Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Nonparametric Density Estimation and Testing

by

Patrick Marsh

Granger Centre Discussion Paper No. 16/03

The University of Nottingham

Nonparametric Density Estimation and Testing.

Patrick Marsh*

School of Economics

University of Nottingham

NG7 2RD

email: patrick.marsh@nottingham.ac.uk

31st October 2016

Abstract

A nonparametric likelihood ratio test for the exponential series density estimator is employed as a goodness-of–fit test in the presence of nuisance parameters. These tests are designed to address the perceived weaknesses of those based on the empirical distribution function, such as the Kolmogorov-Smirnov, Cramer-von Mises and Anderson Darling tests. These tests are often criticized for not being asymptotically pivotal, having low power and offering no direction if the null hypothesis is rejected. Instead the tests of this paper are proven to be asymptotically pivotal and numerical experiments illustrate this. Further experiments suggest the tests are generally more powerful in a variety of testing problems whether bootstrap critical values are used or not. Finally, in the event of rejection, the proposed procedures involve density estimators which can be used directly and accurately to estimate quantiles.

Keywords: Goodness-of-fit, series density estimator, likelihood ratio, nuisance parameters, parametric bootstrap.

*Thanks are due to Francesco Bravo, Giovanni Forchini, Les Godfrey, Peter Phillips and participants at seminars given at the Universities of York, Surrey and the Tinbergen Institute, Amsterdam.

1 Introduction

As well as generating an important literature in its own right the goodness-of-fit problem is fundamental to many diagnostic procedures in Statistics and Econometrics. Primarily this involves checking the adequacy or predictive ability of a fitted model, see Corradi and Swanson (2006) for a comprehensive review. The most commonly used tests are extensions of those based upon the empirical distribution function (edf), such as the Kolmogorov-Smirnov (KS), Cramér-von Mises (CM) and Anderson-Darling (AD) procedures (see Conover (1999)).

Such tests, however, suffer from three shortcomings. First, when nuisance parameters must be estimated, tests based on the edf are not asymptotically pivotal, even in the simplistic testing circumstances considered by, for example, Stephens (1974, 1976) and Babu and Rao (2004). In the diagnostic applications in Corradi and Swanson (2006) the analysis swiftly becomes impractical to implement. Second, in simple testing situations, they lack power when compared to appropriate parametric tests. Thirdly, and perhaps most importantly, if we do reject the null of correct specification, edf based procedures are not particularly helpful in suggesting how to proceed.

This paper instead extends the nonparametric likelihood ratio goodness-of-fit tests developed in Claeskens and Hjort (2004) and Marsh (2007, 2010) to the case where nuisance parameters must be estimated. Similar to those papers we apply the test of Portnoy (1988) via application of the exponential series density estimator of Barron and Sheu (1991).

To address the three shortcomings of edf based procedures, first we prove that the tests derived here have standard asymptotic distributions and are therefore asymptotically pivotal. They are also consistent against fixed alternatives.

Second, it is already known that in standard goodness-of-fit applications tests based on nonparametric likelihood are significantly more powerful than those based on the edf, see Marsh (2007, 2010). This paper offers further evidence to this, although the advantage is lessened when parameters are estimated. The small advantage maintains for testing problems both where the distribution itself is different under the alternative and also when the specification of the moments is different, specifically for breaks in the mean, autoregressions and conditional heteroskedasticity.

Third, the tests of this paper are based on a consistent nonparametric density estimator, with consistency established through convergence of the Kullback-Leibler divergence. As a consequence we are able to consistently estimate the quantile function of the random variable generating the data, whether the null hypothesis of correct specification is true or not.

Given that tests based on the edf are generally not asymptotically pivotal recent attention has been given to providing approximate critical values via bootstrap based procedures. Babu and Rao (2004) prove that the parametric bootstrap provides asymptotically valid critical values under the null hypothesis. Genest and Rémillard (2008) generalize to semi-parametric models, Villaseñor and González-Estrada (2009) consider the special case of the generalized Pareto distribution and Kojadinovic and Yan (2012) provide a more efficient bootstrap.

Since the tests here are asymptotically pivotal then it is trivial to establish asymptotic validity of the bootstrap under the null and consistently under the alternative. Moreover, numerical comparisons reaffirm the theoretical advantages the proposed test has over edf based tests. Under various null hypotheses all tests are shown to be well behaved, as should be expected. The proposed test is not dominated by any other and offers a stable and conservative test for goodness-of-fit.

The plan for the rest of the paper is as follows. The next section first briefly details the exponential series density estimator and its application to the goodness-of-fit problem and then extends to the more problematic case involving nuisance parameters. The asymptotic properties of the test are derived in Section 3. Specifically, under the null hypothesis of correct specification, we establish asymptotic standard normality of the test and its consistency under fixed alternatives. Section 4 implements the procedures in a number of experiments and provides clear comparisons with the edf based tests in cases where the limit theory of the latter is established. Here we also examine the accuracy of the density estimator itself under both null and alternative. Section 5 concludes while all proofs and the tables for the numerical comparisons are placed in the Appendix.

2 Preliminaries

2.1 No Estimated Parameters

The tests developed in this paper derive from those of Marsh (2007). Suppose that our sample $\underline{y}= \{Y_i\}_{i=1}^n$ consists of independent copies of a random variable Y having distribution, $G(y) = \Pr[Y \leq y]$, and we wish to test the goodness-of-fit hypothesis,

$$H_0 : G(y) = F(y;\beta), \tag{1}$$

where both $F(.,.)$ and β, for now, are fully specified.

To proceed let $X_i = F(Y_i;\beta)$ so that, whether H_0 is true or not, the X_i are *iid* copies of the variable X, having distribution and density,

$$U(x) = \Pr[X < x] \quad \& \quad u(x) = \frac{dU(x)}{dx} \quad for \quad x \in (0,1).$$

We first approximate $u(x)$ using the exponential family,

$$p_x(\theta) = \exp\left\{\sum_{k=1}^{m}\theta_k\phi_k(x) - \psi_m(\theta)\right\}, \quad \psi_m(\theta) = \ln\int_0^1 \exp\left\{\sum_{k=1}^{m}\theta_k\phi_k(x)\right\}dx, \tag{2}$$

where the $\phi_k(x)$ are linearly independent functions spanning $\mathbb{R}^m$.

Application requires a choice of basis. Marsh (2007) chose a polynomial basis, i.e. $\phi_k(x) = x^k$. Here we will investigate this but also, for m even, the trigonometric basis, $\phi_k(x) = \{\cos[2k\pi x], \sin[2k\pi x]\}_{k=1}^{m/2}$.

To continue, suppose that $\log[u(x)]$ has $r-1$ absolutely continuous derivatives and that its r^{th} derivative, $d^r lp(x)/dx^r$ is square integrable, i.e. so that $\log[u(x)] \in W_2^r$, the Sobolev space of functions on $(0,1)$. According to Crain (1974) and Barron and Sheu (1991) there then exists a unique $\theta_{(m)} = (\theta_1, .., \theta_m)'$ satisfying

$$\int_0^1 \phi_k(x)p_x\left(\theta_{(m)}\right)dx = \int_0^1 \phi_k(x)u(x)\,dx \quad \text{for } k = 1, 2, ..., m, \tag{3}$$

and, as $m \to \infty$, $p_x\left(\theta_{(m)}\right)$ and $u(x)$ converge in relative entropy, with

$$E_U\left[\ln\left(\frac{u(x)}{p_x\left(\theta_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{u(x)}{p_x\left(\theta_{(m)}\right)}\right)u(x)\,dx = O\left(m^{-2r}\right).$$

Since $\theta_{(m)}$ is unique any hypothesis test on the distribution of X can be tested via a simple hypothesis on $\theta_{(m)}$,

$$H_0 : \Pr[X < x] = U_0(x) \leftrightarrow H_0 : \theta_{(m)} = \theta^0_{(m)}, \tag{4}$$

where $\theta^0_{(m)}$ solves (3) with $u(x) = u_0(x) = dU_0(x)/dx$.

Notice that in the goodness-of-fit problem $u_0(x)$ will be fully specified. Claeskens and Hjort (2004) chose $u_0(x) = 1$, i.e. they test uniformity, which implies that $\theta^0_{(m)} = 0_{(m)}$, the m-vector of zeros. Here, for simplicity, we will also follow this approach, rather than that of Marsh (2007) which effectively applied the test to the square of X_i.

Following Portnoy (1988) H_0 can be tested via a likelihood ratio test in the exponential family (2),

$$\lambda_m = 2\sum_{i=1}^{n} \ln\left[\frac{p_{X_i}\left(\bar{\theta}_{(m)}\right)}{p_{X_i}\left(0_{(m)}\right)}\right] = 2\sum_{i=1}^{n} \ln p_{X_i}\left(\bar{\theta}_{(m)}\right),$$

where $\bar{\theta}_{(m)}$ is the unique maximum likelihood estimator for $\theta_{(m)}$ satisfying

$$\int_0^1 \phi_k(x)\, p_x\left(\bar{\theta}_{(m)}\right) dx = \frac{\sum_{i=1}^{n} \phi_k(X_i)}{n} \quad \text{for } k = 1, 2, ..., m. \tag{5}$$

From Barron and Sheu (1991, Theorem 1), if $m^3/n \to 0$ then, with either choice of basis, $p_x\left(\bar{\theta}_{(m)}\right)$ converges in relative entropy to $u(x)$,

$$E_U\left[\ln\left(\frac{u(x)}{p_x\left(\bar{\theta}_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{u(x)}{p_x\left(\bar{\theta}_{(m)}\right)}\right) u(x)\, dx = O_p\left(\frac{m}{n} + m^{-2r}\right).$$

Applying the results of Portnoy (1988) for tests on infinite dimensional parameters, Marsh (2007, Theorem 1) proves that as $n \to \infty$ with $m^3/n \to 0$, then

$$\Lambda_m = \frac{\lambda_m - m}{\sqrt{2m}} \to_d N(0, 1). \tag{6}$$

In addition it is shown that Λ_m diverges under any fixed alternative (i.e. the test is consistent) and it has power against local alternatives parametrized by $\theta^1_{(m)} - 0_{(m)} = c\sqrt{\frac{\sqrt{m}}{n}}$ with $c'c = 1$ and $\theta^1_{(m)}$ satisfies (3) but with $u(x) = u_1(x)$, the density under the alternative.

