The Relative Efficiency of Method of Moments Estimators [1]

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

George Tauchen
Department of Economics
Duke University
Durham NC 27708-0097 USA

May 1997
Last revised: September 1997

[1]Supported by the National Science Foundation. The corresponding author is A. Ronald Gallant, Department of Economics, University of North Carolina, CB# 3305, 6F Gardner Hall, Chapel Hill NC 27599-3305 USA; Phone 1-919-966-5338; FAX 1-919-966-4986; e-mail ron_gallant@unc.edu. Updated versions of this paper are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/papers as file cmm.psz, which is compressed PostScript.

Abstract

The asymptotic relative efficiency of efficient method of moments when implemented with a seminonparametric auxiliary model is compared to that of conventional method of moments when implemented with polynomial moment functions. Because the expectations required by these estimators can be computed by simulation, these two methods are commonly used to estimate the parameters of nonlinear latent variables models. The comparison is for the models in the Marron-Wand test suite, a scale mixture of normals, and the second largest order statistic of the lognormal distribution. The latter models are representative of financial market data and auction data, respectively, which are the two most common applications of simulation estimators. Efficient method of moments dominates conventional method of moments over these models.

Key words: Simulation estimators, efficient method of moments, conventional method of moments, financial market data, auction data.

1 Introduction

Efficient method of moments (EMM) is a minimum chi-squared estimator that uses the expected score vector of an auxiliary model estimated by quasi maximum likelihood as the moment functions that enter the chi-squared criterion. Because expectations can be computed by simulation, the method is particularly useful for estimating the parameters of partially observed nonlinear models. If the auxiliary model encompasses the true data generating process, then quasi maximum likelihood estimates become sufficient statistics and EMM is fully efficient (Gallant and Tauchen, 1996). If the auxiliary model is a close approximation to the data generating process, then one can expect the efficiency of EMM to be close to that of classical maximum likelihood (Gallant and Long, 1997; Tauchen, 1997).

Minimum chi-squared estimation can also be implemented using polynomials as moment functions rather than using a score vector. When moments are evaluated analytically, the estimator implemented using polynomials is the classical method of moments estimator. When expectations are computed by simulation, it is the simulated method of moments estimator of Ingram and Lee (1991), Duffie and Singleton (1993), and others as it is usually implemented. We use the term conventional method of moments (CMM) to refer collectively to all variants of the minimum chi-squared estimator implemented using polynomial moment functions.

What is true of EMM is true of CMM. If polynomials are sufficient statistics, then CMM is efficient. If the true data generating process has a moment generating function, then the efficiency of CMM approaches the efficiency of maximum likelihood as degree increases. Examples can be found where EMM has better efficiency than CMM and vice versa. Thus it seems that no general theoretical result exists regarding the efficiency of EMM versus CMM.

Nonetheless, in most applications one would expect EMM to be more efficient than CMM whenever the score of quasi maximum likelihood is a good estimator of the true score. There should be some payoff to the effort of finding a close approximation to the sufficient statistics. One way to ensure that the auxiliary model is asymptotically a close approximation is to use a seminonparametric (SNP) model, which is a density formed from a Hermite expansion with positivity enforced. In this case, the asymptotic variance of EMM converges to that of maximum likelihood as the degree of the Hermite expansion grows (Gallant and Long, 1997). Thus, one would expect EMM implemented this way to outperform CMM.

We investigate this supposition by computing the asymptotic efficiency of EMM and CMM estimates for situations that are representative of anticipated applications. Specifically, we compute relative efficiency for the Marron-Wand test suite, a scale mixtures of normals, and the second largest order statistic from the lognormal distribution. The latter two distributions are representative of data from financial markets and data from auctions, respectively, which have been the two principal applications of simulation estimators to date.

Andersen, Chung, and Sorensen (1997), Chumacero (1997), and Ng and Michaelides (1997) report related Monte Carlo computations. Overall, their work supports the conjecture that EMM is more efficient than CMM in typical applications. The work reported here complements their results because it compares the efficiency of EMM and CMM estimators as degree increases. Performing the necessary computations with Monte Carlo methods would be too computationally intensive to be practicable. A more analytical approach, as used here, must be employed to keep computational complexity within feasible limits.

Our results would apply to the method of indirect inference proposed by Smith (1993) and Gourieroux, Monfort, and Renault (1993) if the SNP auxiliary model were used in that context, because EMM and indirect inference are asymptotically equivalent when the auxiliary model is the same. However, in applications of indirect inference one usually restricts attention to auxiliary models for which the quasi maximum likelihood estimator is cheap to compute because it must be computed repeatedly. EMM requires only a single evaluation of the quasi maximum likelihood estimator of the auxiliary model and therefore allows other considerations to influence the choice of auxiliary model. Because of this structural difference, choice of the auxiliary model partially dictates the choice of estimator and therefore the work reported here does not relate to a comparison of indirect inference to other estimators.

2 A Class of Minimum Chi-Squared Estimators

We consider a minimum chi-squared estimator for the parameters $\rho = (\rho_1, \ldots, \rho_{p_\rho})$ of a density $p(y|\rho)$. Special cases of this estimator are the method of maximum likelihood, (Gauss, 1816; Fisher, 1912) classical method of moments (Pearson, 1894), generalized method of moments (Neyman and Pearson, 1928; Hansen, 1982), simulated method of moments (Ingram and Lee, 1991; Duffie and Singleton, 1993), and efficient method of moments (Gallant and Tauchen, 1996). The estimator is most useful in situations where $p(y|\rho)$ is difficult to determine analytically yet the moment operator

$$\mathcal{M}_\rho g = \int g(y)\, p(y|\rho)\, dy$$

can be computed conveniently by means of simulation or quadrature. These are usually models with unobserved latent variables. Included within the scope of the estimator are cross-sectional regression models, time-series regression models, and stationary dynamic models; these models may be either Markovian (Gallant and Tauchen, 1996) or non-Markovian (Gallant and Long, 1997). Although the scope of the estimator is broad enough to encompass complex time series models, for the issues addressed here it suffices to consider a random sample from $p(y|\rho)$ with y scalar. The observed data $\{\tilde{y}_i\}_{i=1}^n$, are presumed iid with common probability density function $p(y|\rho^o)$ where ρ^o denotes the true value.

The minimum chi-squared estimator is a two-step procedure. In the first step, the observed data are summarized by a projection represented as a vector $\tilde{\psi}_n(y)$ that is in the null space of the empirical moment operator $\mathcal{M}_{\tilde{F}_n}$ defined by

$$\mathcal{M}_{\tilde{F}_n} g = \int g(y)\, d\tilde{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} g(\tilde{y}_i),$$

where $\tilde{F}_n$ denotes the empirical distribution of $\{\tilde{y}_i\}_{i=1}^n$. The elements of the vector $\tilde{\psi}_n(y)$ are functions of y and sample statisitics, which are often sample moments. Therefore, we term it the vector of moment functions, or moment function for short. In the second step, the parameters of the density $p(y|\rho)$ are estimated by finding the parameter value $\hat{\rho}_n$ that puts the vector of moment functions $\tilde{\psi}_n(y)$ as nearly as possible in the null space of the operator $\mathcal{M}_\rho$, where distance is measured by a chi-squared criterion function. We explain.

Consider three moment functions $\tilde{\psi}_{p,n}$, $\tilde{\psi}_{c,n}$, and $\tilde{\psi}_{f,n}$ defined as follows:

$$\tilde{\psi}_{p,n}(y) = \frac{\partial}{\partial\rho}\log p(y|\tilde{\rho}_n),$$

$$\tilde{\psi}_{c,n}(y) = \begin{pmatrix} y - \frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i \\ y^2 - \frac{1}{n}\sum_{i=1}^{n} (\tilde{y}_i)^2 \\ \vdots \\ y^K - \frac{1}{n}\sum_{i=1}^{n} (\tilde{y}_i)^K \end{pmatrix},$$

$$\tilde{\psi}_{f,n}(y) = \frac{\partial}{\partial\theta} \log f(y|\tilde{\theta}_n),$$

where $f(y|\theta)$ is a density that closely approximates the true data generating process in a sense made precise later, and

$$\tilde{\rho}_n = \underset{\rho}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \log p(\tilde{y}_i|\rho),$$

$$\tilde{\theta}_n = \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \log f(\tilde{y}_i|\theta).$$

The moment functions $\tilde{\psi}_{p,n}$, $\tilde{\psi}_{c,n}$, and $\tilde{\psi}_{f,n}$ correspond to maximum likelihood, classical method of moments, and efficient method of moments, respectively. Note that each of them is in the null space of $\mathcal{M}_{\tilde{F}_n}$. That is,

$$\mathcal{M}_{\tilde{F}_n}\tilde{\psi}_{p,n} = \mathcal{M}_{\tilde{F}_n}\tilde{\psi}_{c,n} = \mathcal{M}_{\tilde{F}_n}\tilde{\psi}_{f,n} = 0.$$

Define

$$\tilde{\mathcal{I}}_n = \mathcal{M}_{\tilde{F}_n}(\tilde{\psi}_n)(\tilde{\psi}_n)'.$$

Once $\tilde{\psi}_n$ and $\tilde{\mathcal{I}}_n$ have been computed, the data have been summarized, and the projection step is finished.

To distinguish between the projection and estimation steps, we shall adhere to the convention that quantities corresponding to the projection step are indicated by a tilde, e.g. $\tilde{\psi}_{p,n}$, $\{\tilde{y}_i\}_{i=1}^{n}$, and $\tilde{\rho}_n$, and quantities corresponding to the estimation step are indicated by a circumflex, e.g. $\hat{\rho}_n$ to denote an estimator, and $\{\hat{y}_i\}_{i=1}^{N}$ to denote a simulation from $p(y|\rho)$ for computing $\mathcal{M}_\rho$ by simulation, and $\{(\hat{w}_i, \hat{y}_i)\}_{i=1}^{N}$ to denote weights and points for computing integrals by quadrature.

