Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Hausman type tests for nonparametric likelihood

by

Patrick Marsh

Granger Centre Discussion Paper No. 15/03

The University of Nottingham

Hausman type tests for nonparametric likelihood

Patrick Marsh

School of Economics and Granger Centre for Time Series Econometrics

Sir Clive Granger Building

University of Nottingham

University Park

NG7 2RD

email: patrick.marsh@nottingham.ac.uk

September 2015

Abstract

A nonparametric likelihood ratio test is derived via a consistent exponential series density estimator. The null hypothesis (usually a parametric density) neither needs to be a member of the exponential family nor completely specified, leading to a nonparametric likelihood ratio test based on the Hausman-type testing principle; comparing two density estimators both of which are consistent under the null but only one of them being so under the alternative. A comprehensive asymptotic theory for the test is provided, namely provision of a central limit theorem under the null and consistency under a fixed alternative. In a series of experiments, based on mixtures of normal and exponential distributions, the size and power properties of the test are examined.

1 Introduction

Methods for assessing the adequacy, or predictive ability, of a fitted model have predominantly been built upon the goodness-of-fit techniques applied to the empirical distribution of some probability transform of the data. Corradi and Swanson (2006) provide an excellent review of the available techniques. Predominantly the suggested tests are of the Kolmogorov-Smirnov or Cramer von Mises type (see Anderson and Darling (1952)) although an interesting alternative in this literature is the Kernel density estimator approach of Hong and Li (2003).

In the context of the goodness-of-fit and two sample problems Marsh (2007 and 2010) has instead proposed procedures which employ the nonparametric series density estimator of Barron and Sheu (1991). The essential idea is to seek a density estimator which embeds within it a general class of plausible null densities. Within the literature, and noting this constraint, there seem to be two possible approaches to achieving this, either the empirical likelihood approach, see Owen (1988) and Qin and Lawless (1994), or as here, an exponential series estimator. The advantages of this approach are threefold. First, the resultant density estimator is entropy minimising, see also Csiszar (1975) and Robinson (1991). Second, because in the limit the approximating density is an infinite dimensional exponential, it embeds precisely the kind of parametric densities we would wish to test for. Finally, because the estimation routine is (in the limit infinite dimensional) maximum likelihood, the natural test statistic to employ is the likelihood ratio. Consequently, standard optimality properties of the test carry through from the parametric to the nonparametric cases.

The approach of this paper differs from the papers of Marsh (2007 and 2010) in the way in which the null hypothesis (of correct specification) is imposed. In the latter papers imposition of the null involves specification of (in the limit) an infinite dimensional parameter vector. Here, instead, the null is imposed through moment restrictions on the data that the constrained series density must satisfy. The result is a density estimator which is consistent only when the null hypothesis is true, but is inconsistent otherwise.

The proposed test of this paper therefore takes the form of a likelihood ratio of two series density estimators. One of these (the constrained) is consistent (and relatively efficient) under the null of correct specification while the other, unconstrained estimator, is consistent generally, but inefficient under the null. It is in this respect the test is said to be of 'Hausman-type' utilizing exactly the principle introduced in Hausman (1978).

Specifically, the estimation procedure is set up in such a way, so that consistency, albeit weak, is demonstrated through convergence of functions of the sample. This occurs because the estimation routine equates the moments of the approximating density with those from the sample. Moreover, these functions themselves have specific meaning in regard to the properties of the population, being as they are simple functions of the sample moments. As a consequence, any hypothesised null density may be characterised by the number of these functions required for consistency. For example, if the density to be estimated is an infinite dimensional exponential, then infinitely many sample moments are require, but for a finite, of dimension, say m', exponential only the first m' are required, with the remaining moments restricted to be known functions of those first m'. The test statistic is then simply the ratio of the maximised unconstrained likelihood with that of the constrained likelihood.

The asymptotic distribution of the test, upon correct standardisation, is shown to be the standard normal via modification of the results contained in Portnoy (1988). The subtle departure of the analysis here, is that the population distribution is not assumed to be a member of the infinite dimensional exponential family, nor do we have to fully restrict the parameters under the null. Again, the limit theorem for the distribution of the test is demonstrated through that of the standardised functions of the sample moments.

A detailed analysis of the test, in terms of its finite sample size and power properties is given in the main body of this paper (Section 6). In common with other nonparametric tests we find size distortions, although the power of the test (based on simulated critical values) seems quite impressive. Although artificial, in that discrimination between two simple densities (the normal and exponential) and mixtures of the two is considered, the procedure seems reliable in comparison with naive graphical or histogram based discrimination.

At present, the conditions required for the proofs of both consistency and the limit theorem, are somewhat restrictive. Essentially, the sample must be independent and identically distributed copies of a random variable with bounded support. However, the density estimator itself has been successfully applied, see Barron and Sheu (1991), and has proved most useful when a tractable, analytic estimate is required, for example as in Chesher (1991), rather than a kernel based one. Extensions to, for instance, the regression case, although not trivial, should follow, with consistency and asymptotic normality demonstrated via similar techniques used here. However, the results of this paper may be used to test hypotheses on the residuals from some regression, under the maintained assumption of independence. Moreover, since we obtain, as a density estimate, a member of the exponential family, higher-order theory, analogous to that for empirical likelihood (see Diciccio, Hall and Romano (1991)) and for the kernel estimator (see Fan and Linton (1997)), should follow.

The plan for the rest of the paper is as follows. Section 2 outlines the estimation procedure while its consistency is detailed in Section 3. Section 4 provides a central limit theorem, which is utilised in Section 5 to prove asymptotic normality of the nonparametric likelihood ratio test. Section 6 contains details of the numerical experiments, while Section 7 concludes and an appendix contains the proofs of the three main theorems as well as several figures used in the numerical analysis.

2 The Exponential Series Estimator

Here, we present the series estimation technique and detail its properties. We suppose that we wish to estimate the density of a scalar random variable X, with bounded sample space $\Omega_X = [a, b]$ and density $p(x) = dP(x) : \{\Omega_X \to \mathbb{R}\ , . \int_{\Omega_X} dP(x) = 1,$ $p(x) \geq 0\}$. Without loss of generality we assume $a = 0$ and $b = 1$. Following Barron and Sheu (1991), the density estimator is a member of the exponential family

$$p_\theta(x) = p_0(x) \exp\left\{\sum\nolimits_{k=1}^{m} \theta_k \phi_k(x) - \varphi_m(\theta)\right\}, \tag{1}$$

where in (1) the cumulant function is defined by

$$\varphi_m(\theta) = \log \int_{\Omega_X} p_0(x) \exp\left\{\sum\nolimits_{k=1}^{m} \theta_k \phi_k(x)\right\} dx. \tag{2}$$

In (2), $\theta = (\theta_1, .., \theta_m)' \in \mathbb{R}^m$ and $p_0(x)$ is a reference probability density function on $[0, 1]$ and the $\phi_k(x)$ are a set of linearly independent functions, which form an orthogonal basis for some linear space, S_m. Although the choice of S_m is somewhat arbitrary, popular choices being polynomials, trigonometric (and/or exponential) series and splines, we will be consider only the polynomial case. None of the theoretical results derived below depend upon this choice.