2.2 Estimated Parameters

Now assume β in (1) is unknown and must be estimated as a preliminary step prior to application of the likelihood ratio test described above. Let $\hat{\beta}_n$ denote the (quasi) maximum likelihood estimator of β obtained from the sample $\{Y_1, .., Y_n\}$ using the specified likelihood $L = \prod_{i=1}^{n} f(Y_i; \beta)$. Typically the alternative will be the (unspecified) negation of H_0, as in

$$H_1 : G(y) \neq F(y; \beta).$$

We require the following assumptions on both $F(y;\beta)$, $G(y)$ and the respective densities $f(y;\beta)$ and $g(y)$, to ensure the existence of $\hat{\beta}_n$ and under which the asymptotic distribution of the proposed test will be derived.

Assumption 1 :

(i) The density $f(y;\beta)$ is measurable in y for every β in a compact subset of $p-$dimensional Euclidean space, and are continuous in β for every y.

(ii) $G(y)$ is an absolutely continuous distribution function, $E[\log[g(y)]$ *exists and* $|\log f(y,\beta)| < v(y)$ *for all* β *where* $v(.)$ *is integrable with respect to* $G(.)$.

(iii) Let

$$I(\beta) = E\left[\ln\left[\frac{g(y)}{f(y,\beta)}\right]\right] = \int_y \ln\left[\frac{g(y)}{f(y,\beta)}\right] g(y),$$

then $I(\beta)$ has a unique minimum at β_*.

(iv) $H(\beta) = dF(Y_i,\beta)/d\beta$ is finite for all β in a closed ball of radius $n^{-1/2}$ around β_*.

(v) Both $\log[g(y)]$ and $\log[f(y;\beta)]$ have $r \geq 2$ derivatives in y which are absolutely continuous and square integrable.

Immediate from White (1982, Theorems 2.1, 2.2 and 3.2) is that under Assumption 1(i-iii) $\hat{\beta}_n$ exists and

$$\hat{\beta}_n = \beta_* + O(n^{-1/2}).$$

That is $\hat{\beta}_n$ is a $\sqrt{n}$ consistent Quasi maximum likelihood estimator for the pseudo-true value β_*. Note that under H_0 we have $\beta_* = \beta$.

To derive the test, first denote $\hat{X}_i = F\left(Y_i, \hat{\beta}_n\right)$with the mean value expansion,

$$\hat{X}_i = F(Y_i, \beta_*) + \left(\hat{\beta}_n - \beta_*\right)' \frac{dh}{dF} H\left(\beta^+\right),$$

where β^+ lies on a line segment joining $\hat{\beta}_n$ and β_*. As a consequence we can write

$$\hat{X}_i = X_i + e_i, \tag{7}$$

where X_i is exactly as above, i.e. they are IID copies of X with distribution $U(x)$. In addition, by construction and as a consequence of Assumption 1 (iv),

$$e_i \in (-1, 1) \quad \& \quad e_i = O_p\left(n^{-1/2}\right), \tag{8}$$

that is e_i is both bounded and degenerate.

The modification required to deal with the fact that β must be estimated is as follows. We are still testing on the distribution $U(x)$, however we do not observe outcomes on X_i but instead those on $\hat{X}_i$. Notice that under Assumption 1(v) the density $u(x) = dU(x)/dx$ satisfies $\log[u(x)] \in W_2^r$. The maximum likelihood estimator for the parameter in the exponential family (2), say $\hat{\theta}_{(m)}$, based on the likelihood $\hat{L}\left(\theta_{(m)}\right) = \prod_{i=1}^n p_{\hat{X}_i}\left(\theta_{(m)}\right)$ satisfies

$$\int_0^1 \phi_k(x)\, p_x\left(\hat{\theta}_{(m)}\right) dx = \frac{\sum_{i=1}^n \phi_k(X_i)}{n} \quad \text{for } k = 1, 2, ..., m. \tag{9}$$

In the presence of nuisance parameters we must therefore test the null hypothesis $H_0 : X \sim U_0(x)$ (i.e. $H_0 : \theta_{(m)} = 0_{(m)}$) using the likelihood ratio

$$\hat{\lambda}_m = 2\sum_{i=1}^n \ln\left[\frac{p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)}{p_{\hat{X}_i}\left(0_{(m)}\right)}\right] = 2\sum_{i=1}^n \ln p_{\hat{X}_i}\left(\bar{\theta}_{(m)}\right).$$

In the following section we will detail the asymptotic properties of $\hat{\lambda}_m$. It is, upon standardization, asymptotically standard normal - and therefore crucially asymptotically pivotal. In addition it will be shown that tests based on $\hat{\lambda}_m$ remain consistent against fixed alternatives under Assumption 1.

3 Asymptotic Properties

3.1 Density Estimation with estimated parameters

Above, the non-parametric likelihood ratio test for goodness-of-fit with estimated parameters was contrasted with that of the simpler case. Key is that, in (7), we do not observe directly a sample from the random variable upon which the hypothesis is being tested, unlike in the standard goodness-of-fit case. If we did know the true value of β we could observe X_i directly and obtain the maximum likelihood estimator

$\bar{\theta}_{(m)}$via (5). Instead, in the nuisance parameter case, we only observe $\hat{X}_i$ and obtain $\hat{\theta}_{(m)}$ via (9) and apply Portnoy's (1988) test using that.

For a given m, this test is just an application of a likelihood-ratio test in a linear exponential family. For a given choice of basis

$$\phi_{(m)}(x) = (\phi_1(x), \phi_2(x), ..., \phi_m(x))',$$

we construct the m dimensional sufficient statistics,

$$\bar{x}_{(m)} = \left(\frac{\sum_{i=1}^{n} \phi_k(X_i)}{n}\right)_{k=1}^{m},$$

in the standard goodness-of-fit case and

$$\hat{x}_{(m)} = \left(\frac{\sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right)}{n}\right)_{k=1}^{m},$$

in the nuisance parameter case. The respective maximum likelihood estimators, $\bar{\theta}_{(m)}$ and $\hat{\theta}_{(m)}$ then satisfy

$$\int_0^1 \phi_{(m)}(x) p_x\left(\bar{\theta}_{(m)}\right) dx = \bar{x}_{(m)}^{\phi} \quad \text{and} \quad \int_0^1 \phi_{(m)}(x) p_x\left(\hat{\theta}_{(m)}\right) dx = \hat{x}_{(m)}^{\phi}. \tag{10}$$

These will produce a consistent nonparametric density estimator as $m \to \infty$. This follows in the standard case immediately from the analysis of Barron and Sheu (1991) and is proved below in the nuisance parameter case.

Standard properties of the linear exponential family still apply, specifically the duality between the (sufficient statistic) sample space, say Ω_m, and the parameter space, say Θ_m. As in Barndorff-Nielsen (1978), consider arbitrary points in both Ω_m and Ω_m, $\omega_{(m)} = \{\omega_1, .., \omega_m\}' \in \Omega_m$ and $\theta_{(m)} = (\theta_1, .., \theta_m) \in \Theta_m$ then the system of m equations

$$\int_0^1 \phi_k(x) p_x\left(\theta_{(m)}\right) dx = \omega_k, \quad k = 1, .., m, \tag{11}$$

has a unique solution. That is solving (11), and therefore either equation in (10), generates a one-to-one mapping between Ω_m and Θ_m.

Here we will be interested in three points in each space,

$$\theta_{(m)} \in \Theta_m \leftrightarrow \mu_{(m)} = (E_U[\phi_1(x)], E_U[\phi_2(x)], ..., E_U[\phi_m(x)])' \in \Omega_m$$

$$\begin{aligned}
\bar{\theta}_{(m)} &\in \Theta_m \leftrightarrow \bar{x}_{(m)} = \left(\frac{\sum_{i=1}^n \phi_1(X_i)}{n}, \frac{\sum_{i=1}^n \phi_2(X_i)}{n}, ..., \frac{\sum_{i=1}^n \phi_m(X_i)}{n} \right)' \in \Omega_m \\
\hat{\theta}_{(m)} &\in \Theta_m \leftrightarrow \hat{x}_{(m)} = \left(\frac{\sum_{i=1}^n \phi_1\left(\hat{X}_i\right)}{n}, \frac{\sum_{i=1}^n \phi_2\left(\hat{X}_i\right)}{n}, ..., \frac{\sum_{i=1}^n \phi_m\left(\hat{X}_i\right)}{n} \right)' \in \Omega_m.
\end{aligned} \tag{12}$$

Note that although these points also depend on the choice of basis ϕ here we will suppress the dependence for notational brevity. In summary, $\theta_{(m)}$ in Θ_m maps to the mean of the (unobserved) sufficient statistic $\bar{x}_{(m)}$, $\mu_{(m)} = E\left[\bar{x}_{(m)}\right]$. The mle for $\theta_{(m)}$ based on $\bar{x}_{(m)}$ is $\bar{\theta}_{(m)}$, while for the observed sufficient statistic $\hat{x}_{(m)}$, it is $\hat{\theta}_{(m)}$. By exploiting this duality, we are first able to show that the estimated density $p_x\left(\hat{\theta}_{(m)}\right)$ converges in relative entropy at exactly the same rate as $p_x\left(\bar{\theta}_{(m)}\right)$ does in the following theorem, which is proved in the Appendix.

