ADAPTIVE TRUNCATED ESTIMATON APPLIED

TO MAXIMUM ENTROPY

by

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

Paper presented at the
Western Agricultural Economics Association Annual Meetings
Logan, Utah
July, 2001

[a]Assistant Professor, Kansas State University, Manhattan, KS, 66506, 785-532-4913, tlmarsh@agecon.ksu.edu and [b]Professor, Washington State University. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Adaptive Truncated Estimation Applied to Maximum Entropy

Abstract: An adaptive estimator is proposed to optimally estimate unknown truncation points of the error support space for the general linear model. The adaptive estimator is specified analytically to minimize a risk function based on the squared error loss measure. It is then empirically applied to a generalized maximum entropy estimator of the linear model using bootstrapping, allowing the information set of the model itself to determine the truncation points. Monte Carlo simulations are used to demonstrate performance of the adaptive entropy estimator relative to maximum entropy estimation coupled with alternative truncation rules and to ordinary least squares estimation. A food demand application is included to demonstrate practical implementation of the adaptive estimator.

Key Words: doubly truncated errors, bootstrapping, truncation points

1. INTRODUCTION

Researchers are continually expanding their use of regression models in which the support space of the error distribution is truncated. Most commonly, truncated regression has been associated with singly truncated distributions with known truncation points (for example, see Maddala). However, singly truncated error distributions with unknown truncation points, as well as doubly truncated error distributions with known and unknown truncation points, also arise in theory and practice and are receiving attention in the literature (Bhattacharya, Chaturvedi, and Singh; Cohen; Nakamura and Nakamura; Schneider; Golan, Judge, and Miller). Accounting for truncated random variables in regression analysis is important because ordinary least squares estimators can be inefficient, biased, or inconsistent otherwise (Maddala). 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, but rather are contained in a finitely-bounded subset of real space (Malinvaud). 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 statistic/econometric literature that account for truncated distributions. Most prevalent are maximum likelihood estimators, which have been specified for truncated regression models with continuous and discrete distributions that have known truncation points (Cohen; Schneider). For regression models with unknown truncation points, order statistics have been used with limited success (Schneider). More recently, generalized maximum entropy (GME) estimators have been proposed that assume compactness of the coefficient space and finite support of the error distribution (Golan, Judge, and Miller). Here, the approach is semi-nonparameteric in nature in that a discrete error distribution of the regression model is not specified, but rather estimated. In exchange for this flexibility, GME estimation requires user-supplied truncation points. Yet unresolved for GME, and other estimators, is a criterion to optimally estimate unknown truncation points for the error distribution of the regression model. In small and medium sized samples the selection of truncation points is critical to specification of the GME estimator, as different truncation points can lead to quantitatively different coefficient estimates (Fraser).

The objective of this paper is to propose an approach to optimally estimate unknown truncation points of the error support space for the general linear model. Initially, an optimal estimator is specified analytically to minimize a risk function based on the squared error loss measure. Then, an adaptive estimator is specified empirically using bootstrapping, allowing the information set of the model itself to estimate the unknown truncation points. This adaptive approach is applied to a GME estimator of the general linear model.

To evaluate the performance of alternative estimators for a range of finite sample sizes, Monte Carlo sampling experiments are completed. We focus on small- to- medium sized sample performance of the adaptive GME estimator, and its performance relative to ordinary least squares (OLS) and to GME estimators coupled with alternative truncation rules. In performing comparisons, we examine the impacts that the choice of truncation points has on precision of GME coefficient estimates and accuracy of predictive fit. To complete the paper each estimator is applied to a derived demand equation for US wheat.

2. ADAPTIVE TRUNCATED REGRESSION

Consider the general linear model (GLM) with N observations $y_i = x_{i.}\beta^* + u_i^*,\ i = 1,...,N$. In this equation y_i is the ith observation on the dependent variable, x_i is a ($1\times K$) row vector of exogenous values for each observation i, β_i^* is a ($K\times 1$) column vector of true coefficient values, and u_i^* is the unknown model residual associated with observation i. It is assumed that the u_i^* are independent and identically distributed and that the ($N\times K$) matrix of exogenous variables, X, has rank K and consists of nonstochastic elements such that the $\lim_{N\to\infty} \frac{1}{N}\left(X'X\right) = \Omega$ where Ω is some ($K\times K$) positive definite matrix.

Malinvaud asserts that in practice support spaces for regression residuals can always be truncated in a manner that is a good approximation to the true distribution because upper and lower truncation points can be selected sufficiently wide to contain the true residuals of the model. Nevertheless, empirical applications implementing truncated regression models are left with (a) little or no guidance into how to optimally choose unknown truncation parameters and (b) limited insight on the implications of doing so. To select unknown truncation points, Golan, Judge, and Miller suggest using a three-sigma truncation rule (Pukelsheim) for GME estimation. In practice this requires an estimate of the standard error of an individual random variable in order to identify upper and lower truncation points. A limitation is that the three-sigma rule depends only on the empirical distribution of the random variable and it does not account for the full information set provided by the regression model through coefficient restrictions. Moreover, the three-sigma rule is not based explicitly on any goodness of fit criteria relating to parameter estimation, prediction ability, or both. Hence, for empirical application the question remains open regarding how to optimally choose truncation points that depend on the entire information set of the proposed estimator.