Implementation of the density estimator proceeds as follows. Given an independent sample, $(x_1, ..., x_n)$, the estimator, $p_{\hat{\theta}}(x)$ is defined as the Maximum Likelihood Estimator (MLE) in the family (1). That is

$$\hat{\theta} : \max_\theta L(\theta) = \prod\nolimits_{i=1}^{n} p_\theta(x_i),$$

or in terms of the log-likelihood

$$\hat{\theta} : \max_\theta l(\theta) = \ln[p_0(x)] + \sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{m} \theta_k \phi_k(x_i) - n\varphi_m(\theta). \tag{3}$$

From now on, to save notation, all sums over i run from 1 to n, while those over k run from 1 to m and all integrals are over the sample space, $[0, 1]$. From (3) some properties are immediately obtainable. First, the score:

$$l^k(\theta) = \sum\nolimits_i \phi_k(x_i) - n\varphi_m^k(\theta), \tag{4}$$

where the superscript(s) indicates, with respect to which variable(s) we are deriving. From (2), we have

$$\begin{aligned}\varphi_m^k(\theta) &= \frac{\frac{d}{d\theta_k}\int p_0(x)\exp\{\sum_{k=1}^m \theta_k\phi_k(x)\}\,dx}{\int p_0(x)\exp\{\sum_{k=1}^m \theta_k\phi_k(x)\}\,dx} \\ &= \int \phi_k(x)p_{\hat{\theta}}(x)dx,\end{aligned}$$

and hence the MLE is simply the solution to the m estimating equations,

$$\int \phi_k(x)p_\theta(x)dx = \frac{1}{N}\sum\nolimits_i \phi_k(x_i), \quad k = 1, ..., m. \tag{5}$$

Likewise we can calculate the Hessian, the second derivative of the log-likelihood, which is

$$\begin{aligned}l^{j,k}(\theta) &= -n\varphi_m^{j,k}(\theta) = \frac{d}{d\theta_k}\int \phi_j(x)p_\theta(x)dx \\ &= -n\left(\int \phi_j(x)\phi_k(x)p_\theta(x)dx - \int \phi_j(x)p_\theta(x)dx\int \phi_k(x)p_\theta(x)dx\right).\end{aligned} \tag{6}$$

Hence, using standard likelihood results, and if m were fixed, we would obtain the usual asymptotic result

$$\left(n\varphi_m^{j,k}(\theta)\right)^{1/2} l^k(\theta) \to_d N_m(0, I_m). \tag{7}$$

Likewise, from (7) a Central Limit Theorem (CLT) for the MLE, $\hat{\theta}$, itself may be obtained. However, in this paper we require $m \to \infty$ as $n \to \infty$, hence standard asymptotic results are not applicable. Moreover, we will not assume $p(x)$ is an exponential density and so the estimated parameters will not necessarily have any particular statistical meaning. As mentioned above, in this paper we concentrate on the case where the $\phi_k(x)$ are polynomials, specifically the simplest case where,

$$\phi_k(x) = x^k.$$

Considering the estimating equations (5), and defining $E_{p_\theta}[.]$ as expectations with respect to the series exponential family, we have

$$E_{p_{\hat{\theta}}}[X^k] = \frac{1}{N}\sum\nolimits_i x_i,$$

or in general $E_{p_{\hat{\theta}}}[\phi_k(x)] = \frac{1}{N}\sum_i \phi_k(x_i)$. That is (5) equates 'moments' from the family $p_\theta(x)$ with the sample moments, alternatively the non-parametric MLE chooses as the density estimate, the member of $p_\theta(x)$ having moments $\frac{1}{N}\sum_i \phi_k(x_i)$. Moreover, since the sample space is bounded, then, with respect to the dominating measure $dP(x)$,

$$\begin{aligned} E_{dP(x)}[\phi_k(X)] &= \int \phi_k(x)dP(x),\ \forall k \\ &= \mu_k < \infty, \end{aligned}$$

so that the moments of x are themselves bounded. Consequently, for the purposes of the asymptotic analysis to follow, the properties of the density estimator will be related to those of the sample moments, in particular Laws of Large Numbers (LLN) and CLTs. In summary then, letting $\bar{X} = \frac{1}{N}\sum_i (\phi_1(x_i), .., \phi_m(x_i))'$ and $\phi = (\phi_1(x), .., \phi_m(x))'$we have the set of m estimating equations

$$\int \phi\, p_{\hat{\theta}}(x)dx = \bar{X}, \tag{8}$$

and the 'true' moments $\mu = (\mu_1, .., \mu_m)'$. In the following section we demonstrate a weak LLN for $\bar{X}$.

3 Consistency of the Procedure

Unlike in the Barron and Sheu (1991) analysis, we are directly interested in the parameters $\hat{\theta}$, in the sense that any hypothesis we test will take the form of a (profile) likelihood ratio test, and therefore will involve a statistical measure of the distance between $\hat{\theta}$, and any other point in $\mathbb{R}^m$. Here we will demonstrate consistency (in a functional norm sense) via a LLN for the sample moments.

Let $y_{k,i} = \phi_k(x_i) - \mu_k$ determine the deviation of the sample moments from their (true, but unknown) expectations. Obviously, since $k = 1, ..., m$ and $m \to \infty$, strong laws for $\bar{X}$ are precluded. However, a strong law can be established for each $\phi_k(x_i)$, and then letting m become infinite, a weak law will follow. First note the following Lemma.

