

# Nonparametric series density estimation and testing

Patrick Marsh<sup>1</sup> 

Accepted: 13 July 2018 / Published online: 1 August 2018  
© The Author(s) 2018, corrected publication September 2018

**Abstract** This paper first establishes consistency of the exponential series density estimator when nuisance parameters are estimated as a preliminary step. Convergence in relative entropy of the density estimator is preserved, which in turn implies that the quantiles of the population density can be consistently estimated. The density estimator can then be employed to provide a test for the specification of fitted density functions. Commonly, this testing problem has utilized statistics based upon the empirical distribution function, such as the Kolmogorov-Smirnov or Cramér von-Mises, type. However, the tests of this paper are shown to be asymptotically pivotal having limiting standard normal distribution, unlike those based on the edf. For comparative purposes with those tests, the numerical properties of both the density estimator and test are explored in a series of experiments. Some general superiority over commonly used edf based tests is evident, whether standard or bootstrap critical values are used.

**Keywords** Goodness-of-fit · Nonparametric likelihood ratio · Nuisance parameters and series density estimator

**Mathematics Subject Classification** 62G07 · 62G10 · 62E10

## 1 Introduction

Testing whether a sample of data has been generated from a hypothesized distribution is one of the fundamental problems in statistics and econometrics. Traditionally such tests have been constructed from the empirical distribution function (edf). Even under

---

✉ Patrick Marsh  
Patrick.Marsh@nottingham.ac.uk

<sup>1</sup> School of Economics, University of Nottingham, Nottingham NG7 2RD, UK

the simplest of sampling schemes such tests are known to be not asymptotically pivotal, e.g. see [Stephens \(1976\)](#), [Conover \(1999\)](#) and [Babu and Rao \(2004\)](#). Moreover, under more sophisticated sampling schemes such tests can become prohibitively complex, see [Bai \(2003\)](#) and [Corradi and Swanson \(2006\)](#).

Instead, this paper provides tests based on a generalization of the consistent series density estimator of [Crain \(1974\)](#) and [Barron and Sheu \(1991\)](#). Consistency is maintained when nuisance parameters are estimated as a preliminary step. This, when applied to the infinite dimensional likelihood ratio test of [Portnoy \(1988\)](#) generalizes the tests of [Claeskens and Hjort \(2004\)](#) and [Marsh \(2007\)](#) to test for specification.

The proposed procedure offers three advantages over those tests based on the edf. First they are asymptotically pivotal, and numerical experiments are designed and reported in support of this. This also implies automatic validity, including second-order as in [Beran \(1988\)](#), of bootstrap critical values. Valid bootstrap critical values for the non-pivotal edf based tests, e.g. as in [Kojadinovic and Yan \(2012\)](#), do not benefit from this. Second, they are generally more powerful than the most commonly used edf based tests. Again numerical evidence is presented to support this. Lastly, because they are based on a consistent density estimator, in the event of rejection the density estimator itself can be used to, for instance, consistently estimate the quantiles of the underlying variable.

The plan for the paper is as follows. The next section presents the density estimator and demonstrates that it converges in relative entropy to the population density. A corollary provides consistent quantile estimation, with accuracy demonstrated in numerical experiments. Section 3 provides the nonparametric test, establishes that it is asymptotically pivotal and consistent against fixed alternatives. A corollary establishes validity of bootstrap critical values. Numerical experiments are presented in support of these results as well as demonstrating some superiority over edf based tests. Section 4 concludes while two appendices present the proofs of two theorems and tables containing the outcome of the experiments, respectively.

## 2 Consistent nonparametric estimation of possibly misspecified densities

### 2.1 Theoretical results

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 density  $g(y) = dG(y)/dy$ . For this sample we fit the parametric likelihood,  $L = \prod_{i=1}^n f(Y_i; \beta)$  for some chosen density function  $f(y; \beta)$ , where  $\beta$  is an unknown  $k \times 1$  parameter. Denote the (quasi) maximum likelihood estimator for  $\beta$  by  $\hat{\beta}_n$ .

In this context the hypothesis to be tested is:

$$H_0 : G(y) = F(y; \beta_0), \quad (1)$$

where  $F(y; \beta) = \int_{-\infty}^y f(z; \beta) dz$  and for some (unknown) value  $\beta_0$ . Tests for  $H_0$  will be detailed in the next Section. First, however, we assume the following, whether or not  $H_0$  holds:

### Assumption 1

- (i) The density  $f(y; \beta)$  is measurable in  $y$  for every  $\beta \in \mathbf{B}$ , a compact subset of  $p$ -dimensional Euclidean space, and is continuous in  $\beta$  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  $\beta$  where  $v(\cdot)$  is integrable with respect to  $G(\cdot)$ .
- (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 some  $\beta_* \in \mathbf{B}$ . (iv)  $F(Y; \beta)$  is continuously differentiable with respect to  $\beta$ , such that  $H(\beta) = \partial F(Y_i, \beta) / \partial \beta$  is finite, for all  $\beta$  in a closed ball of radius  $\epsilon > 0$ , around  $\beta_*$ . (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  $\beta_*$ . Note that under  $H_0$  we have  $\beta_* = \beta_0$ . To proceed denote  $\hat{X}_i = F(Y_i, \hat{\beta}_n)$  having mean value expansion,

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

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

$$\hat{X}_i = \bar{X}_i + e_i, \quad (2)$$

where  $\bar{X}_i = F(Y_i, \beta_*)$  and by construction and as a consequence of Assumption 1(iv),

$$e_i \in (-1, 1) \quad \& \quad e_i = O_p(n^{-1/2}), \quad (3)$$

that is  $e_i$  is both bounded and degenerate.