To proceed consider an error distribution such that $v_l < u_i^* < v_u$ for $i = 1, \dots, n$ where v_l and v_u are extended real-valued parameters. Technically the parameter v_l is the lower truncation point and v_u is the upper truncation point of the support space for each error term. Truncated estimators are commonly classified by the type of truncation: (a) $v_l = -\infty$ and $v_u = \infty \Rightarrow$ not truncated, (b) $v_l = -\infty$ and $v_u < \infty \Rightarrow$ singly right truncated, (c) $-\infty > v_l$ and $v_u = \infty \Rightarrow$ singly left truncated, and (d) $-\infty > v_l$ and $v_u < \infty \Rightarrow$ doubly truncated. If $v_u = v_l = v < \infty$, then the support space is not only doubly truncated but also symmetric.

The conceptual approach taken in this study is to define a risk function dependent upon the truncation points of the error support space and then choose "optimal" truncation points by minimizing the risk function. Let $\hat{\beta}(v_l, v_u)$ denote a truncated estimator of β^* contingent upon the upper truncation point v_u and the lower truncation point v_l. In the event v_l and v_u are known points of truncation, then $\hat{\beta}(v_l, v_u)$ is itself an obvious choice for the estimator of β^*. Both Cohen and Schneider provide excellent introductions into censored and truncated estimators with known boundary points. In the event that either v_l or v_u are unknown points of truncation, then we define a risk function with squared error loss (SEL) to estimate β^* as

$$\begin{aligned} r(v_l, v_u) &= \mathrm{E}\left[\left(\hat{\beta}(v_l, v_u) - \beta^*\right)' \left(\hat{\beta}(v_l, v_u) - \beta^*\right)\right] \\ &= \sum_{i=1}^{k} var\left(\hat{\beta}_{\mathrm{i}}(v_l, v_u)\right) + \sum_{i=1}^{k}\left[bias\left(\hat{\beta}_{\mathrm{i}}(v_l, v_u)\right)\right]^2 \end{aligned} \tag{1}$$

where $\hat{\beta}_{\mathrm{i}}(v_l, v_u)$ denotes the i$^{\text{th}}$ element of $\hat{\beta}(v_l, v_u)$ and E[] is the expectation operator. An *optimal* truncated estimator

$$\hat{\beta}\left(v_l^*, v_u^*\right) = \left\{\hat{\beta} \,\middle|\, \underset{v_l, v_u}{\arg\min}\left[r\left(v_l, v_u\right)\right]\right\} \tag{2}$$

is obtained by minimizing (1) with respect to either v_l or v_u or both depending upon whether it is singly or doubly truncated regression. In practice the risk function in (1) is not observable because β^* is not observable, and the probability distribution required to calculate the expectations in (1) is unavailable as well. Consequently, empirical estimation of the *optimal* truncated estimator is generally infeasible.

To estimate (1) we follow Hall (1992), Efron and Tibshirani, and others, and empirically approximate the SEL function with bootstrapping. For the purposes of this study the optimization problem in (2) is simplified by assuming $0 < v_u = -v_l = v < \infty$, implying the support space is doubly symmetrically truncated. This assumption is consistent with the estimators considered in the empirical analysis below. The bootstrapped estimate of the risk function in (1) is then

$$\hat{r}_b\left(v\right) = \hat{\mathrm{E}}_{\mathrm{b}}\left[\left(\hat{\beta}\left(v\right) - \mathrm{b}\right)'\left(\hat{\beta}\left(v\right) - \mathrm{b}\right)\right] = \sum_{i=1}^{k} \widehat{var_b}\left(\hat{\beta}_{\mathrm{i}}\left(v\right)\right) + \sum_{i=1}^{k}\left[\widehat{bias_b}\left(\hat{\beta}_{\mathrm{i}}\left(v\right)\right)\right]^2 \tag{3}$$

where the subscript b denotes an expectation taken with respect to the bootstrapped distribution of $\hat{\beta}\left(v\right)$. Explicit details of the bootstrapping approach are discussed later in the empirical section of the paper. The empirical counterpart to the *optimal* truncated estimator is then the estimator associated with the value $\hat{v}$ that minimizes the bootstrap estimate of the risk function given in (2). The truncated estimator is defined empirically as

$$\hat{\beta}\left(\hat{v}\right) = \left\{\hat{\beta} \,\middle|\, \underset{v}{\arg\min}\left[\hat{r}_b\left(v\right)\right]\right\} \tag{4}$$

The *empirical* truncated estimator $\hat{\beta}(\hat{v})$ is an adaptive estimator that chooses the value $\hat{v}$ to minimize the empirical estimated risk function.

Hypothesis testing with the adaptive estimator in (4) is straightforward using either asymptotic standard errors or boot strapped standard errors. In particular, if asymptotic properties of $\hat{\beta}(v)$ are known for any truncation point *v*, then they are known for the specific case $\hat{\beta}(\hat{v})$ with $v = \hat{v}$. Further, given the use of bootstrapping for the empirical adaptive estimator in (4), it is straightforward to calculate bootstrapped standard errors of $\hat{\beta}(v)$ (Hall 1992; Efron and Tibshirani).

3. APPLICATION TO GME

GME estimation is typically motivated in several ways. For example, it is well known that ordinary least squares is especially attractive for Gaussian error models because it is asymptotically efficient. But outside of the Gaussian error model the least squares estimator is no longer efficient and can be grossly inefficient (Koenker, Machado, Skeels, and Welsh). GME estimation has been proposed as an alternative to least squares in the presence of small samples or ill-posed problems, where traditional approaches may provide parameter estimates with high variance and/or bias, or provide no solution at all. Moreover, for empirical problems with inherent uncertainties, restrictions, or truncations, additional appeal of GME estimators is that they offer a systematic framework for incorporating such information into an econometric model.