Theorem 1 *Let $\hat{\theta}_{(m)}$ denote the estimated exponential parameter determined by (9) then under Assumption 1 and for $m, n \to \infty$ with $m^3/n \to 0$,*

$$E_U\left[\ln\left(\frac{u(x)}{p_x\left(\hat{\theta}_{(m)}\right)}\right)\right] = \int_0^1 \ln\left(\frac{u(x)}{p_x\left(\hat{\theta}_{(m)}\right)}\right) u(x)\, dx = O_p\left(\frac{m}{n} + m^{-2r}\right). \quad \blacksquare$$

According to Theorem 1, in terms of the density estimator, at least, the effect of observing $\left\{\hat{X}_1, .., \hat{X}_n\right\}$ rather than $\{X_1, .., X_n\}$ is asymptotically negligible under Assumption 1 and for either choice of basis. Moreover, if the goal were only nonparametric estimation of the density, then the optimal choice of the dimension m is the same as when no parameters are estimated, i.e. $m_{opt} \propto n^{\frac{1}{1+2r}}$ (with a mini-max rate of $m_n^* = O\left(n^{-1/5}\right)$, since $r \geq 2$ by assumption). The optimal rate the rate of convergence of the estimator remains of order $O_p\left(n^{-\frac{2r}{1+2r}}\right)$. It should not be surprising that the rate of convergence is unaffected when parameters are replaced by $\sqrt{n}$ consistent estimators.

3.2 Properties of the Likelihood Ratio Test

Full implementation of the non-parametric likelihood ratio test for goodness-of-fit with estimated parameters proceeds as follows. Given the sample $\{Y_1, .., Y_n\}$ consisting of IID copies of Y having distribution $G(y)$ we wish to test $H_0 : G(y) = F(y, \beta)$, as above.

Letting $\hat{X}_i = F\left(Y_i, \hat{\beta}_n\right)$ and $X_i = F\left(Y_i, \beta\right)$ where X_i has uniform distribution and density $u(x) = 1$, then testing H_0 is equivalent to testing

$$H_0 : \theta_{(m)} = 0_{(m)},$$

in the exponential family (2). The likelihood ratio test of Portnoy (1988) applied via the density estimator of Crain (1974) and Barron and Sheu (1991) obtained from the sample $\left\{\hat{X}_1, .., \hat{X}_n\right\}$ is

$$\hat{\lambda}_m = 2\sum_{i=1}^{n} \log\left[\frac{p_{\hat{X}_i}\left(\hat{\theta}_{(m)}\right)}{p_{\hat{X}_i}\left(0_{(m)}\right)}\right] = 2n\left[\hat{\theta}'_{(m)}\hat{x}_{(m)} - \psi_m\left(\hat{\theta}_{(m)}\right)\right],$$

where $\hat{\theta}_{(m)}$ solves (9). The null hypothesis is rejected for large values of $\hat{\lambda}_m$.

Under any fixed alternative $H_1 : G(y) \neq F(y;\beta)$ the distribution of $X_i = F_i(Y_i;\beta_*)$ will not be uniform. For every fixed alternative distribution for Y there is a unique alternative distribution for X on $(0,1)$ and associated with that distribution will be another consistent density estimator given by say, $p_x(\theta^1_{(m)})$. In practice, of course, $\theta^1_{(m)}$ will be neither specified nor known. The following Theorem, again proved in the Appendix, gives the asymptotic distribution of the likelihood ratio test statistic both under the null hypothesis (4) and also demonstrates consistency against any such fixed alternative.

Theorem 2 *Suppose that we construct* $\left\{\hat{X}_i\right\}_{i=1}^{n}$ *as described above, that the conditions required in Assumption 1 are met and that* $m, n \to \infty$ *with* $m^3/n \to 0$*, then:*
(i) Under the null hypothesis, $H_0 : G(y) = F(y;\beta)$,

$$\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}} \to_d N(0,1).$$

(ii) Under any fixed alternative $H_1 : G(y) \neq F(y;\beta)$ *and for any finite* κ,

$$\Pr\left[\hat{\Lambda}_m \geq \kappa\right] \to 1. \quad \blacksquare$$

4 Implementation and Numerical Properties

The purpose of this section is to illustrate the properties of the nonparametric likelihood tests and estimators described above. First, in the context of Exponentiality and Normality testing, we examine the size properties of tests based on both trigonometric and polynomial bases using asymptotic critical values and as both m and n increase. We then compare the powers of two particular variants against the standard tests in this field, the Kolmogorov-Smirnov (KS), Cramer-von Mises (CM) and Anderson Darling (AD) tests. Results are provided for cases where we do and also do not estimate unknown parameters.

Although the tests of this paper are pivotal, outside of testing for Exponentiality and Normality, the competitor tests are feasible only when bootstrapped. We therefore also compare the finite sample performance of bootstrap critical values for the tests of this paper, repeating experiments of Kojadinovic and Yan (2012).

The last set of experiments concern what we may do if the test rejects the null hypothesis. The tests of this paper are based on a consistent (see Theorem 1) density estimator. Therefore we simulate mean-square errors for the estimators of the quantiles of the distribution, when we mis-specify that distribution. All experiments were performed using Mathematica 8 and are based on 2500 Monte Carlo replications.

4.1 Numerical properties when testing for Exponentiality and Normality

Theorem 2 proves that, for either choice of basis, the likelihood ratio tests $\hat{\Lambda}_m$ are asymptotically pivotal, specifically standard normal, and consistent against fixed alternatives. Competitor tests, such as KS, CM and AD (these tests are mathematically detailed in Stephens (1976) or Conover (1999)) are not pivotal, although asymptotic critical values are readily available for all cases of testing for Exponentiality and Normality.

First we will demonstrate that indeed asymptotic critical values for nonparametric likelihood tests do have close to nominal size for large values of n and m. We are interested in testing the null hypotheses

$$H_0^E : Y \sim Exp\,(1) \quad \& \quad H_0^N : Y \sim N\,(0,1)\,,$$

with nominal significance levels 10%, 5% and 1% and based on sample sizes $n = 25, 50, 100$ and 200. Letting $\bar{y}$ and $\hat{\sigma}^2$ be the estimated mean and variance (i.e. $\hat{\beta}_n = \bar{y}$ for H_0^E and $\hat{\beta}_n = \left(\bar{y}, \hat{\sigma}^2\right)'$ for H_0^N) then the tests are constructed from the mapping to $(0,1)$;

$$\hat{X}_i = 1 - e^{-Y_i/\bar{y}_n}, \tag{13}$$

to test H_0^E, and

$$\hat{X}_i = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{Y_i - \bar{y}_n}{\hat{\sigma}_n}\right)\right], \tag{14}$$

to test H_0^N.

Table 1a in the appendix provides rejection frequencies for tests constructed using the trigonometric basis for $m = 2, 4, 6, 8, 10$ and 12. The left hand panel of numbers correspond to testing H_0^E and the right to H_0^N. Similarly Table 1b provides rejection frequencies for tests based on the polynomial basis for $m = 3, 5, 7, 9, 13$ and 17.

The finite sample performance of the tests clearly improves as both n and m increase. We have deliberately chosen three different significance levels because although there are cases with particular values of m and n where finite sample size is acceptable at a particular chosen significance level its only when all three are close to nominal when the procedure is actually justified. This happens for significantly smaller values of m for the trigonometric basis than the polynomial.

Table 2 compares the 5% size corrected powers of two variants of the tests with the three direct competitors for a single sample size of $n = 100$. For $m = 3$ we denote the test constructed from the polynomial basis by $\hat{\Lambda}_3^P$ and for $m = 10$ denote the test constructed using the trigonometric basis by $\hat{\Lambda}_{10}^T$. Tables 2a through 2b present rejection frequencies for these tests and the KS, CM and AD tests for testing H_0^N under alternatives that the data is instead drawn from,

$$H_1^a : Y \sim t_{(v)}, \quad H_1^b : Y \sim \chi^2_{(v)} - v.$$

Tables 2b, 2c and 2e, consider alternatives where the moments of the data are not correctly specified, i.e.

$$\begin{aligned} H_1^c &: Y_i|Y_{i-1} \sim N\left(vY_{i-1}, 1\right), \\ H_1^d &: Y_i|Y_{i-1} \sim N\left(0, 1 + vY_{i-1}^2\right), \\ H_1^e &: Y_i \sim N\left(v \times 1\left(i > \lfloor n/2 \rfloor\right), 1\right), \end{aligned}$$

where $1(.)$ denotes the indicator function. These latter three alternatives represent simplistic variants of common types of mis-specification in econometric or financial data, i.e. mis-specification of a conditional mean, variance or the possibility of a break in the mean (here half way through the sample). Note that these models trivially satisfy Assumption 1, but X_i as defined in (7) will not be IID on $(0,1)$. Finally, table 2f considers instead testing H_0^E against the alternative

$$H_1^f : Y \sim \Gamma(1, v).$$

In each table the left hand panel corresponds to the case where we construct the test imposing the parameter values specified in the null rather than estimating them (i.e. using the test in (6)). The right hand panel has the rejection frequencies for tests based on estimated values, i.e. using (14) and (13), respectively.

