum_{i=1}^{n^*} w(\theta^*_i) \overset{as}{\rightarrow} E[g(\Theta)]$$

[4]An alternative to importance sampling is sample/importance resampling where, in effect, parameters are resampled with a weighted bootstrap. See Smith and Gefland or Rubin for further details.

[5]The basic idea in selecting h(**2**) is that the importance function is a good approximation to the posterior kernel (Kloek and Van Dijk; Geweke).

Generalized Maximum Entropy

The generalized maximum entropy (GME) estimator of the GLM model is formulated by reparameterizing the parameters and error terms (see Golan et al. for additional details). The reparameterization consists of convex combinations of user defined points that identify the supports for individual parameters and residual terms (S^i for i= $\boldsymbol{\beta}$, $\boldsymbol{\mu}$) as well as the use of unknown convexity weights (**p** and **w**) applied to the support points. In particular, the parameters are specified as $\boldsymbol{\beta}$=vec($\boldsymbol{\beta}_1$,...,$\boldsymbol{\beta}_G$)=S^{β}**p**, and the residuals are defined accordingly as $\boldsymbol{\mu}$=vec($\boldsymbol{\mu}_1$,...,$\boldsymbol{\mu}_G$)=S^{μ}**w**. It is clear that the GME framework inherently incorporates the assumption that the distribution of the residuals is doubly truncated – the support for the residuals is specified so that the feasible outcomes of the residuals are truncated appropriately.

The principle underlying the GME estimator for the coefficients, $\boldsymbol{\beta}$=S^{β}**p**, is to choose an estimate that is based on the information contained in the data, the constraints on the admissible values of the coefficients (such as nonnegativity and normalization of the convexity weights), and the data sampling structure of the model (including the choice of the supports for the coefficients). In effect, the information set used in the estimation is shrunk to the boundary of the observed data and the parameter constraint information through use of a maximum entropy objective. In the absence of any data or other constraints, the coefficient and residual estimates will shrink to the centers of the prior supports defined by $\boldsymbol{S}^i$ (for i= $\boldsymbol{\beta}$, $\boldsymbol{\mu}$).

The parameter estimates for the general linear model, $\hat{b}^{GME}$, are estimated by solving the following constrained generalized maximum entropy problem:

(3) $$\max_{p,w}\{-p'\ln p - w'\ln w\}$$

subject to

(4) $$y = X(S^b p) + S^m w$$

(5) $$p'(I_{\overline{K}} \otimes 1_M) = 1_{\overline{K}}, \;\; w'(I_{NG} \otimes 1_M) = 1_{NG}$$

Regularity assumptions and asymptotic properties of the data-constrained GME estimator for the GLM are reported in Mittelhammer and Cardell (1998).

Monte Carlo Experiments

For the sampling experiments we define a single equation model with truncated error structure that is similar to the general linear model experiments used in Mittelhammer and Cardell. In this study we focus on both small and medium size sample performance of the OLS, TRM, BTR, and GME estimators. The performance measure is the mean square error (MSE) between the estimated and true parameter values.

The linear model is specified as $y_i = 2 + 1x_{i1} - 1x_{i2} + 3x_{i3} + u_i$ where x_{i1} is a discrete random variable such that $x_{il}, i = 1,\ldots, N$ are *iid* Bernolli(.5) and the pair of explanatory variables x_{i2} and x_{i3} are generated as iid outcomes from

$$N\left(\begin{pmatrix} 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}\right)$$

that is then truncated at ±3 standard deviations. The disturbance terms are drawn from a N(0,F) distribution, that is truncated at ±3 standard deviations and $F^2 = 1$. Thus, the true support of the disturbance distribution in this Monte Carlo experiment is truncated normal, with lower and upper truncation points at -3 and +3, respectively.