Lemma 1 *Hajek-Renyi (1950) Inequality*

Let $\{Y_i\}_{i=1}^n$ *denote a sequence of independent random variables, such that* $E[Y_i] = 0$ *and* $var[Y_i] < \infty$. *If* $c_1, ..., c_n$ *is a non-increasing sequence of positive constants, then for any positive integers* r, s, *with* $r < s$ *and some arbitrary* $\varepsilon > 0$

$$\Pr[\max_{r<i<s} c_t|Y_1 + .. + Y_i| \geq \varepsilon] \leq \frac{1}{\varepsilon^2}\left(c_s^2 \sum\nolimits_{i=1}^{r} var[Y_i] + \sum\nolimits_{i=r+1}^{s} c_t^2 var[Y_i]\right). \quad \blacksquare \quad (9)$$

For a proof of the Hajek-Renyi Inequality see, for example, Rao (1973), problem (3.3). Direct application of the inequality leads to the following theorem.

Theorem 1 *Let* $|.|$ *denote the Euclidean norm in* $\mathbb{R}^m$, *and given an* $\epsilon > 0$, *then under the conditions of this paper, and if* $m/n \to 0$, *then*

$$\lim_{n,m\to\infty} \Pr\left[\left|\bar{X} - \mu\right| > \epsilon\right] = 0. \quad \blacksquare \quad (10)$$

Theorem 1 establishes that

$$\bar{X} \to_p \mu,$$

with convergence (in probability) obtained at a rate $O(m/n)$. Alternatively, in light of their definitions and equation (8), we have

$$\int \phi p_{\hat{\theta}}(x)dx - \int \phi dP(x) = O\left(\frac{m}{n}\right),$$

where $\phi = \{\phi_1(x), ..., \phi_m(x)\}'$, if $m/n \to 0$. Consequently, applying Jensen's inequality and the Lebesgue dominated convergence theorem we establish

$$\int \left(\log[p_{\hat{\theta}}(x)] - \log[p(x)]\right) dP(x) = O\left(\frac{m}{n}\right),$$

or, with respect to the dominating measure $dP(x)$,

$$\left(\log[p_{\hat{\theta}}(x)] - \log[p(x)]\right) = O_p\left(\frac{m}{n}\right), \quad (11)$$

that is convergence of the log-densities is obtained at this rate.

4 A Central Limit Theorem for the Sample Moments

Any hypothesis test for the validity of restrictions imposed on any estimator of θ is going to be based on the stochastic difference implied in equation (11). In particular this difference will take the form of a log-likelihood ratio. However, there are some technical aspects which need to be considered before applying any statistical procedure. First, note that (11) is a weak, not a strong law, convergence is obtained in probability only. In fact, unless $p(x)$ is itself a (finite) member of the exponential family (1) then, even asymptotically, the series log-density estimator does not converge point-wise to the true log-density. Second, in the limit the density estimator is a member of an infinite dimensional exponential family, and hence the number of parameters to be estimated diverges. Consequently, standard central limit theorems do not apply. However, Portnoy (1988) details the analysis of estimation and testing in exponential families as the number of parameters diverges. In that paper though, it is tacitly assumed that the true density is itself a member of this family. Here, we are unable to make such an assumption. Therefore, some subtle refinements of the Portnoy (1988) analysis are required. To begin, as in the latter paper, we require the following central limit theorem for martingale differences.

Lemma 2 *(Theorem 9.3.1, Chow and Teicher (1988))*
Let R_n be a sequence of martingale differences, with associated sigma-field $\mathcal{F}_n$, such that $E_{\mathcal{F}_n}[R_{n+1}] = E[R_{n+1}|\mathcal{F}_n] = 0$. Let $S_n = \sum_{i=1}^n R_i$ and $s_n = \sum_i \sigma_i^2$, where $\sigma_i^2 = E[(S_i - S_{i-1})^2] = E[R_i^2]$, then if R_n satisfies the Lindeberg condition

$$\sum_i E\left[|R_i|^3\right] = o(s_n^{-3}), \tag{12}$$

and also that the conditional variances are bounded, so that

$$\sum_i E\left[\left|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right] - \sigma_i^2\right|\right] = o(s_n^{-2}) \tag{13}$$

then

$$\frac{S_n}{s_n} \to_d N(0,1). \quad \blacksquare$$

Returning to the set of estimating equations, we have the score vector

$$l^k(\theta) = \int \phi_k(x) p_\theta(x) dx - \frac{1}{n} \sum\nolimits_i \phi_k(x_i),$$

having variance

$$var[l^k(\theta)] = var[\frac{1}{n} \sum\nolimits_i \phi_k(x_i)] = \left(n\varphi_m^{j,k}(\theta)\right)^{-1}.$$

Consequently we let $T_m = \left(n\varphi_m^{j,k}(x)\right)^{1/2}$, $Y_i = (\phi_1(x_i) - \mu_1, ..., \phi_m(x_i) - \mu_m)'$, and finally define

$$Z_i = T_m Y_i \text{ and } \bar{Z} = \frac{\sum_i Z_i}{n} = T_m(\bar{X} - \mu).$$

In this set up we apply Lemma 2 directly to the sum of squared elements of $\bar{Z}$, giving the following Theorem.

Theorem 2 *Let $C_n = n^2 \bar{Z}'\bar{Z} - nm$, then*

a) C_n is a martingale and

b) if $m/n \to 0$ as $m, n \to \infty$

$$\frac{C_n}{n\sqrt{2m}} = \frac{n\bar{Z}'\bar{Z} - m}{\sqrt{2m}} \to_d N(0,1). \qquad \blacksquare$$

Theorem 2 thus establishes a central limit theorem for the standardised sum of square elements of

$$\bar{Z} = T_m(\bar{X} - \mu),$$

obtained from the estimating equations (5). In the following section, we will relate this quantity to a log-likelihood difference, and hence derive an asymptotic distribution for the likelihood ratio test for the validity of certain restrictions placed on the estimating equations.