The outcomes in Table 2 imply the following broad conclusions. The nonparametric likelihood test based $\hat{\Lambda}_3^P$ is the most powerful almost uniformly, across all alternatives and whether parameters are estimated or not. The observed lack of power of the most commonly used test, KS, is particularly evident, it is consistently the poorest performing test. The other edf based tests and $\hat{\Lambda}_{10}^T$ are broadly comparable in terms of their rejection frequencies, although AD is perhaps on average slightly more powerful and CM less powerful.

From Tables 1 and 2 we can thus conclude that $\hat{\Lambda}_{10}^T$ has size close to nominal and power comparable with the best edf based tests. Its advantage, however, is that it is based on an asymptotically pivotal procedure. It is the same set of critical values being used in Table 1a, equivalent experiments for edf based tests would require different sets of critical values.

Despite this, that test is outperformed both in terms of power and computationally by $\hat{\Lambda}_3^P$. But, of course, this test does not have adequate finite sample size. This inadequacy will be addressed in the next sub-section.

4.2 Bootstrap Critical Values

Here we will compare the performance of bootstrap critical values for $\hat{\Lambda}_3^P$ with those of CM and AD by repeating many of the experiments of Kojadinovic and Yan (2012). Asymptotic justification for the bootstrap is automatic given that $\hat{\Lambda}_m \to_d N(0,1)$ giving the following corollary to Theorem 2.

Corollary 1 Under Assumption 1 and if $n, m \to \infty$ with $m^3/n \to 0$, then

$$\begin{aligned} \text{i)} \quad \Pr\left[\hat{I}_B^{\Lambda} = 1 \,|\, H_0\right] &\to \alpha, \\ \text{ii)} \quad \Pr\left[\hat{I}_B^{\Lambda} = 1 \,|\, H_1\right] &\to 1. \quad \blacksquare \end{aligned}$$

For just the $\hat{\Lambda}_3$ test, the bootstrap procedure is as follows: On obtaining the mle $\hat{\beta}_n$ and calculating $\hat{\Lambda}_3$, as described above;

1. Generate bootstrap samples $Y_i^b \sim IID\, F\left(y; \hat{\beta}_n\right)$ for $i = 1, .., n$.
2. Estimate, via mle, $\hat{\beta}_n^b$ and construct $\hat{X}_i^b = F\left(Y_i^b; \hat{\beta}_n^b\right)$ for $i = 1, .., n$.
3. Repeat 1 and 2 B times, obtaining bootsrap versions of the test $\hat{\Lambda}_3^b$.
4. Order the $\hat{\Lambda}_3^b$ so the bootstrap critical value at size α is $\kappa_B = \hat{\Lambda}_3^{\lfloor(1-\alpha)B/100\rfloor}$.
5. Denote the indicator function $\hat{I}_B^{\Lambda} = \left\{ \begin{array}{c} 1 \text{ if } \hat{\Lambda}_3 > \kappa_\alpha^B \\ 0 \text{ if } \hat{\Lambda}_3 \leq \kappa_\alpha^B \end{array} \right\}$.

We then reject H_0 if $\hat{I}_B^{\Lambda} = 1$.

In this sub-section all experiments described in this sub-section are performed on the basis of $B = 200$ bootstrap replications. All nuisance parameters were estimated via maximum likelihood using Mathematica 8's own numerical optimization algorithm.

The first set of experiments mimic those presented in Kojadinovic and Yan (2012, Table 1). Specifically we define the following Normal, Logistic, Gamma and Weibull Distributions;

$$\begin{aligned} N^* &\sim N(10, 1), \quad L^* \sim L(10, 0.572), \\ \Gamma^* &\sim \Gamma(98.671, 1/9.866) \quad \& \quad W^* \sim W(10.618, 10.452). \end{aligned} \tag{15}$$

The specific parameter values for L^*, Γ^* and W^* are chosen to minimize relative entropy for each family to the distribution of N^*. Sample sizes of $n = 25, 50, 100, 200$ are used in the experiments described below.

Table 3a contains the finite sample size of each test. It is clear that, under H_0, the parametric bootstrap provides highly accurate critical values for all of the tests. On size alone there is nothing to choose between them. It is however, worth reporting, the computational time of each bootstrap critical value. For the $\hat{\Lambda}_3$ test critical values were obtained after 2.0 and 3.2 seconds for sample sizes $n = 100$ and 200, respectively.

The times for the other tests were similar to each other, taking around 0.9 and 2.9 seconds, respectively.

Table 3a and 3b contain the finite sample rejection frequencies under various alternative hypotheses, covering all pairwise permutations of the distributions in (15). As with the finite sample sizes it is not possible to pick a clear winner, moreover where they overlap the results are in line with those of Kojadinovic and Yan (2012). There is, of course, no uniformly most powerful test of goodness-of-fit so it is not surprising that the power of $\hat{\Lambda}_3$ is not always the largest. However its performance over this range of nulls and alternatives is far less volatile and in no circumstance is the test dominated by any of the other two.

4.3 What if the null hypothesis is rejected?

The final criticism of edf based tests of goodness-of-fit, and diagnostics in general, is that rejection of the null hypothesis is not indicative of how the specification could or should be changed. The tests of this paper, however, are based on the consistent nonparametric density estimator of Barron and Sheu (1991). This consistency can readily be extended to the current context of the presence of nuisance parameters, as in the following corollary.

Corollary 2 Let $\hat{T}_{n,m} \in (0,1)$ be a random variable having density function $p_t\left(\hat{\theta}_{(m)}\right)$ and let $X_i = F\left(Y_i;\beta_*\right) \sim iid.X$, as defined under (7), then

$$\hat{T}_{n,m} \underset{\mathcal{L}}{\rightarrow} X,$$

as $n, m \rightarrow \infty, m^3/n \rightarrow 0$. I.e. $\hat{T}_{n,m}$ converges in law to the random variable X. ■

Proof of corollary 2 follows immediately from Theorem 1, i.e. convergence of Kullback-Leibler implies convergence in law. Thus the quantiles associated with $T_{n,m}$ converge to those of Y, i.e. letting $q_A(\pi)$, for $0 < \pi < 1$, denote the quantile function of the random variable A, we have

$$q_{F^{-1}\left(\hat{T}_{n,m};\hat{\beta}_n\right)}(\pi) = q_Y(\pi) + o_p(1). \tag{16}$$

The final set of experiments compare the Mean Square Errors of estimators for the quantiles of Y based on $\hat{T}_{n,m}$ for $n = 25, 50, 100$ and 200 and for $m = 3$ (with the polynomial basis) as well as $m = 10$ (with the trigonometric basis) and for quantiles calculated at the probabilities, $\pi = .05, .25, .50, .75, .95$. We also consider cases where nuisance parameters are, and are not, estimated.

First suppose that $Y_i^a \sim iid Y^a := t_{(4)}$, then define

$$\hat{X}_i = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{Y_i - \bar{y}_n}{\hat{\sigma}_n}\right)\right] \quad \& \quad X_i^* = \frac{1}{2}\left[1 + \mathrm{erf}\left(Y_i\right)\right], \quad i = 1, .., n,$$

i.e. we proceed as if we are testing for Normality. Following the development in Section 2 and 3, as well as that of Barron and Sheu (1991) let $\hat{\theta}_{(m)}$ and $\theta^*_{(m)}$ denote the estimated parameters for the exponential series density estimators for the samples $\left\{\hat{X}_i\right\}_1^n$ and $\{X_i^*\}_1^n$, respectively. Let $\hat{T}_{n,m}$ have density $p_t\left(\hat{\theta}_{(m)}\right)$, as in corollary 2, and also $T^*_{n,m}$ having density $p_t\left(\theta^*_{(m)}\right)$ (note that this is simply the original set-up of Barron and Sheu (1991), i.e. no parameters are estimated). The estimated quantiles for Y^a are then constructed as in

$$\hat{q}_{Y^a}(\pi) = \bar{y} + \hat{\sigma}_n \sqrt{2}\,\mathrm{erf}^{-1}\left(2q_{\hat{T}_{n,m}}(\pi) - 1\right) \quad \text{and} \quad q^*_{Y^a}(\pi) = \sqrt{2}\,\mathrm{erf}^{-1}\left(2q_{T^*_{n,m}}(\pi) - 1\right).$$

The mean square error of these quantiles, for each probability π, are presented in Tables 5a (for $m = 3$ and the polynomial basis) and Table 5b (for $m = 10$ and the trigonometric basis).

Next suppose that $Y_i^b \sim iid Y^b := \Gamma(1.2, 1)$ and define

$$\hat{X}_i = 1 - e^{-Y_i/\bar{y}_n} \quad \& \quad X_i^* = 1 - e^{-Y_i}, \quad i = 1, .., n,$$

i.e. as if we were testing exponentiality. Analogous to above let $\hat{T}_{n,m}$ and $T^*_{n,m}$ have densities $p_t\left(\hat{\theta}_{(m)}\right)$ and $p_t\left(\theta^*_{(m)}\right)$and so estimated quantiles for Y^b are constructed as in,

$$\hat{q}_{Y^b}(\pi) = -\bar{y}_n \ln\left(1 - q_{\hat{T}_{n,m}}(\pi)\right) \quad \text{and} \quad q^*_{Y^b}(\pi) = -\ln\left(1 - q_{T^*_{n,m}}(\pi)\right).$$

The mean square error of these quantiles, for each probability π, are presented in Tables 5c (for $m = 3$ and the polynomial basis) and Table 5d (for $m = 10$ and the trigonometric basis).