Since the  $\bar{X}_i$  are IID denote their common distribution and density function by  $U(x) = \Pr[\bar{X} < x]$  and  $u(x) = dU(x)/dx$ , respectively. Here we will apply the series density estimator of Crain (1974) and Barron and Sheu (1991) to consistently estimate  $u(x)$  and thus quantiles of  $U(x)$ , from which the quantiles of  $G(y)$  can be consistently recovered. Application of the density estimator requires choice of approximating basis, here we choose the simplest polynomial basis, similar to Marsh (2007).

We will approximate  $u(x)$  via the exponential family,

$$p_x(\theta) = \exp \left\{ \sum_{k=1}^m \theta_k x^k - \psi_m(\theta) \right\}, \quad \psi_m(\theta) = \ln \int_0^1 \exp \left\{ \sum_{k=1}^m \theta_k x^k \right\} dx, \quad (4)$$

where  $\psi_m(\theta)$  is the cumulant function, defined so that  $\int_0^1 p_x(\theta) dx = 1$ .

From Assumption 1  $\log[u(x)]$  has, at least,  $r-1$  absolutely continuous derivatives and its  $r^{\text{th}}$  derivative is square integrable. According to [Barron and Sheu \(1991\)](#) there exists a unique  $\theta_{(m)} = (\theta_1, \dots, \theta_m)'$  satisfying

$$\int_0^1 x^k p_x(\theta_{(m)}) dx = \mu_k = \int_0^1 x^k u(x) dx \quad \text{for } k = 1, 2, \dots, m, \quad (5)$$

and, as  $m \rightarrow \infty$ ,  $p_x(\theta_{(m)})$  converges, in relative entropy, to  $u(x)$  at rate  $m^{-2r}$ , meaning that

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

as  $m \rightarrow \infty$ . Moreover, if a sample  $\{\bar{X}_i\}_1^n$  were available then if  $m^3/n \rightarrow 0$  and letting  $\bar{\theta}_{(m)}$  be the unique solution to

$$\int_0^1 x^k p_x(\bar{\theta}_{(m)}) dx = \frac{\sum_{i=1}^n \bar{X}_i^k}{n} \quad \text{for } k = 1, 2, \dots, m, \quad (6)$$

then  $p_x(\bar{\theta}_{(m)})$  converges in relative entropy to  $u(x)$ ,

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

see Theorem 1 of [Barron and Sheu \(1991\)](#).

Here, however, the sample  $\{\bar{X}_i\}_1^n$  is not available, instead we only observe  $\{\hat{X}_i\}_1^n$  and consequently have  $\hat{\theta}_{(m)}$  as the unique solution to

$$\int_0^1 x^k p_x(\hat{\theta}_{(m)}) dx = \frac{\sum_{i=1}^n \hat{X}_i^k}{n} \quad \text{for } k = 1, 2, \dots, m. \quad (7)$$

Note that the Eqs (5), (6) and (7) define one-to-one mappings between the sample space  $\Omega_{(m)} \in \mathbb{R}^m$  and the parameter space  $\Theta_{(m)} \in \mathbb{R}^m$  in the exponential family, see [Barndorff-Nielsen \(1978\)](#). We can therefore define three pairs of  $m$  dimensional parameter and statistics, respectively as  $\{\theta_{(m)} : \mu_{(m)}\}$ ,  $\{\bar{\theta}_{(m)} : \bar{X}_{(m)}\}$  and

$\{\hat{\theta}_{(m)} : \hat{X}_{(m)}\}$ , where  $\mu_{(m)} = \{\mu_k\}_{k=1}^m$ ,  $\bar{X}_{(m)} = \{n^{-1} \sum_{i=1}^n \bar{X}_i^k\}_{k=1}^m$  and  $\hat{X}_{(m)} = \{n^{-1} \sum_{i=1}^n \hat{X}_i^k\}_{k=1}^m$ . Generically these mappings can be expressed via

$$\{\theta_{(m)}^* : \{\mu_k^*\}_1^m\} \quad \text{where} \quad \int_0^1 x^k p_x(\theta_{(m)}^*) dx = \mu_k^*, \quad k = 1, \dots, m. \quad (8)$$

The uniqueness of these mappings can be exploited in the following Theorem, proved in Appendix A, to show that the density estimator  $p_x(\hat{\theta}_{(m)})$  converges in relative entropy at the same rate as  $p_x(\bar{\theta}_{(m)})$ .

