Journal of Economic Theory and Econometrics, Vol. 23, No. 4, Dec. 2012, 261–277

A New Method for Estimating and Simulating Maximum Entropy Densities*

Jae-Young Kim† Joonhwan Lee‡

Abstract The maximum entropy (maxent) approach provides an attractive and appealing method for obtaining a probability distribution in some limited information environments. The maxent approach, however, is not easy to use for practical application since analytic derivation of a maxent density is usually not feasible. This paper proposes a method of estimating and simulating an maxent density. It is a numerically tractable and stable method that is relatively simple in real computation. The proposed method is capable of handling problems for which existing methods are difficult to apply or are subject to occasional failure. Monte Carlo results confirm that the proposed method can be well applied for cases when existing methods are subject to occasional failure.

Keywords Entropy maximization, Cross entropy, Moment conditions, Numerical method

JEL Classification C01, C2

*The authors are thankful to the editor and two anonymous referees for helpful comments. This work was supported by the National Research Foundation of Korea (grant number: KRF-327-2008-2-B00080).

†Correspondence: Department of Economics, Seoul National University, Seoul, 151-746, Korea. +822-880-6390. e-mail: jykim017@snu.ac.kr

‡Department of Economics, MIT

Received November 26, 2012, Accepted December 17, 2012

1. INTRODUCTION

The maximum entropy (maxent) approach is a convenient and attractive method for obtaining a probability distribution known only to satisfy certain moment conditions. A few numerical methods for estimating maxent densities are proposed in the literature. However, in many cases of practical applications the existing methods have difficulty in real computation and are subject to occasional failure. In this paper we propose an alternative method of estimating and simulating a maxent distribution given a set of moment conditions. Our approach is based on a combined method of simulation and direct computation that does not have computational difficulties involved in the existing methods.

The maxent approach has been used widely in the literature. In Bayesian econometrics, Bayesian method of moments (Zellner, 1997, among others) and Bayesian generalized method of moments (Kim, 2002) employed the maxent approach to get (non-likelihood based) posterior densities. In the finance literature, there have been many applications of the maxent approach. Stutzer (1996) and Buchen and Kelly (1996), for example, used the maxent approach to derive risk neutral measures (or state price densities) of asset returns for derivative pricing. Rockinger and Jondeau (2002) and Bera and Park (2004) extended ARCH type models by applying maxent densities.

An important problem involved in a maxent density is that the analytic derivation of it is not feasible in many cases. Instead, some numerical methods based on Newton-Raphson type algorithms are proposed in the literature for computing maxent densities (Zellner and Highfield, 1988, and Ormoneit and White, 1999). However, the existing numerical methods have difficulty for practical use since they require non-trivial numerical integration and iterative nonlinear optimization with computation of gradient values that is highly sensitive to starting values. Wu (2003) tries to resolve this difficulty by proposing a new algorithm to setting up appropriate starting values with sequential updating of moment conditions. However, it also has difficulty when computing maxent densities from extreme moment conditions. Also, Newton numerical optimization usually requires computation of gradient values.

In this paper, we propose an alternative method of computing maxent densities. Our method does not involve numerical integration or computation of gradient values. Our approach is developed based on the following theoretical facts: (1) The maxent problem is a special case of the cross-entropy minimization; (2) The constrained optimization in the maxent problem or in the cross-entropy minimization can be transformed to an unconstained optimization; (3) The solution of the unconstrained optimization can be approximated with arbitrary precision

by the solution of an empirical unconstrained optimization. The solution of the empirical cross-entropy minimization with a uniform reference density is the same as that of the empirical maxent problem.

Our numerical method proceeds as follows: (1) Draw a random sample from a uniform reference density; (2) Evaluate the objective function in an (empirical) unconstrained optimization for the drawn random sample; (3) Get the solution of the empirical unconstrained optimization for the target maxent density. For this purpose, we use a direct search algorithm, Nelder-Mead algorithm, that does not involve computation of gradient values or numerical integration. This solution of an empirical unconstrained optimization is an estimate of the maxent density. Given an estimate of the maxent density we can numerically compute such quantities of distribution characteristics as moments, quantiles and the underlying probability measure. Monte Carlo experiments performed on several examples show that our method works well in cases when the existing methods do not work or are difficult to apply.

Discussion of the paper goes as follows. Section 2 studies some theoretical background of our numerical method. In Section 3 we discuss our numerical method for estimation and simulation of the maxent density. Section 4 evaluates performance of our method by Monte Carlo simulation, and Section 5 concludes the paper.

2. THE MAX-ENT DENSITY

Let x be an independent random variable from a distribution P_0 with a density p_0. We do not know the distribution P_0 or its density p_0 but only know that x satisfies a set of moment conditions:

$$E_{P_0}[c_j(x)] \equiv \int c_j(x)p_0(x)dx = d_j, \quad j = 1,\ldots,m \tag{1}$$

where c_j is a continuous function on $\mathbb{R}$; d_j is a constant. Thus, c_j and d_j characterize moment conditions for x. To save notation we use x both for a random variable and its values. Having the moment conditions (1) with respect to p_0 we can apply the maxent method to get an approximate density for x.