To implement the estimation step, define

$$m_n(\rho) = \mathcal{M}_\rho \tilde{\psi}_n.$$

If the dimensions of ρ and $\tilde{\psi}_n(y)$ are the same, then usually the equations

$$m_n(\rho) = 0$$

can be solved to obtain an estimator $\hat{\rho}_n$. For $\tilde{\psi}_{p,n}$, the solution is the maximum likelihood estimator (Gauss, 1816; Fisher, 1912). For $\tilde{\psi}_{c,n}$ with $K = p_\rho$, it is the classical method of moments estimator (Pearson, 1894). For $\tilde{\psi}_{c,n}$ with $K > p_\rho$, no solution exists and the moment functions $\tilde{\psi}_{c,n}$ are those of generalized method of moments or minimum chi-squared (Neyman and Pearson, 1928; Hansen, 1982) as customarily implemented.

When $K > p_\rho$, then $\tilde{\psi}_n$ cannot be placed in the null space of the operator $\mathcal{M}_\rho$ for any ρ, because the equations $m_n(\rho) = 0$ have no solution. In this case, the minimum chi-squared estimator relies on the fact that, under standard regularity conditions (Gallant and Tauchen, 1996) and choices of $\tilde{\psi}_n$ similar to the above, there is a function ψ^o such that

$$\begin{aligned}
\lim_{n\to\infty} \tilde{\psi}_n(y) &= \psi^o(y) \quad \text{a.s.} \\
\lim_{n\to\infty} \tilde{\mathcal{I}}_n &= \mathcal{E}_{\rho^o}(\psi^o)(\psi^o)' \quad \text{a.s.} \\
\sqrt{n}\, m_n(\rho^o) &\xrightarrow{\mathcal{L}} N\Big[0,\, \mathcal{E}_{\rho^o}(\psi^o)(\psi^o)'\Big]
\end{aligned}$$

where $\mathcal{E}_{\rho^o}$ denotes expectation taken with respect to $p(y|\rho^o)$. For the three choices $\tilde{\psi}_{p,n}$, $\tilde{\psi}_{c,n}$, and $\tilde{\psi}_{f,n}$ of $\psi_n(y)$ above, the functions ψ_p^o, ψ_c^o, and ψ_f^o given by this result are

$$\psi_p^o(y) = \frac{\partial}{\partial\rho} \log p(y|\rho^o)$$

$$\psi_c^o(y) = \begin{pmatrix} y - \mathcal{E}_{\rho^o}(y) \\ y^2 - \mathcal{E}_{\rho^o}(y^2) \\ \vdots \\ y^K - \mathcal{E}_{\rho^o}(y^K) \end{pmatrix}$$

and

$$\psi_f^o(y) = \frac{\partial}{\partial\theta} \log f(y|\theta^o),$$

where

$$\theta^o = \operatorname*{argmin}_\theta \mathcal{E}_{\rho^o} \log f(\cdot|\theta).$$

With these results in hand, ρ may be estimated by minimum chi-squared, viz.,

$$\hat{\rho}_n = \underset{\rho}{\operatorname{argmin}}\, m_n'(\rho)\,(\tilde{\mathcal{I}}_n)^{-1} m_n(\rho)$$

and

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\Big[0,\ (V^o)^{-1}\Big],$$

where

$$V^o = \Big[\mathcal{E}_{\rho^o}\,(\psi_p^o)\,(\psi^o)'\Big]\Big[\mathcal{E}_{\rho^o}\,(\psi^o)\,(\psi^o)'\Big]^{-1}\Big[\mathcal{E}_{\rho^o}\,(\psi^o)\,(\psi_p^o)'\Big].$$

Note that for any $a \neq 0$,

$$\min_b \mathcal{E}_{\rho^o}\Big[a'\psi_p^o - (\psi^o)'b\Big]^2 = \mathcal{E}_{\rho^o}\Big(a'\psi_p^o\Big)^2 - a'V^o a \geq 0.$$

This expression implies that $a'V^o a$ cannot exceed

$$\mathcal{E}_{\rho^o}\Big(a'\psi_p^o\Big)^2 = a'\Big[\mathcal{E}_{\rho^o}\,(\psi_p^o)\,(\psi_p^o)'\Big]a$$

and therefore the best achievable asymptotic variance of the estimator $\hat{\rho}_n$ is

$$(\mathcal{I}_p^o)^{-1} = \Big[\mathcal{E}_{\rho^o}\,(\psi_p^o)\,(\psi_p^o)'\Big]^{-1},$$

which is the variance of the maximum likelihood estimator of ρ. It is also apparent from this expression that if $\{\psi_i^o\}_{i=1}^{\infty}$ spans the $L_{2,p}$ probability space

$$L_{2,p} = \{g : \mathcal{E}_{\rho^o} g^2 < \infty\}$$

and

$$\psi^o = (\psi_1^o, \ldots, \psi_K^o),$$

then $\hat{\rho}_n$ has good efficiency relative to the maximum likelihood estimator for large K. The polynomials span $L_{2,p}$ if $p(y|\rho)$ has a moment generating function (Gallant, 1980). Therefore, one might expect good asymptotic efficiency from $\tilde{\psi}_{c,n}$ for large K.

Rather than just spanning $L_{2,p}$, EMM requires, in addition, that the moment functions actually be the score vector $\psi_{f,n}(y)$ of some density $f(y|\tilde{\theta}_n)$ that closely approximates $p(y|\rho^o)$. Possible choices of $f(y|\tilde{\theta}_n)$ are discussed in Gallant and Tauchen (1996). Of them, one commonly used in applications is the SNP density, which was proposed by Gallant and

Nychka (1987) in a form suited to cross-sectional applications and by Gallant and Tauchen (1989) in a form suited to time-series applications. This density, in the present notation, is

$$f_K(y|\theta) = \frac{\left\{\sum_{i=0}^{K} c_i w_i\left[(y-\mu)/\sigma\right]\right\}^2}{\sigma \sum_{i=0}^{K} c_i^2},$$

where

$$w_0(y), w_1(y), \ldots, w_K(y) \tag{1}$$

are the Hermite polynomials, and $\theta = (\mu, \sigma, c_0, \ldots, c_K)$ with the normalization rule $c_0 \equiv 1$ imposed. The Hermite polynomials are obtained by orthonormalizing the sequence

$$(\sqrt{\phi}, y\sqrt{\phi}, y^2\sqrt{\phi}, \cdots, y^K\sqrt{\phi}), \tag{2}$$

where ϕ is the standard normal density function. Recursive formulas for the orthonormalized sequence (1) are given in Chapter 22 of Abramowitz and Stegun (1972), which when related to the present application have the form

$$\begin{array}{ll} i = 0 & w_0(x) = (\sqrt{2\pi})^{-1/2} e^{-x^2/4} \\ i = 1 & w_1(1) = (\sqrt{2\pi})^{-1/2} x e^{-x^2/4} \\ i \geq 2 & w_i(x) = [x w_{i-1}(x) - \sqrt{i-1}\, w_{i-2}(x)]/\sqrt{i} \end{array}$$

(Fenton and Gallant, 1996). Gallant and Long (1997) have shown that

$$\psi_f^o(y) = \frac{\partial}{\partial\theta} \log f_K(y|\theta^o),$$

with

$$\theta^o = \operatorname*{argmin}_{\theta} \mathcal{E}_{p^o} \log f_K(\cdot|\theta)$$

spans $L_{2,p}$.

While a spanning argument can be used to show that high efficiency obtains for large K, it gives no indication as to what might be the best choice of moment functions with which to span $L_{2,p}$. Moreover, if ψ_p is in the span of ψ^o some finite K, then full efficiency obtains at once (Gallant and Tauchen, 1996). For instance, the score of the normal density is in the span of both $\tilde{\psi}_{c,n}$ and $\tilde{\psi}_{f,n}$ for $K \geq 2$. These considerations seem to rule out any hope of general results showing that one moment function should, in general, be better than another.

Whether one moment function is better than another might not be of much importance if one were about as cheap to compute as another. However, computing $\tilde{\psi}_{f,n}$ according established protocol (Gallant, Rossi, and Tauchen, 1992; Andersen and Lund, 1997a, 1997b) involves nonlinear optimizations and is potentially costly in terms of both researcher and machine time. One would like some assurance that there is a payoff to this effort.

With general results unattainable, the best one can do is compare efficiencies over a class of densities designed to stress-test an estimator and over some densities thought to be representative of situations likely to be encountered in practice to see if any conclusions seem to be indicated. That is what we do next, and we find evidence that the extra effort to compute $\tilde{\psi}_{f,n}$ has a payoff.

Our standard of comparison will be the volume of a confidence region on the parameters of the density $p(y|\rho)$ computed using the asymptotic distribution of $\hat{\rho}_n$. This region has the form $\{\rho : (\rho - \rho^o)'(V^o)^{-1}(\rho - \rho^o) \leq \mathcal{X}_p^2/n\}$ with volume

$$\frac{2\pi^{p/2}(\mathcal{X}_p^2/n)^p}{p\Gamma(p/2)\det(V^o)}.$$

As small volumes are to be preferred, and the region $\{\rho : (\rho - \rho^o)'\mathcal{I}_p^o(\rho - \rho^o) \leq \mathcal{X}_p^2/n\}$ has the smallest achievable volume, we use

$$\mathrm{RE} = \frac{\det(V^o)}{\det(\mathcal{I}_p^o)}$$

as our measure of relative efficiency in comparisons.