**Theorem 1** *Let  $\hat{\theta}_{(m)}$  denote the estimated exponential parameter determined by (7) then under Assumption 1 and for  $m, n \rightarrow \infty$  with  $m^3/n \rightarrow 0$ ,*

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

According to Theorem 1, in terms of the density estimator, at least, the effect of observing  $\{\hat{X}_1, \dots, \hat{X}_n\}$  rather than  $\{X_1, \dots, X_n\}$  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 minimax rate of  $m_n^* = O(n^{-1/5})$ , since  $r \geq 2$  by assumption). The optimal rate the rate of convergence of the estimator remains of order  $O_p(n^{-\frac{2r}{1+2r}})$ . It should not be surprising that the rate of convergence is unaffected when parameters are replaced by  $\sqrt{n}$  consistent estimators. Theorem 1 thus generalises the results of [Crain \(1974\)](#) and [Barron and Sheu \(1991\)](#), as summarized in Lemma 1 of [Marsh \(2007\)](#), by permitting estimation of nuisance parameters as a preliminary step.

Additionally, we may recover the quantiles of  $Y$  from those implied by the approximating series density estimator. This is captured in the following Corollary, which follows immediately since convergence in relative entropy implies convergence in law.

**Corollary 1** *Let  $\hat{T}_{n,m} \in (0, 1)$  be a random variable having density function  $p_t(\hat{\theta}_{(m)})$  where  $\hat{\theta}_{(m)}$  is defined by (7), then*

$$\hat{T}_{n,m} \xrightarrow{\mathcal{L}} \bar{X},$$

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

## 2.2 Numerical application of a quantile estimator

The consequence of Corollary 1 is that 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

$$q_{F^{-1}(\hat{T}_{n,m}; \hat{\rho}_n)}(\pi) = q_Y(\pi) + o_p(1). \quad (9)$$

The following set of experiments compare the Mean Square Errors (MSE) of estimators for the quantiles of  $Y$  based on those of  $\hat{T}_{n,m}$  for  $m = 3, 9$  and for quantiles calculated at the probabilities,  $\pi = .05, .25, .50, .75, .95$ . We also compare the accuracy of estimated quantiles when unknown parameters are estimated against cases where they are not.

First suppose that  $Y_i \sim IID Y := t_{(4)}$  but we estimate the Gaussian likelihood implied by  $N(\mu, \sigma^2)$ . Define

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

the first obtained from the (misspecified) Gaussian model imposing zero mean and unit variance, and the second from the Gaussian model with estimated mean and variance.

Following the development above, as well as that of [Barron and Sheu \(1991\)](#), let  $\theta_{(m)}^*$  and  $\hat{\theta}_{(m)}$  and denote the estimated parameters for the exponential series density estimators for the samples  $\{X_i^*\}_1^n$  and  $\{\hat{X}_i\}_1^n$ , respectively. Let  $T_{n,m}^*$  have density  $p_t(\theta_{(m)}^*)$  (note that this is just straight forward application of the original set-up of [Barron and Sheu 1991](#)) and let  $\hat{T}_{n,m}$  have density  $p_t(\hat{\theta}_{(m)})$ , as in Corollary 1. The pairs of estimated quantiles for  $Y$  are then constructed as in

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

The MSE of these quantiles, for each probability  $\pi$ , are presented in Appendix B, Tables 1a for  $m = 3$  and 1b for  $m = 9$ .

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

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

Analogous to above let  $T_{n,m}^*$  and  $\hat{T}_{n,m}$  have densities  $p_t(\theta_{(m)}^*)$  and  $p_t(\hat{\theta}_{(m)})$  and so pairs of estimated quantiles for  $Y$  are constructed via,

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

The MSE of these quantiles, for each probability  $\pi$ , are presented in Table 1c ( $m = 3$ ) and 1d ( $m = 9$ ).

The consistency of the quantiles obtained from, in particular,  $\hat{T}_{m,n}$  is illustrated clearly in Table 1. More relevant, however, is that estimating the parameters of the fitted model as a preliminary step produces quantile estimators that can be superior, as the sample size becomes large, to those obtained by simply imposing parameter values, as can be clearly seen by comparing the right and left panels in Table 1. Note also that although the larger value of  $m$  yields more accurate quantile estimates in these cases, this is at some computational cost and, in other cases, potential numerical instability. Although this latter possibility is greatly mitigated, since the  $\{\hat{X}_i\}_{i=1}^n$  are bounded.

### 3 Consistent, asymptotically pivotal tests for goodness of fit

#### 3.1 Main results

Here we provide a test of the null hypothesis that the fitted likelihood is correctly specified as in (1). The previous section generalized the Barron and Sheu (1991) series density estimator and the resulting nonparametric likelihood ratio test then generalizes the test of Marsh (2007).

To proceed note that when  $H_0$  is true then in Assumption 1,  $\beta_* = \beta_0$  and in (2)  $\bar{X}_i = F(Y_i, \beta_0) \sim IIDU[0, 1]$ . Direct generalization of the principle in Marsh (2007) means that (1) can be tested via,

$$H_0 : \lim_{m \rightarrow \infty} \theta_{(m)} = 0_{(m)}, \quad (10)$$

in the exponential family (4), where  $\theta_{(m)}$  is the solution to (5) and  $0_{(m)}$  is an  $m \times 1$  vector of zeros.