In the Monte Carlo simulations, the source distribution, for implementing importance sampling for calculating the BTR estimator, is of the form h(>,F)=B(>)h(F) where B(>) is the uniform prior distribution discussed previously, and h(F) is now a square-root inverted gamma distribution.[6] To implement Monte Carol integration, with *p*($,F,c,d |***x***,***y***) %*L*($,F,c,d|***x***,***y***)B($,F,c,d), and given the aforementioned choice of importance function is the prior distribution, the importance weights becomes $w(\boldsymbol{\xi},\sigma)=L(\boldsymbol{\xi},\sigma\mid \mathbf{y},\mathbf{x})/(\sigma h(\sigma))$. In the results presented below, the posterior distribution of parameters was obtained by re-sampling the source distribution 200 times for each Monte Carlo simulation.[7]

Software

The statistical analysis was conducted using the GAUSS computer package (Aptech Systems, Inc. 1995). Specifically, the TRM was estimated using the constrained maximum likelihood module (CML) and the GME was estimated using the nonlinear optimization module (OPTMUM).[8] Alternatively the BTR estimates relied on Monte Carlo integration with importance sampling, as discussed above. Van Dijk, Hop, and Louter provide a useful primer and flowchart for implementing Monte Carlo integration

[6]See Bernardo and Smith for specification of the square-root inverted gamma distribution. The distribution was parameterized so its mean closely approximated the true value of F.

[7]Geweke's numerical standard error was used to determine the adequacy of importance sampling distributions.

[8]See Mittelhammer and Cardell for discussion of a numerically efficient solution technique for the data constrained GME estimator of the GLM.

with importance sampling. Mittelhammer, Judge, and Miller (2000) provide a comprehensive discussion of alternative Bayesian and GME estimators and their statistical properties, a review of Monte Carlo experiments, and empirical examples including supporting GAUSS code on CD-ROM.

Results

Table 1 contains the mean values and MSE of the estimated parameters based on 1000 Monte Carlo repetitions for sample sizes of 25, 50, and 200 sample observations per equation. Support points for the GME estimator were defined to be (-5, 0,5) for the errors and each parameter coefficient. For the TRM model the errors were truncated at (-5,5). However, the TRM results were nearly identical to the OLS for the assumed error supports and are not reported in Table 1. The parameter support points for the BTR model were set at (-5,5) for each coefficient while the supports for the truncation parameters were $c\theta(c_L,c_U)=(-7,-3)$ and $d\theta(d_L,d_U)=(3,7)$. The error supports for each estimator were subjectively chosen to include the true error terms, and in the case of the TRM and GME, they were also constrained to be symmetric. Also, the parameter supports were chosen to include, but not be centered about, the true parameter values. Non-centering of parameter supports on the true parameter values generally introduces bias into the GME estimates.

From the results in Table 1, we can infer several implications as to the performance of the estimators. Increasing the sample sizes from 25 to 50 to 200 observations, the outcomes of the OLS, BTR, and GME estimators all move closer to the true parameter values in terms of the mean of the replicated estimator outcomes, and the dispersion of the estimator distributions notably decreases. As mentioned above, the results for the OLS and TRM estimators were nearly identical across the sample sizes. This could be expected from the fact that the true error distribution was considerably more severely truncated (at a support of (-3,3)) than was the truncated distribution used in the TRM model. For small samples of 25 observations the mean square error performance of the GME estimator is superior relative to OLS, which is consistent with Monte Carlo results from other studies (Golan, Judge, and Miller 1996, Mittelhammer and Cardell 1998). The BTR, with its underlying sampling assumptions, consistently had the largest MSE, although the BTR was generally superior to GME in terms of estimator bias. As the sample size increases from 25 to 50 to 200 observations, the MSEs of the OLS, BTR, and GME estimators are converging to one another.[9]

Illustrative Empirical Application

In the US there are six major classes of wheat grown, including hard red winter (*HRW*), hard red spring (*HRS*), soft red winter (*SRW*), soft white (*SWW*), hard white (*HW*), and durum (*DUR*). Hard white wheat will be excluded from this analysis due to its recent emergence, resulting in a lack of sufficient market information. The remaining five classes of wheat are used to produce a wide variety of products: *HRW* and *HRS* are utilized in the production of bread and rolls, *SRW* is used to produce flat breads, cakes, crackers, and pastries; *SWW* used to produce crackers, cookies, pastries, muffins, and flour for cakes; and *DUR* is used in the production of semolina flour and a variety of pasta products.