Two general conclusions can be drawn from the results presented in Tables 5a through 5d. First there is no advantage, at any sample size, of using a larger value of m. Although not reported, additional experiments show no advantage for one basis over another. On the other hand there is a slight advantage of estimating the mean (and variance) prior mapping to the interval $(0, 1)$ to construct the density estimator. This is particularly the case when m is large and the quantile is in the extreme right tail of a positively skewed random variable - see the last row of results in Table 5d.

4.4 Application Recommendation

The nonparametric likelihood based tests and estimators described above come in a variety of specifications. These depend on whether m is large, to achieve accuracy of critical values under the null, or small, to achieve higher power under a variety of alternatives, and also upon the choice of basis, whether polynomial or trigonometric.

The results of this section, taken together, offer the possibility of a clear procedure. The computational cost of a large value of m (for a given sample size) overrides the improved numerical performance under the null. Since the tests are asymptotically pivotal the bootstrap for such tests is rigorously justified, simple to implement and offers good size and power properties.

Although the theoretical results demand $m \to \infty$ as $n \to \infty$, in practice we want m to grow slowly, e.g. implementing the procedures with the mini-max rate, with

$$m = m_n^* = 1 + \lfloor n^{1/5} \rfloor ,$$

where $\lfloor . \rfloor$ denotes the integer part, in conjunction with the polynomial basis. Note that this simple rule implies $m = 3$ for sample sizes of $n = 25$ through 200, exactly as in the experiments described above.

5 Conclusions

This paper has extended the nonparametric likelihood ratio based tests of Marsh (2007) to cover specifications involving estimated parameters. Having to estimate unknown parameters affects neither the rate of convergence of the density estimator of Barron and Sheu (1991) nor the asymptotic distribution of the associated likelihood ratio test of Portnoy (1988).

The general aim has been to provide a test procedure which overcomes the three main criticisms of edf based tests, i.e. that they are not pivotal, have low power, and offer no direction in case of rejection. Instead the tests of this paper are shown to be asymptotically standard normal, they have power advantages generally, and are often considerably more powerful than the most commonly used Kolmogorov-Smirnov test. Finally, in case of rejection, they offer a quantile approximation which is numerically superior to that obtained from the original set-up of Barron and Sheu (1991).

In a thorough set of experiments, the test applied with the polynomial basis a small value of m, consistent with a mini-max rule, demonstrates both numerical and computational superiority (even when bootstrapped) over competitor edf based tests, and also those based on the same estimator but applied with a trigonometric basis and much larger m.

References.

Babu, G. and C.R. Rao (2004), 'Goodness of fit tests when parameters are estimated' *Sankhya*, 66, 63–74.

Barron, A.R. and C-H. Sheu (1991) 'Approximation of density functions by sequences of exponential families' *Annals of Statistics*, 19, 1347-1369.

Barndorff-Nielsen O. (1978), *Information and Exponential Families in Statistical Theory*, Wiley, New York.

Claeskens, G. and N.L. Hjort (2004) 'Goodness of fit via nonparametric likelihood ratios', *Scandinavian Journal of Statistics*, 31, 487-513.

Conover, W.J. (1999) *Practical Nonparametric Statistics*, John Wiley and Sons, New York.

Corradi, V. and N. R. Swanson (2006) Predictive Density Evaluation, in: G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting,* Vol. 1, Elsevier, 197-284.

Crain, B.R. (1974) 'Estimation of distributions using orthogonal expansions', *Annals of Statistics*, 2, 454–463.

Genest, C and B Rémillard (2008) 'Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models', *Annales de L'Institute Henri Poincaré: Probabilités et Statistiques*, 44, 1096-1127.

Kojadinovic, I. and J. Yan (2012), 'Goodness-of-fit testing based on a weighted bootstrap: A fast large-sample alternative to the parametric bootstrap', *The Canadian Journal of Statistics*, 40, 480-500.

Marsh, P. (2007) 'Goodness of fit tests via exponential series density estimation', *Computational Statistics and Data Analysis*, 51, 2428–2441.

Marsh, P. (2010), 'A two-sample nonparametric likelihood ratio test' *Journal of Nonparametric Statistics*, 22, p. 1053-1065.

Portnoy, S. (1988) 'Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity', *Annals of Statistics*, 16, 356-366.

Stephens, M.A. (1974), 'EDF statistics for goodness of fit and some comparisons', *Journal of the American Statistical Association*, 69, 730-736.

Stephens, M.A. (1976), ‘Asymptotic results for goodness of fit statistics with unknown parameters’, *Annals of Statistics*, 4, 357–369.

Villaseñor, J.A. and E. González-Estrada (2009), ‘A bootstrap goodness of fit test for the generalized Pareto distribution’, *Computational Statistics and Data Analysis*, 53, 3835-3841.

White, H. (1982), ‘Maximum likelihood estimation of misspecified models’, *Econometrica*, 50, 1-25.

Appendix I: Proofs

In order to avoid any ambiguity throughout this appendix the order of magnitude symbol $O(.)$ is defined by,

$$a_{n,m} = O\left(b_{n,m}\right) \Longleftrightarrow \lim_{m,n\to\infty\ ;\ m^3/n\to 0} \frac{a_{n,m}}{b_{n,m}} \leq c_1 < \infty,$$

and analogously for the probabilistic versions $O_p(.)$ and $o_p(.)$. If the quantity under scrutiny does not depend upon the dimension m then the condition $m^3/n \to 0$ becomes redundant.

Proof of Theorem 1:

We will consider the two choices of basis separately using a superscript P or T to denote the polynomial or trigonometric basis. First consider the polynomial basis, $\phi_k(x) = x^k$ and denote ,

$$\hat{x}^P_{(m)} = \left(\frac{\sum_{i=1}^n \hat{X}_i^k}{n}\right)_{k=1}^m, \quad \bar{x}^P_{(m)} = \left(\frac{\sum_{i=1}^n X_i^k}{n}\right)' \quad \text{and} \quad \mu^P_{(m)} = E\left[\bar{x}^P_{(m)}\right].$$

The Euclidean distance between the two polynomial sufficient statistics is,

$$\begin{aligned}\left|\hat{x}^P_{(m)} - \bar{x}^P_{(m)}\right| &= \left|\frac{1}{n}\left(\sum_{i=1}^n \left(\hat{X}_i - X_i\right), ..., \sum_{i=1}^n \left(\hat{X}_i^m - X_i^m\right)\right)'\right| \\ &\leq \sum_{j=1}^m \left|\frac{1}{n}\sum_{i=1}^n \left(\hat{X}_i^j - X_i^j\right)\right|.\end{aligned}$$

Taking the j^{th} element and noting $\hat{X}_i = X_i + e_i$,then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \left(\hat{X}_i^j - X_i^j\right) &= \frac{1}{n}\sum_{i=1}^n \left((X_i + e_i)^j - X_i^j\right) = \frac{1}{n}\sum_{i=1}^n\sum_{s=0}^j \left(\frac{j!}{s!(j-s!)}X_i^{j-s}e_i^s - X_i^j\right) \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{s=1}^j \frac{j!}{s!(j-s)!}X_i^{j-s}e_i^s.\end{aligned}$$

Since $X_i \in (0,1)$ while, as in (8), $e_i = O_p(n^{-1/2})$ and $e_i \in (-1,1)$ then,

$$\frac{j!}{s!(j-s)!}X_i^{j-s}e_i^s \leq \frac{j^s}{s!}c_1^{j-s}e_i^s = \frac{j^s}{s!}c^{j-s}O_p\left(n^{-s/2}\right), \tag{17}$$

where $c_1 < 1$. For finite j (17) is $O_p\left(n^{-s/2}\right)$ while as $j \to \infty$ (17) is $o(1)\,O_p\left(n^{-s/2}\right)$ and so,

$$\sup_{j\in\mathbb{N}} \frac{j!}{s!(j-s)!}X_i^{j-s}e_i^s = O_p\left(n^{-s/2}\right),$$

implying that

$$\sum_{s=1}^{j} \frac{j!}{s!(j-s)!} X_i^{j-s} e_i^s = O_p(n^{-1/2}),$$

uniformly in j, and hence,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i^j - X_i^j\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{s=1}^{j}\frac{j!}{s!(j-s!)} X_i^{j-s} e_i^s\right) = O_p(n^{-1/2}).$$

Consequently, and also from the definition of Euclidean distance, we have,

$$\left|\hat{x}_{(m)}^P - \bar{x}_{(m)}^P\right| = \sqrt{\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i^j - X_i^j\right)\right)^2} = O_p\left(\sqrt{\frac{m}{n}}\right). \tag{18}$$

Consider now $\mu_{(m)} = \left(E_{U_o}[x], E_{U_0}[x^2], ..., E_{U_0}[x^m]\right)'$, then from the triangle inequality,

$$\left|\hat{x}_{(m)}^P - \mu_{(m)}^P\right| \leq \left|\bar{x}_{(m)}^P - \mu_{(m)}^P\right| + \left|\hat{x}_{(m)}^P - \bar{x}_{(m)}^P\right| = O_p\left(\sqrt{\frac{m}{n}}\right), \tag{19}$$

which follows from (18) and noting the same order of magnitude applies for the first distance, as in Barron and Sheu (1991, eq. 6.5), which represents the distance in the case that the sequence $\left(X_i^j\right)_1^n$ were observed directly.