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

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

The null hypothesis is rejected for large values of  $\hat{\lambda}_m$ .

Under any fixed alternative  $H_1 : G(y) \neq F(y; \beta_0)$  the distribution of  $\bar{X}_i = F_i(Y_i; \beta_*)$  will not be uniform, i.e.  $\theta_{(m)} \neq 0_{(m)}$ . 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,  $p_x(\theta_{(m)}^1)$ . In practice, of course,  $\theta_{(m)}^1$  will be neither specified nor known. The following Theorem, again proved in Appendix A, gives the asymptotic distribution of the likelihood ratio test statistic both under the null hypothesis (10) and also demonstrates consistency against any such fixed alternative.

**Theorem 2** Suppose that Assumption 1 holds, we construct  $\{\hat{X}_i\}_{i=1}^n$  as described in (2), and we let  $m, n \rightarrow \infty$  with  $m^3/n \rightarrow 0$ , then:

(i) Under the null hypothesis,  $H_0 : G(y) = F(y; \beta_0)$ ,

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

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

$$\Pr[\hat{\Lambda}_m \geq \kappa] \rightarrow 1.$$

Theorem 2 generalizes the test of Marsh (2007) establishing asymptotic normality and consistency against fixed alternatives when  $\beta$  has to be estimated. Via Claeskens and Hjort (2004) it is demonstrated that as  $n \rightarrow \infty$  with  $m^3/n \rightarrow 0$ , then the test  $\bar{\Lambda}_m$  (i.e. the, here, unfeasible test based on the notional sample  $\{\bar{X}_i\}_1^n$ ) has power against local alternatives parametrized by  $\theta_{(m)} - 0_{(m)} = c\sqrt{\frac{\sqrt{m}}{n}}$  with  $c'c = 1$ . Heuristically, implicit from the proof of Theorem 2 the properties of the test follow from:  $\hat{\Lambda}_m - \bar{\Lambda}_m = O_p\left(\sqrt{\frac{m}{n}}\right)$ , and so  $\hat{\Lambda}_m$  has power against that same rate of local-alternatives.

### 3.2 Testing for normality or exponentiality

The likelihood ratio test  $\hat{\Lambda}_m$  is asymptotically pivotal, specifically standard normal. 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

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

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

$$\hat{X}_i = 1 - e^{-Y_i/\bar{y}_n}, \tag{11}$$

to test  $H_0^E$ , and

$$\hat{X}_i = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{Y_i - \bar{y}_n}{\hat{\sigma}_n} \right) \right], \tag{12}$$

to test  $H_0^N$ .



Table 2 in Appendix B provides rejection frequencies for the tests constructed for values of  $m = 3, 5, 7, 9, 11, 17$ . The left hand panel of numbers correspond to testing  $H_0^E$  and the right to  $H_0^N$ , critical values at the 1, 5 and 10% significance level from the standard normal distribution are used throughout.

The purpose of these experiments is only to demonstrate that the finite sample performance of the tests clearly improves as both  $n$  and  $m$  increase, as predicted by Theorem 2(i). Note the use of three significance levels to better illustrate convergence for large values of both  $m$  and  $n$ .

Although competitor tests are not asymptotically pivotal (and therefore no comparisons under the null are made) instead Table 3 compares the 5% size corrected powers of two variants of the tests, with  $m = 3$  and  $m = 9$  with the three direct competitors for a single sample size of  $n = 100$ . Table 3a and b 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,

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

Table 3c, d and e, consider alternatives where the moments of the data are not correctly specified, i.e.

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

where  $1(\cdot)$  denotes the indicator function. These latter three alternatives represent simplistic variants of common types of misspecification in econometric or financial data, i.e. misspecification of a conditional mean, variance or the possibility of a break in the mean (here half way through the sample). Note that these models imply that (2) will not be IID on  $(0, 1)$ , but ergodicity implies the sample moments will still converge. Finally, Table 3f 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, unfeasible, test of Marsh 2007)). The right hand panel has the rejection frequencies for tests based on estimated values, i.e. using (11) and (12), respectively.

The outcomes in Table 3 imply the following broad conclusions. The nonparametric likelihood test based  $\hat{\Lambda}_3$  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  $\hat{\Lambda}_9$  are broadly comparable in terms of their rejection frequencies, although AD is perhaps on average slightly more powerful and CM less powerful.

### 3.3 Bootstrap critical values

The proposed tests require a choice of dimension,  $m$ . The results presented in Tables 2 and 3 suggest an inevitable compromise, larger values of  $m$  imply tests having size closer to nominal, while smaller values of  $m$  imply tests having greater power. In order to overcome this compromise we can instead consider the properties of these tests when bootstrap critical values are instead employed.