5 Non-Parametric Likelihood Ratios

In the previous two sections, we have first shown that the sample 'moments' generated by the estimating equations converge in probability to their true but unknown expectations. This, as a consequence, implies probabilistic convergence of the log-likelihood series estimator to the 'true' log density. It therefore seems obvious to construct any testing procedure for the validity of any restrictions placed on the estimating equations to be based on the stochastic difference between the two. Second, a central limit theorem for the sample 'moments' was established, which in this section will be used to derive the asymptotic distribution of any resultant test statistic. Indeed, under the assumption that $p(x)$ is a member of (1) and moreover that the hypothesis under consideration was simple, i.e.

$$H_0 : p(x) = p_{\theta_0},$$

for some fixed sequence, $\theta_0 = \lim_{m\to\infty}\{\theta_k\}_{k=1}^{m}$, this was precisely what was achieved in Portnoy (1988)

However, here we do not necessarily wish to place such restrictions on the problem. In particular we will consider two kinds of null hypotheses. First suppose that H_0 does completely specify $p(x)$, although not assuming that $p(x)$ is a member of (1). Complete specification implies, amongst other things, that the sequence

$$\lim_{m\to\infty} \quad \mu_k = \int \phi_k(x)dP(x), \ \ k = 1, ..., m$$

is known. In this case we let $\bar{\theta}$ the unique (see Barron and Sheu (1991)) solution to

$$\lim_{m\to\infty} \quad \mu_k = \int \phi_k(x)p_{\bar{\theta}}(x)dx, \ \ k = 1, ..., m. \tag{14}$$

That is $p_{\bar{\theta}}(x)$ is a density, in the family of (1), with $E_{p_{\bar{\theta}}}[\phi_k(x)] = \mu_k = E_{dP(x)}[\phi_k(x)]$, for all k. That is $p_{\bar{\theta}}(x)$ is not an estimator of $p(x)$, but an approximation to it, in fact, Barron and Sheu (1991), it is the unique L_∞ approximation.

Alternatively, suppose that $p(x)$ is specified only in that the number of estimation equations in (5) is fixed, that is under the null hypothesis the estimation programme is in fact standard finite dimensional maximum likelihood. Hence, let θ^* be the solution to the set of equations

$$\begin{aligned} \frac{1}{N}\sum\nolimits_i \phi_k(x_i) &= \int \phi_k(x) p_{\theta^*}(x)dx,\ \ k=1,...,m', \\ \lim_{m\to\infty} \quad f_k\left(\frac{1}{N}\sum\nolimits_i \phi_k(x_i)\right) &= \int \phi_k(x) p_{\theta^*}(x)dx,\ \ k=m'+1,...,m \end{aligned} \tag{15}$$

where m' is fixed and finite, and the $\{f_k(.)\}_{k=1}^m$ are known functions of the first m' sample moments. Estimation programme (15) thus allows free estimation with respect the first m' 'moments', but restricts all further moments to be known functions of those first m' (as an illustration, for the exponential distribution the k^{th} raw moment is proportional to the first moment rased to the power k). For this restricted case, we also define a $\bar{\theta}$ as in (14), although in this case it remains an unknown quantity.

To summarise, we have unconstrained estimation gives the parameter $\hat{\theta}$, constrained estimation gives θ^*, which in the fully parametric case equals the L_∞ approximating parameter, $\bar{\theta}$. Respectively, these parameters give densities $p_{\hat{\theta}}(x)$, $p_{\theta^*}(x)$ and $p_{\bar{\theta}}(x)$, each of which are members of (1). Finally, we have the 'true' density $p(x)$. Essentially it is the relationship between these densities which determines the likelihood ratio test, and its rate of convergence.. Moreover here we base a test on the difference between the density estimators $p_{\hat{\theta}}(x)$ and $p_{\theta^*}(x)$, where the former is always consistent while the latter is consistent only if the restrictions in (15) are true, so the test is in the same spirit (albeit applied to density estimators) as the principle introduced in Hausman (1978). Note that this approach differs to that of Marsh (2007) where the tests are based on $p_{\hat{\theta}}(x)$ and $p_{\bar{\theta}}(x)$.

For a given sample, we wish to test whether restrictions imposed, as in (15) are valid. A natural choice of test is the likelihood ratio

$$-2\log\Lambda = -2\log\left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)}\right], \tag{16}$$

which forms a profile log-likelihood ratio test. We decompose the likelihood ratio,

$$\log\left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)}\right] = \log\left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)}\right] + \log\left[\frac{p_{\theta^*}(x)}{p_{\bar{\theta}}(x)}\right]$$

$$= \quad \log\left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)}\right] + \log\left[\frac{p_{\theta^*}(x)}{p(x)}\right] + \log\left[\frac{p(x)}{p_{\bar{\theta}}(x)}\right], \tag{17}$$

and examine convergence in the last two terms in (17). Supposing the restrictions in (15) are valid, then from Theorem 1, we have

$$\Pr[|\phi - \mu| > \varepsilon] = O(n^{-1}),$$

because the event $|\phi - \mu| > \varepsilon$ may be decomposed into a finite union, rather than the infinite (31). Applying Jensen's inequality, we have

$$\log\left[\frac{p_{\theta^*}(x)}{p(x)}\right] = O_p(n^{-1}), \tag{18}$$

if H_0 is true. Moreover, let C denote the hyper-plane of density functions dP, with the property

$$C = \left\{dP : \int \phi dP = \mu\right\},$$

then trivially both $p_{\bar{\theta}}(x)$ and $p(x)$ lie in C, and in particular $p_{\bar{\theta}}(x)$ is the unique member of (1) in C, see Barron and Sheu (1991), Lemma 3. The rate of convergence of $\log[p_{\bar{\theta}}(x)]$ to $\log[p(x)]$ is determined by the analyticity of $\log[p(x)]$. Define W_2^r to be the Sobolev space of functions, $f(x)$, for which $f^{(r-1)} = \frac{d^{r-1}f}{dx^{r-1}}$ is continuous and $\left\|f^{(r/2)}\right\|_2$ is finite. Then, again as a consequence of Barron and Sheu (1991), Theorem 1

$$\log\left[\frac{p(x)}{p_{\bar{\theta}}(x)}\right] = O_p\left(m^{-2r}\right),$$

so that if either $p(x)$ is analytic ($p(x) \in W_2^\infty$), or $m \propto n^{\frac{1}{2r+1}}$ and on account of (18) then

$$E_{dP(x)}\left[\log\left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)}\right] - \log\left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)}\right]\right] = O_p(n^{-d}),$$

where $d \in (1/2, 1)$. Hence, if $m^2/n \to 0$,

$$-2\log\Lambda = -2\log\bar{\Lambda} + O_p\left(\frac{m^2}{n}\right), \tag{19}$$

where $\bar{\Lambda} = \log\left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)}\right]$, and $-2\log\bar{\Lambda}$ is a likelihood ratio test for testing the simple hypothesis,

$$H_0' : \theta = \bar{\theta}, \tag{20}$$

in the family of exponential densities (1).