A maxent density is obtained by maximizing Shannon's entropy measure subject to some moment conditions as in (1). That is, a maxent density p^* is the

solution of

$$\max_{p(x)} \int -p(x)\log p(x)dx \quad \text{subject to} \tag{2}$$
$$\int p(x)dx = 1, \quad \int c_j(x)p(x)dx = d_j,\ j = 1,\ldots,m$$

where the condition $\int p(x)dx = 1$ is to ensure that $p(x)$ is a probability density. The above maxent problem is a special case of the cross-entropy minimization:

$$\min_{p(x)} \int p(x)\log\frac{p(x)}{q(x)}dx \quad \text{subject to} \tag{3}$$
$$\int p(x)dx = 1, \quad \int c_j(x)p(x)dx = d_j,\ j = 1,\ldots,m$$

where q is a specific reference density.

The problem (3) can be transformed to an unconstrained optimization problem:

Proposition 1. *Assume that $\int \exp\{\lambda' c(x)\}q(x)dx$ exists for $\lambda \in \Lambda$, an open subset of $\mathbb{R}^m$. The problem (3) is equivalent to the following*

$$\min_{\lambda} \int \exp\{\sum_{j=1}^{m} \lambda_j(c_j(x) - d_j)\}q(x)dx \tag{4}$$

and the solution of the problem is

$$p^*(x) = \frac{1}{\mu^*}\exp\{\lambda^{*\prime} c(x)\}q(x) \ \ \textit{where}\ \mu^* = \int \exp\{\lambda^{*\prime} c(x)\}q(x)dx$$

where λ^* is the solution of (4).

Proof. We follow Golan, et. al. (1996) for the proof. Thus, the Lagrangian function for the problem (3) is

$$\mathscr{L}(p) = \int p(x)\log\frac{p(x)}{q(x)}dx + (1+\lambda_0)(1 - \int p(x)dx)$$
$$+ \sum_{j=1}^{m} \lambda_j(d_j - \int c_j(x)p(x)dx).$$

The Lagrangian function $\mathscr{L}(p)$ is minimized when its Frechet derivative equals zero. The first order condition for the minimization problem is

$$D\mathscr{L} = \int [1 + \log\frac{p(x)}{q(x)} - (1+\lambda_0) - \sum_{j=1}^{m} \lambda_j c_j(x)]\delta p(x)dx = 0.$$

This leads to

$$\log p(x) = \log q(x) + \lambda_0 + \sum_{j=1}^{m} \lambda_j c_j(x).$$

Letting $\mu \equiv \exp\{-\lambda_0\}$ as defined above, we have

$$p(x) = \frac{1}{\mu} \exp\{\sum_{j=1}^{m} \lambda_j c_j(x)\} q(x).$$

The objective function now can be written in unconstrained form

$$\int p(x) \log \frac{p(x)}{q(x)} dx = -\log \mu + \sum_{j=1}^{m} \lambda_j d_j.$$

Therefore, the problem is equivalent to minimizing $\int \exp\{\lambda'(c(x)-d)\}q(x)dx$ with respect to λ's. Note that the Hessian matrix of the object function is the covariance matrix of moment conditions. That is,

$$\frac{\partial^2 \int \exp\{\lambda'(c(x)-d)\}q(x)dx}{\partial \lambda_i \partial \lambda_j} = \int c_i(x)c_j(x)q(x)dx.$$

Therefore, the objective function is convex and the solution is unique. □

Now, we consider a sample analogue of the problem (3). Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from the reference density q. Also, let $q_i = q(X_i) = \frac{1}{n}$. Then, we can construct the cross-entropy minimization problem with respect to the empirical measure for the random sample:

$$\begin{aligned} \mathbf{p}_n^* = &\underset{\mathbf{p}}{\operatorname{argmin}} \sum_{i=1}^{n} p_i \log \frac{p_i}{1/n} \quad \text{s.t.} \\ &\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} c_j(X_i) p_i = d_j, \ j = 1, \ldots, m \end{aligned} \tag{5}$$

where $\mathbf{p} = (p_1, \ldots, p_n)$ is an n-vector of the choice variable in the above minimization problem, and accordingly $\mathbf{p}_n^* = (p_{1,n}^*, \ldots, p_{n,n}^*)$. Notice that the above problem is the same as the following sample version of the entropy maximization:

$$\begin{aligned} \mathbf{p}_n^* = &\underset{\mathbf{p}}{\operatorname{argmax}} -\sum_{i=1}^{n} p_i \log p_i \quad \text{s.t.} \\ &\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} c_j(X_i) p_i = d_j, \ j = 1, \ldots, m. \end{aligned}$$

We can transform the minimization problem (5) to an unconstrained problem as in Propositon 1:

Proposition 2. *The problem (5) is equivalent to the following*

$$\min_{\lambda} \sum_{i=1}^{n} \exp\{\sum_{j=1}^{m} \lambda_j (c_j(X_i) - d_j)\} \frac{1}{n} \tag{6}$$

and the solution of the problem is as follows

$$\hat{p}^*_{i,n} = \frac{1}{\hat{\mu}^*_n} \exp\{\hat{\lambda}^{*\prime}_n c(X_i)\} \frac{1}{n} \quad \textit{where } \hat{\mu}^*_n = \sum_{i=1}^{n} \exp\{\hat{\lambda}^{*\prime}_n c(X_i)\} \frac{1}{n} \tag{7}$$

where $\hat{\lambda}^{*\prime}_n$ is the solution of (6).