The trigonometric basis case is more straight forward. The relevant Euclidean distance satisfies

$$\begin{aligned}\left|\hat{x}_{(m)}^T - \bar{x}_{(m)}^T\right| &\leq \left|\hat{x}_{(m)}^T - \mu_{(m)}^T\right| + \left|\bar{x}_{(m)}^T - \mu_{(m)}^T\right| \\ &= \sqrt{\sum_{k=1}^{m/2}\left(\frac{\sum_{i=1}^n\left(\cos\left[2\pi k\hat{X}_i\right]\right)}{n} - \mu_{k,c}^T\right)^2 + \left(\frac{\sum_{i=1}^n \sin\left[2\pi k\hat{X}_i\right]}{n} - \mu_{k,s}^T\right)^2} \\ &\quad + \sqrt{\sum_{k=1}^{m/2}\left(\frac{\sum_{i=1}^n \cos\left[2\pi kX_i\right]}{n} - \mu_{k,c}^T\right)^2 + \left(\frac{\sum_{i=1}^n \sin\left[2\pi kX_i\right]}{n} - \mu_{k,s}^T\right)^2},\end{aligned} \tag{20}$$

where

$$E_U\left(\cos\left[2\pi kX_i\right]\right) = \mu_{k,c}^T \quad \text{and} \quad E_U\left(\sin\left[2\pi kX_i\right]\right) = \mu_{k,s}^T.$$

Since, for any k,

$$\frac{\sum_{i=1}^n\left(\cos\left[2\pi kX_i\right]\right)}{n} - \mu_{k,c}^T = O_p\left(n^{-1/2}\right) \quad \& \quad \frac{\sum_{i=1}^n\left(\sin\left[2\pi kX_i\right]\right)}{n} - \mu_{k,s}^T = O_p\left(n^{-1/2}\right),$$

and that cos [.] and sin [.] are bounded and differentiable then from Assumption 1

$$\frac{\sum_{i=1}^{n}\left(\cos\left[2\pi k\hat{X}_i\right]\right)}{n}-\mu_{k,c}^{T}=O_p\left(n^{-1/2}\right) \quad \& \quad \frac{\sum_{i=1}^{n}\left(\sin\left[2\pi k\hat{X}_i\right]\right)}{n}-\mu_{k,s}^{T}=O_p\left(n^{-1/2}\right),$$

also. Consequently,

$$\left|\hat{x}_{(m)}^{T}-\bar{x}_{(m)}^{T}\right|\le 2\sqrt{\sum_{k=1}^{m/2}(d^*)^2}=O_p\left(\sqrt{\frac{m}{n}}\right),$$

where $d^*=O_p\left(n^{-1/2}\right)$ is

$$\begin{aligned} d^* &= \max_k\left\{\frac{\sum_{i=1}^{n}\left(\cos\left[2\pi kX_i\right]\right)}{n}-\mu_{k,c}^{T},\frac{\sum_{i=1}^{n}\left(\sin\left[2\pi kX_i\right]\right)}{n}-\mu_{k,s}^{T}\right.\\ &\quad \left.,\frac{\sum_{i=1}^{n}\left(\cos\left[2\pi k\hat{X}_i\right]\right)}{n}-\mu_{k,c}^{T},\frac{\sum_{i=1}^{n}\left(\sin\left[2\pi k\hat{X}_i\right]\right)}{n}-\mu_{k,s}^{T}\right\}.\end{aligned}$$

Generally, i.e. for either basis, we thus have $\left|\hat{x}_{(m)}-\bar{x}_{(m)}\right|=O_p\left(\sqrt{\frac{m}{n}}\right)$ and $\left|\hat{x}_{(m)}-\mu_{(m)}\right|=O_p\left(\sqrt{\frac{m}{n}}\right)$, so that utilizing the respective mles and extending the decomposition of the Kullback-Leibler divergence of Barron and Sheu (1991, eq. 6.9) we obtain,

$$\begin{aligned} E_U\left[\ln\left(\frac{u(x)}{p_x(\hat{\theta}_{(m)})}\right)\right] &= E_U\left[\ln\left(\frac{u(x)}{p_x(\theta_{(m)})}\right)\right]+E_U\left[\ln\left(\frac{p_x(\theta_{(m)})}{p_x(\bar{\theta}_{(m)})}\right)\right]\\ &\quad +E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right]. \end{aligned} \tag{21}$$

Given that Assumption 1(v) implies that $u(x)\in W_2^r$ then from Barron and Sheu (1991, Theorem 1) the first two terms in (21) are, respectively, $O(m^{-2r})$ and $O_p(m/n)$, noting that under Assumption 1, $\log[u(x)]\in W_2^r$. Application of Barron and Sheu (1991, Lemma 5), which holds for any two values in $\Omega_m\subset\mathbb{R}^m$, here uniquely defined by equations (5) and (9), implies that

$$O\left(E_U\left[\ln\left(\frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})}\right)\right]\right)=O_p\left(\left|\hat{x}_{(m)}-\bar{x}_{(m)}\right|^2\right)=O_p\left(\frac{m}{n}\right),$$

and hence

$$\begin{aligned} E_U\left[\ln\left(\frac{u(x)}{p_x(\hat{\theta}_{(m)})}\right)\right] &= O(m^{-2r})+O_p\left(\frac{m}{n}\right)+O_p\left(\frac{m}{n}\right)\\ &= O_p\left(m^{-2r}+\frac{m}{n}\right), \end{aligned}$$

as required. ■

Proof of Theorem 2:

Consider the problem of testing $H_0 : \theta_{(m)} = 0_{(m)}$ against the alternative $H_1 : \theta_{(m)} \neq 0_{(m)}$ when $n, m \to \infty$, but $m^3/n \to 0$.

Part (i): To proceed we have defined,

$$\hat{\lambda}_m = 2n\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)'\hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(0_{(m)}\right)\right)\right] = 2n\hat{\theta}'_{(m)}\hat{x}_{(m)},$$

where $\hat{\theta}_{(m)}$ solves (9), or equivalently,

$$\psi'_m\left(\hat{\theta}_{(m)}\right) = \left.\frac{\partial\psi_m\left(\theta_{(m)}\right)}{\partial\theta_{(m)}}\right|_{\theta_{(m)}=\hat{\theta}_{(m)}} = \hat{x}_{(m)}.$$

Similarly the value $\theta^0_{(m)}$ is defined by,

$$\psi'_m\left(\theta^0_{(m)}\right) = \mu_{(m)} = E(\bar{x}_{(m)}).$$

Since the exponential log-likelihood is strictly convex then the mapping,

$$\theta_{(m)}(\eta) : \psi'_m(\theta) = \mu_{(m)}$$

is one-to-one between the parameter space $\Theta_m \subset \mathbb{R}^m$ and sample space $\Omega_m \subset \mathbb{R}^m$ and application of Barron and Sheu (1991, eq. 5.6) and also (19) gives,

$$O_p\left(\left|\hat{\theta}_{(m)} - 0_{(m)}\right|\right) = O_p\left(\left|\hat{x}_{(m)} - \mu_{(m)}\right|\right) = O_p\left(\sqrt{\frac{m}{n}}\right). \qquad (22)$$

As a consequence of both (22) and (19) we have that,

$$O_p\left(\left|\hat{\theta}_{(m)} - 0_{(m)}\right|\right) = O_p\left(\left|\bar{\theta}_{(m)} - 0_{(m)}\right|\right) \quad \& \quad O_p\left(\left|\hat{x}_{(m)} - \mu_{(m)}\right|\right) = O_p\left(\left|\bar{x}_{(m)} - \mu_{(m)}\right|\right),$$

and note that the expansions provided in the provided in the proofs of Theorems 3.1 and 3.2 of Portnoy (1988) apply for any two pairs of values, i.e. $\left(\bar{\theta}_{(m)}, 0_{(m)}\right)$ and $\left(\bar{x}_{(m)}, \mu_{(m)}\right)$.

To continue, noting expectations under the null hypothesis can be written here as $E_{U_0}[.]$ since U_0 is the uniform distribution with density $p_{0_{(m)}}(x) - 1$, we have expansions analogous to Portnoy (1988, eq. 3.5 and 3.6),

$$|\hat{\theta}_{(m)} - 0_{(m)}|^2 = \left(\hat{\theta}_{(m)} - 0_{(m)}\right)'\hat{x}_{(m)} - \frac{1}{2}E_{U_0}\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)'U\right]^2 + O_p\left(\frac{m^2}{n^2}\right),$$

and (23)

$$\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} = |\hat{x}_{(m)}|^2 - \frac{1}{2} E_{U_0}\left[\left(\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' U\right)^2 \hat{x}'_{(m)} U\right] + O_p\left(\frac{m^2}{n^2}\right). \tag{24}$$

Subtracting (24) from (23) and applying arguments identical to those given below Portnoy (1988, Theorem 3.1, eq. 3.7) yields,

$$|\hat{\theta}_{(m)} - \theta_{(m)} - \hat{x}_{(m)}| = O_p\left(\frac{m}{n}\right).$$

From the definition of the likelihood ratio test we therefore have,

$$\begin{aligned} \hat{\lambda}_m &= 2n\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right)\right)\right] \\ &= n\left[|\hat{x}_{(m)}|^2 - |\hat{\theta}_{(m)} - \theta^0_{(m)} - \hat{x}_{(m)}|^2 + \frac{1}{6} E_{\theta_0}\left(\left(\hat{\theta}_{(m)} - \theta^0_{(m)}\right)' U\right)^3\right] + O_p\left(\frac{m^2}{n}\right), \end{aligned} \tag{25}$$

as in Portnoy (1988, eq. 3.12). Let $\bar{e} = \hat{x}_{(m)} - \bar{x}_{(m)}$ then from the proof of Theorem 1, we have

$$|\hat{x}_{(m)}|^2 = |\bar{x}_{(m)} + \bar{e}|^2 = |\bar{x}_{(m)}|^2 + O_p\left(\frac{m}{n}\right). \tag{26}$$