Having established that the log-likelihood ratio test for the imposed restrictions is asymptotically proportional to $-2\log\bar{\Lambda}$, in we can relate the asymptotic distribution of this approximate criterion with that derived in Theorem 2. Letting $\bar{\theta}$ be defined as in (14), we let the random variable $V \in \mathbb{R}^m$, have density

$$p_{\bar{\theta}}(v) = p_0 \exp\left\{\bar{\theta}' v - \varphi_m(\bar{\theta})\right\},$$

i.e. $V \sim p_{\bar{\theta}}$. Note that, by definition

$$E_{p_{\bar{\theta}}}[V] = \frac{d\varphi_m(\bar{\theta})}{d\bar{\theta}} = \left.\frac{d\varphi_m(\theta)}{d\theta}\right|_{\theta=\bar{\theta}} = \mu = E_{dP(x)}[\phi],$$

and hence let $U = V - \mu$. Since we have decomposed the log-likelihood ratio as in (19), it is asymptotically equivalent, to order $O(m^2/n)$, to the criterion in Portnoy (1988). As a consequence, the asymptotic distribution of $-2\log\bar{\Lambda}$ is given by the following Theorem.

Theorem 3 *Let $-2\log\bar{\Lambda}$ be the likelihood ratio test for the simple hypothesis (20), then if $m^{3/2}/n \to 0$ as $m, n \to \infty$*

$$\frac{-2\log\bar{\Lambda} - m}{\sqrt{2m}} = \frac{n\tilde{Z}'\tilde{Z} - m}{\sqrt{2m}} + o_p(1), \tag{21}$$

where $\tilde{Z} = \left(\varphi_m''(\bar{\theta})\right)^{-1/2}(\phi - \mu)$and, as a consequence, if $m^2/n \to 0$ as $m, n \to \infty$, and the restrictions implied by (15) are true, then

$$\frac{-2\log\Lambda - m}{\sqrt{2m}} \to_d N(0,1), \tag{22}$$

where $\Lambda = \log\left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)}\right]$, otherwise the criterion diverges and the test is consistent under any fixed alternative. ■

Notice that the effect of not assuming that the true $p(x)$ is a fully specified member of (1) is a slower rate of convergence of the criterion to its asymptotic result, than in Portnoy (1988).

6 Numerical Properties

In this section, the numerical performance of the density estimator and the non-parametric likelihood ratio test are examined through a series of experiments. All experiments are based on estimating the density of the random variable, X distributed according to the following law;

$$X \sim \begin{cases} N(\mu, \sigma^2), \text{ with probablilty } \rho \\ Exponential[\lambda], \text{ with probability } 1-\rho \end{cases}. \tag{23}$$

In particular, two null hypotheses are considered,

$$\begin{aligned} H_0^0 &: \rho = 0 \Rightarrow X \sim Exponential[\lambda] \\ H_0^1 &: \rho = 1 \Rightarrow X \sim N(\mu, \sigma^2), \end{aligned}$$

which were tested against varying alternatives with ρ. In order to ease the computational burden, the particular values, $\mu = 0.5$, $\sigma^2 = 0.15$ and $\lambda = 8$, were chosen, which for the purposes of these experiments, ensured no values were recorded outside the unit interval. Although the probability of $\{x < 0 \cup x > 1\}$ is positive, it is of exponential order, and thus swamped by the asymptotic orders of the procedure.

6.1 Implementation

Practical application of the procedure described in Section 2, proceeded as follows. By definition the MLE satisfies (3), which we rewrite as

$$\begin{aligned} l(\theta) - \ln[p_0(x)] &= \sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{m} \theta_k(\phi_k(x_i) - \bar{\phi}_k) \\ &\quad -n \log \int_0^1 \exp\left\{\sum\nolimits_{k=1}^{m} \theta_k(\phi_k(x_i) - \bar{\phi}_k)\right\} dx, \end{aligned} \tag{24}$$

and since, at the MLE, the contribution of the first term of (24) is zero, $\hat{\theta}$ formally minimises,

$$\lim_{m,n,J\to\infty} R_u(\theta) = \frac{1}{J+1} \sum_{j=1}^{J} \exp\left\{\sum\nolimits_{k=1}^{m} \theta_k(\phi_k(\beta_j) - \bar{\phi}_k)\right\} \tag{25}$$

subject to $m/n \to 0$, where $\beta_j = (j-1)/J$. Reasonable density estimators are then found by setting m and J to large positive integers. In the experiments which follow the values $J = 150$ and $m = 7$ were employed in (25), which resulted in no convergence problems for the procedure. Although not reported, increasing J and m beyond these values added no explanatory power to the procedure.

Equation (25) delivers the unconstrained density estimator for x, given an independent sample $\{x_1, .., x_n\}$ taken from (23). Following the analysis outlined in the previous section, under either of the null hypotheses, we estimate the density using the restrictions as in (15). As above, in this case, the restricted MLE, minimises

$$\lim_{m,n,J\to\infty} R_r(\theta) = \frac{\sum_{j=1}^{J} \exp\left\{\sum_{k=1}^{m'} \theta_k(\phi_k(\beta_j)) - \bar{\phi}_k) + \sum_{k=m'}^{m} \theta_k(\phi_k(\beta_j) - f_k(\bar{\phi}))\right\}}{J+1}, \tag{26}$$

where $f(.)$ is a known function of $\phi = \{\phi_1, .., \phi_{m'}\}'$, and $\bar{\phi}$ is its sample average. Again, the values $J = 150$ and $m = 7$ were used, while the value m' depends upon the particular null hypothesis under consideration.

Letting, for the purpose of this section, $\phi_k(x) = x^k$, and considering $H_0^0 : \rho = 0$, then under H_0^0,

$$E[x^k] = k!E[x]^k,$$

which implies imposing the restrictions

$$f_k(\bar{\phi}) = k!\bar{\phi}_1{}^k = k!\left(\bar{x}\right)^k, \text{ for } k > 1,$$

where $\bar{\phi}_1 = \bar{x} = \frac{\sum x_i}{n}$ is the only unrestricted moment, i.e. in this case $m' = 1$ Under H_0^1, and also defining $s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$, then the restrictions to be imposed are

$$\begin{aligned}
f_3(\bar{\phi}) &= \bar{x}^3 + 3\bar{x}s^2 \quad ; \quad f_4(\bar{\phi}) = \bar{x}^4 + 6\bar{x}^2s^2 + 3s^4 \\
f_5(\bar{\phi}) &= \bar{x}^5 + 10\bar{x}^3s^2 + 15\bar{x}s^4 \quad ; \quad f_6(\bar{\phi}) = \bar{x}^6 + 15\bar{x}^4s^2 + 15\bar{x}^2s^4 + 15s^6 \\
f_7(\bar{\phi}) &= \bar{x}^7 + 21\bar{x}^5s^2 + 105\bar{x}^3s^4 + 15\bar{x}s^6,
\end{aligned}$$

while $\bar{\phi}_1 = \bar{x}$ and $\bar{\phi}_2 = \bar{x}^2 + s^2$ are the unrestricted moments, giving $m' = 2$. Details of an extensive simulation study of the properties of both the estimation procedure and the non-parametric test are given in the following section.