Proof. The proof is virtually the same as that of Proposition 1. □

Let $\hat{P}^*_n$ be the probability measure corresponding to the solution $\mathbf{p}^*_n$:

$$\hat{P}^*_n(t) = \sum_{i=1}^{n} \hat{p}^*_{i,n} \mathbf{1}_{(t \geq X_i)}(t). \tag{8}$$

Also, let P^* be the probability measure corresponding to the solution of (3) or (4), p^*:

$$P^*(t) = \int_{-\infty}^{t} p^*(x)dx = \int_{-\infty}^{t} \frac{1}{\mu^*} \exp\{\lambda^{*\prime} c(x)\} dx. \tag{9}$$

We can show that the empirical probability measure $P^*_n(t)$ converges to $P^*(t)$. To do it we first show that the Lagrange multiplier λ is consistently estimated by the empirical cross-entropy minimization (5):

Theorem 3. *Let λ^* and $\hat{\lambda}^*_n$, respectively, be the solutions of the problems (4) and (6). Then, we have*

$$\hat{\lambda}^*_n \xrightarrow{p} \lambda^*.$$

Proof. Let $\rho(x,\lambda) \equiv \exp\{\lambda'(c(x) - d)\}$. The object functions of (4) and (6), respectively, are rewritten as in the following:

$$E_Q[\rho(x,\lambda)] = \int \rho(x,\lambda) q(x) dx = \int \exp\{\lambda'(c(x) - d)\} q(x) dx$$

$$\bar{\rho}_n(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \rho(X_i,\lambda) = \sum_{i=1}^{n} \exp\{\lambda'(c(X_i) - d)\} \frac{1}{n}$$

where Q is the probability measure corresponding to the density q, $Q(t) = \int_{-\infty}^{t} q(x)dx$ for each $t \in \mathbb{R}$. Note that the function $\rho(\cdot)$ is continuous since $c(\cdot)$ is continuous. Then, by a weak law of large numbers we have

$$\bar{\rho}_n(\lambda) \xrightarrow{p} E_Q[\rho(x,\lambda)].$$

Similar to the proof of Proposition 1, we can show that $\bar{\rho}_n(\lambda)$ is a convex function. Then the convexity lemma (See Pollard (1991) or Andersen and Gill (1982)) implies that for each compact subset K of $\mathbb{R}^m$,

$$\sup_{\lambda \in K} |\bar{\rho}_n(\lambda) - E_Q[\rho(x,\lambda)]| \longrightarrow 0.$$

Then, by a standard arguement used to prove consistency of a minimum contrast estimator (for example, in Van der Vaart (1998)) we get the result. □

Now, we can estabilish the uniform convergence of $\hat{P}_n^*(t)$ to $P^*(t)$.

Theorem 4. *Let $\hat{P}_n^*(t)$ and $P^*(t)$ be as defined in (8) and (9). Then, it is true that*

$$\lim_{n\to\infty} Q\big(\sup_{t\in\mathbb{R}} |\hat{P}_n^*(t) - P^*(t)| > \varepsilon\big) = 0 \tag{10}$$

where Q is the probability measure for the reference density q.

Proof. For $\hat{\mu}_n^* = \frac{1}{n}\sum \exp\{\hat{\lambda}_n^{*\prime} c(X_i)\}$, we have

$$\begin{aligned}
|\hat{\mu}_n^* - \mu^*| &= \left| \left(\frac{1}{n}\sum \exp\{\hat{\lambda}_n^{*\prime} c(X_i)\} - \frac{1}{n}\sum \exp\{\lambda^{*\prime} c(X_i)\} \right) \right. \\
&\quad \left. + \left(\frac{1}{n}\sum \exp\{\lambda^{*\prime} c(X_i)\} - \int \exp\{\lambda^{*\prime} c(x)\} q(x)dx \right) \right| \\
&\leq \left| \frac{1}{n}\sum \exp\{\hat{\lambda}_n^{*\prime} c(X_i)\} - \frac{1}{n}\sum \exp\{\lambda^{*\prime} c(X_i)\} \right| \\
&\quad + \left| \frac{1}{n}\sum \exp\{\lambda^{*\prime} c(X_i)\} - \int \exp\{\lambda^{*\prime} c(x)\} q(x)dx \right| \\
&\xrightarrow{p} 0
\end{aligned}$$

where the convergence in the last step is due to Theorem 3, a weak law of large numbers and the continuous mapping theorem. Now, write $\hat{P}_n^*(t)$ as

$$\hat{P}_n^*(t) \left(= \sum_{i=1}^{n} \hat{p}_{i,n}^* \mathbf{1}_{(t \geq X_i)}(t) \right) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\hat{\mu}_n^*} \exp\{\hat{\lambda}_n^{*\prime} c(X_i)\} \mathbf{1}_{(t \geq X_i)}(t).$$

