THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 24-07

THANASIS STENGOS
University Of Guelph
and
Rimini Center for Economic Analysis

XIMING WU
Texas A&M University

"INFORMATION-THEORETIC DISTRIBUTION TEST WITH APPLICATION TO NORMALITY"

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

Information-Theoretic Distribution Test with Application to Normality

Thanasis Stengos* and Ximing Wu†

Abstract

We derive general distribution tests based on the method of Maximum Entropy density. The proposed tests are derived from maximizing the differential entropy subject to moment constraints. By exploiting the equivalence between the Maximum Entropy and Maximum Likelihood estimates of the general exponential family, we can use the conventional Likelihood Ratio, Wald and Lagrange Multiplier testing principles in the maximum entropy framework. In particular we use the Lagrange Multiplier method to derive tests for normality and their asymptotic properties. Monte Carlo evidence suggests that the proposed tests have desirable small sample properties.

*Department of Economics, University of Guelph and Rimini Center for Economic Analysis. Email: tstengos@uoguelph.ca.

†Corresponding author. Department of Agricultural Economics, Texas A&M University and Department of Economics, University of Guelph. Email: xwu@ag.tamu.edu.

We want to thank seminar participants at Penn State University, the 2004 European Meetings of the Econometric Society and the 2004 Canadian Econometrics Study Group for comments. We also thank the associate editor and two referees for comments and suggestions that helped us improve this paper. Financial support from SSHRC of Canada is gratefully acknowledged.

JEL code: C1, C12, C16; Key words: distribution test, maximum entropy, normality.

1 Introduction

Testing that a given sample comes from a particular distribution is one of the most important topics in inferential statistics and can be dated back to as early as Pearson (1900)'s χ^2 goodness-of-fit test. In particular, testing for normality, thanks to the prominent role of the central limit theorem in statistics, has received an extensive treatment in the literature, see Thode (2002) who provides a comprehensive review on this topic. In this paper we present alternative tests for a given distribution, in particular the normal distribution, based on the method of maximum entropy (ME) density. They are derived from maximizing differential entropy subject to known moment constraints. By exploiting the equivalence between ME and maximum likelihood estimates for the exponential family, we can use the conventional likelihood ratio (LR), Wald and Lagrange Multiplier (LM) testing principles in the maximum entropy framework. Hence, our tests share the optimality properties of the standard maximum likelihood based tests. Using the LM method, we show that the ME approach leads to simple yet powerful tests for normality. Our Monte Carlo simulations show that the proposed tests compare favorably and often outperform the commonly used tests in the literature, such as the Jarque-Bera test and the Kolmogorov-Smirnov-Lillie test for normality, especially when the sample size is small. In addition, we show that the proposed method can be easily extended to: i) other distributions in addition to the normal; ii) regression residuals; iii) dependent and/or heteroskedastic data. We apply the proposed tests to test the normality of residuals from a stochastic production frontier model using a benchmark dataset.

The paper is organized as follows. In the next section we present the information theoretic framework on which we base our analysis. We then proceed to derive our normality tests and discuss their properties. In the following section we present some simulation results. Finally, before we conclude, we discuss some possible extensions and an empirical application. The appendix collects the proofs of the main results.

2 Information-theoretic distribution test

The information entropy, the central concept of the information theory, was introduced by Shannon (1949). Entropy is an index of disorder and uncertainty. The maximum entropy (ME) principle states that among all the distributions that satisfy certain information constraints, one should choose the one that maximizes Shannon's information entropy. According to Jaynes (1957), the ME distribution is "uniquely determined as the one which is maximally noncommittal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters."

The ME density is obtained by maximizing the entropy subject to some moment constraints. Let x be a random variable distributed with a probability density function (pdf) $f(x)$, and $X_1, X_2, ..., X_n$ be an i.i.d. random sample of size n generated according to $f(x)$. The unknown density $f(x)$ is assumed to be continuously differentiable, positive on the interval of support (usually the real line if there is no prior information on the support of the density) and bounded. We maximize the entropy

$$\max_{f(x)} : \; W = -\int f(x) \log f(x) \, dx,$$

subject to

$$\int f(x) \, dx = 1,$$
$$\int g_k(x) f(x) \, dx = \hat{\mu}_k, \; k = 1, 2, \ldots, K,$$

where $g_k(x)$ is continuously differentiable and $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} g_k(X_i)$. The solution takes the form

$$f\left(x, \hat{\boldsymbol{\theta}}\right) = \exp\left(-\hat{\theta}_0 - \sum_{k=1}^{K} \hat{\theta}_k g_k(x)\right), \tag{1}$$

where $\hat{\theta}_k$ is the Lagrangian multiplier associated with the kth moment constraint in the optimization problem (see Zellner and Highfield, 1988, for details). To ensure $f\left(x, \hat{\boldsymbol{\theta}}\right)$

integrates to one, we set

$$\hat{\theta}_0 = \log\left(\int \exp\left(-\sum_{k=1}^{K}\hat{\theta}_k g_k(x)\right)dx\right).$$

The maximized entropy $W = \hat{\theta}_0 + \sum_{k=1}^{K}\hat{\theta}_k\hat{\mu}_k$.

The ME density is of the generalized exponential family and can be completely characterized by the moments $Eg_k(x), k = 1, 2, \ldots, K$. We call these moments "characterizing moments", whose sample counterparts are the sufficient statistics of the estimated ME density $f(x, \hat{\boldsymbol{\theta}})$. A wide range of distributions belong to this family. For example, the Pearson family and its extensions described in Cobb et al. (1982), which nest the normal, beta, gamma and inverse gamma densities as special cases, are all ME densities with simple characterizing moments.