3 Marron-Wand Test Suite

Marron and Wand (1992) proposed a test suite as a battery for use in evaluating density estimators. It is designed to evaluate the ability of an estimator to track complex behavior at the center of the distribution. The densities in the suite are mixtures of normals; the mixing proportions $\{\pi_j\}_{j=1}^J$ and the corresponding means $\{\mu_j\}_{j=1}^J$ and standard deviations $\{\sigma_j\}_{j=1}^J$ of the constituent normal densities are given in Marron and Wand (1992). Here, we treat the means and variances as unknown parameters to be estimated and assume that the mixing proportions are known. In addition, we assume that the means $\{\mu_j\}$ are separated, that is, contained in disjoint open intervals. Under these assumptions, the maximum likelihood estimator of $\rho = (\mu_1, \sigma_1, \ldots, \mu_J, \sigma_J)$ is consistent and asymptotically normally distributed with a score that is easy to compute analytically. Figure 1 displays these densities in order of increasing challenge to the method of moments.

For each density $p(\cdot|\rho^o)$ in the test suite we compute

$$\begin{aligned}
\theta^o &= \underset{\theta}{\operatorname{argmin}}\, \mathcal{E}_{\rho^o} \log f_K(\cdot|\theta) \\
\sqrt{\text{mse}} &= \sqrt{\mathcal{E}_{\rho^o}[p(\cdot|\rho^o) - f_K(\cdot|\theta^o)]^2} \\
\text{KL} &= \mathcal{E}_{\rho^o}[\log f_K(\cdot|\theta^o) - \log p(\cdot|\rho^o)] \\
\mathcal{I}_p^o &= \mathcal{E}_{\rho^o} (\psi_p^o)(\psi_p^o)' \\
V_\alpha^o &= \left[\mathcal{E}_{\rho^o} (\psi_p^o)(\psi_\alpha^o)'\right]\left[\mathcal{E}_{\rho^o} (\psi_\alpha^o)(\psi_\alpha^o)'\right]^{-1}\left[\mathcal{E}_{\rho^o} (\psi_\alpha^o)(\psi_p^o)'\right]
\end{aligned}$$

for $\alpha = f, p, c$ and $K = 2, \ldots, 95$ using numerical methods discussed below. From these quantities we compute relative efficiencies

$$\text{RE}_\alpha = \frac{\det(V_\alpha^o)}{\det(\mathcal{I}_p^o)}$$

for $\alpha = f, p, c$. These, together with $\sqrt{\text{mse}}$ are displayed in Tables 1 through 5 in order of increasing difficulty of estimation by the method of moments, as is apparent from the tables.

Over the Marron-Wand test suite, EMM completely dominates CMM as seen by comparing columns emm and cmm of Tables 1 through Tables 5.

The SNP estimator uses linear and quadratic information in the data twice: once to compute the sample mean and variance to center and scale the data and once again to compute the first and second degree terms of an expansion in Hermite polynomials. Technically, this would allow an EMM estimator on $K = 2$ to identify a four parameter model. As one might expect, this redundant sample information is not utilized very effectively, which may be seen by inspecting the the columns emm of Tables 1 through 5 that have entries in row $K = 2$. We have aligned columns emm and cmm of Tables 1 through 5 on degree K rather than on the number of identifying restrictions because it seems unreasonable to compare, for instance, an estimator that uses quadratic sample information with an estimator that uses quartic information. Nonetheless, if one wishes to compare EMM and CMM on the basis of the number of identifying restrictions, it can be done by shifting column cmm up two rows in Tables 1 through 5. This would not affect conclusions: Other than a small number of instances where CMM has better efficiency than EMM in a few initial rows of Tables 1 and 2, EMM dominates CMM.

In inspecting Tables 1 through 5, be aware that theory does not imply that RE_f is nondecreasing in K. Theory does imply that RE_c is nondecreasing, whereas there are some slight violations of this condition in the tables. This is due to the effect of rounding error in computing $\left[\mathcal{E}_{\rho^o}\left(\psi_\alpha^o\right)\left(\psi_\alpha^o\right)'\right]^{-1}$ as discussed below. No conclusions are affected by these slight violations of monontonicity.

Figure 1 about here

Table 1 about here

Table 2 about here

Table 3 about here

Table 4 about here

Table 5 about here

As can be seen from columns $\sqrt{\text{mse}}$ and emm of Tables 1 through 5, once $f_K(\cdot|\theta^o)$ begins to approximate $p(\cdot|\rho^o)$ accurately, efficiency begins to increase rapidly. Figure 2 illustrates this phenomenon. Once $f_K(\cdot|\theta^o)$ has detected the third mode of the trimodal density, efficiency increases rapidly. Figure 3 illustrates the same phenomenon but also indicates that extraordinary accuracy is required for spiked densities. Excessive spikiness will defeat both EMM and CMM as seen in Table 5.

———————————

Figure 2 about here

———————————

———————————

Figure 3 about here

———————————

The numerical methods employed are as follows. All code is in C++. The major software components are available by anonymous ftp from host ftp.econ.duke.edu in directories pub/arg/libcpp and pub/arg/npe. Throughout, the operator $\mathcal{E}_{\rho^o}$ is computed by simulation, averaging over realizations of length 50,000. The random number generator employed, and other simulation details are described in Section 4.3 of Fenton and Gallant (1996). Also described there are the details of the computation of $f_K(\cdot|\theta^o)$ and θ^o. As mentioned above, accurate computation of $\left[\mathcal{E}_{\rho^o}(\psi_\alpha^o)(\psi_\alpha^o)'\right]^{-1}$ proved a nettlesome problem for large K. We computed the inverse using a translation to C++ of the singular value decomposition algorithm given by Businger and Golub (1969). Because the singular values (eigen-values in this instance) of $\mathcal{E}_{\rho^o}(\psi_\alpha^o)(\psi_\alpha^o)'$ decline slowly to near machine zero for large K, one must assign a pseudo-rank to this matrix by declaring all singular values past some point to be zero. The consequence of insisting that all singular values are positive is ludicrously inaccurate computed relative efficiencies. Pseudo-rank was determined by declaring all singular values S[i] less than S[1]*eps to be zero. For all pseudo-ranks between those suggested by eps=1.0e-8 and eps=1.0e-15 that rank giving the largest computed value of RE_α was the one selected for inclusion in the tables. Determinants were computed as the product of the singular values of V_α^o. All remaining computations are straightforward matrix algebra implemented though a C++ matrix class found in pub/arg/libcpp.

4 Scale Mixture of Normals

The stochastic volatility model is one of the standard models for statistical analysis of financial market data. A recent summary of the literature is Ghysels, Harvey, and Renault (1995). Although specialized estimators for various specifications of the stochastic volatility model have been proposed, simulation estimators are often used instead because they permit one to change specifications with little effort and they provide informative diagnostic tests of a tentative specification. A recent example is Gallant, Hsieh, and Tauchen (1997). Andersen, Chung, and Sorensen (1997) have compared EMM to CMM for the stochastic volatility model using Monte Carlo methods. They studied the expected mean squared error in estimation and the accuracy of asymptotic rejection probabilities of related test statistics. On the whole, they find that EMM is an improvement over CMM according to these criteria. The work reported here complements their results because it addresses questions of relative efficiency that require computations that are too computationally intensive for the Monte Carlo method. A more analytic approach, as used here, must be employed to keep computational complexity within feasible limits. The model we study is the classic lognormal scale mixture of normals considered by Clark (1973) and Tauchen and Pitts (1983).

A lognormal scale mixture of normals has density function

$$\begin{aligned} p(y|\rho) &= \int_{-\infty}^{\infty} n(y|\rho_1, e^{2u})\, n(u|\rho_2, \rho_3^2)\, du & (3) \\ &= \int_{-\infty}^{\infty} n[y|\rho_1, \exp^2(\sqrt{2}\rho_3 v + \rho_2)]\, (\pi)^{-1/2} e^{-v^2}\, dv & (4) \end{aligned}$$

where $n(u|\alpha, \beta^2)$ denotes the normal density with mean α and variance β^2. Expression (4) is suited to Gaussian quadrature which we implement using a 15 point Gauss-Hermite rule

$$p(y|\rho) = \sum_{i=-7}^{7} n[y|\rho_1, \exp^2(\sqrt{2}\rho_3 \hat{v}_i + \rho_2)]\, (\pi)^{-1/2} \hat{w}_i \tag{5}$$

with weights $\{\hat{w}_i\}_{i=-7}^{7}$ and ordinates $\{\hat{v}_i\}_{i=-7}^{7}$ computed using an algorithm due to Golub and Welsch (1969).

The density defined by (5) is a scale mixture of normals with respect to the discrete density with support points $\{\hat{v}_i\}_{i=-7}^{7}$ and probabilities $\{\hat{p}_i = (\pi)^{-1/2}\hat{w}_i\}_{i=-7}^{7}$. Therefore, we can use the numerical methods described in Section 3 to compute relative efficiency.

The mean μ, variance σ^2, and kurtosis κ of $p(y|\rho)$ as given by (3) determine ρ as follows

$$\begin{aligned} \rho_1 &= \mu \\ \rho_2 &= \log\sigma - \rho_3^2 \\ \rho_3 &= \sqrt{[\log(\kappa/3)/4]}. \end{aligned}$$

We find that RE_α depends only on κ and is invariant to choice of μ and σ^2. However, rank determination difficulties, as discussed in Section 3, are not invariant to (μ, σ^2) and we find that putting $(\mu, \sigma^2) = (0, 1/4)$ leads to reasonably stable computations.