Then, by the Slutsky lemma and a weak law of large numbers, we have

$$\hat{P}_n^*(t) \longrightarrow P^*(t) \quad \forall t \in \mathbb{R}$$

in Q-probability. Now, since both $P_n^*(t)$ and $P^*(t)$ are increasing and bounded, the uniform convergence is established by the same arguement for the Glivenko-Cantelli lemma. □

3. A NEW METHOD FOR ESTIMATING MAXENT DENSITIES

This section discusses a new method for computing and simulating a maxent density based on the theoretical results of the previous section. Several authors studied computation methods for the maxtent density such as Zellner and Highfield (1988), Ormoneit and White (1999) and Wu (2003). The existing methods, however, have difficulty for practical use in two aspects. First, they involve numerical integration that is complex in many cases of actual computation. Second, these methods employ the standard Newton-Rhapson algorithm which uses explicit gradient values. Usually, however, in most of the search region the Jacobian matrix is near-singular, which leads to frequent failure of the algorithm unless the starting value is set close to the solution (Ormoneit and White (1999)).

The method we propose in this paper is based on Monte-Carlo draws to estimate the maxent density. We consider the discrete minimization of (5) with the reference density q being the density of a uniform distribution. Denote by $q_M(x)$ the uniform density over an interval $[-M,M] \subset \mathbb{R}$ for some $M \in \mathbb{R}$. That is, $q_M(x) = \frac{1}{2M}\mathbf{1}_{[-M,M]}(x)$. Let $\hat{\lambda}_{n,M}^*$ be the solution of (6) where X_i, $i = 1,\ldots,n$ are drawn from q_M.

Our method proceeds as follows. We first draw random numbers $\{X_i\}_{i=1}^n$ from q_M. Then, we get the solution $\hat{\lambda}_{n,M}^*$ in (6) and resulting solutions $\hat{\mathbf{p}}_{n,M}^*$ and $\hat{P}_{n,M}^*$ from $\hat{\lambda}_{n,M}^*$:

$$\hat{\lambda}_{n,M}^* = \operatorname*{argmin}_{\lambda} \sum_{i=1}^n \exp\{\sum_{j=1}^m \lambda_j (c_j(X_i) - d_j)\}\frac{1}{n},$$

$$\hat{p}_{i,n,M}^* = \frac{1}{\hat{\mu}_{n,M}^*} \exp\{\hat{\lambda}_{n,M}^{*\prime} c(X_i)\}\frac{1}{n},$$

$$\hat{P}_{n,M}^*(t) = \sum_{i=1}^n \hat{p}_{i,n,M}^* \mathbf{1}_{(t \geq X_i)}(t).$$

where $\hat{\mu}_{n,M}^* = \sum_{i=1}^n \exp\{\hat{\lambda}_{n,M}^{*\prime} c(X_i)\}\frac{1}{n}$. These solutions are obtained by a numerical method known as the Nelder-Mead simplex search algorithm. See explanations after Proposition 5 below.

The solution $\hat{\lambda}^*_{n,M}$ is a good approximate to its target value, and so is $\hat{p}^*_{i,n,M}$:

Proposition 5. *The maxent density p^* can be approximated with an arbitrary precision by $\hat{p}^*_{n,M}$ with M sufficiently large.*

Proof. Let λ^*_M be the solution of

$$\min_{\lambda} \int \exp\{\sum_{j=1}^{m} \lambda_j (c_j(x) - d_j)\} q_M(x) dx,$$

where $q_M(x) = \frac{1}{2M}\mathbf{1}_{[-M,M]}(x)$. By the monotone convergence theorem and Theorem 3 λ^*_M converges to the maxent solution λ^* as $M \to \infty$. Also, by Theorem 3,

$$\hat{\lambda}^*_{n,M} \xrightarrow{p} \lambda^*_M.$$

Therefore, as $n \to \infty$ and $M \to \infty$, $\hat{\lambda}_{n,M} \xrightarrow{p} \lambda^*$. The conclusion follows from the continuous mapping theorem. □

In practice, one confronts the problem of how to choose M. The choice of M depends on the distribution of x. For example, suppose that $c_j(x) = x^j$ for $j = 1, 2$ and $d_1 = 0$. Then, with a known value of d_2, we get the standard deviation of x, $\sigma_x = d_2^{1/2}$. We can choose M large enough, say $M = 4\sigma_x$, to cover almost 100% (probability 0.9999) of the support of x.

We employ the Nelder-Mead algorithm (Lagarias *et al.*, 1997) for numerical solution of our problem. The Nelder-Mead algorithm is a direct search algorithm which does not involve computation of gradient values. The algorithm first sets a simplex of $n+1$ vertices for an n-dimensional problem around the initial values. Then the algorithm transforms the simplex along the surface by reflection, expansion, contraction or shrinkage for each iteration step. Only several additional function evaluations are needed for each iteration step, so that the computational load is relatively small. The algorithm continues until the diameter of the simplex becomes smaller than specified threshold. For the detailed explanation of the algorithm, see Lagarias *et al.* (1997).