6.2 Analysis

The practical analysis of the procedure as a whole simply consists of a comparison of the performance of the restricted estimator, given by minimising (26), and the unrestricted estimator, given by minimising (25) as ρ varies. First, Appendix B compares the 'fit' of the restricted and unrestricted density estimates, based on a sample size of $n = 150$, with $m' = 1$ and $\rho = 0, 0.15, 0.3$, with the empirical density obtained through a Monte Carlo study involving 50,000 replications. Likewise Appendix C compares the restricted and unrestricted density estimates, with $n = 150$, but with $m' = 2$ and $\rho = 1, 0.85, 0.7$. Clearly, as the distribution moves away from the respective null hypotheses, the ability of the restricted density estimator to approximate declines. However, for values of ρ 'close' to these nulls, graphically, the difference between the restricted and unrestricted is difficult to discern, in particular for smaller sample sizes. In fact, in terms of the distribution (or quantiles) this becomes even more difficult, which is why the comparison is made in terms of the density.

The second part of the analysis concerns the ability of the likelihood ratio test, given in (16), to discriminate between the null densities and alternatives lying between. Formally, for (23), we test

$$\begin{aligned} H_0^0 &: \rho = 0, \text{ against} \\ H_1^0 &: \rho = \rho^*, \end{aligned}$$

for values of ρ^* from 0.025 to 0.15, in steps of 0.025, and

$$\begin{aligned} H_0^1 &: \rho = 0, \text{ against} \\ H_1^1 &: \rho = \rho^*, \end{aligned}$$

for values of ρ^* from 0.975 to 0.85, in steps of -0.025, for sample sizes of 75 and 150, using the standardised log-likelihood ratio test

$$\frac{-2\log\Lambda - m}{\sqrt{2m}} \tag{27}$$

Although, the limiting distribution of (27) is standard normal, we in fact reject for large values of $\log\Lambda$, and hence small values of (27), giving the one-sided rejection rule, reject H_0 if

$$\frac{-2\log\Lambda - m}{\sqrt{2m}} < c_\alpha, \tag{28}$$

where c_α is chosen so that the size of the test is α. The simulations contained in this section were performed according to the following algorithm. First, for some value of ρ an independent sample of size $n = 75$, or $n = 150$, was taken from a random variable distributed as (23). Then the unrestricted and restricted density estimates were found through minimising (25) and (15), respectively, from which (27) was constructed. Repeating the process gives the empirical distribution of the test statistic.

The critical values for the test, obtained via a 10,000 replication Monte Carlo study, along with their nominal values are given in the following table.

Table 1: Nominal and actual critical values.

Size	Nominal	H_0^0, $n = 75$	H_0^0, $n = 150$	H_0^1, $n = 75$	H_0^1, $n = 150$
5%	-1.645	-1.521	-1.593	-1.546	-1.607
10%	-1.282	-1.155	-1.207	-1.164	-1.219

As should be expected, since convergence to the asymptotic density is of order m^2/n, the size properties of the test are not impressive. However, for any null hypothesis accurate critical values may be obtained through simulation, or parametric bootstrap. The power properties of the test were examined through further simulations of (28) with 5,000 replications, and are reported in the following two tables.

Table 2: The power of (28) for testing H_0^0 against varying values of ρ^*

$n = 75$			$n = 150$		
ρ^*	5%	10%	ρ^*	5%	10%
0.025	0.079	0.138	0.025	0.086	0.152
0.050	0.151	0.234	0.050	0.204	0.327
0.075	0.280	0.381	0.075	0.435	0.585
0.100	0.432	0.561	0.100	0.641	0.757
0.125	0.579	0.685	0.125	0.818	0.885
0.150	0.705	0.798	0.150	0.905	0.949

Table 3: The power of (28) for testing H_0^1 against varying values of ρ^*

$n = 75$			$n = 150$		
ρ^*	5%	10%	ρ^*	5%	10%
0.9875	0.086	0.176	0.9875	0.122	0.219
0.9750	0.171	0.297	0.9750	0.257	0.403
0.9625	0.267	0.400	0.9625	0.432	0.582
0.9500	0.352	0.508	0.9500	0.593	0.726
0.9375	0.471	0.612	0.9375	0.712	0.836
0.9250	0.548	0.698	0.9250	0.816	0.925

For either null hypothesis, the test yields reasonable power in moderately large sample sizes. In particular, considering that for values of ρ^* 'close' to the respective null hypotheses the unrestricted and restricted are graphically indistinguishable, the criterion provides acceptable levels of discrimination.

7 Conclusions

As in Marsh (2007) the purpose of this paper has been foundational. That is to explore the properties, both analytic and numerical, of a new test for goodness-of-fit.

Specifically the requisite asymptotic theory has been formally established, asymptotic normality under the null and consistency under a fixed alternative. In addition the finite sample numerical performance of both the nonparametric density estimator are analysed in a series of Monte Carlo experiments.

Ultimately, however, what will be required is the extension of these methods to circumstances under which additional parameters are required to be estimated. Only then can such tests be applied to the kinds of model adequacy and predictive ability problems detailed in Corradi and Swanson (2006). The numerical performance indicated by the experiments of this paper suggests that such an extension is worth pursuing.

8 Bibliography

Anderson, T.W. and D.A. Darling (1952). Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193-212.

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

Chesher, A. (1991). The effect of measurement error. *Biometrika*, **78**, 451-462.

Chow, Y.S. and Teicher, H. (1988). *Probability theory,* 2^{nd} ed. Springer-Verlag, 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.

Csiszar, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, **3**, 146-158.

Diciccio, T.J, Hall, P. and J. P. Romano (1991). Empirical likelihood is Bartlett-correctable. *Annals of Statistics*, **19**, 1053-1061.

Fan, Y. and Linton, O. (1997). Some higher order theory for a consistent nonparametric model specification test. *Mimeo, Yale University, U.S.A.*

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica,* **46**, 1251-1271.

Hong, Y.M. and H.F. Li (2003). Nonparametric specification testing for continuous time models with application to term structure of interest rates. *Review of Financial Studies*, **18**, 37-84.

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.

Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-249.

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

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, **22**, 300-325.