The densities $p(\cdot|\rho)$ for the choices $(\mu, \sigma^2) = (0, 1/4)$ and $\kappa = 4, 6, 8, 10$ are plotted in Figure 4. Relative efficiencies together with the root mean squared error of the SNP auxiliary model for these choices are displayed in Table 6 for $K = 2, \ldots, 29$.

As seen by comparing the columns labeled emm and cmm of Table 6, EMM dominates CMM. In Table 6, estimators are aligned by equating the degree of polynomial used in estimation. If, instead, one equates the number of identifying restrictions by shifting the columns labeled cmm in Table 6 up two rows, then EMM dominates CMM except for $K < 4$ and $\kappa = 4, 6$.

Table 6 about here

Figure 4 about here

---

5 Second Largest Order Statistic of the Lognormal

An area of current interest in simulation estimation is the analysis of auction data; see for instance, Laffont, Ossard, and Vuong (1995). Under an independent private value assumption, the values assigned to an object by n bidders are a random sample of size n from a common valuation distribution. In either a single object second price auction or a single object English (oral ascending bid) auction, the observed selling price is the second largest order statistic in this sample. In an sealed bid first price auction, and other auction formats, the winning bid is a functional of the order statistics of the valuation distribution. These observations suggest that the relative performance of the EMM estimator in estimating the second largest order statistic in a sample of size n has relevance to the application of simulation estimators to auction data. Here we consider the case when the common valuation distribution is lognormal.

The second largest order statistic from a sample of size $\rho_1 = n$ from the lognormal distribution has density function

$$p(y|\rho) = \frac{n(n-1)}{y-\rho_2}\left\{\Phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right]\right\}^{n-2}\left\{1-\Phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right]\right\}\phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right] \quad (6)$$

where $y > \rho_2$ and ϕ and Φ denote the standard normal density and distribution functions, respectively. This density has mean $\mu = \rho_2 + e^{1/2}$ and variance $\sigma^2 = \rho_3^2[e^2 - e]$. The numerical methods described in Section 3 can be used to compute relative efficiency with $p(y|\rho)$ and $(\partial/\partial\rho)\log p(y|\rho)$ evaluated analytically using (6). For the Monte Carlo integration, one can generate a simulated value $\hat{y}_i$ from (6) by (i) generating a random sample $\{z_j\}_{j=1}^{n}$ from ϕ with a normal random number generator, (ii) sorting the $\{z_j\}$ and selecting the second largest to get $\hat{z}_i$, and (iii) putting $\hat{y}_i = \rho_3 \log(\hat{z}_i) + \rho_2$. The specific random number and heapsort algorithms that we employ are in directory pub/arg/libcpp on ftp.econ.duke.edu. Similarly to Section 4, RE_α is invariant to μ and σ^2 but the problem of rank determination is not. We find that putting $(\mu, \sigma^2) = (0, 1/8)$ leads to reasonably stable computations.

The densities $p(\cdot|\rho)$ for the choices $(\mu, \sigma^2) = (0, 1/8)$ and $n = 5, 50, 100, 200$ are plotted in Figure 5. Relative efficiency together with the root mean squared error of the SNP auxiliary model for these choices are displayed in Table 7 for $K = 2, \ldots, 29$.

As seen by comparing the columns labeled emm and cmm of Table 7, EMM dominates CMM. This is true whether one equates the degree of polynomial used in estimation, as in the table, or equates the number of identifying restrictions by shifting the columns labeled cmm in Table 7 up two rows.

———————————————

Table 7 about here

———————————————

———————————————

Figure 5 about here

———————————————

6 Conclusion

Over the models considered, which are the Marron-Wand test suite, a scale mixture of normals, and the second largest order statistic from the lognormal distribution, efficient method of moments dominates conventional method of moments. The relative efficiency of EMM is uniformly higher than the relative efficiency of CMM.

This finding requires that the EMM and CMM estimators be aligned for the purpose of comparison on the basis of the highest degree monomial used by each estimator in forming the moment functions of the minimum chi-squared criterion. This seems the most reasonable basis of comparison because it allows both estimators access to the same sample information. However, the SNP auxiliary model used to form the EMM moment function uses linear and quadratic information twice: Once to determine the location and scale of the data and once again to form Hermite polynomials. Therefore, at given degree of polynomial, EMM has two more moment functions available to identify parameters than CMM.

If one feels that the number of identifying restrictions is a more reasonable alignment, then EMM dominates CMM except for a few instances at degree less than four for the four of the easiest to estimate models in the Marron-Wand test suite and two low kurtosis scale mixture models. Therefore, the basic conclusion is not altered by alignment on equivalent identifying restrictions.

7 References

Abramowitz, M., and I. A. Stegun (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

Andersen, T. G., and J. Lund (1997a) "Estimating Continuous Time Stochastic Volatility Models of the Short Term Interest Rate", *Journal of Econometrics* 77, 343–378.

Andersen, T. G., and J. Lund (1997b) "Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Sources of Steepness, level, and Curvature of the Yield Curve," Manuscript, Northwestern University, Kellogg School of Management.

Andersen, T. G., H-J Chung, and B. E. Sorensen (1997) "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," Manuscript, Northwestern University, Kellogg School of Management.

Businger, P. A., and G. H. Golub (1969) "Singular Value Decomposition of a Complex Matrix," *Communications of the ACM* 12, 564–565.

Clark, P. K. (1973) A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41, 135–156.

Chumacero, R. (1997) "Finite Sample Properties of the Efficient Method of Moments," Manuscript, University of Chile, Department of Economics.

Duffie, D., and K. J. Singleton (1993) "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica* 61, 929–952.

Fenton, V. M., and A. R. Gallant (1996) "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics* 74, 77–118.

Fisher, R. A. (1912) "On an Absolute Criterion for Fitting Frequency Curves," *Messages of Mathematics* 41, 155–157.

Gallant, A. R. (1980). "Explicit Estimators of Parametric Functions in Nonlinear Regression," *Journal of the American Statistical Association* 75, 182–93.

Gallant, A. R., D. A. Hsieh and G. Tauchen (1997) "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, forthcoming.

Gallant, A. R., and J. Long (1997) Estimating Stochastic Differential Equations Efficiently by Minimum Chi-square. *Biometrika* 84, 125–141.

Gallant, A. R., and D. W. Nychka (1987) Semi-nonparametric Maximum Likelihood Estimation. *Econometrica* 55, 363–390.

Gallant, A. R., P. E. Rossi and G. Tauchen (1992) "Stock Prices and Volume," *Review of Financial Studies* 5, 199–242.

Gallant, A. R., and G. Tauchen (1989) Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications. *Econometrica* 57, 1091–1120.

Gallant, A. R., and G. Tauchen (1996) Which Moments to Match. *Journal of Econometric Theory* 12, 657–681.

Gauss, C. F. (1816) "Bestimmung der Genauigkeit der Beobachtungen," *Zeitschrift für Astronomie und verwandte Wissenschaften* 1, 15–25.

Ghysels, E., A. Harvey, and E. Renault (1995) "Stochastic Volatility," in ed. G. S. Maddala, *Handbook of Statistics, Vol. 14, Statistical Methods in Finance*, Amsterdam: North Holland.

Golub, G. H., and J. H. Welsch (1969) "Calculation of Gaussian Quadrature Rules," *Mathematics of Computation* 23, 221–230.

Gourieroux, C., A. Monfort, and E. Renault (1993) "Indirect Inference," *Journal of Applied Econometrics* 8, S85–S118.

Hansen, L. P. (1982) "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica* 50, 1029–1054.

Ingram, B. F. and B. S. Lee (1991) "Simulation Estimation of Time Series Models," *Journal of Econometrics* 47, 197–250.

Laffont, J.-J., H. Ossard, and Q. Vuong (1995) "The Econometrics of First Price Auctions," *Econometrica* 63, 953–980.

Marron, J. S., and M. P. Wand (1992) "Exact mean integrated squared error," *The Annals of Statistics* 20, 712–736.

Neyman, J., and E. S. Pearson (1928) "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika* 20A, 175–240, 263–294.

Ng, S., and A. Michaelides (1997) "Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of Three Simulation Estimators," Manuscript, Boston College, Department of Economics.

Pearson, K. (1894) "Contributions to the Mathematical Theory of Evolution," *Philosophical Transactions of the Royal Society of London, Series A* 185, 71–78.

Smith, A. A. (1993) "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Tauchen, G. (1997b) "The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space," *Review of Economics and Statistics,* forthcoming.

Tauchen, G., and M. Pitts (1983) The Price Variability-Volume Relationship on Speculative Markets. *Econometrica* 51, 485–505.

Tables and Figures

Table 1. Marron-Wand Test Suite (a).

	gaussian			bimodal			skewed_unimodal		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.001	1.000	1.000	0.071	0.282		0.058		
3				0.069	0.328		0.055		
4				0.008	0.872	0.533	0.016	0.010	
5				0.007	0.977	0.597	0.007	0.620	
6				0.004	0.992	0.773	0.006	0.763	0.065
7				0.004	0.995	0.820	0.005	0.889	0.329
8				0.003	0.997	0.892	0.003	0.947	0.548
9				0.003	0.997	0.918	0.003	0.963	0.687
10				0.002	0.999	0.949	0.004	0.985	0.815
11				0.002	0.999	0.962	0.003	0.994	0.862
12				0.001	0.999	0.976	0.003	0.998	0.918
13							0.003	0.999	0.947
14							0.002	0.999	0.973

Notes: K is the degree of the Hermite polynomial in the case of SNP and EMM estimators and is both the number of moments and the the degree of the highest monomial in the case of the CMM estimator. $\sqrt{\text{mse}}$ is root mean squared error of the SNP estimator of degree K. emm is the efficiency of the EMM estimator relative to the maximum likelihood estimator using an SNP auxiliary model of degree K. cmm is the efficiency of the CMM estimator relative to the maximum likelihood estimator using K moments. Initial blanks in a column indicate that the model being estimated is not identified for that value of K. Columns emm and cmm are aligned on degree, which roughly represents equivalent sample information for EMM and CMM estimators. To align columns on equivalent number of identifying restrictions, shift column cmm up two rows.