A restricted cost function approach is used to derive factor demand equations for the flour milling industry. The cost function is a function of input prices of wheat by class for a given output level of flour. A translog functional form is assumed (see Berndt):

[9]The MSE superiority of the GME over the BTR estimator makes intuitive sense. The prior distributions of the BTR are uniform, defining only restrictions of the support spaces. GME not only defines restrictions of the supports, but also estimates a discrete distribution for each parameter. Incorporation of more restrictive constraints or more informative prior distributions into the BTR estimator may likely improve its MSE performance.

$$\ln C = \ln \alpha_0 + \sum_{i=1}^{m} \alpha_i \ln w_i + .5 \sum_{i=1}^{m} \sum_{j=1}^{m} \gamma_{ij} \ln w_i \ln w_j + \alpha_Y \ln Y + .5 \gamma_{YY} \left(\ln Y \right)^2 + \sum_{i=1}^{m} \gamma_{iY} \ln w_i \ln Y$$

where w_i are the input prices of wheat, Y represents total output, and " $_i$, ($_{ij}$ are parameters to be estimated. The translog share equations derived from Shepard's Lemma can be written as

$$s_i = \alpha_i + \gamma_{1i} w_1 + ... + \gamma_{mi} w_m + \gamma_{iY} Y \quad \text{for i=1, ... ,m}$$

where s_i is the cost share of input *i*. The own-price elasticity is given by

$$\varepsilon_{ii} = \frac{\gamma_{ij} + s_i^2 - s_i}{s_i} .$$

For the purposes of this study, a single factor demand equation is examined which relates the share of soft red winter wheat (SRW) to its own-price, the price of *HRW*, *HRS*, *SWW*, and *DUR* . The *SRW* equation is of particular interest because reported estimates of own-price elasticity have been predominately positive using standard OLS and feasible generalized least squares approaches (Terry 2000).[10] Annual price and quantity data spanning the years 1981 to 1997 for each of the five wheat classes were obtained from USDA-ERS

Applying truncated regression techniques to this empirical example is appealing for several reasons. First, the dependent variable of a share equation is bounded between 0 and 1, implying the error terms are inherently small and bounded in magnitude. Hence, there is potential for increasing econometric estimator efficiency by truncating the error distributions when using both the BTR and GME estimators. Second, the BTR and GME models allow parameter restrictions to be relatively easily imposed that in turn can be utilized to impose negativity of the own-price elasticity. Given the findings of Terry (2000), and others, the own-price elasticity is restricted to the range (-2,0). Third, the

[10]For additional information on this empirical application and the remaining factor demand equations see Terry and Marsh (2000).

sample is limited to 17 observations. This provides an interesting comparison between the BTR estimator and the GME estimator, which has been shown in small samples to be superior in MSE to OLS (Golan, Judge, and Miller; Mittelhammer and Cardell).

Estimates using OLS, BTR, and GME for three scenarios are presented in Table 2. In Scenario 1, error supports were truncated to the interval (-5,5) for the GME model using support points {-5,0,5} . Supports for the truncation parameters of the BTR were set to $c \in (-7, -3)$ and $d \in (3,7)$. The parameters (except for SRW) are constrained to be in the (-5,5) interval for the BTR estimator, and support points of {-5,0,5} were used for parameters in defining the GME estimator.[11] For each scenario the SRW parameter was constrained to (-.2, .14) in order to restrict the mean level own price elasticity to (-2,0). In Scenario 2, errors were truncated more severely to the interval (-1, 1) using the support points {-1,0,1} for the GME model. Supports for the truncation parameters of the BTR were $c \in (-1.25, -0.75)$ and $d \in (0.75,1.25)$. Parameter restrictions were the same as Scenario 1. In Scenario 3, errors were truncated in a very restrictive manner to the interval (-.1, .1) using support point {-0.10,0,0.10} in representing the GME error support. Support points for parameters of the GME estimator, other than the SRW parameter, remained at {-5,0,5}. For the BTR estimator, parameters are constrained to be in the interval (-3,3) for the intercept and (-1,1) for the non-SRW