For these tests 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  $Y_i^b \sim IID F(y; \hat{\beta}_n)$  for  $i = 1, \dots, n$ .
2. Estimate, via ML,  $\hat{\beta}_n^b$  and construct  $\hat{X}_i^b = F(Y_i^b; \hat{\beta}_n^b)$  for  $i = 1, \dots, n$ .
3. Repeat 1 and 2  $B$  times, obtaining bootstrap versions of the test  $\hat{\Lambda}_3^b$ .
4. Order the  $\hat{\Lambda}_3^b$  so the bootstrap critical value at size  $\alpha$  is  $\kappa_B = \hat{\Lambda}_3^{[(1-\alpha)B/100]}$ .
5. Denote the indicator function  $\hat{I}_B^A = \begin{cases} 1 & \text{if } \hat{\Lambda}_3 > \kappa_\alpha^B \\ 0 & \text{if } \hat{\Lambda}_3 \leq \kappa_\alpha^B \end{cases}$ .

We then reject  $H_0$  if  $\hat{I}_B^A = 1$ . First, however, the required asymptotic justification for the bootstrap is automatic given that  $\hat{\Lambda}_m \rightarrow_d N(0, 1)$  giving the following corollary to Theorem 2.

**Corollary 2** *Under Assumption 1 and if  $n, m \rightarrow \infty$  with  $m^3/n \rightarrow 0$ , then*

$$\begin{aligned} i) \Pr \left[ \hat{I}_B^A = 1 \mid H_0 \right] &\rightarrow \alpha, \\ ii) \Pr \left[ \hat{I}_B^A = 1 \mid H_1 \right] &\rightarrow 1. \end{aligned}$$

Here we will compare the performance of bootstrap critical values for  $\hat{\Lambda}_3$  with those of CM and AD by repeating many of the experiments of Kojadinovic and Yan (2012). 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;

$$\begin{aligned} N^* &\sim N(10, 1), \quad L^* \sim L(10, 0.572), \\ \Gamma^* &\sim \Gamma(98.671, 1/9.866) \quad \& \quad W^* \sim W(10.618, 10.452). \end{aligned} \quad (13)$$

The specific parameter values for  $L^*$ ,  $\Gamma^*$  and  $W^*$  are chosen to minimize relative entropy ( $I(\beta)$  in Assumption 1(iii)) for each family to the distribution of  $N^*$ . Sample sizes of  $n = 25, 50, 100, 200$  are used in the experiments described below.

Table 4a 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 4b and c contain the finite sample rejection frequencies under various alternative hypotheses, covering all pairwise permutations of the distributions in (13). 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 Conclusions

This paper has generalized the series density estimator of Barron and Sheu (1991) to cover the case where parameters are estimated in the context of misspecified models. The nonparametric likelihood ratio tests of Marsh (2007) can be thus extended to cover the case of estimated parameters. The general aim has been to provide a testing 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 and they have power advantages over edf tests, whether critical values are size corrected or obtained by a consistent bootstrap. This suggests the proposed tests will be much simpler to generalize to the settings of Bai (2003) or Corradi and Swanson (2006). Finally, in the event of rejection, the series density estimator upon which the tests are built may be employed to consistently estimate the quantiles of the density from which the sample is taken.

**Acknowledgements** This article was funded by University of Nottingham.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## A Appendix A: Proofs

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

$$a_{n,m} = O(b_{n,m}) \iff \lim_{m,n \rightarrow \infty} \frac{a_{n,m}}{b_{n,m}} \leq c_1 < \infty, \quad ; \quad m^3/n \rightarrow 0$$

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

**Proof of Theorem 1:**

First recall the definitions note ,

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

The Euclidean distance between the two polynomial sufficient statistics satisfies,

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

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

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

Since  $\bar{X}_i \in (0, 1)$  while, as in (3),  $e_i = O_p(n^{-1/2})$  and  $e_i \in (-1, 1)$  then,

$$\frac{j!}{s!(j-s)!} \bar{X}_i^{j-s} e_i^s \leq \frac{j^s}{s!} c_1^{j-s} e_i^s = \frac{j^s}{s!} c_1^{j-s} O_p(n^{-s/2}), \quad (14)$$

where  $c_1 < 1$ . For finite  $j$  (14) is  $O_p(n^{-s/2})$  while as  $j \rightarrow \infty$  (14) is  $o(1) O_p(n^{-s/2})$  and so,

$$\sup_{j \in \mathbb{N}} \frac{j!}{s!(j-s)!} \bar{X}_i^{j-s} e_i^s = O_p(n^{-s/2}),$$

implying that

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

uniformly in  $j$ , and hence,

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

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

$$\left| \hat{X}_{(m)} - \bar{X}_{(m)} \right| = \sqrt{\sum_{j=1}^m \left( \frac{1}{n} \sum_{i=1}^n (\hat{X}_i^j - \bar{X}_i^j) \right)^2} = O_p \left( \sqrt{\frac{m}{n}} \right). \quad (15)$$

Consider now  $\mu_{(m)}$ , then from the triangle inequality,

$$\left| \hat{X}_{(m)} - \mu_{(m)} \right| \leq \left| \bar{X}_{(m)} - \mu_{(m)} \right| + \left| \hat{X}_{(m)} - \bar{X}_{(m)} \right| = O_p \left( \sqrt{\frac{m}{n}} \right), \quad (16)$$

which follows from (15) 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  $(\bar{X}_i^j)_1^n$  were observed directly.