In general, there is no analytical solution for the ME density problem, and nonlinear optimization methods are required (Zellner and Highfield (1988), Ornermite and White (1999) and Wu (2003)). We use Lagrange's method to solve this problem by iteratively updating $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{(t+1)} = \hat{\boldsymbol{\theta}}_{(t)} - \mathcal{H}^{-1}\boldsymbol{b},$$

where for the $(t+1)$th stage of the updating, $b_k = \int g_k(x) f\left(x, \hat{\boldsymbol{\theta}}_{(t)}\right)dx - \hat{\mu}_k$ and the Hessian matrix $\mathcal{H}$ takes the form

$$\mathcal{H}_{k,j} = \int g_k(x)\, g_j(x)\, f\left(x, \hat{\boldsymbol{\theta}}_{(t)}\right)dx,\ \ 0 \leq k, j \leq K.$$

The positive-definitiveness of the Hessian ensures the existence and uniqueness of the solution.[1]

[1]Let $\gamma' = [\gamma_0, \gamma_1, \ldots, \gamma_K]$ be a non-zero vector and $g_0(x) = 1$, we have

$$\begin{aligned}\gamma'\mathcal{H}\gamma &= \sum_{k=0}^{K}\sum_{j=0}^{K}\gamma_k\gamma_j \int g_k(x)\, g_j(x)\, f(x, \boldsymbol{\theta})\, dx \\ &= \int \left(\sum_{k=0}^{K}\gamma_k g_k(x)\right)^2 f(x, \boldsymbol{\theta})\, dx > 0.\end{aligned}$$

Hence, $\mathcal{H}$ is positive-definite.

Given Equation (1), we can also estimate $f(x, \boldsymbol{\theta})$ using the MLE. The maximized log-likelihood

$$
\begin{aligned}
l = \sum_{i=1}^{n} \log f\left(x_i, \hat{\boldsymbol{\theta}}\right) &= -\sum_{i=1}^{n}\left(\hat{\theta}_0 + \sum_{k=1}^{K} \hat{\theta}_k g_k(x_i)\right) \\
&= -n\left(\hat{\theta}_0 + \sum_{k=1}^{K} \hat{\theta}_k \hat{\mu}_k\right) = -nW.
\end{aligned}
$$

Therefore, when the distribution is of the generalized exponential family, MLE and ME are equivalent. Moreover, they are also equivalent to a method of moments (MM) estimator. This ME/MLE/MM estimator only uses the sample characterizing moments.

Although the MLE and ME are equivalent in our case, there are some conceptual differences. For the MLE, the restricted estimates are obtained by imposing certain constraints on the parameters. In contrast, for the ME, the dimension of the parameter is determined by the number of moment restrictions imposed: the more moment restrictions, the more complex and at the same time the more flexible the distribution is. To reconcile these two methods, we note that a ME estimate with the first m moment restrictions has a solution of the form

$$
f(x, \boldsymbol{\theta}) = \exp\left(-\theta_0 - \sum_{k=1}^{m} \theta_k g_k(x)\right),
$$

which implicitly sets θ_j, $j = m+1, m+2, \ldots$, to be zero. When we impose more moment restrictions, say, $\int g_{m+1}(x) f(x, \boldsymbol{\theta})\, dx = \hat{\mu}_{m+1}$, we let the data choose the appropriate value of θ_{m+1}.[2] In this sense, the estimate with more moment restrictions is in fact less restricted, or more flexible. The ME and MLE share the same objective function (up to a proportion) which is determined by the moment restrictions of the maximum entropy problem. Therefore, one can regard the ME approach as a method of model selection, which generates a MLE solution.

We can use the ME approach for distribution tests. Consider a M dimension parameter space Θ_M. Suppose we want to test the hypothesis that $\boldsymbol{\theta} \in \Theta_m$, a subspace of Θ_M, where

[2] Denote $\boldsymbol{\theta}_m = [\theta_1, \ldots, \theta_m]$. The only case that $\theta_{m+1} = 0$ is when the moment restriction $\int g_{m+1}(x) f(x, \boldsymbol{\theta}_m)\, dx = \hat{\mu}_{m+1}$ is not binding, or the $(m+1)$th moment is identical to its prediction based on the ME density $f(x, \boldsymbol{\theta}_m)$ from the first m moments. In this case, the $(m+1)$th moment does not contain any additional information that can further reduce the entropy.

$m \leq M$. Because of the equivalence between the ME and MLE, we can use the traditional LR, Wald and LM principles to construct test statistics.[3] For $j = m, M$, let $\boldsymbol{\theta}_j$ be the MLE estimates in Θ_j, l_j and W_j be their corresponding log-likelihood and maximized entropy, we have

$$\begin{aligned}
& -\int f(x, \boldsymbol{\theta}_m) \log f(x, \boldsymbol{\theta}_m)\, dx \\
&= \int \left(\sum_{k=0}^{m} \theta_{m,k} g_k(x) \right) f(x, \boldsymbol{\theta}_m)\, dx \\
&= \sum_{k=0}^{m} \theta_{m,k} \int g_k(x) f(x, \boldsymbol{\theta}_m)\, dx \\
&= \sum_{k=0}^{m} \theta_{m,k} \int g_k(x) f(x, \boldsymbol{\theta}_M)\, dx \\
&= \int \left(\sum_{k=0}^{m} \theta_{m,k} g_k(x) \right) f(x, \boldsymbol{\theta}_M)\, dx \\
&= -\int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_m)\, dx.
\end{aligned}$$

The fourth equality follows because the first m moments of $f(x, \boldsymbol{\theta}_m)$ are identical to those of $f(x, \boldsymbol{\theta}_M)$. Consequently, the log-likelihood ratio

$$\begin{aligned}
\mathrm{LR} &= -2(l_m - l_M) \\
&= 2n(W_m - W_M) \\
&= -2n\left(\int f(x, \boldsymbol{\theta}_m) \log f(x, \boldsymbol{\theta}_m)\, dx - \int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_M)\, dx \right) \\
&= 2n\left(\int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_M)\, dx - \int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_m)\, dx \right) \\
&= 2n \int f(x, \boldsymbol{\theta}_M) \log \frac{f(x, \boldsymbol{\theta}_M)}{f(x, \boldsymbol{\theta}_m)} dx,
\end{aligned}$$

which is the Kullback-Leibler distance between $f(x, \boldsymbol{\theta}_M)$ and $f(x, \boldsymbol{\theta}_m)$ multiplied by twice

[3] Imbens et al. (1998) discussed similar tests in the information-theoretic generalized empirical likelihood framework. The proposed tests differ from their tests, which minimize the discrete Kullback-Leibler information criterion (cross entropy) or other Cressie-Read family of discrepancy indices subject to moment constraints.

of the sample size. Consequently, if the true model $f(x, \boldsymbol{\theta}_M)$ nests $f(x, \boldsymbol{\theta}_m)$, the quasi-MLE estimate $f(x, \boldsymbol{\theta}_m)$ minimizes the Kullback-Leibler statistic between $f(x, \boldsymbol{\theta}_M)$ and $f(x, \boldsymbol{\theta}_m)$, as shown in White (1982).