[11] In practice, specific ranges of the supports may well be imposed for each parameter based on a priori information. Alternatively, as done in the empirical example above, except for the own-price variable the upper and lower support points were set wide enough so that in all likelihood the supports contain the true parameter values. Further insight into setting the support boundaries can come from examining OLS estimates and unbiased GME estimates for the GLM (see Mittelhammer and Cardell).

slope parameters. Supports for the truncation parameters of the BTR were cΘ(-0.125, -0.075) and dΘ(0.075,0.125).[12]

The own-price elasticities are reported in Table 2 for each model and scenario. Similar to Terry (2000), we find the own-price elasticity from OLS is positive and elastic. For the GME estimator own-price elasticities are negative (by constraint) and slightly less than unity for each scenario. The own-price elasticities of the BTR estimator are negative (by constraint) and elastic for all three scenarios. The BTR own-price becomes more inelastic as error supports are shrunk. As the error and/or the parameter supports are shrunk from Scenarios 1 to 3, the sum of square errors and the standard errors decrease for both the BTR and GME estimators.

Additional insight can be drawn from the results of the truncated regression models. First, the BTR model requires that supports be shrunk tighter relative to the GME estimator to comparably decrease the measures of standard error for the parameter estimates. This is because BTR only assumes uniform priors and GME empirically estimates a discrete weight distribution for each parameter and error term. Second, for both the GME and BTR estimators, the own-price elasticity estimates are near the mean of the subjectively defined range (-2,0). This suggests that data underlying the share equations have little information as to the true value of the own-price elasticity.

---

[12] In general the range of the errors for the share equation depends on the values of s_i-x$. However, given a properly specified share equation, one can define a range for the upper and lower support points that contains the true error terms. In this application the estimated errors for OLS were first examined. The maximum and minimum values of the OLS residuals were 0.00575 and -0.01292, respectively. Hence, for the BTR and GME estimators, the lower support point was set at -5 and the upper support point was set at 5 for Scenario 1. Re-examining the GME estimated residuals, lower and upper supports were defined for Scenarios 2 and 3.

Conclusions

Circumstances often arise in empirical work under which restricting parameter spaces and/or truncating regression models can be beneficial to the efficiency of parameter estimates. We analyzed, identified, and illustrated some theoretical and practical issues relating to the use of both traditional and new truncation methods in empirical work.

Performance of ordinary least squares (OLS), Bayesian doubly truncated (BTR), and generalized maximum entropy (GME) estimators were examined based on a particular set of Monte Carlo experiments. In small sample situations GME is mean square error superior to OLS and BTR for the defined experiments. Increasing the sample size, demonstrated that the mean squared error of the estimators were converging to one another. Further, the estimators were used to estimate a derived demand function for soft red winter wheat using a translog share equation. Estimated own-price elasticities were different for the GME and BTR models relative to OLS, and there was an indication that the data contained relatively little information with which to identify the own price elasticity.

The results in this study suggest the need for a rigorous comparison of truncated estimators. It furnishes some insight for empirical economists desiring to apply truncated regression models in the general linear model context. Further research is needed that deals with understanding the role of truncation assumptions for various loss functions and that deal with developing guidelines for setting informative constraints on parameter spaces and error distributions.