We thus have  $\left| \hat{X}_{(m)} - \mu_{(m)} \right| = O_p \left( \sqrt{\frac{m}{n}} \right)$  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,

$$\begin{aligned} E_U \left[ \ln \left( \frac{u(x)}{p_x(\hat{\theta}_{(m)})} \right) \right] &= E_U \left[ \ln \left( \frac{u(x)}{p_x(\theta_{(m)})} \right) \right] + E_U \left[ \ln \left( \frac{p_x(\theta_{(m)})}{p_x(\bar{\theta}_{(m)})} \right) \right] \\ &\quad + E_U \left[ \ln \left( \frac{p_x(\bar{\theta}_{(m)})}{p_x(\hat{\theta}_{(m)})} \right) \right]. \end{aligned} \quad (17)$$

Given that Assumption 1 assures the required conditions of Barron and Sheu (1991, Theorem 1) are met then the first two terms in (17) 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 Eqs (6) and (7), 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.  $\square$

### 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  $n, m \rightarrow \infty$ , but  $m^3/n \rightarrow 0$ . For notational convenience and comparisons with [Portnoy \(1988\)](#) and [Barron and Sheu \(1991\)](#), expressions involving  $\theta_{(m)}$  will not be immediately resolved.

*Part (i):* To proceed we have defined,

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

where  $\hat{\theta}_{(m)}$  solves (7), 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  $0_{(m)}$  defines,

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

The exponential log-likelihood is strictly convex so that the mapping,  $\psi'_m \left( \theta_{(m)} \right) = \mu_{(m)}$  is one-to-one between the parameter space  $\Theta_m \subset \mathbb{R}^m$  and sample space  $\Omega_m \subset \mathbb{R}^m$ , similar to (8). Application of [Barron and Sheu \(1991, eq. 5.6\)](#) and also (16) thus gives,

$$O_p \left( \left| \hat{\theta}_{(m)} - 0_{(m)} \right| \right) = O_p \left( \left| \hat{X}_{(m)} - \mu_{(m)} \right| \right) = O_p \left( \sqrt{\frac{m}{n}} \right). \quad (18)$$

As a consequence of both (18) and (16) we have that,

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

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, here  $(\bar{\theta}_{(m)}, 0_{(m)})$  and  $(\bar{X}_{(m)}, \mu_{(m)})$ .

To continue, noting expectations under the null hypothesis can be written here as  $E_U[\cdot]$  since  $\bar{X} \sim U := U[0, 1]$ , the uniform distribution with density  $p_{0(m)}(x) = 1$ , we then 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 (19)

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

(20)

Subtracting (20) from (19) 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,

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

(21)

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

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

(22)

Now define the  $m \times 1$  random variable  $V_m = \psi_m''(0_{(m)})^{-1/2} (\bar{x} - \psi_m'(0_{(m)}))$ , having density  $p_V(\theta_{(m)}^V)$ , so that  $E[V] = 0_{(m)}$  and  $\text{Var}[V_m] = I_m$ . Since the likelihood ratio statistic is parameterization invariant the likelihood ratio test based on observations on  $V_m$  would be identical to that based on  $\bar{X}_{(m)}$ . Rather than defining a new triple of values, analogous to those in (5), (6) and (7), in both the parameter space  $\Theta_m$  (note that in particular the hypothesized value would no longer satisfy  $\theta_{(m)} = 0_{(m)}$ ) and sample space  $\Omega_m$  we will instead, and without any loss of generality assume a parameterization in which both  $E[\bar{X}_{(m)}] = 0$  and  $\text{Var}[\bar{X}_{(m)}] = I_m$ . Note, however, that it is the unobserved  $\bar{X}$  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 (22)) and then via (21) for  $\hat{\Lambda}_m = \frac{\hat{\lambda}_m - m}{\sqrt{2m}}$  follows exactly as in Portnoy (1988, Theorem 4.1).  $\square$

Part (ii): Under any fixed alternative the density of  $\bar{X}_i = F(Y_i; \beta_*)$  is

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

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

$$\int_0^1 x^j p_h(\theta_{(m)}^1) dx = \int_0^1 h^j u_1(x) dx \quad ; \quad j = 1, \dots, m. \quad (23)$$

The uniqueness of solutions to (23) imply  $\theta_{(m)}^1 \neq 0_{(m)}$ .