Other numerical methods such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm may be used. It is an approximate Newton's method which usually use the first and second derivatives of the objective function, requiring continuous differentiability. BFGS, however, is known to have good performance even for non-smooth optimizations.

With a large number n of random draws (500,000 or more), the estimation of λ is accurate to at least 4 decimal places in our simulation study (Section 4). Our

method is relatively simple and stable so that there is less concern of failure than existing methods. Monte Carlo experiments in Section 4 show that the proposed method works well in cases with large numbers of moment conditions as well as with moments whose values are close to bounds (extreme values).

With the computed discrete probabilities $\hat{p}^*_{i,n,M}$ for X_i for $i = 1,\ldots,n$, we can draw random numbers from the estimated maxent density by weighted resampling on the original uniform draws:

Algorithm 1. *We can draw a random number from the estimated maxent density by the following algorithm.*

1. *Draw* $u \sim U(0,1)$.
2. *Let* $X^{(1)},\ldots,X^{(n)}$ *be the order statistic of the random sample.*
3. *Let*
$$k = \inf_{1\le\tilde{k}\le n} \{\tilde{k} | \sum_{i=1}^{\tilde{k}} \hat{p}^*_{i,n,M} \ge u,\ \tilde{k} \in \mathbb{N}\}.$$
3. *Let* $Y = X^{(k)}$.

Then $Y \overset{A}{\sim} \hat{P}^*_{n,M}(t)$.

4. MONTE CARLO EXPERIMENTS

In this section, we examine performance of the numerical method studied in the previous section. First, we examine performance of our method for the examples considered in Ormoneit and White(1999) and Wu(2003). Similar to them, the moment conditions[1] imposed are

$$\begin{aligned} EX &= 0 & EX^2 &= 1 \\ EX^3 &\equiv \mu_3 \in [0,3] & EX^4 &\equiv \mu_4 \in [(EX^3)^3 + 1.1, 10]. \end{aligned}$$

Ormoneit and White (1999) reported that their algorithm failed for some ranges of (μ_3,μ_4). For example, Ormoneit and White (1999) reported that their algorithm failed when $\mu_3 = 0$ and $\mu_4 > 3$. Also, they encountered numerical error when $\mu_4 > 10$. Figures 1-2 show the estimated results of our method for $\lambda_i,\ i = 1,2,3,4$, for various values of (μ_3,μ_4). The results resemble those in

[1]For the covariance matrix to be positive definite the restriction of $\mu_3^2 + 1 < \mu_4$ must hold (Ormoneit and White, 1999).

Ormoneit and White (1999) and Wu (2003) in the range of (μ_3, μ_4) where the algorithms in these existing papers work. Figure 3 shows some examples of estimated maxent densities for some extreme moment conditions that cause failure of the algorithm proposed by Ormoneit and White (1999). Our method does not have such difficulty involved in the existing algorithms and successfully obtains solutions in the cases of extreme moment conditions, such as $\mu_3 = 5, \mu_4 = 26.1$ or $\mu_3 = 0, \mu_4 = 12$. The plotted densities are obtained by the weighted kernel smoothing where the weights are given as p_i's.

$$\hat{p}(x) = \sum_{i=1}^{n} p_i \frac{1}{h} K(\frac{x - X_i}{h})$$

We also perform Monte Carlo simulation in the case when a large number of moment conditions are available. We impose ten moment conditions which are close to but different from those of the standard normal distribution. Figure 4 shows various maxent densities with ten moment conditions imposed, comparing them with the standard normal density. $M1$ has moments such that each $(2k)^{th}$ moment for $k = 2, \ldots, 5$ is larger than that of the standard normal distribution, while $M2$ has each $(2k)^{th}$ moment smaller than that of the standard normal.[2] We also have an estimate of the maxent density for the daily return of KOSPI200 with ten (standardized) empirical moments. The standardized moments of KOSPI 200 daily returns are estimated as $\mu_3 = -0.34$, $\mu_4 = 6.17$, $\mu_5 = -14.42$, $\mu_6 = 139.47$, $\mu_7 = -733.67$, $\mu_8 = 5763.1$, $\mu_9 = -37815$, $\mu_{10} = 279256$ for data in the period of 2005.1.3-2008.4.30.

5. CONCLUSION

We propose a new method of estimating and simulating a maximum entropy distribution given a set of moment conditions. Our approach is based on a combined method of simulation and direct computation while existing methods are based on first-order approximation and numerical integration which are very sensitive to initial values or are subject to occasional failure. Monte Carlo experiments show that the proposed method works well in the cases where the existing methods either does not work or is difficult to apply.

[2] The $(2k)^{th}$th moment of standard normal distribution is $\frac{(2k)!}{2^k k!}$

REFERENCES

Andersen, P. K., and R. Gill (1982). Cox's regression model for counting processes: A large sample study, Annals of Statistics 10, 1100–1120.