Two GME estimators are considered that account explicitly for the truncated nature of the error term, wherein the truncation points themselves are unknown. The first is the data-constrained GME estimator where the error distribution is doubly truncated in a symmetric manner, which is labeled GME-D (Golan, Judge, and Miller). Finite support of the coefficient space is assumed in addition to a truncated symmetric error distribution. The parameters for the GME-D model are estimated via nonlinear optimization by maximizing an entropy function subject to data and other constraints. Truncation points are selected using the three-sigma rule discussed below. The second estimator is defined by applying the adaptive estimator in (4) to GME-D, which is labeled as GME-A. In this way a quadratic risk function with bootstrapping is used to estimate unknown truncation points and, hence, coefficient values.

3.1 GME Estimation

Following the maximum entropy principle, the entropy of a distribution of probabilities $p=(p_1,..,p_M)'$ is defined as

$$H(p) = -\sum_{m=1}^{M} p_m \ln p_m$$

with adding up condition $\sum_{m=1}^{M} p_m = 1$. This entropy measure is based on an axiomatic approach that defines a unique objective function to measure uncertainty of a collection of events (Shannon; Jaynes). Generalizations of the entropy function that have been examined elsewhere include the Cressie-Read statistic (Imbens, Spady, and Johnson), Kullback-Leibler Information Criterion (Kullback; Kullback and Leibler), and α-entropy measure (Pompe). We restrict the analysis to maximum entropy, in part due to its efficiency and robustness properties (Imbens, Spady, and Johnson). In the above equation, H(p) reaches a maximum when $p_m=1/M$ for $m=1,...,M$, which is the uniform distribution.

Golan, Judge, and Miller extended the principle of maximum entropy to inverse problems with noise, or the principle of generalized maximum entropy. The principle underlying the GME estimator of the GLM 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).

The GME estimator of the GLM model is formulated by reparameterizing coefficients and error terms. The reparameterization parameters consist of convex combinations of user-defined points that identify discrete support values for individual coefficients and residual terms (s^i for i= β, μ) as well as the use of unknown vectors of convexity weights (p and w) applied to the support points. Using M support points, the coefficient space is represented by the identity β=(β$_1$,...,β$_G$) where $\beta_k = \sum_{m=1}^{M} s_{km}^{\beta} p_{km}$ for k=1,…,K. Note that for the uniform distribution, $p_{km} = 1/M$ with the coefficients shrinking to the mean of the support points, or $\overline{\beta}_k = \sum_{m=1}^{M} s_{km}^{\beta} / M$. Residuals are defined accordingly as μ=(μ$_1$,...,μ$_N$) with identity $\mu_i = \sum_{m=1}^{M} s_{im}^{\mu} w_{im}$ for i=1,…,N. It is clear that the GME framework inherently incorporates the assumption that the distribution of the residuals is doubly truncated with lower truncation point s_{i1}^{μ} and upper truncation point s_{iM}^{μ} .

Given the above identities, parameters for the data-constrained GME model are estimated by solving the following constrained optimization problem:

$$\underset{b,p,w}{Max}\left(\sum_{k=1}^{K} \sum_{m=1}^{M} p_{km} ln(p_{km}) - \sum_{i=1}^{N} \sum_{m=1}^{M} w_{im} \ln(w_{im}) \right) \tag{5}$$

subject to

$$\sum_{m=1}^{M} s_{km}^{\beta} p_{k\ell} = b_k ,\ \beta_{kL} = s_{k1}^{\beta} \leq ... \leq s_{kM}^{\beta} = \beta_{kH} , k = 1,..., K \tag{6a}$$

$$\sum_{m=1}^{M} s_{im}^{\mu} w_{im} = e_i = y_i - X_{i\cdot} b = e_i(b),\ \mathrm{v}_l = s_{i1}^{\mu} \leq ... \leq s_{iM}^{\mu} = v_u , i = 1,..., N \tag{6b}$$

$$\sum_{\ell=1}^{J_k} p_{k\ell} = 1, k = 1,\ldots, K \text{ and } \sum_{\ell=1}^{J} w_{i\ell} = 1, i = 1,\ldots, N \quad (7)$$

It's assumed that the error supports are doubly truncated and symmetric about the origin with truncation points $0 < v_u = -v_l = v < \infty$.

The GME-D model consists of the entropy objective function defined in (5), the parameter and data constraints in (6a,b), and the required adding up conditions in (7). The objective function defined in (5) is optimized to balance the entropy of parameters in the coefficient space with those in the error space for the GLM in (6a,b). Balanced objective functions provide alternative estimation procedures to deal with explanatory variables that are ill-conditioned where traditional estimators are highly unstable and serve as an unsatisfactory basis for estimation and inference (Zellner 1994). Asymptotic properties for the GME-D estimator are derived by Mittelhammer and Cardell.

3.2 Parameter Restrictions

For the purposes of this study, it is assumed that truncation points of the coefficient and error support spaces are selected independent of one another. In other words, we assume the truncation points of the coefficient space are known or specified in a Bayesian manner in that prior information or uncertainties guide the selection of truncation points on support spaces of coefficient values. Effectively, the coefficient truncation points are held constant while widening or narrowing the truncation points of the error support space. See Preckel for a discussion of simultaneously adjusting the truncation points of the coefficient and error support spaces.

Imposing doubly truncated support spaces on coefficients has several implications for GME estimation. It provides an effective and computationally efficient way to restrict coefficients of the GME-D estimator. If knowledge is limited about unknown coefficients, wider truncation points may be used in an effort to ensure the support space contains the true β^*. If knowledge is known about, say, the sign of a specific coefficient from economic theory, then narrower support points may be more confidently imposed. Perhaps most important is a bias-efficiency tradeoff that arises when coefficient support spaces are specified. Truncated coefficient support spaces not centered on the true coefficient values introduce bias into the GME-D estimator. However, in extensive Monte Carlo analysis (Mittelhammer and Cardell), it has been demonstrated that bias introduced by truncation in the GME-D estimator can be offset by substantial decreases in efficiency. Although the impact of truncating coefficient space plays an important role in estimation, the primary focus of the current paper is to examine the implications of truncating the error support space given the truncation points of the coefficient space are known.