To take the least favorable case, define

$$\theta_{(m)}^1 = (\theta_1^1, \theta_2^1, \dots, \theta_m^1)'$$

and suppose that  $\theta_k^1 \neq 0$  for some finite  $k$  but that  $\theta_j^1 = 0$  for all  $j \neq k$ . The series density estimator is consistent for  $\theta_{(m)}^1$ , under  $H_1$ , in that  $|\hat{\theta}_{(m)} - \theta_{(m)}^1| = O_p\left(\sqrt{\frac{m}{n}}\right)$ , analogous to (18) above, and so we can write,

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

We can therefore write the likelihood ratio as

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

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

Thus, under  $H_1$ , we can write

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

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

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



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

$$\psi_m(\theta_{(m)}^1) - \psi_m(\theta_{(m)}^0) \neq 0,$$

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,

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

since  $m^3/n \rightarrow 0$  and hence  $\Pr[\hat{A}_m > \kappa] \rightarrow 1$ , as required.  $\square$

## B Appendix B: Tables

See Tables 1, 2, 3, and 4.

**Table 1** Mean Square Errors of Quantiles

$n$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$
$\pi$	25	50	100	200	25	50	100	200
<b>a:</b> MSE of estimated quantiles for $Y \sim t_{(4)}, m = 3$ .								
.05	.1266	.0906	.0774	.0718	.1309	.0731	.0506	.0389
.25	.0450	.0227	.0127	.0066	.0446	.0218	.0119	.0058
.50	.0397	.0183	.0101	.0049	.0348	.0159	.0087	.0042
.75	.0442	.0222	.0126	.0074	.0445	.0215	.0117	.0066
.95	.1293	.0976	.0768	.0693	.1333	.0806	.0505	.0375
<b>b:</b> MSE of estimated quantiles for $Y \sim t_{(4)}, m = 9$ .								
.05	.1408	.1270	.1179	.1159	.1060	.0551	.0354	.0240
.25	.0308	.0187	.0139	.0116	.0513	.0244	.0138	.0064
.50	.0175	.0089	.0053	.0023	.0411	.0182	.0105	.0052
.75	.0313	.0194	.0138	.0113	.0512	.0253	.0123	.0065
.95	.1441	.1271	.1186	.1155	.1077	.0599	.0344	.0228
<b>c:</b> MSE of estimated quantiles for $Y \sim \Gamma(1.2, 1), m = 3$ .								
.05	.0025	.0014	.0012	.0012	.0035	.0012	.0005	.0004
.25	.0157	.0083	.0046	.0027	.0202	.0099	.0049	.0022
.50	.0451	.0235	.0127	.0077	.0563	.0246	.0120	.0061
.75	.0890	.0487	.0279	.0188	.1381	.0601	.0348	.0172
.95	.3088	.2109	.1644	.1449	.0538	.2364	.1378	.0709

**Table 1** continued

$n$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$q_{Y\Gamma}^*$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$	$\hat{q}_{Y\Gamma}$
$\pi$	25	50	100	200	25	50	100	200
<b>d:</b> MSE of estimated quantiles for $Y \sim \Gamma(1.2, 1)$ , $m = 9$ .								
.05	.0006	.0004	.0003	.0003	.0002	.0001	.0001	.0000
.25	.0094	.0065	.0051	.0042	.0033	.0017	.0009	.0004
.50	.0494	.0375	.0335	.0310	.0156	.0073	.0037	.0019
.75	.2085	.1676	.1583	.1532	.0626	.0276	.0146	.0078
.95	.4456	.3352	.2380	.1395	.2180	.1093	.0527	.0267

**Table 2** Sizes of tests for both  $H_0^E$  and  $H_0^N$  for different  $m$  and  $n$ 

$\alpha$	$H_0^E$			$H_0^N$		
	.10	.05	.01	.10	.05	.01
$m$						
$n = 25$						
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
$n = 50$						
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
$n = 100$						
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
$n = 200$						
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 3** Rejection frequencies under various alternatives. The left hand panels corresponds parameter values imposed, while for the right had panels they are estimated

<b>a</b> 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
$\hat{\Lambda}_3$	.935	.705	.386	.267	.114	.605	.294	.166	.127	.097
$\hat{\Lambda}_9$	.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
<b>b</b> Power $H_0 : Y_i \sim N(0, 1)$ vs. $H_1 : Y_i \sim \chi_{(v)}^2 - v$ .										
$v$	12	20	28	36	44	12	20	28	36	44
$\hat{\Lambda}_3$	.859	.660	.577	.476	.422	.572	.274	.189	.146	.114
$\hat{\Lambda}_9$	.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
<b>c</b> 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
$\hat{\Lambda}_3$	.694	.592	.386	.161	.093	.902	.736	.510	.271	.101
$\hat{\Lambda}_9$	.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
<b>d</b> Power $H_0 : Y_i \sim N(0, 1)$ vs. $H_1 : Y_i \sim N(0, 1 + vY_{i-1}^2)$ .										
$v$	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
$\hat{\Lambda}_9$	.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
<b>e</b> Power $H_0 : Y_i \sim N(0, 1)$ vs. $H_1 : Y_i \sim N(v1_{t> T/2 }, 1)$ .										
$v$	0.9	0.7	0.5	0.3	0.1	0.9	0.7	0.5	0.3	0.1
$\hat{\Lambda}_3$	.754	.563	.349	.196	.079	.653	.495	.274	.141	.080
$\hat{\Lambda}_9$	.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 3** continued