Table 2. Marron-Wand Test Suite (b).

	skewed_bimodal			separated_bimodal			trimodal		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.079	0.004		0.168	0.011		0.081		
3	0.047	0.042		0.023	0.001		0.080		
4	0.044	0.096	0.017	0.021	0.847	0.031	0.080	0	
5	0.040	0.174	0.046	0.015	0.727	0.294	0.079	0	
6	0.040	0.243	0.092	0.014	0.944	0.356	0.023	0.009	0
7	0.025	0.348	0.203	0.013	0.944	0.555	0.023	0.021	0.001
8	0.019	0.683	0.279	0.012	0.980	0.591	0.023	0.039	0.004
9	0.010	0.791	0.370	0.011	0.987	0.714	0.023	0.049	0.007
10	0.009	0.792	0.544	0.004	0.993	0.736	0.022	0.071	0.016
11	0.009	0.902	0.609	0.002	0.994	0.809	0.022	0.073	0.025
12	0.008	0.934	0.696	0.002	0.994	0.824	0.022	0.099	0.047
13	0.006	0.973	0.824	0.002	0.995	0.869	0.022	0.147	0.068
14	0.006	0.979	0.844	0.002	0.995	0.881	0.014	0.414	0.106
15	0.006	0.987	0.897	0.002	0.995	0.909	0.007	0.716	0.143
16	0.005	0.997	0.946	0.002	0.996	0.918	0.004	0.863	0.197
17	0.004	0.998	0.948	0.003	0.996	0.935	0.003	0.939	0.251
18	0.004	0.999	0.970	0.002	0.997	0.942	0.003	0.972	0.318
19	0.004	0.999	0.977	0.003	0.997	0.953	0.003	0.975	0.383
20				0.003	0.998	0.959	0.003	0.987	0.455
21				0.003	0.998	0.961	0.003	0.988	0.503
22				0.003	0.998	0.964	0.003	0.995	0.525
23				0.003	0.998	0.967	0.003	0.996	0.547
24				0.003	0.999	0.968	0.003	0.999	0.613
25							0.003	0.998	0.618
26							0.003	0.999	0.634
27							0.003	0.999	0.700

Notes: See notes for Table 1.

Table 3. Marron-Wand Test Suite (c).

	outlier			kurtotic_unimodal			claw		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	1.340	0		0.522	0		0.144		
3	1.338	0		0.509	0		0.144		
4	1.338	0.002	0	0.506	0	0	0.144		
5	0.901	0.018	0	0.393	0.001	0	0.135		
6	0.902	0.022	0.001	0.364	0.002	0	0.133		
7	0.901	0.022	0.002	0.364	0.002	0	0.133		
8	0.902	0.037	0.005	0.277	0.012	0	0.133		
9	0.500	0.157	0.008	0.277	0.012	0	0.133		
10	0.499	0.169	0.015	0.277	0.019	0	0.133	0	
11	0.499	0.205	0.020	0.248	0.028	0.001	0.133	0	
12	0.498	0.216	0.033	0.195	0.062	0.001	0.133	0	0
13	0.276	0.446	0.043	0.195	0.062	0.002	0.133	0	0
14	0.276	0.486	0.065	0.195	0.086	0.002	0.133	0	0
15	0.276	0.506	0.080	0.174	0.105	0.004	0.133	0	0
16	0.163	0.659	0.114	0.139	0.172	0.005	0.133	0	0
17	0.163	0.668	0.136	0.136	0.189	0.007	0.133	0	0
18	0.163	0.710	0.176	0.136	0.190	0.010	0.128	0	0
19	0.163	0.723	0.204	0.120	0.251	0.014	0.128	0	0
20	0.106	0.799	0.258	0.106	0.321	0.017	0.125	0	0
21	0.106	0.805	0.288	0.091	0.371	0.024	0.125	0	0
22	0.106	0.810	0.306	0.091	0.411	0.029	0.125	0	0
23	0.106	0.846	0.350	0.080	0.431	0.029	0.125	0	0
24	0.074	0.883	0.364	0.073	0.481	0.039	0.015	0.001	0
25	0.074	0.887	0.394	0.068	0.502	0.041	0.015	0.001	0
26	0.074	0.909	0.409	0.068	0.540	0.046	0.015	0.001	0
27	0.074	0.914	0.409	0.062	0.568	0.049	0.015	0.001	0
28	0.055	0.929	0.457	0.053	0.644	0.050	0.015	0.001	0
29	0.055	0.932	0.457	0.046	0.693	0.071	0.015	0.002	0
30	0.055	0.946	0.482	0.044	0.734	0.071	0.015	0.002	0
32	0.030	0.975	0.515	0.037	0.790	0.072	0.014	0.002	0
34	0.029	0.982	0.543	0.025	0.865	0.080	0.014	0.004	0
36	0.019	0.988	0.581	0.024	0.872	0.102	0.014	0.005	0
38	0.019	0.991	0.599	0.017	0.932	0.106	0.014	0.011	0
40	0.017	0.992	0.638	0.017	0.948	0.120	0.012	0.031	0
42	0.018	0.995	0.652	0.015	0.955	0.129	0.008	0.257	0
44	0.017	0.993	0.675	0.014	0.974	0.153	0.008	0.348	0
46	0.017	0.996	0.688	0.011	0.986	0.160	0.007	0.475	0
48	0.016	0.999	0.715	0.011	0.988	0.172	0.007	0.641	0
50				0.011	0.990	0.180	0.007	0.700	0
52				0.011	0.990	0.209	0.007	0.785	0
54				0.011	0.994	0.215	0.007	0.837	0
56				0.011	0.995	0.223	0.006	0.876	0
58				0.011	0.997	0.238	0.007	0.903	0
60				0.011	0.998	0.248	0.007	0.926	0
65				0.011	0.999	0.289	0.007	0.965	0
70							0.005	0.990	0

Notes: See notes for Table 1.

Table 4. Marron-Wand Test Suite (d).

	smooth_comb			discrete_comb			strongly_skewed		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.171			0.176			0.490		
3	0.128			0.164			0.400		
4	0.127			0.164			0.400		
5	0.110			0.127			0.346		
6	0.100			0.126			0.345		
7	0.086			0.085			0.333		
8	0.086			0.084			0.293		
9	0.086			0.079			0.284		
10	0.081	0		0.069	0		0.284		
11	0.081	0		0.069	0		0.254		
12	0.078	0	0	0.066	0	0	0.253		
13	0.076	0	0	0.066	0	0	0.224		
14	0.068	0	0	0.066	0	0	0.224	0	
15	0.068	0	0	0.066	0	0	0.217	0	
16	0.068	0	0	0.066	0	0	0.194	0	0
17	0.063	0	0	0.066	0	0	0.193	0	0
18	0.063	0	0	0.066	0	0	0.193	0	0
19	0.063	0	0	0.066	0	0	0.170	0	0
20	0.059	0	0	0.066	0	0	0.169	0	0
21	0.059	0	0	0.066	0	0	0.169	0	0
22	0.059	0	0	0.064	0	0	0.145	0	0
23	0.057	0	0	0.063	0	0	0.145	0	0
24	0.057	0	0	0.061	0	0	0.145	0	0
25	0.057	0	0	0.061	0	0	0.124	0	0
26	0.054	0	0	0.061	0	0	0.125	0	0
27	0.054	0	0	0.060	0	0	0.125	0	0
28	0.054	0	0	0.060	0	0	0.113	0	0
29	0.052	0	0	0.060	0	0	0.114	0	0
30	0.052	0	0	0.060	0	0	0.114	0	0
32	0.049	0.001	0	0.060	0	0	0.099	0	0
34	0.049	0.001	0	0.060	0.001	0	0.089	0	0
36	0.046	0.002	0	0.060	0.003	0	0.085	0	0
38	0.043	0.003	0	0.056	0.013	0	0.077	0	0
40	0.042	0.007	0	0.049	0.033	0	0.070	0	0
42	0.041	0.012	0	0.046	0.066	0	0.067	0	0
44	0.039	0.015	0	0.039	0.086	0	0.064	0	0
46	0.038	0.020	0	0.030	0.195	0	0.063	0	0
48	0.038	0.031	0	0.030	0.204	0	0.051	0	0
50	0.037	0.038	0	0.022	0.305	0	0.049	0.001	0
52	0.037	0.059	0	0.013	0.426	0	0.046	0.001	0
54	0.037	0.066	0	0.014	0.475	0	0.043	0.001	0
56	0.036	0.091	0	0.011	0.516	0	0.041	0.004	0
58	0.036	0.127	0	0.009	0.611	0	0.037	0.014	0
60	0.036	0.139	0	0.009	0.612	0	0.035	0.018	0
65	0.035	0.224	0	0.008	0.717	0	0.033	0.042	0
70	0.034	0.272	0	0.008	0.771	0	0.027	0.294	0
75	0.032	0.357	0	0.008	0.825	0	0.018	0.701	0
80	0.032	0.423	0	0.008	0.878	0	0.017	0.765	0
85	0.030	0.561	0	0.008	0.926	0	0.016	0.822	0
90	0.027	0.635	0	0.008	0.952	0	0.014	0.923	0
95	0.024	0.689	0	0.008	0.959	0	0.010	0.971	0

Notes: See notes for Table 1.