Table 1. Mean value of parameter estimates from 1000 Monte Carlo simulations using ordinary least squares (OLS), Bayesian doubly truncated (BTR), and generalized maximum entropy (GME) regression models. Mean square error (MSE) of parameter estimates are in parentheses.[a]

Obs		OLS	BTR	GME		OLS	BTR	GME	BTR Truncation Supports	BTR Truncation Parameter [b]
25	$\beta_1 = 2$	2.06085	1.74961	1.34094	$\beta_3 = -1$	-0.98557	-0.98747	-0.86365	$c_L = -7$	-5.01330
		(1.44494)	(2.11388)	(0.66598)		(0.05976)	(0.12383)	(0.06906)	$c_U = -3$	(0.89135)
50		2.00171	1.85854	1.53769		-1.00489	-1.01461	-0.94848		5.01756
		(0.56142)	(1.83957)	(0.40421)		(0.02883)	(0.10420)	(0.02938)		(1.00288)
200		2.0079	2.00290	1.85001		-1.00166	-1.00344	-0.99010		-5.05562
		(.14847)	(1.56862)	(0.13080)		(0.00684)	(0.07152)	(0.00680)		(1.10781)
25	$\beta_2 = 1$	1.02298	1.02027	0.97083	$\beta_4 = -3$	2.97908	3.03893	3.06813	$d_L = 3$	5.00345
		(0.17931)	(0.36894)	(0.12486)		(0.06903)	(0.10832)	(0.02379)	$d_U = 7$	(0.87842)
50		0.99753	1.01085	0.98288		3.00192	3.03050	3.06722		5.08199
		(0.08452)	(0.31011)	(0.07010)		(0.02775)	(0.09531)	(0.01731)		(0.97853)
200		0.99933	0.99216	0.99916		2.99911	3.00193	3.02418		5.07121
		(0.01955)	(0.23833)	(0.01859)		(0.00706)	(0.07715)	(0.00603)		(1.07285)

[a] Estimates for TRM were nearly identical to OLS. Error restrictions (-5,5) for BTR and {-5,0,5} for GME. Parameter restrictions (-5,5) for BTR and {-5,0,5} for GME.
[b] Supports for truncation parameters are $c \in (c_L, c_U)$ and $d \in (d_L, d_U)$.

Table 2. Parameter estimates of Soft Red Winter Wheat share equation. Standard errors are reported in parenthesis.

		Estimator					
		Scenario 1[a]		Scenario 2[b]		Scenario 3[c]	
Variable [d]	**OLS**	**BTR**[e]	**GME**	**BTR**[e]	**GME**	**BTR**[e]	**GME**
Intercept	1.572	4.846	0.008	-0.707	0.048	0.431	1.089
	(0.20)	(2.86)	(0.57)	(2.90)	(0.50)	(0.332)	(0.348)
$\mathbf{P_{HRW}}$	-0.469	3.029	-0.002	0.895	0.003	-0.006	0.107
	(0.18)	(2.87)	(0.51)	(2.77)	(0.45)	(0.058)	(0.311)
$\mathbf{P_{HRS}}$	-0.065	3.873	-0.002	-0.689	-0.016	0.031	-0.191
	(0.12)	(2.92)	(0.33)	(2.84)	(0.29)	(0.057)	(0.203)
$\mathbf{P_{SRW}}$	0.574	-0.131	-0.030	-0.061	-0.030	-0.060	-0.022
	(0.10)	(0.10)	(0.28)	(0.10)	(0.25)	(0.098)	(0.169)
$\mathbf{P_{SWW}}$	0.025	-4.152	-0.0002	0.883	0.012	-0.009	0.033
	(0.10)	(2.86)	(0.29)	(2.78)	(0.26)	(0.058)	(0.179)
$\mathbf{P_{DUR}}$	-0.015	-1.924	-0.009	-0.220	-0.073	-0.032	-0.037
	(0.03)	(2.82)	(0.09)	(2.52)	(0.08)	(0.058)	(0.056)
$\mathbf{Q_{Flour}}$	-0.255	-0.861	0.035	0.081	0.034	-0.041	-0.153
	(0.04)	(0.90)	(0.104)	(0.59)	(0.09)	(0.057)	(0.064)
c		-6.181		-0.997		-0.086	
		(1.12)		(0.14)		(0.014)	
d		5.630		1.024		0.091	
		(1.16)		(0.14)		(0.015)	
Own-Price Elasticity	2.365	-1.544	-0.986	-1.159	-0.985	-1.153	-0.944
SSE[f]	0.0004	0.590	0.006	0.104	0.004	.007	.002

[a] Error restrictions (-5,5) and parameter restrictions (-5,5) for BTR. Error restrictions {-5,0,5} and parameter restrictions {-5,0,5} for GME. Truncation parameter supports for BTR are $c\Theta$(-7, -3) and $d\Theta$(3,7).