3.3 Truncating Error Supports

Selecting truncation points for the error support space plays a key role in specifying the GME-D estimator. In the limit, wider truncation points for GME estimators imply the discrete distribution of the errors tends to an empirical distribution that is more uniform. Alternatively, narrowing the width of truncation points to zero will lead to an infeasibility. Further, appropriately specifying error truncation points for the GME regression model can lead to least squares like behavior. We emphasize that given the nature of the objective function and constraints in (5)-(7), the choice of truncation points for the error support depends on the entire information set including the underling data and user-supplied support points for the coefficients.

Given ignorance regarding the error distribution, Golan, Judge, and Miller suggest using the three-sigma rule to determine error bounds. The three-sigma rule for random variables states that the probability for a unimodal random variable falling away from its mean by more than three standard deviations is at most 5% (Vysochanskii and Petunin; Pukelsheim). The three-sigma rule is a special case of Vysochanskii and Petunin's bound for unimodal distributions. Let Y be a real random variable with mean μ and variance σ^2, then the bound is given as

$$\Pr\left(\left|Y-\mu\right|\geq r\right)\leq\frac{4}{9}\frac{\sigma^2}{r^2} \tag{8}$$

where $r>0$ is the radius of an interval centered at μ. For $r=3\sigma$ it yields the three-sigma rule and more than halves the Chebyshev bound. In more general terms the above bound can yield a j-sigma rule with $r=j\sigma$ for $j=\{1,2,3, \ldots\}$. In the empirical applications below the Vysochanskii and Petunin bound is used in estimation of the GME-D model.

Before turning to the empirical analysis, it is useful to examine the gradient of the GME-D estimator. This provides important insight into properties of GME-D estimation and the impact of truncating the support spaces. Mittelhammer and Cardell derive the gradient of the GME estimator defined by (5)-(7) (see Appendix), which is given by

$$G(b)=-\eta(b)+X'\gamma(e(b)) \tag{9}$$

where $\eta(b)$ and $\gamma(e(b))$ are $K\times1$ and $N\times1$ vectors representing optimal values of the Lagrangian multipliers defined in the GME optimization problem, $e(b)$ is a $N\times1$ vector of estimated residuals of the GLM, and b is a $K\times1$ element of the coefficient space. The first term on the right hand side of (9), $\eta(b)$, is the component of the entropy objective function linked to coefficient support points. As a result, this term plays an integral role in estimation of small and medium sized samples. However, because $\eta(b)$ is independent of N, it is asymptotically uninformative. The second term on the right hand side, $\gamma(e_i(b))$, is the component of the entropy objective function linked to error support points. It depends on the sample size N and is asymptotically informative, playing a key role in GME estimation of all sample sizes and in establishing asymptotic properties.

The terms of the GME gradient, and its relationship to the OLS gradient, provide guidance in setting truncation points for support spaces on the error terms in empirical applications. An important insight is that the OLS estimator can be thought of in some sense as a special case of GME estimation. That is, in the event that $\gamma(e(b)) \approx e(b) = \mathbf{y} - X\,b$, the gradient of the GME estimator is effectively that of OLS. Because the GME and OLS gradients can be approximately equivalent, then so are the coefficient estimates.

3.4 Adaptive GME Estimation

The adaptive estimator of the GME-D model is based on the empirical SEL function in (4) with bootstrapping. A balanced bootstrap procedure is used for generating an estimate of (4) as a function of *v*. Balanced bootstrapping can reduce simulation error without necessarily increasing the observed sample size (Davison, Hinkley, and Schechtman; Gleason; Graham, Hinkley, John, and Shi; Hall 1990). To implement this approach we collect n_b copies of the sample observation indices 1, ... , *N*, and then draw n_b size-n random samples of these indices, without replacement, from the assembled set of $\mathrm{N} \times \mathrm{n_b}$ indices. The actual sampling scheme can be accomplished by randomly permuting the $\mathrm{N} \times \mathrm{n_b}$ indices, and then grouping the resulting list into n_b sequential blocks of n indices each. In this permutation sampling scheme, each sample observation is guaranteed to appear across the n_b resamples exactly n_b times. Then the value of $\hat{\beta}(v)$ is recalculated n_b times from the resamples of the original data identified by the n_b samples of indices. The result is a balanced bootstrap distribution of outcomes for the estimator $\hat{\beta}(v)$, which can then be used to form bootstrapped empirical bias and variance estimates based on empirical moments relating to the bootstrapped distribution of $\hat{\beta}(v)$.

Given the discussion above we also calculate the bootstrapped standard errors of $\hat{\beta}(v)$. For a specific truncation point $\hat{v}$, the bootstrapped estimate of the standard error for $\hat{\beta}(\hat{v})$ is

$$\hat{s}_b = \left\{ \sum_{i=1}^{n_b} \left[\hat{\beta}^i(\hat{v}) - \overline{\beta}(\hat{v}) \right]^2 / (n_b - 1) \right\}^{1/2}$$

where $\hat{\beta}^i(\hat{v})$ is the ith bootstrap resample and $\overline{\beta}(\hat{v}) = \sum_{i=1}^{n_b} \hat{\beta}^i(\hat{v}) / n_b$.