Table 5. Marron-Wand Test Suite (e).

	asymmetric_claw			double_claw			asymmetric_double_claw		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.086			0.074			0.085		
3	0.084			0.072			0.039		
4	0.084			0.023			0.038		
5	0.084			0.023			0.038		
6	0.082			0.022			0.038		
7	0.083			0.022			0.038		
8	0.083			0.022			0.038		
9	0.073			0.022			0.038		
10	0.072	0		0.022			0.038		
11	0.070	0		0.022			0.038		
12	0.069	0	0	0.022			0.038		
13	0.042	0	0	0.022			0.038		
14	0.040	0	0	0.022			0.037	0	
15	0.040	0	0	0.022			0.037	0	
16	0.040	0	0	0.022	0		0.037	0	0
17	0.040	0	0	0.022	0		0.037	0	0
18	0.040	0	0	0.022	0	0	0.037	0	0
19	0.040	0	0	0.022	0	0	0.037	0	0
20	0.039	0	0	0.022	0	0	0.037	0	0
21	0.038	0	0	0.022	0	0	0.037	0	0
22	0.037	0	0	0.022	0	0	0.037	0	0
23	0.029	0	0	0.022	0	0	0.037	0	0
24	0.028	0	0	0.022	0	0	0.037	0	0
25	0.027	0	0	0.022	0	0	0.037	0	0
26	0.027	0	0	0.022	0	0	0.037	0	0
27	0.026	0	0	0.022	0	0	0.037	0	0
28	0.026	0	0	0.022	0	0	0.037	0	0
29	0.026	0	0	0.022	0	0	0.037	0	0
30	0.026	0	0	0.022	0	0	0.037	0	0
32	0.025	0	0	0.022	0	0	0.036	0	0
34	0.025	0	0	0.022	0	0	0.019	0	0
36	0.024	0	0	0.022	0	0	0.019	0	0
38	0.024	0	0	0.022	0	0	0.019	0	0
40	0.024	0	0	0.022	0	0	0.019	0	0
42	0.024	0	0	0.022	0	0	0.019	0	0
44	0.018	0	0	0.022	0	0	0.018	0	0
46	0.018	0	0	0.022	0	0	0.018	0	0
48	0.017	0	0	0.022	0	0	0.019	0	0
50	0.017	0	0	0.022	0	0	0.019	0	0
52	0.017	0	0	0.022	0	0	0.018	0	0
54	0.017	0.001	0	0.022	0	0	0.016	0	0
56	0.017	0.001	0	0.022	0	0	0.016	0	0
58	0.014	0.002	0	0.021	0	0	0.016	0	0
60	0.014	0.002	0	0.021	0	0	0.016	0	0
65	0.014	0.005	0	0.020	0	0	0.015	0	0
70	0.013	0.004	0	0.020	0	0	0.015	0	0
75	0.012	0.008	0	0.021	0	0	0.015	0	0
80	0.010	0.011	0	0.021	0	0	0.016	0	0
85	0.009	0.039	0	0.021	0	0	0.015	0	0
90	0.009	0.053	0	0.021	0	0	0.015	0	0
95	0.009	0.057	0	0.021	0	0	0.015	0	0

Notes: See notes for Table 1.

Table 6. Scale Mixture of Normals.

	$\kappa = 4$			$\kappa = 6$			$\kappa = 8$			$\kappa = 10$		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.054	0.550		0.111	0.220		0.158	0.137		0.202	0.104	
3	0.042	0.671		0.107	0.194		0.159	0.134		0.203	0.099	
4	0.017	0.899	0.597	0.048	0.669	0.197	0.086	0.513	0.080	0.203	0.188	0.041
5	0.010	0.909	0.611	0.048	0.669	0.214	0.086	0.513	0.105	0.122	0.428	0.063
6	0.010	0.965	0.862	0.048	0.747	0.418	0.086	0.531	0.221	0.122	0.464	0.132
7	0.010	0.965	0.875	0.040	0.779	0.451	0.073	0.625	0.271	0.101	0.522	0.176
8	0.002	0.989	0.951	0.027	0.858	0.609	0.053	0.728	0.353	0.085	0.631	0.219
9	0.002	0.989	0.961	0.022	0.868	0.636	0.050	0.739	0.427	0.078	0.636	0.293
10	0.002	0.993	0.984	0.022	0.893	0.742	0.050	0.766	0.482	0.078	0.662	0.331
11	0.002	0.993	0.988	0.019	0.909	0.762	0.044	0.792	0.564	0.072	0.679	0.413
12	0.003	0.995	0.994	0.018	0.926	0.830	0.042	0.826	0.615	0.068	0.702	0.452
13	0.003	0.997	0.996	0.014	0.941	0.847	0.032	0.859	0.670	0.053	0.769	0.517
14	0.003	0.999	0.998	0.014	0.949	0.895	0.032	0.874	0.708	0.050	0.801	0.553
15	0.003	0.999	0.999	0.012	0.957	0.904	0.027	0.886	0.752	0.046	0.809	0.609
16				0.011	0.971	0.934	0.029	0.871	0.782	0.046	0.823	0.639
17				0.009	0.979	0.938	0.025	0.892	0.809	0.041	0.831	0.685
18				0.009	0.981	0.958	0.025	0.906	0.839	0.038	0.855	0.711
19				0.008	0.982	0.960	0.025	0.909	0.857	0.035	0.867	0.750
20				0.008	0.984	0.974	0.015	0.949	0.883	0.034	0.881	0.770
21				0.008	0.986	0.974	0.017	0.938	0.892	0.028	0.894	0.770
22				0.009	0.987	0.975	0.018	0.945	0.901	0.028	0.906	0.810
23				0.009	0.989	0.976	0.018	0.944	0.901	0.027	0.899	0.817
24				0.007	0.993	0.977	0.012	0.965	0.916	0.026	0.908	0.822
25				0.006	0.995	0.977	0.013	0.967	0.916	0.026	0.911	0.823
26				0.006	0.995	0.985	0.013	0.969	0.926	0.024	0.914	0.853
27				0.007	0.997	0.985	0.012	0.971	0.926	0.021	0.934	0.854
28				0.007	0.999	0.985	0.012	0.975	0.930	0.021	0.937	0.858
29							0.012	0.974	0.938	0.023	0.929	0.863

Notes: κ is raw kurtosis, for excess kurtosis subtract 3. Efficiency is invariant to the mean and variance of the scale mixture; shown here is efficiency for $(\mu, \sigma^2) = (0, 1/4)$. See notes for Table 1.

Table 7. Second Largest Order Statistic From the Lognormal Distribution.

	$n=5$			$n=50$			$n=100$			$n=200$		
K	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm	$\sqrt{\text{mse}}$	emm	cmm
2	0.840	0.523	0.065	0.257	0.820	0.157	0.204	0.834	0.140	0.164	0.884	0.204
3	0.552	0.705	0.163	0.143	0.943	0.331	0.107	0.955	0.343	0.085	0.973	0.405
4	0.512	0.918	0.306	0.129	0.999	0.580	0.095	0.997	0.562	0.077	0.995	0.647
5	0.514	0.918	0.476	0.129	0.999	0.760	0.095	0.997	0.725	0.076	0.994	0.803
6	0.497	0.969	0.611	0.087	0.996	0.878	0.063	0.993	0.849	0.045	0.990	0.899
7	0.497	0.969	0.720	0.089	0.996	0.943	0.065	0.996	0.925	0.046	0.996	0.958
8	0.348	0.987	0.799	0.086	0.999	0.973	0.062	0.999	0.965	0.046	0.997	0.984
9	0.354	0.986	0.859	0.082	1.000	0.988	0.060	1.000	0.965	0.026	1.000	0.984
10	0.345	0.995	0.903									
11	0.343	0.995	0.905									
12	0.343	0.996	0.905									
13	0.247	0.999	0.929									
14	0.269	0.999	0.931									
15	0.266	0.999	0.931									
16	0.176	1.000	0.931									

Notes: n is the size of the sample from the lognormal distribution. Efficiency is invariant to the mean and variance of the lognormal; shown here is efficiency for $(\mu, \sigma^2) = (0, 1/8)$. See notes for Table 1.