Rao, C.R. (1973). *Linear statistical inference and its applications.* Wiley, New York.

Robinson, P.M. (1991). Consistent nonparametric entropy-based testing. *Review of Economic Studies*, **58**, 437-453.

Royden, H.L. (1988). *Real Analysis*, 3^{rd} ed. Macmillan, New York.

Appendix A

Proof of Theorem 1: We apply Lemma 1 to the $y_{k,i}$, letting $c_t = t^{-1}$, $s = n$ and defining $\bar{y}_{i,k} = i^{-1}\sum_{j=1}^{i} y_{k,j}$, then

$$\Pr[\max_{r\leq i\leq n} |\bar{y}_{i,k}| > \epsilon_k] \leq \frac{1}{\epsilon_k^2}\left(\frac{1}{n^2}\sum_{i=1}^{r} var[y_{k,i}] + \sum_{i=r+1}^{n} i^{-2} var[y_{k,i}]\right). \quad (29)$$

Since x is bounded, then

$$\begin{aligned} var[y_{k,i}] &= E_{dP(x)}[\phi_k(x_i)^2] - \mu_k^2 \\ &\leq \int (\phi_k(x))^2 \, dP(x) = O(1), \end{aligned}$$

and hence

$$\sum_{i=1}^{r} var[y_{k,i}] = O(r).$$

Substituting into (29), we then have

$$\lim_{n\to\infty} \Pr[\max_{i\leq n} |\bar{y}_{i,k}| > \epsilon_k] = O(n^{-1}),$$

which immediately establishes a LLN for $\phi_k(x_i)$, namely

$$\frac{1}{n}\sum_i \phi_k(x_i) - \mu_k \to_{a.s} 0,$$

where the subscript $a.s.$ denotes convergence almost surely. Equally, direct application of Chebychev's Theorem yields,

$$\Pr[\left|\frac{1}{n}\sum_i \phi_k(x_i) - \mu_k\right| > \epsilon_k] = O(n^{-1}). \quad (30)$$

Now consider the m terms in the estimating equations (5), $\left\{\frac{1}{n}\sum_i \phi_k(x_i)\right\}_{k=1}^{m}$ as $m, n \to \infty$, while $m/n \to 0$. Let a_k be the event that $\left|\frac{1}{n}\sum_i \phi_k(x_i) - \mu_k\right| > \epsilon_k$, and so from (30),

$$\Pr[a_k] = O(n^{-1}).$$

Let the m vector $A_m = (a_1, .., a_m)$, so that the probability statement in (10) may be written

$$\lim_{n,m\to\infty} \Pr[|A_m| > \epsilon],$$

and convergence is established if this limiting probability is zero. We let A be the event that $|A_m| > \epsilon$, then for suitably chosen $\epsilon_1, ..., \epsilon_m$, we may write

$$\lim_{m\to\infty} A = \lim_{m\to\infty} \bigcup_{k=1}^{m} a_k. \tag{31}$$

Since

$$\begin{aligned}\lim_{m\to\infty} \Pr(A) &\leq \lim_{m\to\infty} \sum_{k=1}^{m} \Pr(a_k) \\ &\leq \lim_{m\to\infty} m \sup_{k\leq m} P(a_k),\end{aligned}$$

and noting $P(a_k) = O(n^{-1}) = b_1 n^{-1}$, say for some constant b_1, then

$$\lim_{m\to\infty} \Pr(A) \leq \lim_{m\to\infty} b_1 \frac{m}{n},$$

and so since $m/n \to 0$, then the theorem is proved. ■

Proof of Theorem 2: Let $\mathcal{F}_n$ be the sigma-field $\mathcal{F}\{\phi(x_1), ..., \phi(x_n)\} = \mathcal{F}\{C_1, ..., C_n\}$ generated by the set of estimating equations (5), and define for any positive integer $n^* \leq n$, $\bar{Z}_{(n^*)} = (n^*)^{-1} \sum_i^{n^*} Z_i$. For the purposes of this proof, all expectations, unless indicated otherwise, are taken with respect to the dominating measure and to save on notation we write $E[.] = E_{dP(x)}[.]$.

Consider the difference

$$\begin{aligned} R_n &= C_n - C_{n-1} = n^2 \bar{Z}'\bar{Z} - nm - (n-1)^2 \bar{Z}'_{(n-1)}\bar{Z}_{(n-1)} - (n-1)m \\ &= 2(n-1)Z'_n\bar{Z}_{(n-1)} + (Z'_n Z_n - m), \end{aligned} \tag{32}$$

then that R_n is a martingale difference, and hence C_n is a martingale follows from

$$E[Z'_n\bar{Z}_{(n-1)}] = 0, \quad \text{and} \quad E[Z_n Z'_n] = I_m,$$

by definition, and so from (32), we have

$$\begin{aligned} E[R_n \,|\, \mathcal{F}_{n-1}] &= 0, \\ E[C_n \,|\, \mathcal{F}_{n-1}] &= C_{n-1}. \end{aligned}$$

As far as the limiting distribution is concerned, in order to apply Lemma 2 we merely have to check that conditions (12) and (13) are satisfied. For the first, letting σ_i^2 and s_n^2 be defined as in the statement of Lemma 2, we have

$$\begin{aligned} \sigma_i^2 &= E\left[R_i^2\right] = 4(i-1)^2 E\left[Z_i'\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}'Z_i\right] + 4(i-1)E\left[Z_i'\bar{Z}_{(i-1)}Z_i'Z_i\right] \\ &\quad + E\left[(Z_i'Z_i - m)^2\right] \\ &= E\left[R_i^2\right] = 4(i-1)^2 E\left[Z_i'\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}'Z_i\right] + E\left[(Z_i'Z_i - m)^2\right], \end{aligned}$$

since R_i is a martingale difference, therefore,

$$\begin{aligned} s_n^2 &= \sum\nolimits_i \sigma_i^2 = 4\sum\nolimits_i (i-1)^2 E\left[Tr(\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}'Z_iZ_i')\right] \\ &\quad + \sum\nolimits_i E\left[Tr[(Z_iZ_i')^2]\right] + nm^2. \end{aligned}$$

Again $\bar{Z}_{(i-1)}$ and Z_i are independent, so

$$E\left[Tr(\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}'Z_iZ_i')\right] = Tr\left[E[\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}']E[Z_iZ_i']\right],$$

and further since $E\left[Tr[(Z_iZ_i')^2]\right] = O(m^2)$, we have

$$\begin{aligned} s_n^2 &= 4\sum\nolimits_i (i-1)^2 Tr\left[E[\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}']I_m\right] + O(nm^2) \qquad (33) \\ &= 2n(n-1)m + O(nm^2) \\ &= 2n^2 m\,(1 + O(1)), \end{aligned}$$

which establishes the rate of divergence of the right hand side of (12).