Now define the $m \times 1$ random variable $V_m = \psi''_m\left(0_{(m)}\right)^{-1/2}\left(\bar{x} - \psi'_m\left(0_{(m)}\right)\right)$, having density $p_V\left(\theta^V_{(m)}\right)$, so that $E[V] = 0_{(m)}$ and $Var[V_m] = I_m$. Since the likelihood ratio statistic is parameterization invariant the likelihood ratio test based on observations on V_m will be identical to that based on $\bar{x}_{(m)}$. Rather than defining a new triple of values, analogous to those in (12), in both the parameter space Θ_m (note that in particular the hypothesized value would no longer satisfy $\theta^0_{(m)} = 0_{(m)}$) and sample space Ω_m we will instead, and without any loss of generality assume a parameterization in which both $E\left[\bar{x}_{(m)}\right] = 0$ and $V\left[\bar{x}_{(m)}\right] = I_m$. Note, however, that it is the unobserved $\bar{x}_{(m)}$ which is assumed to be standardized not the observed $\hat{x}_{(m)}$.

In this parameterization the asymptotic distribution of first $|\bar{x}_{(m)}|^2$ and hence $|\hat{x}_{(m)}|^2$ (via (26)) and then via (25) for $\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}}$ follows exactly as in Portnoy (1988, Theorem 4.1). ■

Part (ii): Under any fixed alternative the density of $X_i = F\left(Y_i; \beta_*\right)$ is

$$u_1(x) = \frac{g\left(F^{-1}(x; \beta_*)\right)}{f\left(F^{-1}(x; \beta_*)\right)},$$

and so let $\theta^1_{(m)}$ be the unique solution to,

$$\int_0^1 x^j \, p_h \left(\theta^1_{(m)}\right) dx = \int_0^1 h^j \, u_1 (x) \, dx \quad ; \quad j = 1, .., m. \tag{27}$$

The uniqueness of solutions to (27) imply $\theta^1_{(m)} \neq 0_{(m)}$.

To take the least favorable case, define

$$\theta^1_{(m)} = \left(\theta^1_1, .\theta^1_2, .., \theta^1_m\right)'$$

and suppose that $\theta^1_k \neq 0$ for some finite k but that $\theta^1_j = 0$ for all $j \neq k$. The series density estimator is consistent for $\theta^1_{(m)}$, under H_1, in that $\left|\hat{\theta}_{(m)} - \theta^1_{(m)}\right| = O_p\left(\sqrt{\frac{m}{n}}\right)$, analogous to (22) above, and so we can write,

$$n\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} = n\left[\left(\hat{\theta}_{(m)} - \theta^1_{(m)}\right)' \hat{x}_{(m)} + \left(\theta^1_k\right) \frac{1}{n} \sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right)\right].$$

We can therefore write the likelihood ratio as

$$\begin{aligned}
\hat{\lambda}_m &= 2n\left[\left(\hat{\theta}_{(m)} - 0_{(m)}\right)' \hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(0_{(m)}\right)\right)\right] \\
&= 2n\left[\left(\hat{\theta}_{(m)} - \theta^1_{(m)}\right)' \hat{x}_{(m)} - \left(\psi_m\left(\hat{\theta}_{(m)}\right) - \psi_m\left(\theta^1_{(m)}\right)\right)\right] \\
&\quad +2n\left[\left(\theta^1_k - \theta^0_k\right) \frac{1}{n} \sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right) - \left(\psi_m\left(\theta^1_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right)\right)\right] \\
&= \hat{\lambda}^1_m + 2n\left[\left(\theta^1_k - \theta^0_k\right) \frac{1}{n} \sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right) - \left(\psi_m\left(\theta^1_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right)\right)\right],
\end{aligned}$$

where $\hat{\lambda}^1_m$ is the likelihood ratio for testing $H_1 : \theta_{(m)} = \theta^1_{(m)}$.

Thus, under H_1, we can write

$$\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}} = \frac{\hat{\lambda}^1_m - m}{\sqrt{2m}} + \frac{2n\left[\left(\theta^1_k - \theta^0_k\right) \frac{1}{n} \sum_{i=1}^{n} \phi_k\left(\hat{X}_i\right) - \left(\psi_m\left(\theta^1_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right)\right)\right]}{\sqrt{2m}}.$$

Immediate from Part (i) of this theorem is that as $m, n \to \infty$, with $m^3/n \to 0$,

$$\frac{\hat{\lambda}^1_m - m}{\sqrt{2m}} \to_d N(0, 1),$$

i.e. $\left(\hat{\lambda}^1_m - m\right)/\sqrt{2m}$ is $O_p(1)$. However, since $\psi_m(.)$ is a uniquely defined cumulant function then

$$\psi_m\left(\theta^1_{(m)}\right) - \psi_m\left(\theta^0_{(m)}\right) \neq 0,$$

and since $0 < \hat{X}_i < 1$ then $\frac{1}{n}\sum_{i=1}^{n} \hat{X}_i^k = O_p(1)$ and positive. Consequently,

$$\hat{\Lambda}_m = O_p(1) + O_p\left(\frac{n}{\sqrt{m}}\right) \to \infty,$$

since $m^3/n \to 0$ and hence $\Pr\left[\hat{\Lambda}_m > \kappa\right] \to 1$, as required. ■

Appendix II Tables

Table 1a: Sizes of tests for both H_0^E and H_0^N for different m and n using the trigonometric basis.

$n = 25$

	H_0^E			H_0^N		
α / m	.10	.05	.01	.10	.05	.01
2	.053	.033	.017	.012	.007	.002
4	.071	.042	.020	.039	.023	.011
6	.073	.037	.007	.046	.026	.011
8	.081	.056	.020	.054	.027	.004
10	.092	.057	.018	.058	.030	.005
12	.091	.055	.014	.065	.034	.005

$n = 50$

	H_0^E			H_0^N		
α / m	.10	.05	.01	.10	.05	.01
2	.059	.040	.017	.012	.005	.001
4	.067	.044	.021	.042	.029	.009
6	.072	.045	.016	.045	.027	.009
8	.077	.047	.020	.056	.031	.008
10	.096	.055	.020	.065	.037	.009
12	.094	.055	.018	.063	.038	.007

$n = 100$

	H_0^E			NH_0^N		
α / m	.10	.05	.01	.10	.05	.01
2	.055	.037	.017	.013	.006	.001
4	.070	.042	.008	.050	.035	.009
6	.088	.051	.016	.063	.041	.011
8	.093	.055	.016	.078	.045	.013
10	.098	.054	.014	.080	.047	.015
12	.096	.053	.014	.092	.045	.018

$n = 200$

	H_0^E			H_0^N		
α / m	.10	.05	.01	.10	.05	.01
2	.065	.038	.020	.012	.006	.002
4	.070	.041	.015	.049	.031	.010
6	.085	.047	.017	.065	.041	.011
8	.093	.048	.016	.077	.043	.010
10	.098	.055	.012	.085	.045	.017
12	.096	.052	.013	.094	.046	.016

Table 1b: Sizes of tests for both H_0^E and H_0^N for different m and n using the polynomial basis.

$n = 25$

α / m	H_0^E .10	H_0^E .05	H_0^E .01	H_0^N .10	H_0^N .05	H_0^N .01
3	.035	.016	.003	.030	.013	.003
5	.050	.025	.004	.041	.019	.002
7	.062	.033	.006	.049	.024	.004
9	.064	.034	.006	.051	.023	.004
13	.069	.037	.006	.050	.028	.009
17	.063	.031	.005	.055	.023	.003

$n = 50$

α / m	H_0^E .10	H_0^E .05	H_0^E .01	H_0^N .10	H_0^N .05	H_0^N .01
3	.044	.019	.003	.034	.017	.005
5	.047	.023	.005	.041	.023	.004
7	.063	.030	.005	.051	.027	.004
9	.067	.032	.006	.059	.028	.004
13	.074	.035	.004	.065	.031	.005
17	.069	.029	.006	.066	.029	.006

$n = 100$

α / m	H_0^E .10	H_0^E .05	H_0^E .01	H_0^N .10	H_0^N .05	H_0^N .01
3	.051	.026	.004	.035	.019	.003
5	.056	.028	.006	.043	.021	.004
7	.068	.035	.008	.056	.028	.005
9	.073	.040	.007	.065	.031	.005
13	.085	.047	.008	.075	.038	.007
17	.091	.043	.009	.081	.041	.009

$n = 200$

α / m	H_0^E .10	H_0^E .05	H_0^E .01	H_0^N .10	H_0^N .05	H_0^N .01
3	.051	.023	.005	.045	.021	.004
5	.061	.037	.006	.053	.029	.007
7	.071	.043	.008	.063	.031	.006
9	.081	.045	.011	.078	.040	.006
13	.095	.047	.009	.086	.045	.009
17	.097	.048	.011	.095	.049	.011

Table 2: Rejection frequencies under various alternatives. The left hand panels corresponds to cases where unknown parameters are not estimated, while for the right had panels parameters are estimated.