If we partition $\boldsymbol{\theta}_u = (\boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) = (\boldsymbol{\theta}_{1u}, \boldsymbol{\theta}_{2u})$ for the unrestricted model and similarly $\boldsymbol{\theta}_r = (\boldsymbol{\theta}_{1r}, 0)$ for the restricted model, then the score function

$$\mathcal{S}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) = \begin{pmatrix} \frac{\partial \ln f}{\partial \boldsymbol{\theta}_m}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) \\ \frac{\partial \ln f}{\partial \boldsymbol{\theta}_{M-m}}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) \end{pmatrix},$$

and the Hessian

$$\mathcal{H}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) = \begin{pmatrix} \frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}_m \partial \boldsymbol{\theta}'_m}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) & \frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}_m \partial \boldsymbol{\theta}'_{M-m}}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) \\ \frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}_{M-m} \partial \boldsymbol{\theta}'_m}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) & \frac{\partial^2 \ln f}{\partial \boldsymbol{\theta}_{M-m} \partial \boldsymbol{\theta}'_{M-m}}(x, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{M-m}) \end{pmatrix}.$$

If we partition similarly the inverse of the information matrix $\mathcal{I} = -E(H)$ as

$$\mathcal{I}^{-1} = \begin{pmatrix} \mathcal{I}^{11} & \mathcal{I}^{12} \\ \mathcal{I}^{21} & \mathcal{I}^{22} \end{pmatrix},$$

then the Wald test statistic is defined as

$$\text{WALD} = n\hat{\boldsymbol{\theta}}'_{2u} \left(\hat{\mathcal{I}}^{22}\right)^{-1} \hat{\boldsymbol{\theta}}_{2u},$$

whereas the Lagrange Multiplier test statistic is defined as

$$\text{LM} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathcal{S}}\left(x_i, \hat{\boldsymbol{\theta}}_{1r}, 0\right)' \hat{\mathcal{I}}^{-1} \sum_{i=1}^{n} \hat{\mathcal{S}}\left(x_i, \hat{\boldsymbol{\theta}}_{1r}, 0\right).$$

All three tests are asymptotically equivalent and distributed as χ^2 with $(M - m)$ degrees of freedom under the null hypothesis (see for example, Engle, 1984).

3 Tests of Normality

In this section, we use the proposed ME method to derive tests for normality. Since the LR and Wald procedures require the estimation of the unrestricted ME density, which in general has no analytical solution and can be computationally involved, we focus on the LM test, which reduces surprisingly to a test statistic with a simple closed form.

3.1 Flexible ME Density Estimators

Suppose a density can be rewritten as or approximated by a sufficiently flexible ME density

$$f_0(x) = \exp\left(-\sum_{k=0}^{2}\theta_k x^k - \sum_{k=3}^{K}\theta_k g_k(x)\right).$$

Two conditions are required to ensure that $f_0(x)$ is integrable over the real line. First, the dominant term in the exponent must be an even function; otherwise, $f_0(x)$ will explode at either tail as $x \to \infty$ or $x \to -\infty$. The second condition is that the coefficient associated with the dominant term, which is an even function by the first condition, must be positive; otherwise $f_0(x)$ will explode to ∞ at both tails as $|x| \to \infty$.

The LM test of normality amounts to testing $\theta_k = 0$ for $k = 3, \ldots, K$. In practice, only a small number of moments $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} g_k(x_i)$ are used for the test, especially when the sample size is small. In this paper we consider three simple, yet flexible functional forms. If we approximate $f_0(x)$ using the ME density subject to the first four arithmetic moments, the solution takes the form

$$f_1(x) = \exp\left(-\sum_{k=0}^{4}\theta_k x^k\right).$$

This classical exponential quartic density was first discussed by Fisher (1922) and studied in the maximum entropy framework in Zellner and Highfield (1988), Ornermite and White (1999) and Wu (2003).

In practice, it is well known that the third and fourth sample moments can be sensitive to outliers. In addition to the robustness consideration, Dalén (1987) shows that sample moments are restricted by sample size, which makes higher order moments unsuitable for small sample problem. A third problem with the quartic exponential form is that this specification does not admit $\mu_4 > 3$ if $\mu_3 = 0$. To see this point denote $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_4]$. Stohs (2003) shows that for the one-to-one mapping $\boldsymbol{\theta} = M(\boldsymbol{\mu})$, the gradient matrix $\mathcal{H}$ with $\mathcal{H}_{ij} = \mu_{i+j} - \mu_i\mu_j$, $1 \leq i, j \leq 4$, is positive definite and so is $\mathcal{H}^{-1}$. Denote $\mathcal{H}^{(4,4)}$ the lower-right-corner entry of $\mathcal{H}^{-1}$. It follows $\mathcal{H}^{(4,4)} > 0$. Consider a distribution with $\boldsymbol{\mu} = [0, 1, 0, 3]$, which are identical to the first four moments of the standard normal distribution. Clearly, $\theta_2 = 1/2$ and $\theta_1 = \theta_3 = \theta_4 = 0$. Suppose we introduce a small disturbance $d\boldsymbol{\mu} = [0, 0, 0, \delta]$, where $\delta > 0$. Since $d\boldsymbol{\theta} = -\mathcal{H}^{-1}d\boldsymbol{\mu}$, we have $d\theta_4 = -\mathcal{H}^{(4,4)}\delta < 0$. It then follows that $\theta_4 < 0$, which renders the approximation $f_1(x)$ nonintegrable.

Although $f_1(x)$ is rather flexible, the restriction discussed above precludes the applicability of the ME density to symmetric fat-tailed distributions, which occur frequently in practice. Hence, we consider an alternative specification which is motivated by the fat-tailed Student's t distribution. We note that the t distribution with r degrees of freedom has the density

$$T_r(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma\left(\frac{r}{2}\right)\left(1 + \frac{x^2}{r}\right)^{(r+1)/2}} = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\Gamma\left(\frac{r}{2}\right)} \exp\left\{-\frac{r+1}{2}\log\left(1 + \frac{x^2}{r}\right)\right\},$$

which can be characterized as an exponential distribution with the general moment of $\log\left(1 + \frac{x^2}{r}\right)$. Accordingly, we can modify the normal density by adding the extra moment condition that $E\log\left(1 + \frac{x^2}{r}\right)$ be equal to its sample estimate. The resulting general ME density

$$f_1'(x) = \exp\left(-\sum_{k=0}^{2}\theta_k x^k - \theta_3\log\left(1 + \frac{x^2}{r}\right)\right),$$

where $r > 0$. Since $\log\left(1 + \frac{x^2}{r}\right) = o(x)$, x^2 is the dominant term for all r, which implies that $\theta_2 > 0$ to ensure the integrability of $f_1'(x)$ over the real line. The presence of $\log(1 + \frac{x^2}{r})$ allows the ME density to accommodate symmetric fat-tailed distributions.