Bera, A. K., and S. Y. Park (2004). Maximum entropy autoregressive conditionial heteroskedasticity model, Working paper, University of Illinois at Urbana-Champaign.

Buchen, P., and M. Kelly (1996). The maximum entropy distribution of an asset inferred from option prices, Journal of Financial and Quantitative Analysis 31, 143–159.

Golan, A., G. Judge, and D. Miller (1996). Maximum Entropy Econometrics: Robust Estimation with Limited Data, John Wiley and Sons.

Jaynes, E. T. (1957). Information theory and statistical mechanics, Physics Review 106, 620–630.

Kim, J. Y. (2002). Limited information likelihood and Bayesian analysis, Journal of Econometrics 107, 175–193.

Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright (1998). Convergence properties of the Nelder-Mead simplex algorithm in low dimensions, SIAM Journal on Optimization 9, 112–147.

Ormoneit, D., and H. White (1999). An efficient algorithm to compute maximum entropy densities, Econometrics Reviews 18, 127–140.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators, Econometric Theory 7, 186–199.

Rockinger, M., and E. Jondeau (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis, Journal of Econometrics 106, 119–142.

Stutzer, M. (1996). A simple nonparametric approach to derivative security valuation, Journal of Finance 51, 1633–1652.

Van der Vaart, A. W. (1998). Asymptotic Statistics, Cambridge University Press.

Wu, X. (2003). Calculation of maximum entropy densities with application to income distribution, Journal of Econometrics 115, 347–354.

Zellner, A. (1997). The Bayesian method of moments (BMOM): Theory and applications, in Advances in Econometrics (Applying Maximum Entropy to Econometric Problems), Vol. 12, 85–105, eds., R. C. Hill, Emerald Group Publishing Limited.

Zellner, A., and R. A. Highfield (1988). Calculation of maximum entropy distribution and approximation of marginal posterior distributions, Journal of Econometrics 37, 195–209.