Table 2a: Power $H_0 : Y \sim N(0,1)$ vs. $H_1 : Y \sim t_{(v)}$.

v	4	6	8	10	12	4	6	8	10	12
$\hat{\Lambda}_3^P$	.935	.705	.386	.267	.114	.605	.294	.166	.127	.097
Λ_{10}^T	.856	.563	.254	.159	.087	.494	.241	.133	.111	.081
KS	.614	.206	.091	.055	.049	.217	.114	.075	.059	.052
CM	.722	.309	.165	.092	.061	.296	.132	.087	.075	.066
AD	.767	.361	.182	.115	.065	.530	.240	.139	.103	.090

Table 2b: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim \chi^2_{(v)} - v$.

v	12	20	28	36	44	12	20	28	36	44
$\hat{\Lambda}_3^P$	.859	.660	.577	.476	.422	.572	.274	.189	.146	.114
Λ_{10}^T	.796	.641	.546	.427	.377	.388	.189	.158	.111	.096
KS	.717	.568	.443	.388	.350	.238	.151	.106	.093	.075
CM	.837	.663	.563	.463	.403	.274	.176	.131	.100	.091
AD	.843	.647	.529	.439	.388	.286	.165	.117	.098	.083

Table 2c: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N(vY_{i-1}, 1)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_3^P$	.694	.592	.386	.161	.093	.902	.736	.510	.271	.101
Λ_{10}^T	.688	.483	.351	.141	.071	.847	.683	.461	.235	.091
KS	.592	.458	.254	.091	.053	.579	.359	.207	.122	.058
CM	.690	.585	.362	.140	.066	.648	.448	.273	.162	.083
AD	.691	.580	.371	.138	.057	.866	.704	.471	.242	.089

Table 2d: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N\left(0, 1 + vY_{i-1}^2\right)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_3^P$	.722	.514	.276	.116	.079	.869	.729	.493	.225	.106
Λ_{10}^T	.704	.503	.263	.113	.074	.864	.740	.460	.225	.094
KS	.568	.361	.161	.063	.052	.509	.350	.201	.112	.080
CM	.709	.497	.255	.109	.075	.511	.352	.185	.115	.073
AD	.708	.494	.246	.088	.054	.849	.721	.451	.215	.088

Table 2e: Power $H_0 : Y_i \sim N(0,1)$ vs. $H_1 : Y_i \sim N\left(v1_{t>\lfloor T/2 \rfloor}, 1\right)$.

v	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_3^P$	.754	.563	.349	.196	.079	.653	.495	.274	.141	.080
Λ_{10}^T	.738	.525	.311	.173	.064	.592	.442	.256	.139	.066
KS	.256	.189	.127	.088	.052	.542	.349	.185	.078	.059
CM	.362	.291	.164	.103	.066	.601	.445	.260	.130	.078
AD	.750	.539	.321	.185	.075	.625	.467	.258	.111	.067

Table 2f: Power $H_0 : Y_i \sim Exp[1]$ vs. $H_1 : Y_i \sim \Gamma(v,1)$.

v	1.10	1.15	1.20	1.25	1.30	1.10	1.15	1.20	1.25	1.30
$\hat{\Lambda}_3^P$	.115	.121	.238	.305	.428	.191	.298	.585	.769	.866
Λ_{10}^T	.103	.106	.179	.277	.398	.177	.285	.550	.712	.825
KS	.066	.069	.136	.200	.252	.096	.193	.404	.616	.747
CM	.094	.099	.179	.237	.343	.174	.280	.551	.732	.853
AD	.097	.109	.227	.303	.419	.182	.299	.589	.770	.884

Table 3a: Rejection Frequencies at 5% level under the respective null hypotheses

i) $H_0^N : Y \sim N^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.064	.065	.058	.044
CM	.064	.058	.057	.054
AD	.060	.059	.062	.058

ii) $H_0^\Gamma : Y \sim \Gamma^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.062	.056	.049	.046
CM	.068	.060	.065	.061
AD	.065	.055	.052	.061

iii) $H_0^W : Y \sim W^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.067	.055	.055	.047
CM	.063	.058	.056	.058
AD	.055	.066	.065	.057

iv) $H_0^L \sim L^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.065	.062	.050	.042
CM	.071	.066	.059	.055
AD	.062	.054	.055	.055

Table 3b: Rejection Frequencies at 5% level under various alternatives

i) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim \Gamma^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.069	.088	.116	.175
CM	.078	.094	.123	.185
AD	.069	.090	.108	.161

ii) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim N^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.068	.085	.099	.129
CM	.055	.066	.079	.088
AD	.076	.085	.092	.113

iii) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim W^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.196	.364	.584	.897
CM	.094	.192	.465	.776
AD	.183	.315	.550	.806

iv) $H_0 : Y \sim W^*$ vs. $H_1 : Y \sim N^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.101	.164	.351	.690
CM	.107	.233	.388	.580
AD	.098	.164	.334	.602

v) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim L^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.173	.249	.393	.458
CM	.111	.152	.212	.358
AD	.131	.190	.246	.417

vi) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim N^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.046	.055	.065	.101
CM	.036	.054	.073	.109
AD	.041	.046	.070	.108

Table 3c: Rejection Frequencies at 5% level under various alternatives

i) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim L^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.091	.122	.188	.253
CM	.078	.081	.122	.193
AD	.105	.128	.174	.257

ii) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim W^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.285	.448	.709	.937
CM	.320	.476	.781	.970
AD	.155	.306	.638	.938

iii) $H_0 : Y \sim W^*$ vs. $H_1 : Y \sim \Gamma^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.197	.355	.719	.945
CM	.200	.315	.534	.836
AD	.117	.219	.482	.851

iv) $H_0 : Y \sim W^*$ vs. $H_1 : Y \sim L^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.172	.327	.620	.867
CM	.215	.343	.542	.797
AD	.159	.277	.500	.816

v) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim \Gamma^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.059	.082	.120	.152
CM	.059	.081	.130	.161
AD	.051	.059	.101	.148

vi) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim W^*$

n	25	50	100	200
$\hat{\Lambda}_m$	.243	.343	.592	.892
CM	.124	.241	.519	.882
AD	.204	.325	.583	.912

Table 5a: MSE of estimated quantiles for $Y \sim t_{(4)}$,

polynomial basis, $m = 3$.

	$q^*_{Y^T}$	$q^*_{Y^T}$	$q^*_{Y^T}$	$q^*_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$
n / π	25	50	100	200	25	50	100	200
0.05	.1266	.0906	.0774	.0718	.1309	.0731	.0506	.0389
0.25	.0450	.0227	.0127	.0066	.0446	.0218	.0119	.0058
0.50	.0397	.0183	.0101	.0049	.0348	.0159	.0087	.0042
0.75	.0442	.0222	.0126	.0074	.0445	.0215	.0117	.0066
0.95	.1293	.0976	.0768	.0693	.1333	.0806	.0505	.0375

Table 5b: MSE of estimated quantiles for $Y \sim t_{(4)}$,

trigonometric basis, $m = 10$.

	$q^*_{Y^T}$	$q^*_{Y^T}$	$q^*_{Y^T}$	$q^*_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$	$\hat{q}_{Y^T}$
n / π	25	50	100	200	25	50	100	200
0.05	.1408	.1270	.1179	.1159	.1060	.0551	.0354	.0240
0.25	.0308	.0187	.0139	.0116	.0513	.0244	.0138	.0064
0.50	.0175	.0089	.0053	.0023	.0411	.0182	.0105	.0052
0.75	.0313	.0194	.0138	.0113	.0512	.0253	.0123	.0065
0.95	.1441	.1271	.1186	.1155	.1077	.0599	.0344	.0228

Table 5c: MSE of estimated quantiles for $Y \sim \Gamma(1.2, 1)$,

polynomial basis, $m = 3$.

	$q^*_{Y^\Gamma}$	$q^*_{Y^\Gamma}$	$q^*_{Y^\Gamma}$	$q^*_{Y^\Gamma}$	$\hat{q}_{Y^\Gamma}$	$\hat{q}_{Y^\Gamma}$	$\hat{q}_{Y^\Gamma}$	$\hat{q}_{Y^\Gamma}$
n / π	25	50	100	200	25	50	100	200
0.05	.0002	.0001	.0001	.0000	.0002	.0001	.0001	.0000
0.25	.0022	.0011	.0006	.0003	.0022	.0012	.0006	.0003
0.50	.0113	.0057	.0028	.0014	.0115	.0059	.0029	.0015
0.75	.0379	.0175	.0090	.0047	.0368	.0172	.0091	.0049
0.95	.2007	.1045	.0529	.0286	.2069	.1137	.0637	.0423

Table 5d: MSE of estimated quantiles for $Y \sim \Gamma(1.2, 1)$,

polynomial basis, $m = 10$.

π \ n	$q^*_{Y^\Gamma}$ 25	$q^*_{Y^\Gamma}$ 50	$q^*_{Y^\Gamma}$ 100	$q^*_{Y^\Gamma}$ 200	$\hat{q}_{Y^\Gamma}$ 25	$\hat{q}_{Y^\Gamma}$ 50	$\hat{q}_{Y^\Gamma}$ 100	$\hat{q}_{Y^\Gamma}$ 200
0.05	.0006	.0004	.0003	.0003	.0002	.0001	.0001	.0000
0.25	.0094	.0065	.0051	.0042	.0033	.0017	.0009	.0004
0.50	.0494	.0375	.0335	.0310	.0156	.0073	.0037	.0019
0.75	.2085	.1676	.1583	.1532	.0626	.0276	.0146	.0078
0.95	.7456	.7352	.7380	.7395	.2180	.1093	.0527	.0267