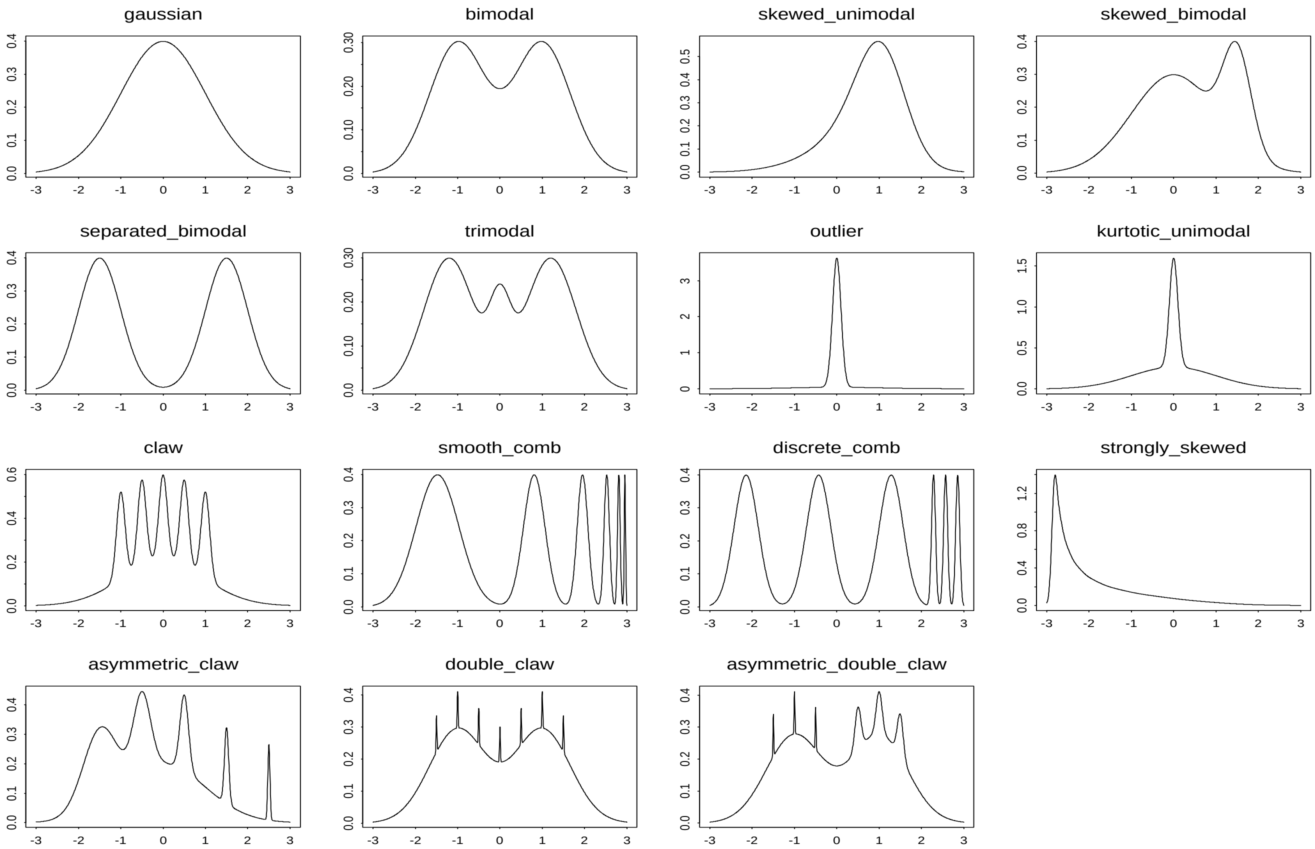

Figure 1. The Marron-Wand Test Suite. In each panel is plotted one of the fifteen densities proposed by Marron and Wand (1992) as a battery for use in evaluating density estimators. The densities are in order of increasing difficulty of estimation by method of moments. The labels at the top of each panel are as in Marron and Wand.

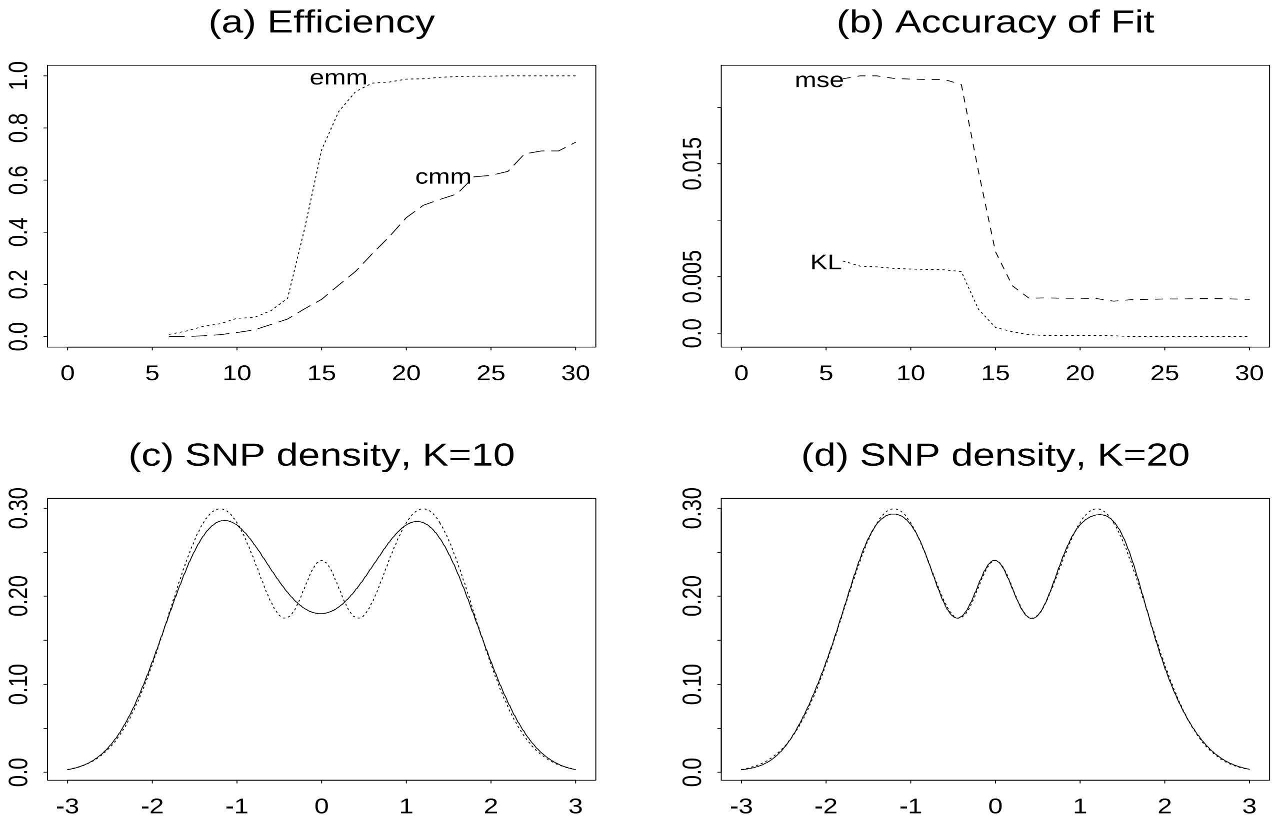

Figure 2. Relative Efficiency for the Trimodal Density. Panel (a) plots the relative effiency of the EMM and CMM estimators against degree K, for the trimodal density of the Marron-Wand test suite, which is a plot of the last two columns of Table 2 against the first. As seen, the efficiency of the EMM estimator increases rapidly when the degree K of the SNP auxilary model is between 10 and 20. Panel (b) plots the root mean squared error and Kullback-Leibler divergence of the SNP approximation to the trimodal density against K, labeled mse and KL, respectively. As seen, the region $10 \leq K \leq 20$ is the region where the error in the SNP approximation to the trimodal density decreases rapidly. Panel (c) plots the SNP approximation at $K = 10$, shown as a solid line, to the trimodal density, shown as a dotted line. As seen, at $K = 10$ the SNP density approximates a trimodal density by a bimodal density. Panel (d) is the same at $K = 20$. As seen, at $K = 20$ the SNP density has correctly determined the number of modes.