To make the specification more flexible, we further introduce a term to capture skewness and asymmetry. One possibility is to use $\tan^{-1}(x)$ which is an odd function and bounded between $(-1, 1)$. Formally, Lye and Martin (1993) derive the generalized t distribution from the generalized Pearson family defined by

$$\frac{df}{dx} = -\frac{\left(\sum_{k=1}^{2}\theta_k x^k\right) f(x)}{(r^2+x^2)}.$$

The solution takes the form

$$f_2(x) = \exp\left(-\sum_{k=0}^{2}\theta_k x^k - \theta_3 \tan^{-1}\left(\frac{x}{r}\right) - \theta_4 \log\left(r^2+x^2\right)\right),\ r>0.$$

Since the "degrees of freedom" r is unknown, we set $r = 1$, which allows the maximum degree of fat-tailedness.[4] Therefore, the alternative ME density is defined as

$$f_2(x) = \exp\left(-\sum_{k=0}^{2}\theta_k x^k - \theta_3 \tan^{-1}(x) - \theta_4 \log\left(1+x^2\right)\right).$$

We further notice an "asymmetry" between $\tan^{-1}(x)$ and $\log(1+x^2)$ in the sense that the former is bounded while the latter is unbounded. Therefore, we consider yet another alternative wherein we replace $\log(1+x^2)$ by $\tan^{-1}(x^2)$.[5] Hence, our third ME density is defined as

$$f_3(x) = \exp\left(-\sum_{k=0}^{2}\theta_k x^k - \theta_3 \tan^{-1}(x) - \theta_4 \tan^{-1}\left(x^2\right)\right).$$

Since $\partial \tan^{-1}(x)/\partial x = 1 - [\tan^{-1}(x)]^2 > 0$, $\tan^{-1}(x)$ is monotonically increasing in x. Similarly, $\partial \tan^{-1}(x^2)/\partial x = 2x\left\{1 - [\tan^{-1}(x^2)]^2\right\}$, implying $\tan^{-1}(x^2)$ monotonically increasing in $|x|$. Therefore, $\tan^{-1}(x)$ and $\tan^{-1}(x^2)$ are able to mimic the behavior of x^3 and x^4 yet at the same time remain bounded such that $f_3(x)$ is able to accommodate distributions with exceptionally large skewness and kurtosis. Note that $f_3(x)$ is in spirit close to Gallant (1981)'s flexible Fourier transformation where low-order polynomials are combined with trigonometric series to achieve a balance of parsimony and flexibility. In Wu and Stengos (2005), we also consider $\sin(x)$ and $\cos(x)$ for flexible ME densities. Generally, using periodic functions like $\sin(x)$ and $\cos(x)$ requires rescaling the data to be within $[-\pi, \pi]$.

[4] A t distribution with one degree of freedom is the Cauchy distribution, which has the fattest tail within the family of t distribution.

[5] We also tried $\left[\tan^{-1}(x)\right]^2$. The performance was essentially the same as that with $\tan^{-1}\left(x^2\right)$.

Although in principle they are equally suitable for density approximation, we do not consider specifications with $\sin(x)$ and $\cos(x)$ in this study as rescaling the data to be within $[-\pi, \pi]$, rather than standardizing them, requires us to calculate the asymptotic variance under normality for each dataset.

The introduction of general moments offers a considerably higher degree of flexibility as we are not restricted to polynomials. Generally, by choosing general moments appropriately from distributions that are known to accommodate the given moment conditions, we make the ME density more robust and at the same time more flexible. As an illustration, Figure 1 shows the ME approximation to a χ^2 distribution with five degrees of freedom by $f_1(x)$, $f_2(x)$ and $f_3(x)$. Although they have relatively simple functional forms, all three ME densities are shown to capture the general shape of the χ_5^2 density rather well.

3.2 Normality Tests

In this section we derive tests for normality based on the ME densities $f_1(x)$, $f_2(x)$ and $f_3(x)$ presented in the previous section. When $\theta_3 = \theta_4 = 0$, all three densities reduce to the standard normal density.[6] The information matrix of $f_1(x)$ under standard normality is

$$\mathcal{I}_1 = \begin{bmatrix} 1 & 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 3 & 0 \\ 1 & 0 & 3 & 0 & 15 \\ 0 & 3 & 0 & 15 & 0 \\ 3 & 0 & 15 & 0 & 105 \end{bmatrix},$$

and the score function under normality is $\hat{\mathcal{S}}_1 = n[0, 0, 0, \hat{\mu}_3, \hat{\mu}_4 - 3]$. It follows that the LM test statistic

$$t_1 = \frac{1}{n}\hat{\mathcal{S}}_1'\mathcal{I}_1^{-1}\hat{\mathcal{S}}_1 = n\left(\frac{\hat{\mu}_3^2}{6} + \frac{(\hat{\mu}_4 - 3)^2}{24}\right).$$

[6]Shannon (1949) shows that among all distributions that possess a density function $f(x)$ and have a given variance σ^2, the entropy $W = -\int f(x)\log f(x)\,dx$ is maximized by the normal distribution. The entropy of the normal distribution with variance σ^2 is $\log\left(\sqrt{2\pi e}\sigma\right)$. Vasicek (1976) uses this property to test a composite hypothesis of normality, based on a nonparametric estimates of sample entropy.

This the familiar JB test of normality. Bera and Jarque (1981) derived this test as a Lagrange Multiplier test for the Pearson family of distributions and White (1982) derived it as an information matrix test. More recently, Bontemps and Meddahi (2005) applied the Stein Equation to the mean of Hermite polynomials to arrive at the same test. However, Bai and Ng (2005) notes that the convergence of $\frac{(\hat{\mu}_4-3)^2}{24}$ to its asymptotic distribution could be rather slow and the sample kurtosis can deviate substantially from its true value even with a large number of observations. Since the kurtosis test has very low power, the power of normality test based on skewness and kurtosis largely reflect the results of the skewness tests.