4. 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, GME-D, and GME-A estimators. Performance measures used are the prediction squared error (PSE) between the estimated and actual observations of the dependent variable and 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 u_i is a discrete random variable, $x_{il}, i = 1,..., N$ 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 $N(0,\sigma^2)$ distribution, that is truncated at ±3 standard deviations and $\sigma^2 = 1$. Thus, the true support of the disturbance distribution in this Monte Carlo experiment is truncated normal, with lower and upper truncation points located at -3 and +3, respectively.

To investigate the impact of coefficients restrictions on the choice of error truncation points, three different sets of supports for the coefficients are defined by

$$Z1 = \begin{bmatrix} -2 & 2 & 6 \\ -3 & 1 & 5 \\ -5 & -1 & 3 \\ -1 & 3 & 7 \end{bmatrix}, \; Z2 = \begin{bmatrix} -3 & 1 & 5 \\ -4 & 0 & 4 \\ -4 & 0 & 4 \\ 0 & 4 & 8 \end{bmatrix}, \text{ and } Z3 = \begin{bmatrix} -10 & 0 & 10 \\ -10 & 0 & 10 \\ -10 & 0 & 10 \\ -10 & 0 & 10 \end{bmatrix}$$

In the support matrices above, Z1 defines supports that are symmetrically centered about the true values of the coefficients and relatively narrow in width, Z2 defines supports that are asymmetric but with the same width as Z1, and Z3 represents supports that are asymmetric and wider than the two previous support sets with an upper and lower truncation points at –10 and 10 respectively. The supports given by Z3 are used to illustrate the case where less information is known about the true parameter values.

To estimate the GME models additional information is necessary. The GME-D estimator is coupled with the three-sigma rule, which we term the GME-3σ_y estimator. The truncation value for the error support based on the three-sigma rule is defined as $v = 3\hat{\sigma}_y$ where $\hat{\sigma}_y$ is the sample standard deviation of the *dependent* variable. The GME-A model is estimated by searching over a grid of truncation points $V = \{v \,|\, v = 4, 6, 10, 12, 14\}$ to find the value $\hat{v}$ that minimizes the risk function. The grid effectively nests the two-, three-, four-, and five-sigma rules defined in the set $\Sigma = \{v \,|\, v = j\hat{\sigma}_y,\, j = 2,3,4,5\}$. Finally, we choose the unbiased and consistent OLS estimator, $\hat{b}$, to provide a common estimate of the coefficient b specified in (3).

The statistical analysis was conducted using the GAUSS computer package (Aptech Systems, Inc. 1995). Specifically, GME estimators were estimated using the nonlinear optimization module (OPTMUM). Mittelhammer, Judge, and Miller (2000) provide a comprehensive discussion of alternative information theoretic estimators and their statistical properties, a review of Monte Carlo experiments, and empirical examples including supporting GAUSS code on CD-ROM.

4.1 Point Estimate Results

The mean square error (MSE), prediction squared error (PSE), and mean truncation points are presented in Table 1 based on 25, 50, and 100 observations for OLS, GME-3σ_y, and GME-A. In the analysis, we used 1000 Monte Carlo simulations for each sample size. For the GME-A estimator 100 bootstrap samples were generated at each grid point for each Monte Carlo simulation.

Several general implications are apparent from the Monte Carlo analysis. First, OLS is predominately larger in MSE than either GME-A or GME-3σ_y. In contrast, OLS is lower in PSE than GME-A that is in turn lower than GME-3σ_y. This reflects a tradeoff between precision of coefficient estimates and accuracy of predictive fit among the three estimators. Second, GME-A is often a compromise between OLS and GME-3σ_y. This is because the truncation point for GME-A is chosen by minimizing a risk function with an incentive for "closeness" to the OLS estimator. As a result, GME-A is lower in PSE and higher in MSE than GME-3σ_y for supports defined by Z1 and Z3. However, this is not always the case.

Focusing further on MSE, as the sample size increases, the MSE decreases for each estimator. This is not surprising as each is an asymptotically consistent estimator of β^*. For Z1 the MSE is lower for both GME-A and GME-3σ_y relative to OLS, as expected. For Z2 the MSE is lowest for GME-A, followed by OLS and GME-3σ_y respectively. For Z3 the MSE is lowest for GME-3σ_y, followed by GME-A and OLS. These results demonstrate that GME-3σ_y

outperformed OLS and GME-A in MSE for two out of the three experiments. However, they also demonstrated that GME-3σ_y is less flexible relative to narrow and asymmetric support points of Z2 on the coefficients.

Mean truncation points are also reported in Table 1 for GME-A and GME-3σ_y. The reported value for the GME-A estimator is the mean of the truncation points from the grid defined by $\{v|v = 4, 6, 8, 10, 12\}$ that minimize (6) for 1000 Monte Carlo simulations. The reported value for the GME-3σ_y estimator is the mean of $v = 3\hat{\sigma}_y$ from the empirical distribution generated with 1000 Monte Carlo simulations. Truncation points are somewhat similar for experiments with Z1 and Z3, which is consistent with the MSE results. In contrast the truncation point for GME-A is almost one half of that for GME-3σ_y when supports are defined by Z2, reflecting “tighter” truncation points better minimized the loss function. Hence, this reflects that the optimal truncation rule is sensitive to support points as well as data of the regression model.

The Monte Carlo experiments demonstrated that the performance of the GME-D estimator coupled with the three-sigma rule has important limitations. Its performance diminished relative to the other estimators in the presence of asymmetric coefficient supports. In practical applications true coefficient values will always be unknown, leading to asymmetric supports. Moreover, it is easy to envision circumstances under which coefficient supports are narrowly specified. In contrast, these issues highlight the advantage of the adaptive GME estimator. It optimizes the choice of error truncation points in a manner that balances the precision of coefficient estimates and accuracy of predictive fit while taking into account the entire information set of the model.