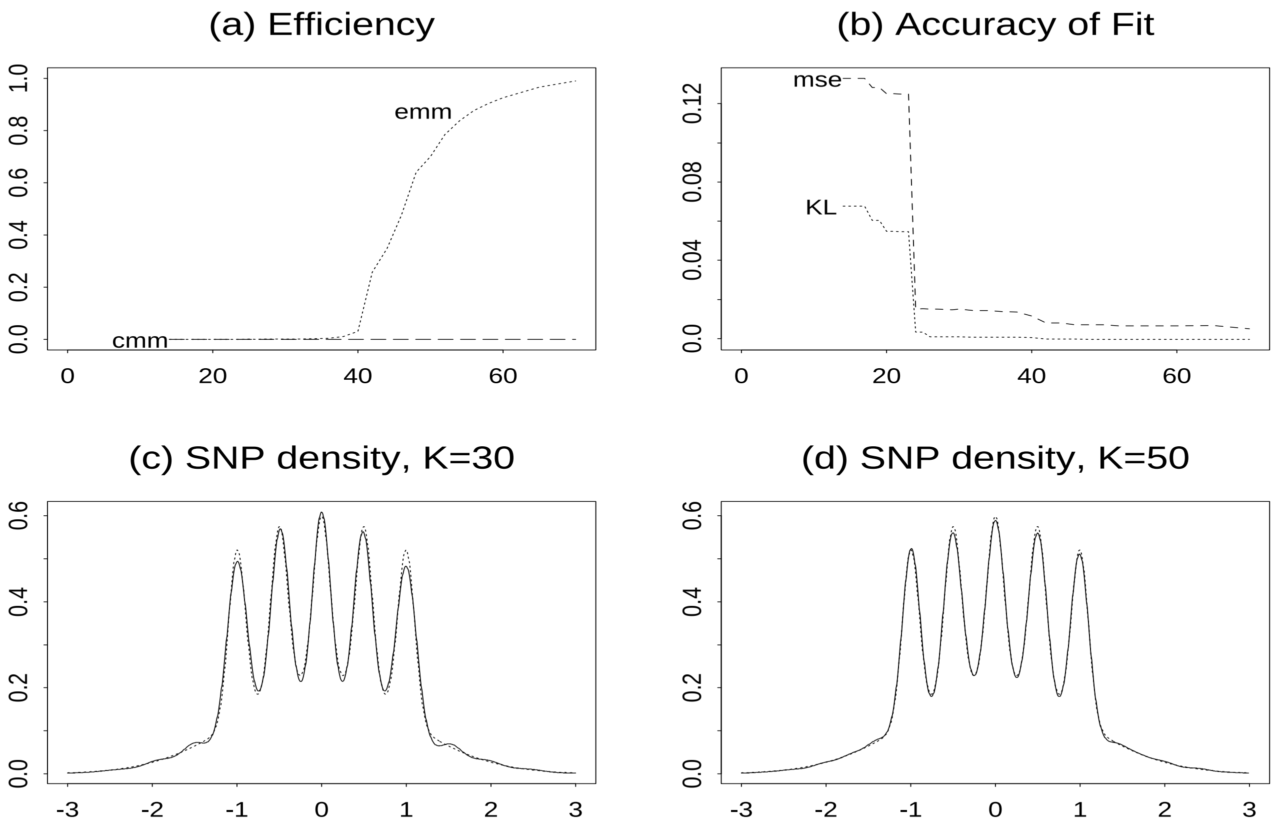

Figure 3. Relative Efficiency for the Claw Density. Panel (a) plots the relative effiency of the EMM and CMM estimators against degree K, for the claw density of the Marron-Wand test suite, which is a plot of the last two columns of Table 3 against the first. As seen, the efficiency of the EMM estimator increases rapidly when the degree K of the SNP auxilary model is between 30 and 50. Panel (b) plots the root mean squared error and Kullback-Leibler divergence of the SNP approximation to the claw density against K, labeled mse and KL, respectively. As seen, the region $30 \leq K \leq 50$ is a region where the error in the SNP approximation to the trimodal density decreases slightly. Panel (c) plots the SNP approximation at $K = 30$, shown as a solid line, to the claw density, shown as a dotted line. As seen, at $K = 30$ the SNP identifies the number of modes correcly but misses the height of the two outer modes. Panel (d) is the same at $K = 50$. As seen, at $K = 50$ the SNP density has more closely approximated the height of the two outer modes.

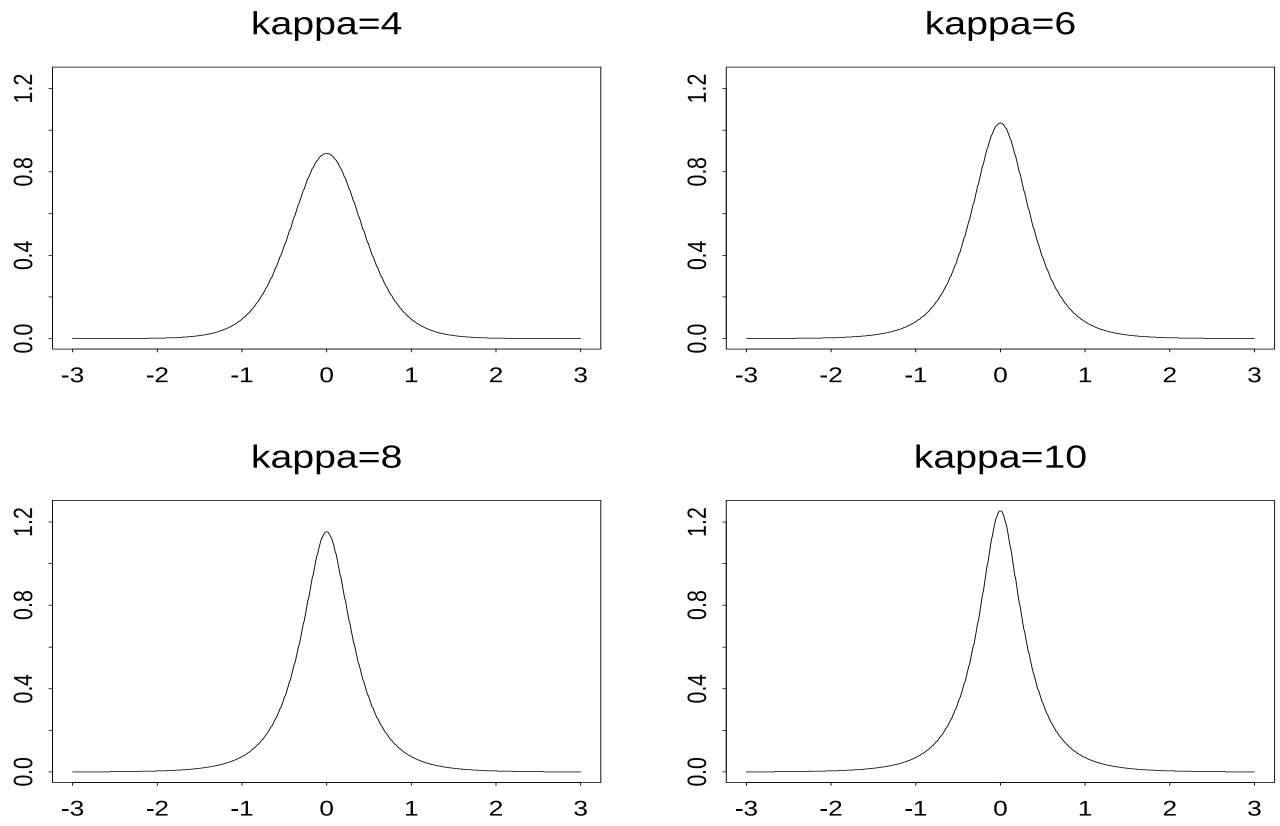

Figure 4. Scale Mixture of Normals. Plotted is the density $p(y|\rho) = \int_{-\infty}^{\infty} n(y|\rho_1, e^{2u})\, n(u|\rho_2, \rho_3^2)\, du$ with ρ chosen so that the density has mean 0, variance 1/4, and raw kurtosis κ as shown above each panel.

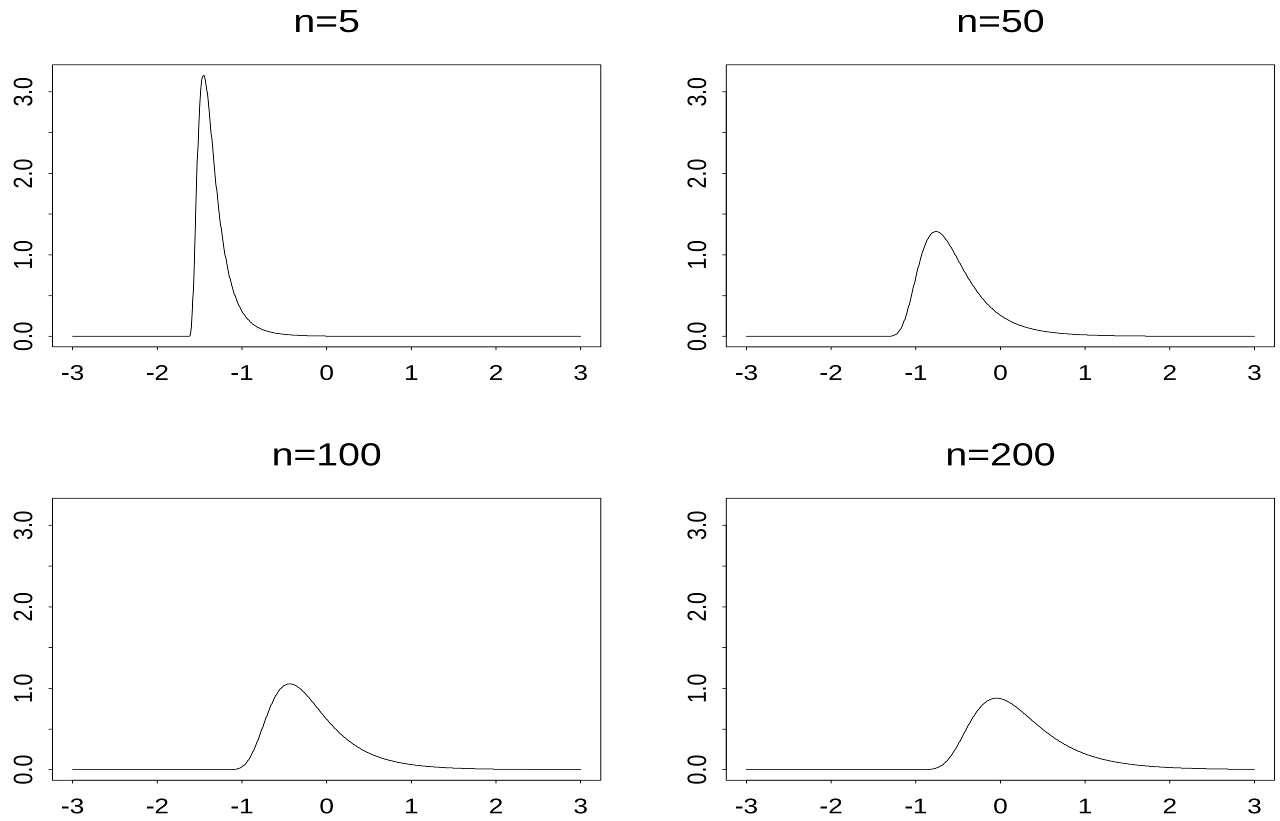

Figure 5. Second Largest Order Statistic of the Lognormal. Plotted is the density $p(y|\rho) = \frac{n(n-1)}{y-\rho_2}\left\{\Phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right]\right\}^{n-2}\left\{1-\Phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right]\right\}\phi\left[\log\left(\frac{y-\rho_2}{\rho_3}\right)\right]$ where $y > \rho_2$, $n = \rho_1$, and ϕ and Φ denote the standard normal density and distribution functions, respectively; ρ is chosen so that the density has mean 0, variance 1/8, and sample size n as shown above each panel.