Instead of using the coefficients of skewness and kurtosis, whose small sample properties are unsatisfactory, we consider tests based on alternative ME densities $f_2(x)$ and $f_3(x)$. Under normality, the information matrix of $f_2(x)$ takes the form

$$\mathcal{I}_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0.5334532 \\ 0 & 1 & 0 & 0.6057055 & 0 \\ 1 & 0 & 3 & 0 & 1.2220941 \\ 0 & 0.6057055 & 0 & 0.3942945 & 0 \\ 0.5334532 & 0 & 1.2220941 & 0 & 0.5529086 \end{bmatrix},$$

and the score

$$\hat{\mathcal{S}}_2 = n\left[0, 0, 0, \hat{\mu}_a, \hat{\mu}_b - 0.5334532\right],$$

where $\hat{\mu}_a = \frac{1}{n}\sum_{i=1}^{n}\tan^{-1}(X_i)$ and $\hat{\mu}_b = \frac{1}{n}\sum_{i=1}^{n}\log\left(1+X_i^2\right)$. The alternative LM test is given by

$$t_2 = \frac{1}{n}\hat{\mathcal{S}}_2'\mathcal{I}_2^{-1}\hat{\mathcal{S}}_2 = n\left(36.47595\hat{\mu}_a^2 + 32.027545\left(\hat{\mu}_b - 0.5334532\right)^2\right).$$

Similarly, the information matrix for $f_3(x)$ under normality is

$$\mathcal{I}_3 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0.4741131 \\ 0 & 1 & 0 & 0.6057055 & 0 \\ 1 & 0 & 3 & 0 & 0.8692134 \\ 0 & 0.6057055 & 0 & 0.3942945 & 0 \\ 0.4741131 & 0 & 0.8692134 & 0 & 0.3620107 \end{bmatrix},$$

and the score

$$\hat{\mathcal{S}}_3 = n\left[0, 0, 0, \hat{\mu}_a, \hat{\mu}_c - 0.4741131\right],$$

where $\hat{\mu}_c = \frac{1}{n}\sum_{i=1}^{n}\tan^{-1}(X_i^2)$. The LM test statistic is then computed as

$$t_3 = \frac{1}{n}\hat{\mathcal{S}}_3'\mathcal{I}_3^{-1}\hat{\mathcal{S}}_3 = n\left(36.47595\hat{\mu}_a^2 + 16.898926\left(\hat{\mu}_c - 0.4741131\right)^2\right).$$

The following theorem shows that all three tests are asymptotically distributed as χ^2 with two degrees of freedom under normality.

Theorem 1. *Under the assumption that $E|x|^{4+\delta} < \infty$ for $\delta > 0$, the test statistics t_1, t_2 and t_3 are distributed asymptotically as χ^2 with two degrees of freedom under normality.*

The proof is presented in the appendix.

Under normality, the correlation between $\hat{\mu}_3$ and $\hat{\mu}_4$ is practically zero, so is that between $\hat{\mu}_a$ and $\hat{\mu}_b$, and between $\hat{\mu}_a$ and $\hat{\mu}_c$. However, we note that the correlation of $|\hat{\mu}_3|$ and $\hat{\mu}_4$ is 0.65 and 0.53 from 10,000 repetitions of random normal samples with $n = 20$ and 50, while the correlation of $|\hat{\mu}_a|$ and $\hat{\mu}_b$ is -0.25 and -0.14, the correlation between $|\hat{\mu}_a|$ and $\hat{\mu}_c$ is -0.19 and -0.13 for the same sample size. Therefore, we expect that t_2 and t_3 to have better small sample performance than t_1.

4 Simulations

In this section, we use Monte Carlo simulations to assess the size and power of the proposed tests. Following Bai and Ng (2005), we consider well known distributions such as the normal, the t and the χ^2, as well as distributions from the generalized lambda family. The generalized lambda distribution, denoted F_λ, is defined in terms of the inverse of the cumulative distribution $F^{-1}(u) = \lambda_1 + \left[u^{\lambda_3} - (1-u)^{\lambda_4}\right]/\lambda_2, 0 < u < 1$. This family nests a wide range of symmetric and asymmetric distributions. In particular, we consider the following symmetric and asymmetric distributions:

S1: $N(0,1)$

S2: t distribution with 5 degrees of freedom

S3: $e_1 I(z \leq 0.5) + e_2 I(z > 0.5)$, where $z \sim U(0,1)$, $e_1 \sim N(-1,1)$, and $e_2 \sim N(1,1)$

S4: $F_\lambda, \lambda_1 = 0, \lambda_2 = 0.19754, \lambda_3 = 0.134915, \lambda_4 = 0.134915$

S5: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.8, \lambda_4 = -0.8$

S6: $F_\lambda, \lambda_1 = 0, \lambda_2 = -.397912, \lambda_3 = -.16, \lambda_4 = -.16$

S7: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -.24, \lambda_4 = -.24$

A1: lognormal: $\exp(e)$, $e \sim N(0,1)$

A2: χ^2 distribution with 3 degrees of freedom

A3: exponential: $-\ln(e)$, $e \sim U(0,1)$

A4: $F_\lambda, \lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 1.4, \lambda_4 = 0.25$

A5: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.0075, \lambda_4 = -0.03$

A6: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -.1, \lambda_4 = -.18$

A7: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -.001, \lambda_4 = -.13$

A8: $F_\lambda, \lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -.0001, \lambda_4 = -.17$

The first seven distributions are symmetric and the next eight are asymmetric, which have a wide range of skewness and kurtosis as shown in Table 1. For each distribution, we draw 10,000 random samples of size $n = 20, 50, 100$ respectively and compute the normality test statistics discussed above. For the sake of comparison, we also compute the commonly used Kolmogorov-Smirnov (KS) test. We note that the general-purpose KS test has very low power. Instead, we use the Lillie test, which is a special version of KS test tailored for the case of normality (see Thode, 2002).

Table 1 reports the results of the normality tests at the 5% significance level. The first row reflects the size and the rest show the power of the tests. All the four tests have similar desirable size, except that the KS test is slightly oversized when $n = 20$.