5. EMPIRICAL APPLICATION

To illustrate implementation of the GME-A estimator, we examine a derived demand relationship for US soft red winter (SRW) wheat. The SRW equation is of particular interest because reported estimates of own-price and cross-price elasticities are not consistent in magnitude nor sign (Terry 2000; Barnes and Shields 1998). Robust measurements of elasiticities are important to USDA food policy (Barnes and Shields 1998), university wheat research programs (Boland, Johnson, and Schumaker), and the milling industry (Milling and Baking News). Annual price and quantity data that span 1981 to 1997 for each of the wheat classes were obtained from USDA-ERS and the Milling and Baking News. Price data are average prices at the farm level ($/bu). Quantity data represent domestic disappearance of wheat by class for food use (million bu). Limited data observations, high correlation among price variables, and lack of robustness of the estimates from standard regression methods make this an ideal example for GME estimation.

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):

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

where w_i are the input prices of wheat ($/bu), Y represents total flour output (10,000 lbs), and α_i, γ_{ij} are parameters to be estimated. The translog share equations derived from Shepard's Lemma can be written as

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

where s_i is the cost share of input i. The own-price and cross-price elasticities are given by

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

The single factor demand equation examined in this application relates the share of soft red winter wheat (SRW) to its own-price, the price of other wheat, and quantity of flour output. The other wheats considered are hard red winter (HRW), hard red spring (HRS), soft white (SWW), hard white (HW), and durum (DUR).

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 GME-D and GME-A estimators. Second, the 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 be negative. Third, the sample is limited to 17 observations and the price data are highly correlated. This provides an interesting comparison between the OLS estimator and the GME estimators, which have been shown in small samples to be superior in MSE to OLS (Golan, Judge, and Miller; Mittelhammer and Cardell).

The information set for the GME estimators expands that of OLS. The three-sigma truncation rule is imposed for GME-D and coefficient restrictions are imposed for both GME-D and GME-A estimates. Coefficient restrictions were {-5,0,5}, {-.2,0,.14},{-5,0,5}, {-5,0,5}, {-5,0,5}, and {-5,0,5} respectively for the intercept, SRW, HRW, HRS, SWW, and DUR variables. This effectively restricted the own-price elasticity estimates to be between (-2, 0) for SRW. For the purposes of this example, cross-price elasticities for HRW, HRS, SWW, and DUR variables were not constrained other than by the relatively wide truncation points of the coefficient space. The grid used for the truncation point was defined by $\{v|v = .03, .05, \ldots,.43\}$, which is incremented by .02 and covered the truncation point for the three-sigma rule $v = 3\hat{\sigma}_y = .055$.

Coefficient estimates, t-values, and summary statistics for GME-A, GME-3σ_y, and OLS are presented in Table 2. Not surprisingly, the sum squared error between the observed and predicted shares is lowest for OLS, followed by GME-A and GME-3σ_y respectively. The R^2 goodness of fit measures are consistent with these results. At the .05 level the significant variables in the OLS equation are the intercept, prices of HRW and SRW, and quantity of output. In the GME-3σ_y equation the intercept and quantity of output are significant. And, in the GME-A equation the significant variables are the intercept, price of SRW, and output quantity. The truncation point selected by the adaptive estimator for this empirical example was $\hat{v} = .03$.

The price elasticities statistics for GME-A, GME-3σ_y, and OLS are also presented in Table 2. Similar to Terry (2000) the own-price elasticity from OLS is positive and elastic, which conflicts with economic theory. For the GME estimators own-price elasticities are negative (by constraint) and inelastic. The HRW cross-price elasticities is negative for OLS, while HRS is negative for GME-D. HRS is the only negative cross-price elasticity for GME-A.

In all, the empirical example illustrates advantages that GME estimators have over OLS in imposing restrictions on coefficients and in small sample situations. Furthermore, the example demonstrates the improvement of the GME-A estimator over GME-3σ_y. That is, GME-A has a higher R^2, lower sum square error between observed and predicted shares, and higher significance on the own-price coefficient relative to GME-D. The inferior performance of GME-3σ_y clearly illustrates the information limitation of the three-sigma rule for truncation.

6. CONCLUSIONS

Circumstances often arise in empirical applications 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 optimally truncated regression models and provided an example using generalized maximum entropy.

Performance of ordinary least squares (OLS), data-constrained generalized maximum entropy (GME-D) estimators, and adaptive generalized maximum entropy (GME-A) estimators were examined based on a particular set of Monte Carlo experiments. The Monte Carlo analysis demonstrated the linkage of truncation points with tradeoffs between prediction squared error (PSE) among the estimated and actual observations of the dependent variable and the mean square error (MSE) among the estimated and true parameter values. In small sample situations GME-A was MSE superior to OLS, while OLS was superior in PSE. Size and power of t-tests were also examined in the Monte Carlo analysis.

Overall, GME-A was quite flexible when estimating truncation points that inherently depend on the entire information set of the model. GME-A provided a systematic means to optimally estimate truncation points in a manner that balanced the precision of coefficient estimates with the accuracy of predictive fit. GME-D coupled with the three-sigma truncation rule was not as flexible. It improved the MSE measure relative to OLS for wide truncation points. But, in circumstances with narrow and asymmetric coefficient supports, it exhibited stark limitations relative to GME-A. This illustrates the advantages of the GME-A approach over the "hit or miss" of the GME-D approach in practical applications, wherein narrow or asymmetric coefficient supports are likely the rule and not the exception.