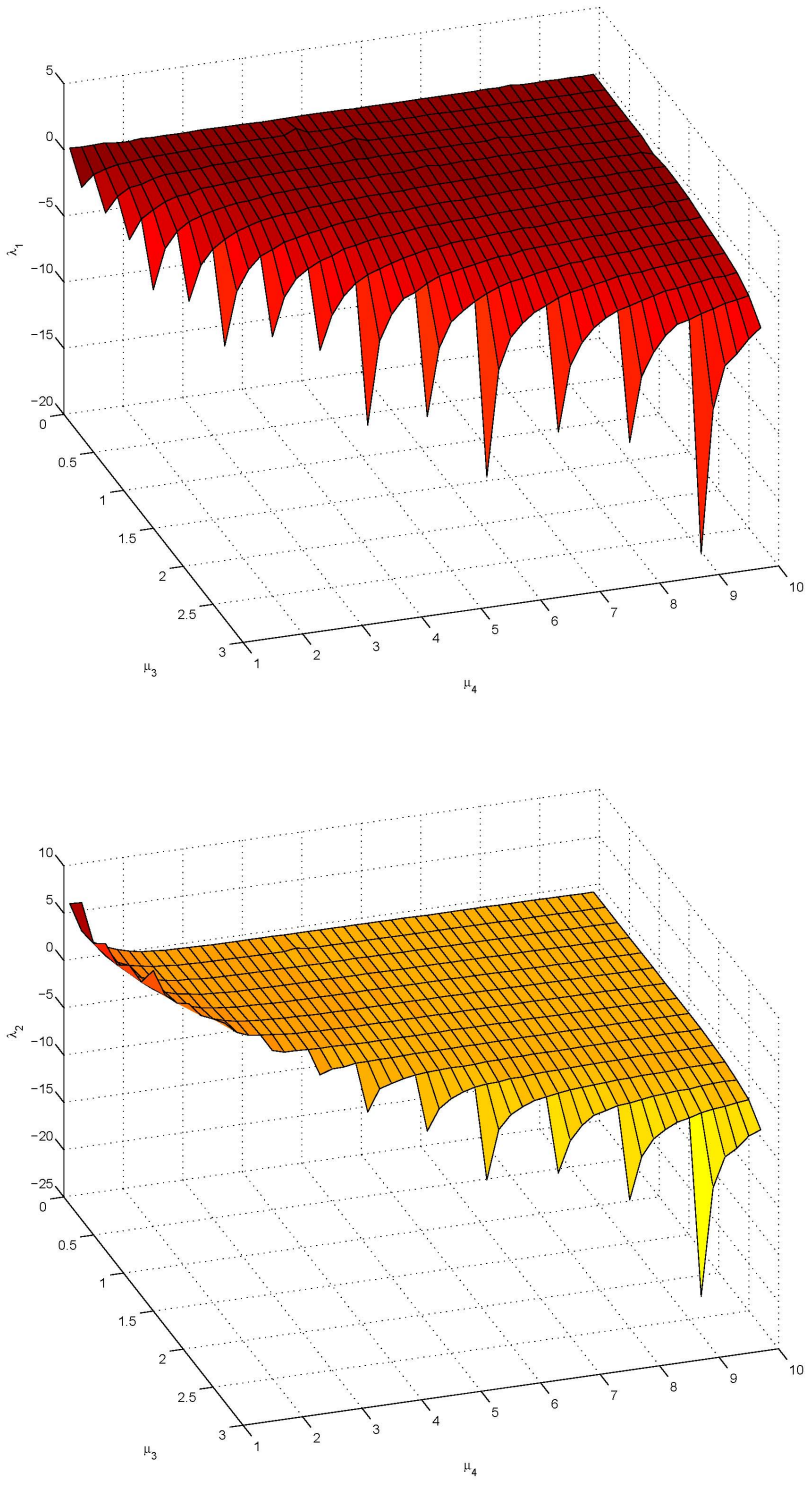

Figure 1: $\hat{\lambda}_1$ (top) and $\hat{\lambda}_2$ (bottom) for given μ_3 and μ_4, $(n = 10,000,\ M = 8)$

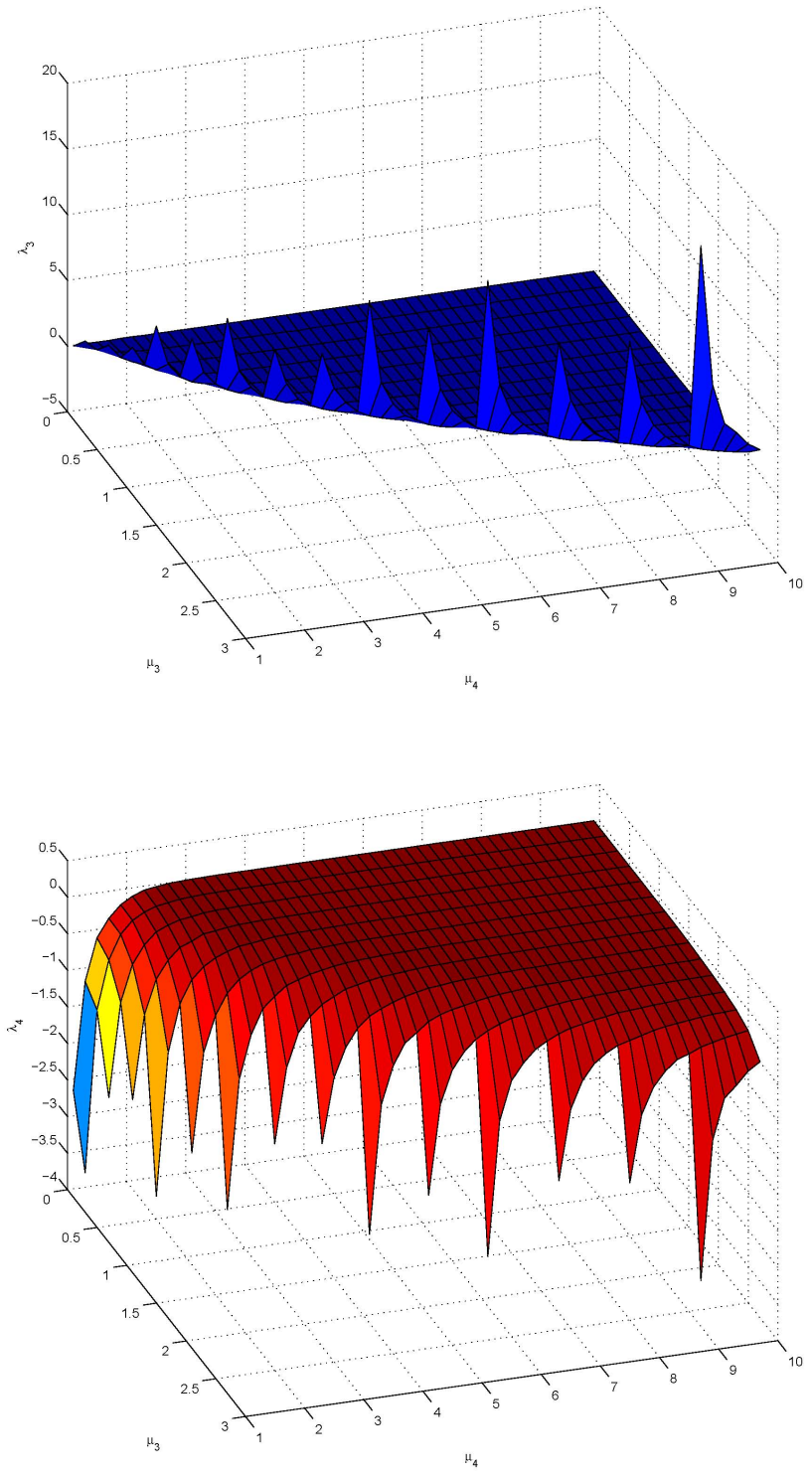

Figure 2: $\hat{\lambda}_3$ (top) and $\hat{\lambda}_4$ (bottom) for given μ_3 and μ_4, $(n = 10,000,\ M = 8)$