For $n = 20$, t_2 and t_3 have higher power than t_1 for all distributions considered in the simulation. The powers of t_2 and t_3 are similar except for $S3$ and $A4$. For the thin-tailed $S3$, the power of t_2 and t_3 is respectively three and five times of that of t_1. For the distribution $A4$ with a relatively large skewness and a thin tail, the power of t_2 and t_3 is respectively ten and fourteen times of that of t_1. On the other hand, the power of the KS test is generally lower than that of t_2 and especially t_3, except for $S4$, where the KS test has a power of 0.06, larger than the power of t_3 test, which is 0.04.

As n increases, the size of all tests converge to the theoretical level and their powers generally increase. For the distribution $S3$, where the t_3 and KS test have comparable powers when $n = 20$, we note that the power of t_3 test is 43% higher than that of the KS test when $n = 50$ and 90% higher when n increases to 100. Also, for the distribution $S4$, the t_3 and KS test have comparable powers when $n = 50$ and 100. Since this distribution shares the same first four moments with the standard normal distribution, the powers of all four tests are similar to their respective size, reflecting the difficulty in distinguishes $S4$ from the normal distribution. This difficulty is also reported in Bai and Ng (2005). We also note that for $A4$, the power of t_2 and t_3 increases rapidly, while the power of t_1 remains considerably lower than other tests when $n = 50$.

Overall, our results suggest that the proposed tests are comparable to and often outperform the commonly used JB test and KS test, especially when the sample size is small.

5 Extensions

In addition to their simplicity, a major advantage of the proposed tests is its generality. In this section, we briefly discuss some easy-to-implement extensions of the tests.

Firstly, we note that we can incorporate higher order polynomials x^k for $k > 4$ and higher order trigonometric terms such as $\tan^{-1}\left(x^k\right)$ for $k > 2$. Usually, addition of higher order terms will improve the approximation to the underlying distribution. However, we note that it does not necessarily improve the test. We experimented with adding x^5 and x^6 to $f_2(x)$ and $\tan^{-1}(x^3)$ and $\tan^{-1}(x^4)$ to $f_3(x)$ and derived tests based on four instead of two moment conditions.[7] These alternative tests are distributed asymptotically according to a χ^2_4 distribution under normality. However, we note that their powers are generally lower than tests based on two moment conditions. This is to be expected as the test statistics are distributed

[7] The first two moments are zero and one by standardization.

according to a non-central χ^2 distribution under alternative non-normal distributions. For a given non-centrality parameter there is an inverse relationship between degrees of freedom and power, see Das Gupta and Perlman (1974). One reason behind the loss in power in our case is that with four moment conditions, the two even moments and two odd moments are each correlated, which lowers the power of the tests when the sample size is small.

Secondly, we can use the proposed method for other distributions than the normal. For example, the gamma distribution can be characterized as a ME distribution

$$f(x) = \exp(-\theta_0 - \theta_1 x - \theta_2 \log x), x > 0.$$

Because Ex and $E \log x$ are the characterizing moments for gamma distribution, the presence of any additional terms in the exponent of $f(x)$ rejects the hypothesis that x is distributed according to a gamma distribution. Let $f_K(x) = \exp\left(-\theta_0 - \theta_1 x - \theta_2 \log x - \sum_{k=3}^{K} \theta_k g_k(x)\right)$, the test of $\theta_k = 0$ for $k \geq 3$ is then the LM test for gamma distribution. The discussions in previous section suggest that the natural candidates for $g_k(x)$ may include polynomials of x and $\log x$, and trigonometric terms of x and $\log x$.

Thirdly, we can generalize our tests to regression residuals within the framework of White and McDonald (1980). Consider a classical linear model

$$Y_i = Z_i \beta + \varepsilon_i, i = 1, \ldots, n. \tag{2}$$

Since the error term ε_i is not observed, one has to replace it with estimated $\hat{\varepsilon}_i$. The following theorem ensures that the test statistics computed from estimated $\hat{\varepsilon}_i$ share the same asymptotic distribution as those from true ε_i.

Theorem 2. *Assume the following assumptions hold:*

1. $\{Z_i\}$ *is a sequence of uniformly bounded fixed* $1 \times K$ *vectors such that* $Z'Z/n \to M_Z$, *a positive definite matrix,* $\{\varepsilon_i\}$ *is a sequence of iid random variables with* $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma_i^2 < \infty$, *and* β *is an unknown* $K \times 1$ *vector.*

2. $E|\varepsilon_i|^{4+\delta} < \infty$ *for* $\delta > 0$.

3. *The density of* ε_i, $f(\varepsilon)$, *is uniformly continuous, positive on the interval of support and bounded.*

Define $\hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^3$, $\hat{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4$, $\hat{\mu}_a = \frac{1}{n}\sum_{i=1}^{n}\tan^{-1}(\hat{\varepsilon}_i)$, $\hat{\mu}_b = \frac{1}{n}\sum_{i=1}^{n}\log(1+\hat{\varepsilon}_i^2)$ *and* $\hat{\mu}_c = \frac{1}{n}\sum_{i=1}^{n}\tan^{-1}(\hat{\varepsilon}_i^2)$. *Then under normality, the test statistics*

$$\hat{t}_1 = n\left(\hat{\mu}_3^2/6 + (\hat{\mu}_4 - 3)^2/24\right) \sim \chi_2^2,$$
$$\hat{t}_2 = n\left(36.47595\hat{\mu}_a^2 + 32.027545\,(\hat{\mu}_b - 0.5334532)^2\right) \sim \chi_2^2,$$
$$\hat{t}_3 = n\left(36.47595\hat{\mu}_a^2 + 16.898926\,(\hat{\mu}_c - 0.4741131)^2\right) \sim \chi_2^2.$$

The proof is presented in the appendix.

Furthermore, for time series or heteroskedastic data, we can use the approach of Bai and Ng (2005) or Bontemps and Meddahi (2005). In general, for non-iid data, to test the Lagrange Multipliers associated with sample moments of $g_k(x)$ in the ME density being zero, we need to estimate a Heteroskedastic-Autocorrelation-Consistent (HAC) covariance matrix for those moments.