An empirical example was included to demonstrate practical application of the GME-A estimator. Factor share functions for soft red winter wheat were reported for each estimator. GME-A outperformed GME-D coupled with the three-sigma truncation rule. As expected, OLS predicted better the shares values of the dependent variable than either GME-A or GME-D. Estimated own-price elasticity results from GME-A were strikingly different to those from OLS.

The results in this study highlight the importance of rigorous examination of truncated estimators. It furnishes insight for empirical economists desiring to apply truncation rules to GME estimators in the general linear model context. Further research is needed that provides deeper understanding of 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.

APPENDIX

Define the maximized entropy function, conditional on $b = \tau$, as

$$F(\tau) = \underset{\substack{p,w:b=\tau \\ (6)-(7)}}{Max} \left(-\sum_{k=1}^{K} \sum_{m=1}^{M} p_{km} \ln(p_{km}) - \sum_{i=1}^{N} \sum_{m=1}^{M} w_{im} \ln(w_{im}) \right). \qquad \text{(A.1)}$$

The optimal value of $w_i = (w'_{i1}, \ldots, w'_{iM})'$ in the conditionally-maximized entropy function is given by

$$w_i(\tau) = \underset{w_i:6b,\ \sum_{m=1}^{M} s^{\mu}_{im} w_{im} = e_i(\tau)}{\arg\max} \left(-\sum_{m=1}^{M} w_{im} \ln(w_{im}) \right),$$

which is the maximizing solution to the Lagrangian

$$L_{w_i} = -\sum_{m=1}^{M} w_{im} \ln(w_{im}) + \lambda_i^{w} \left(\sum_{m=1}^{M} w_{im} - 1 \right) + \gamma_i \left(\sum_{m=1}^{M} s^{\mu}_{im} w_{im} - e_i(\tau) \right).$$

The optimal value of w_{im} is then

$$w_{i\ell}(\gamma(e_i(\tau))) = w_{\ell}(\gamma(e_i(\tau))) = \frac{e^{\gamma(e_i(\tau)) s^{\mu}_{i\ell}}}{\sum_{m=1}^{M} e^{\gamma(e_i(\tau)) s^{\mu}_{im}}}, \ell = 1, \ldots, M,$$

where $\gamma(e_i(\tau))$ is the optimal value of the Lagrangian multiplier γ_i under the condition

$$b = \tau, \text{ and } w_{\ell}(\gamma) \equiv \frac{e^{\gamma s^{\mu}_{i\ell}}}{\sum_{m=1}^{M} e^{\gamma s^{\mu}_{im}}}.$$

Similarly, the optimal value of $p_k = (p_{k1}, \ldots, p_{kM})'$ in the conditionally-maximized entropy function is given by

$$p_k(\tau_k) = \underset{p_k:6a,\ \sum_{m=1}^{M} s^{\beta}_{km} p_{km} = \tau_k}{\arg\max} \left(-\sum_{m=1}^{M} p_{km} \ln(p_{km}) \right),$$

which is the maximizing solution to Lagrangian

$$L_{p_k} = -\sum_{m=1}^{M} p_{km}\ln(p_{km}) + \lambda_k^p\left(\sum_{m=1}^{M} p_{km} - 1\right) + \eta_k\left(\sum_{m=1}^{M} s_{km}^{\beta} p_{km} - \tau_k\right).$$

The optimal value of $p_{k\ell}$ is then

$$p_{k\ell}(\tau_k) = \frac{e^{\eta_k(\tau_k)s_{k\ell}^{\beta}}}{\sum_{m=1}^{M} e^{\eta_k(\tau_k)s_{km}^{\beta}}}, k = 1,\ldots,K,$$

where $\eta_k(\tau_k)$ is the optimal value of the Lagrangian multiplier η_k under the condition $b_k = \tau_k$.

Substituting the optimal solutions for the $p_{k\ell}$'s and $w_{i\ell}$'s into (A.1) obtains the conditional maximum value function

$$F(\tau) = -\sum_{k=1}^{K}\left(\eta_k(\tau_k)\tau_k - \ln\left(\sum_{m=1}^{M} e^{\eta_k(\tau_k)s_{km}^{\beta}}\right)\right) - \sum_{i=1}^{N}\left(\gamma(e_i(\tau))e_i(\tau) - \ln\left(\sum_{m=1}^{M} e^{\gamma(e_i(\tau))s_{im}^{\mu}}\right)\right) \quad \text{(A.2)}$$

Define the gradient vector of $F(\tau)$ as $G(\tau) = \frac{\partial F(\tau)}{\partial \tau}$ so that

$$G_k(\tau) = \frac{\partial F(\tau)}{\partial \tau_k} = -\eta_k(\tau_k) + \sum_{i=1}^{N}\gamma(e_i(\tau))X_{ik}, k = 1,\ldots,K,$$

and thus

$$G(\tau) = -\eta(\tau) + X'\gamma(e(\tau)) \quad \text{(A.3)}$$

where $\eta(\tau)$ and $\gamma(e(\tau))$ are $K \times 1$ and $N \times 1$ vectors of Lagrangian multipliers.

REFERENCES

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

Barnes, J. N., and D. A. Shields. 1998. The Growth in U.S. Wheat Food Demand. Wheat Yearbook. USDA, Economic Research Service. (March): 21-29.

Berndt, E. R. 1990. The Practice of Econometrics: Classic and Contemporary. New York: Addison-Wesley.

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.

Boland, M., M. Johnson, and S. Schumaker. Hard White vs. Hard Red Wheat: What's Going on in Kansas? Forthcoming Choices.

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

Cressie, N. and T. Read. "Multinomial Goodness of Fit Tests," Journal of the Royal Statistical Society, Series B, 46, 440-446.