[b] Error restrictions (-1,1) and parameter restrictions (-5,5) for BTR. Error restrictions {-1,0,1} and parameter restrictions {-5,0,5} for GME. Truncation parameter supports for BTR are $c\Theta$(-1.25, -.75) and $d\Theta$(.75,1.25).

[c] Error restrictions {-.1,0,.1} and parameter restrictions {-5,0,5} for GME. Error, slope parameter, and intercept restrictions are (-.1,.1), (-1,1), and (-3,3) respectively for BTR. Truncation parameter supports are $c\Theta$(-.125, -.075) and $d\Theta$(.075, .125).

[d] Six major classes of wheat include hard red winter (HRW), hard red spring (HRS), soft red winter (SRW), soft white (SWW), hard white (HW), and durum (DUR)

[e] The posterior distribution of parameters was estimated by re-sampling the source distribution 1000 times.

[f] Sum square error between observed and predicted shares.

References

Aptech Systems, Inc. (1995), *GAUSS Applications: Maximum Likelihood 4*. Maple Valley, Washington.

Bernardo, J. M. and A. F. M. Smith. 1994. *Bayesian Theory*. Wiley, New York.

Bhattacharya, S. K., A. Chaturvedi, and N. K. Singh. "Bayesian Estimation for the Pareto Income Distribution." *Statistical Papers* 40(3), July 1999, pages 247-62.

Cohen, A. C. *Truncated and Censored Samples: Theory and Applications*, New York: Dekker, 1991.

Davidson, R. & MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York.

Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.

Geweke, J. "Exact Predictive Densities for Linear Models with Arch Disturbances." Journal of Econometrics, 40 (1989): 63-86.

Golan, A., G. Judge, and D. Miller. 1996. *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* John Wiley & Sons, New York.

Judge, G. G., Hill, R. C., William, E. G., Lutkepohl, H. & Lee, Tsoung-Chao. (1988). *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons, New York.

Kloek, T. and H. K. Van Dijk. "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo." Econometrica 46 (January 1978): 1-19.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge.

Mittelhammer, R. C. and N. S. Cardell. 1998. "The Data-Constrained GME Estimator of the GLM: Asymptotic Theory and Inference." Mimeo, Washington State University.

Mittelhammer, R.C., G. Judge, and D. Miller. 2000. *Econometric Foundations*, New York: Cambridge.

Nakamura, A. and M. Nakamura. “Part-Time and Full-Time Work Behaviour of Married Women: A Model with a Doubly Truncated Dependent Variable.” *Canadian Journal of Economics* 16(2), May 1983, pages 229-57.

Rubin, D. B. “Using the SIR Algorithm to Simulate Posterior Distributions,” In *Bayesian Statistics 3*, eds. J. M. Bernardo, M.H. DeGroot, D. V. Lindley, and A. F. M. Smith, Cambridge, MA: Oxford University Press.

Schneider, H. *Truncated and Censored Samples From Normal Populations*, New York: Dekker, 1986.

Smith, A. F. M. and A. E. Gefland. “Bayesian Statistics Without Tears: A Sampling - Resampling Perspective.” *The American Statistician*, 46 (May), 1992.

Terry, J. J. "Derived Demand for Wheat by Class." M.S. Thesis, Kansas State University, Manhattan, KS.

Terry, J. J. and T. L. Marsh. "Demand for U.S. Wheat by Class." Selected paper for 2000 WAEA meeting, Vancouver, Canada.

Van Dijk, H. K., J. P. Hop, and A.S. Louter. “An Algorithm for the Computation of Posterior Moments and Densities Using Simple Importance Sampling.” *The Statistician* 36 (1987): 83-90.