As an illustration, we apply the proposed normality tests to regression residuals. We use data on the production cost of American electricity generating companies from Christensen and Greene (1976). We estimate a flexible cost function with 123 observations:

$$c = \beta_0 + \beta_1 q + \beta_2 q^2 + \beta_3 p_f + \beta_4 p_l + \beta_5 p_k + \beta_6 q p_f + \beta_7 q p_k + \beta_8 q p_l + \varepsilon,$$

where c is total cost, q is total output, p_f, p_l and p_k is the price of fuel, labor and capital respectively, and ε is the error term. All variables are in logarithm. It is expected that the distribution of error terms from this stochastic production frontier model is skewed to the right due to the presence of a firm specific non-negative efficiency component in the error terms. Nonetheless, the KS test fails to reject the normality hypothesis. On the other hand, all three LM tests reject the normality hypothesis with $p-$value at 0.03, 0.01 and 0.02 respectively.

6 Conclusion

In this paper we derive general distribution tests based on the method of maximum entropy density. The proposed tests are derived from maximizing differential entropy subject to moment constraints. By exploiting the equivalence between ME and ML estimates for the exponential family, we can use the conventional LR, Wald and LM testing principles in the maximum entropy framework. Hence, our tests share the optimality properties of the standard ML based tests. In particular, we show that the ME approach leads to simple yet powerful LM tests for normality. We derive the asymptotic properties of the proposed tests and show that they are asymptotically equivalent to the popular Jarque-Bera test. Our Monte Carlo simulations show that the proposed tests have desirable small sample properties. They are comparable and often outperform the JB test and Kolmogorov-Smirnov-Lillie test for normality. Lastly, we show that the proposed method can be generalized to tests for other distributions than the normal. Also, extensions to regression residuals and non-iid data are immediate. We apply the proposed method to the residuals from a stochastic production frontier model and reject the normality hypothesis.

Appendix

Proof of Theorem 1.

Proof. The assumption that $\mathrm{E}|x|^{4+\delta} < \infty$ for $\delta > 0$ ensures the existence of $E\hat{\mu}_3$ and $E\hat{\mu}_4$. One can easily show that $\sqrt{n}\hat{\mu}_3 \sim N(0,6)$ and $\sqrt{n}(\hat{\mu}_4 - 3) \sim N(0,24)$ if x_i is iid and normally distributed (see for example, Stuart et al., 1994). Since $\mathrm{cov}(\hat{\mu}_3, \hat{\mu}_4) = 0$, it follows that under normality

$$t_1 = n\left(\frac{\hat{\mu}_3^2}{6} + \frac{(\hat{\mu}_4 - 3)^2}{24}\right) \sim \chi_2^2.$$

Similarly, since $\tan^{-1}(x) = o(x)$, $\tan^{-1}(x^2) = o(x)$ and $\log(1 + x^2) = o(x)$ as $|x| \to \infty$, their expectations also exist under the assumption that $\mathrm{E}|x|^{4+\delta} < \infty$ for $\delta > 0$. We then have $\sqrt{n}\hat{\mu}_a \sim N(0, 1/36.47595)$, $\sqrt{n}(\hat{\mu}_b - 0.5334532) \sim N(0, 1/32.027545)$ and $\sqrt{n}(\hat{\mu}_c - 0.4741131) \sim N(0, 1/16.898926)$ under normality. In addition, since $\mathrm{cov}(\hat{\mu}_a, \hat{\mu}_b) = 0$ and $cov(\hat{\mu}_a, \hat{\mu}_c) = 0$, it follows that under normality

$$t_2 = n\left(36.47595\hat{\mu}_a^2 + 32.027545\left(\hat{\mu}_b - 0.5334532\right)^2\right) \sim \chi_2^2,$$
$$t_3 = n\left(36.47595\hat{\mu}_a^2 + 16.898926\left(\hat{\mu}_c - 0.4741131\right)^2\right) \sim \chi_2^2. \qquad \blacksquare$$

Proof of Theorem 2.

Proof. Assumption 1 sets forth the classical linear model (except for the normality of ε_i) and ensures that $\hat{\beta}_n \overset{as}{\to} \beta_0$. Given assumption 1 and 2, one can show that $|\hat{\mu}_3 - \mu_3| \overset{as}{\to} 0$ and $|\hat{\mu}_4 - \mu_4| \overset{as}{\to} 0$ using Lemma 1 and Lemma 2 of White and McDonald (1980). Using Corollary A of Serfling (1980, p.19), one can show that since $\hat{t}_1 \overset{as}{\to} t_1$, $\hat{t}_1 \overset{d}{\to} t_1$ given Assumption 3. Since $t_1 \sim \chi_2^2$ by Theorem 1 in Section 3, we have $\hat{t}_1 \sim \chi_2^2$. Similarly, since $\tan^{-1}(x) = o(x)$, $\tan^{-1}(x^2) = o(x)$ and $\log(1 + x^2) = o(x)$ as $|x| \to \infty$, Assumption 1 and 2 ensure that $|\hat{\mu}_a| \overset{as}{\to} 0$, $|\hat{\mu}_b - \mu_b| \overset{as}{\to} 0$ and $|\hat{\mu}_c - \mu_c| \overset{as}{\to} 0$. Using the similar arguments as the proof for $\hat{t}_1$, one can show that $\hat{t}_2 \overset{d}{\to} \chi_2^2$ and $\hat{t}_3 \overset{d}{\to} \chi_2^2$. $\blacksquare$

References

Bai, J. and S. Ng, 2005, Tests for skewness, kurtosis and normality for time series data, Journal of Business and Economic Statistics, 23(1), 49-60.

Bera, A. and C. Jarque, 1981, Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: monte carlo evidence, Economics Letters, 7, 313-318.

Bontemps, C. and N. Meddahi, 2005, Testing normality: a gmm approach, Journal of Econometrics, 124(1), 149-186.

Christensen L.R. and W.H. Greene, 1976, Economies of scale in u.s. electric power generation, Journal of Political Economy, 84, 655-76.

Cobb, L., P. Koppstein and N. Chen, 1983, Estimation and moment recursion relations for multimodal distributions of the exponential family, Journal of American Statistical Association, 8(381), 124-30.

Dalén, J., 1987, Bounds on standardized sample moments, Statistics and Probability Letters, 5, 329-31.

Das Gupta, S. and M.D. Perlman, 1974, Power of the noncentral f test: effect of additional variates in hotelling's t^2 test, Journal of the American Statistical Association, 69, 174-180.

Engle, R., 1984, Wald, likelihood ratio and lagrange multiplier tests in econometrics, in: Grilliches and Intrilligator, (Eds.), Handbook of econometrics, Vol.3, Elsevier, North Holland.