<b>f:</b> Power $H_0 : Y_i \sim \text{Exp}[1]$ 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
$\hat{A}_3$	.115	.121	.238	.305	.428	.191	.298	.585	.769	.866
$\hat{A}_9$	.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 4** Rejection frequencies at 5% level under **a** the respective null hypotheses **b, c** various alternatives

<b>a:</b> Rejection frequencies at 5% level under the respective null hypotheses				
$n$	25	50	100	200
(i) $H_0^N : Y \sim N^*$				
$\hat{A}_3$	.064	.065	.058	.044
CM	.064	.058	.057	.054
AD	.060	.059	.062	.058
(ii) $H_0^\Gamma : Y \sim \Gamma^*$				
$\hat{A}_3$	.062	.056	.049	.046
CM	.068	.060	.065	.061
AD	.065	.055	.052	.061
(iii) $H_0^W : Y \sim W^*$				
$\hat{A}_3$	.067	.055	.055	.047
CM	.063	.058	.056	.058
AD	.055	.066	.065	.057
(iv) $H_0^L : Y \sim L^*$				
$\hat{A}_3$	.065	.062	.050	.042
CM	.071	.066	.059	.055
AD	.062	.054	.055	.055

**Table 4** continued

<b>b: Rejection frequencies at 5% level under various alternatives</b>				
<i>n</i>	25	50	100	200
(i) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim \Gamma^*$				
$\hat{A}_m$	.069	.088	.116	.175
CM	.078	.094	.123	.185
AD	.069	.090	.108	.161
(ii) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim N^*$				
$\hat{A}_m$	.068	.085	.099	.129
CM	.055	.066	.079	.088
AD	.076	.085	.092	.113
(iii) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim W^*$				
$\hat{A}_m$	.196	.364	.584	.897
CM	.094	.192	.465	.776
AD	.183	.315	.550	.806
(iv) $H_0 : Y \sim W^*$ vs. $H_1 : Y \sim N^*$				
$\hat{A}_m$	.101	.164	.351	.690
CM	.107	.233	.388	.580
AD	.098	.164	.334	.602
(v) $H_0 : Y \sim N^*$ vs. $H_1 : Y \sim L^*$				
$\hat{A}_m$	.173	.249	.393	.458
CM	.111	.152	.212	.358
AD	.131	.190	.246	.417
(vi) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim N^*$				
<i>n</i>	25	50	100	200
$\hat{A}_m$	.046	.055	.065	.101
CM	.036	.054	.073	.109
AD	.041	.046	.070	.108

**Table 4** continued

c: Rejection frequencies at 5% level under various alternatives				
$n$	25	50	100	200
(i) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim L^*$				
$\hat{A}_m$	.091	.122	.188	.253
CM	.078	.081	.122	.193
AD	.105	.128	.174	.257
(ii) $H_0 : Y \sim \Gamma^*$ vs. $H_1 : Y \sim W^*$				
$\hat{A}_m$	.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
$\hat{A}_m$	.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
$\hat{A}_m$	.172	.327	.620	.867
CM	.215	.343	.542	.797
AD	.159	.277	.500	.816
(v) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim \Gamma^*$				
$\hat{A}_m$	.059	.082	.120	.152
CM	.059	.081	.130	.161
AD	.051	.059	.101	.148
(vi) $H_0 : Y \sim L^*$ vs. $H_1 : Y \sim W^*$				
$\hat{A}_m$	.243	.343	.592	.892
CM	.124	.241	.519	.882
AD	.204	.325	.583	.912

## References

- Babu G, Rao CR (2004) Goodness of fit tests when parameters are estimated. *Sankhya* 66:63–74
- Bai J (2003) Conditional distributions of dynamic models. *Rev Econ Stat* 85:531–549
- Barron AR, Sheu C-H (1991) Approximation of density functions by sequences of exponential families. *Ann Stat* 19:1347–1369
- Barndorff-Nielsen O (1978) Information and exponential families in statistical theory. Wiley, New York
- Beran R (1988) Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J Am Stat Assoc* 83:687–697
- Claeskens G, Hjort NL (2004) Goodness of fit via nonparametric likelihood ratios. *Scand J Stat* 31:487–513
- Conover WJ (1999) Practical nonparametric statistics. Wiley, New York
- Corradi V, Swanson NR (2006) Predictive Density Evaluation. In: G. Elliott, C. Granger and A. Timmermann (eds.) *Handbook of Economic Forecasting*, Vol. 1, Elsevier, pp 197–284
- Crain BR (1974) Estimation of distributions using orthogonal expansions. *Ann Stat* 2:454–463

- Kojadinovic I, Yan J (2012) Goodness-of-fit testing based on a weighted bootstrap: a fast large-sample alternative to the parametric bootstrap. *Can J Stat* 40:480–500
- Marsh P (2007) Goodness of fit tests via exponential series density estimation. *Comput Stat Data Anal* 51:2428–2441
- Portnoy S (1988) Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann Stat* 16:356–366
- Stephens MA (1976) Asymptotic results for goodness of fit statistics with unknown parameters. *Ann Stat* 4:357–369
- White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25