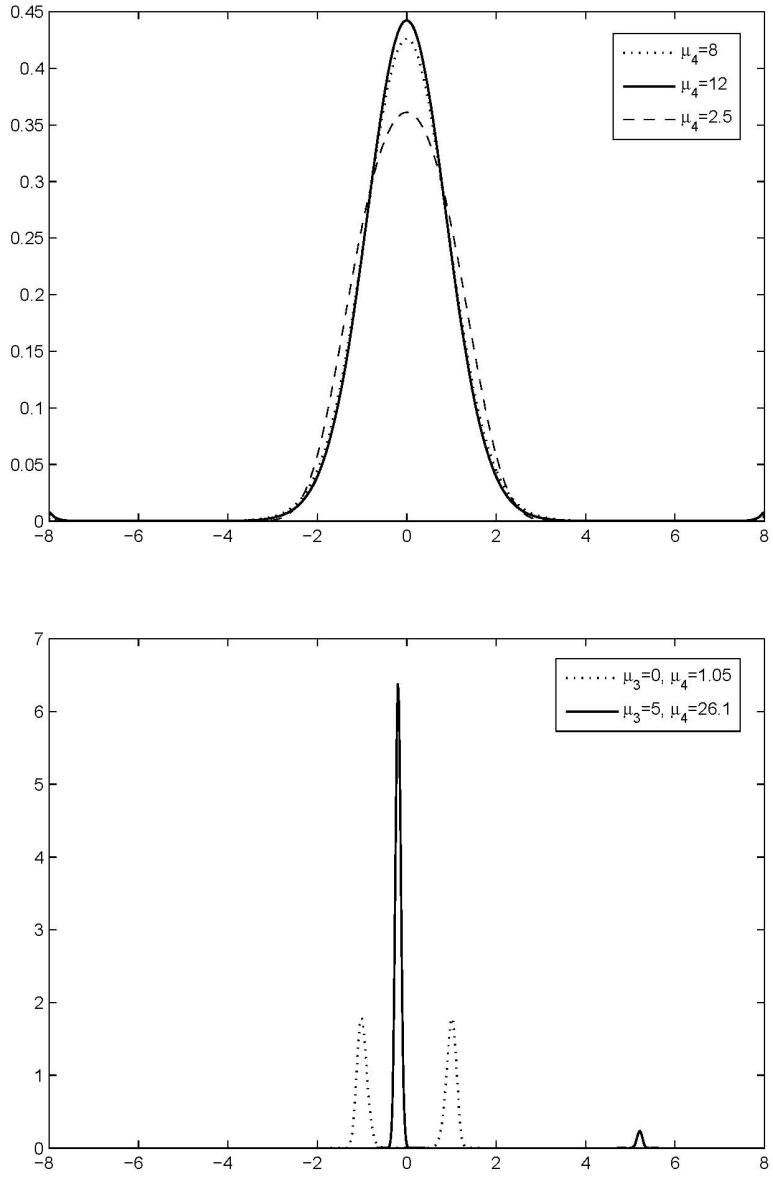

Figure 3: Estimated Maxent Density with Extreme Moment Conditions ($n = 30000$, $M = 8$)

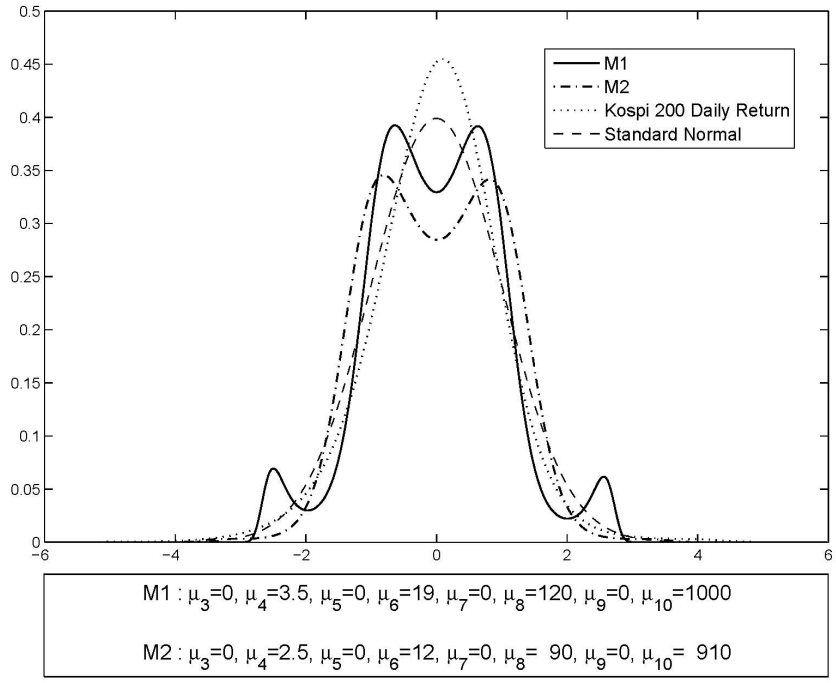

Figure 4: Various Maxent Densities with Ten Moment Conditions ($n = 30000$, $M = 8$)