Davison, A.C., Hinkley, D.V., and Shechtman, E., 1986. Efficient Bootstrap Simulations. Biometrika 73: 555-566.

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

Fraser, I. "An Application of Maximum Entropy Estimation: The Demand For Meat in the United Kingdom." Applied Economics 32, January 2000, 45-59.

Gleason, J.R. 1988. Algorithms for Balanced Bootstrap Simulation. American Statistician 42: 262-266.

Golan, A., Judge, G. G. and Miller, D., 1996. Maximum Entropy Econometrics. New York: John Wiley and Sons.

Graham, R.L., Hinkley, D.V., John, P.W.M., and Shi, S. 1990. Balanced Design of Bootstrap Simulations. Journal of the Royal Statistical Society, B52: 185-202.

Hall, P. 1990. Performance of Bootstrapped Balanced Resampling in Distribution Function and Quantile Problems. Probability Theory Rel. Fields 85: 239-267.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag: New York.

Imbens, G. W., Spady, R. H. and Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66:333-357.

Jaynes, E. T. "Information Theory and Statistical Mechanics," Physics Rev.k 106(1957a):620-630.

Kitamura, Y. and Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65:861-874.

Koenker, R., Machado, J. A. F., Keels, C. L. S., and Welsh, A. H. (1994), "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimation." Econometric Theory 10, 172-97.

Kullback, S., 1959. Information Theory and Statistics. New York: John Wiley and Sons.

Kullback, S. and Leibler, R. A., 1951. On information and sufficiency, The Annals of Mathematical Statistics 22:79-86.

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

Malinvaud, E. Statistical Methods of Econometrics. Third Ed Amsterdam, North-Holland.

Milling & Baking News. Sosland Companies Inc. Selected years. Merriam, KS: Sosland Publishing Company.

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., Judge, G. and Miller, D., 2000. Econometric Foundations. New York: Cambridge University Press.

Mittelhammer, R. C. and G. Judge. 2000. Robust Empirical Exponential Likelihood Estimation of Models with Non-Orthogonal Noise Components, Forthcoming.

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.

Pompe, B. (1994). "On Some Entropy Measures in Data Analysis," Chaos, Solitons, and Fractals 4, 83-96.

Pukelsheim, F. 1994. "The Three Sigma Rule." The American Statistician 48: 88-91.

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

Shannon, C. E. 1948. “A Mathematical Theory of Communication.” Bell System Technical Journal, 27.

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

Vysocahnskii, D. F. and Petunin, Y. I. 1980. “Justification of the 3σ Rule for Unimodal Distributions,” Theory of Probability and Mathematical Statistics, 21, 25-36.

Zellner, A., 1994. Bayesian and non-bayesian estimation using balanced loss functions. In S. Gupta and J. Berger, ed. Statistical Decision Theory and Related Topics. New York: Springer Verlag.

Table 1. Results from Monte Carlo experiments for OLS[a], GME with three-sigma rule (GME-3σ_y)[a], and adaptive GME estimation (GME-A)[b].

	Mean Square Error			Prediction Squared Error			Mean Truncation Point	
Sample Size	OLS	GME-A	GME-3σ_y	OLS	GME-A	GME-3σ_y	GME-A	GME-3σ_y
Z1								
25	1.748	0.389	0.129	21.046	21.569	22.062	7.254	8.516
50	0.735	0.274	0.111	45.489	45.815	46.104	8.912	8.490
100	0.377	0.183	0.099	96.429	96.669	96.813	10.162	8.598
Z2								
25	1.556	0.852	1.209	21.048	21.935	24.544	4.522	8.597
50	0.726	0.540	0.826	45.904	46.596	48.937	4.828	8.590
100	0.360	0.328	0.504	95.369	95.840	97.878	4.848	8.559
Z3								
25	1.588	1.202	0.922	21.051	21.471	21.793	7.390	8.496
50	0.729	0.703	0.549	46.113	46.446	46.556	8.132	8.552
100	0.356	0.406	0.315	95.675	95.897	95.924	8.426	8.565

[a] 1000 Monte Carlo simulations.

[b] 1000 Monte Carlo simulations with 100 bootstrap samples generated at each grid point.

Table 2. Coefficient and elasticity estimates of soft red winter wheat share equation for OLS, GME, and GME-A. Standard errors are reported in parenthesis.

	Coefficient			Input Price Elasticities		
Variable	**OLS**	**GME-3σ_y**	**GME-A** [b]	**OLS**	**GME-3σ_y**	**GME-A** [b]
Intercept	1.572*	1.367*	1.473*	---	---	---
	(0.201)	(0.301)	(0.411)			
PHRW	-0.203*	0.033	-0.030	-0.724	0.589	0.237
	(0.078)	(0.117)	(0.139)			
PHRS	-0.028	-0.098	-0.082	0.096	-0.290	-0.202
	(0.050)	(0.077)	(0.098)			
PSRW	0.249*	-0.019	0.079*	0.563	-0.713	-0.382
	(0.043)	(0.062)	(0.028)			
PSWW	0.011	0.012	0.014	0.139	0.142	0.155
	(0.045)	(0.067)	(0.091)			
PDUR	-0.007	-0.008	-0.005	0.047	0.042	0.056
	(0.014)	(0.021)	(0.026)			
QFlour	-0.111*	-0.089*	-0.099*	---	---	---
	(0.016)	(0.024)	(0.032)			
R^2	0.923	0.695	0.800			
SSE[a]	0.00042	0.00165	0.00108			

[a] Sum square error between observed and predicted shares.
[b] Standard error was calculated from 300 bootstrap samples with truncation point *v*=0.03.
* Significant at the .05 level.