Fisher, R. A., 1922, On the mathematical foundations of theoretical statistics, Philosophical Transactions of the Royal Society of London, Series A, 222, 309-68.

Gallant, A.R., 1981, On the bias in flexible functional forms and an essentially unbiased form, Journal of Econometrics, 15, 211-241.

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

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

Lye, J.N. and V.L., Martin, 1993, Robust estimation, nonnormalities, and generalized exponential distribution, Journal of American Statistical Association, 88(421), 261-267.

Ornermite, D. and H. White, 1999, An efficient algorithm to compute maximum entropy densities, Econometric Reviews, 18(2), 141-67.

Pearson, K., 1900, On a criterion that a given system of deviations from the probable in the case of correlated systems of variables is such that it can be reasonably supposed to have arisen in random sampling, Philosophical Magazine, 50(5), 157-175.

Serfling, R.J., 1980, Approximation theorems of mathematical statistics, John Wiley & Sons, Inc.

Shannon, C. E., 1949, The mathematical theory of communication, University of Illinois Press, Urbana.

Stohs, S., 2003, A bayesian updating approach to crop insurance ratemaking, Ph.D. thesis, University of California at Berkeley.

Stuart, A., K. Ord and S. Arnold, 1994, Kendall's advanced theory of statistics, Vol. 2A, Oxford University Press, New York.

Thode, H., 2002, Testing for normality, Marcel Dekker, New York.

Vasicek, O., 1976, A test for normality based on sample entropy, Journal of the Royal Statistical Society, Series B, 38, 54-59.

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica, 50, 1-26.

White, H. and G. M. McDonald, 1980, Some large-sample tests for nonnormality in the linear regression model, Journal of American Statistical Association, 75, 16-28.

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

Wu, X. and T. Stengos, 2005, Partially adaptive estimation via maximum entropy densities, Econometrics Journal, 8, 352-366.

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

Table 1: Size and Power of Normality Test (τ: skewness; κ: kurtosis)

			$n=20$				$n=50$				$n=100$			
	τ	κ	t_1	t_2	t_3	KS	t_1	t_2	t_3	KS	t_1	t_2	t_3	KS
S1	0	3.0	0.03	0.04	0.04	0.06	0.04	0.04	0.05	0.05	0.04	0.05	0.05	0.05
S2	0	9.0	0.16	0.18	0.17	0.14	0.40	0.37	0.33	0.21	0.63	0.59	0.51	0.34
S3	0	2.5	0.01	0.03	0.05	0.05	0.01	0.07	0.10	0.07	0.01	0.16	0.19	0.10
S4	0	3.0	0.02	0.03	0.04	0.06	0.03	0.04	0.05	0.05	0.04	0.04	0.05	0.05
S5	0	6.0	0.16	0.18	0.17	0.14	0.39	0.37	0.33	0.20	0.63	0.61	0.54	0.35
S6	0	11.6	0.24	0.27	0.26	0.21	0.55	0.56	0.52	0.36	0.82	0.82	0.78	0.60
S7	0	126.0	0.33	0.38	0.37	0.29	0.71	0.71	0.68	0.54	0.92	0.93	0.91	0.80
A1	6.18	113.9	0.72	0.87	0.87	0.81	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
A2	2.0	9.0	0.49	0.69	0.70	0.60	0.95	0.99	0.99	0.96	1.00	1.00	1.00	1.00
A3	2.0	9.0	0.48	0.68	0.69	0.60	0.96	0.99	0.99	0.96	1.00	1.00	1.00	1.00
A4	5.0	2.2	0.02	0.19	0.28	0.20	0.08	0.74	0.79	0.52	0.80	0.99	0.99	0.90
A5	0.5	7.5	0.30	0.37	0.36	0.32	0.73	0.80	0.79	0.63	0.97	0.98	0.98	0.91
A6	2.0	21.2	0.31	0.33	0.32	0.29	0.68	0.70	0.67	0.53	0.91	0.92	0.91	0.81
A7	3.16	23.8	0.60	0.78	0.79	0.71	0.99	1.00	1.00	0.98	1.00	1.00	1.00	1.00
A8	3.8	40.7	0.64	0.82	0.82	0.75	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00

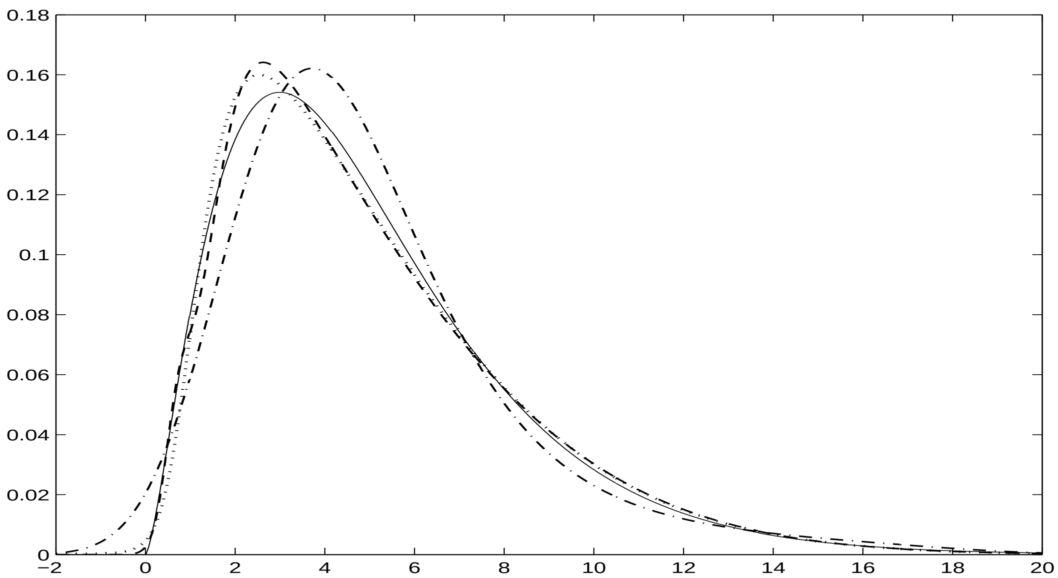

Figure 1: Approximation of χ^2_5 distribution: true distribution (solid), f_1 (dash-dotted), f_2 (dotted), f_3 (dashed)