As for the left hand side of (12), from (32), we have

$$\begin{aligned} \sum\nolimits_i E\left[|R_i|^3\right] &= \sum\nolimits_i E\left[\left|2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right|^3\right] \\ &= \sum\nolimits_i E\left[\left\{\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^6\right\}^{1/2}\right]. \qquad (34) \end{aligned}$$

For any random variable $D = D(x)$, where x has support over $[0,1]$, we may define the L_2 norm $||D||_2$ of that random variable by

$$||D||_2 = \left(\int_0^1 |D|^2 dP(x)\right)^2,$$

see also Royden (1988), Chapter 6. Expectations with respect to $dP(x)$ then satisfy

$$\left(E\left[|D|^2\right]\right)^2 = ||D||_2,$$

and so direct application of Jensen's inequality implies

$$\left(E\left[|D|\right]\right)^2 \leq E\left[|D|^2\right]. \tag{35}$$

Applying inequality (35) to (34) we find

$$\begin{aligned}
\sum\nolimits_i E\left[|R_i|^3\right] &\leq \sum\nolimits_i \left(E\left[\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^6\right]\right)^{1/2} \\
&\leq \sum\nolimits_i \left\|\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^3\right\|_2, \qquad (36)
\end{aligned}$$

in terms of the L_2 norm. Thus applying the Minkowski inequality to (36)

$$\sum\nolimits_i E\left[|R_i|^3\right] \leq \sum\nolimits_i \left(E\left[\left(2(i-1)Z_i'\bar{Z}_{(i-1)}\right)^6\right]\right)^{1/2} + \sum\nolimits_i \left(E\left[\left((Z_i'Z_i - m)\right)^6\right]\right)^{1/2},$$

which on account of $E\left[(Z_i'Z_i)^6\right] = O(m^6)$, gives, for some constant b_1,

$$\sum\nolimits_i E\left[|R_i|^3\right] \leq 8\sum\nolimits_i \left(E\left[\left(2(i-1)Z_i'\bar{Z}_{(i-1)}\right)^6\right]\right)^{1/2} + b_1 n m^3.$$

Finally, writing $(i-1)Z_n'\bar{Z}_{(i-1)} = Z_n' \sum_i^{n-1} Z_i$, and noting the independence of the two terms, Proposition (A.3) of Portnoy (1988) is applicable to the remaining expectation, giving for some constant b_2

$$\begin{aligned}
\sum\nolimits_i E\left[|R_i|^3\right] &\leq b_2 n^{5/2} m^2 \left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right) \\
&= o(s_n^{-3}),
\end{aligned}$$

so that the Lindeberg condition holds.

Considering now the expectations of the conditional variances, and utilising both Jensen's and the Minkowski's inequalities, similar bounds may be found, in particular

$$\begin{aligned}
\sum\nolimits_i E\left[\left|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right] - \sigma_i^2\right|\right] &\leq \sum\nolimits_i \left\|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right] - \sigma_i^2\right\|_2 \\
&= \sum\nolimits_i \left\|4(i-1)^2 \bar{Z}_{(i-1)}\bar{Z}_{(i-1)}' - (i-1)m \right. \\
&\quad \left. + 4(i-1)\bar{Z}_{(i-1)}' E\left[Z_i'\left(Z_i'Z_i - m\right)\right] + O(m^2)\right\|_2 \\
&\leq \sum\nolimits_i \left\|4(i-1)^2 \bar{Z}_{(i-1)}\bar{Z}_{(i-1)}' - (i-1)m\right\|_2 \\
&\quad + \sum\nolimits_i \left\|4(i-1)\bar{Z}_{(i-1)}' E\left[Z_i\left(Z_i'Z_i - m\right)\right]\right\|_2 + O(m^2).
\end{aligned}$$

Again Proposition (A.3) of Portnoy (1988) is applicable to the remaining expectation, which yields, for some constant b_3

$$\begin{aligned}\sum_i E\left[\left|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right] - \sigma_i^2\right|\right] &= b_3 n^2 m \left(1 + O\left(\sqrt{\frac{m}{n}}\right)\right) \\ &= o(s_n^{-2}),\end{aligned}$$

so that (13) holds, and finally noting, from (33),

$$s_n \sim n\sqrt{2m},$$

the theorem is established. ■

Proof of Theorem 3: Consider the family of densities $p_\theta(v)$, then by the intermediate value theorem, there exists some $\tilde{\theta}$ lying between $\hat{\theta}$ and $\bar{\theta}$, such that the following expansions hold,

$$\begin{aligned}\varphi_m(\hat{\theta}) &= \varphi_m(\bar{\theta}) + (\hat{\theta}-\bar{\theta})'\varphi_m'(\bar{\theta}) + \frac{1}{2}(\hat{\theta}-\bar{\theta})'\varphi_m''(\bar{\theta})(\hat{\theta}-\bar{\theta}) \\ &\quad + \frac{1}{6}(\hat{\theta}-\bar{\theta})'\left((\hat{\theta}-\bar{\theta})'\varphi_m'''(\bar{\theta})(\hat{\theta}-\bar{\theta})\right) \\ &\quad + \frac{1}{24}\left\{E_{p_{\tilde{\theta}}}\left[\left((\hat{\theta}-\bar{\theta})'U\right)^4\right] - 3E_{p_{\tilde{\theta}}}\left[\left((\hat{\theta}-\bar{\theta})'U\right)^2\right]^2\right\}, \\ \varphi_m'(\hat{\theta}) &= \varphi_m'(\bar{\theta}) + (\hat{\theta}-\bar{\theta})'\varphi_m''(\bar{\theta}) + \frac{1}{2}E_{p_{\tilde{\theta}}}\left[\left((\hat{\theta}-\bar{\theta})'U\right)^2 U\right].\end{aligned}$$

Since,

$$-2\log\bar{\Lambda} = 2n\left\{(\hat{\theta}-\bar{\theta})'\bar{X} - \left(\varphi_m(\hat{\theta}) - \varphi_m(\bar{\theta})\right)\right\} \tag{37}$$

and on account of

$$\varphi_m'(\bar{\theta}) = \mu \quad \text{and} \quad \varphi_m'(\hat{\theta}) = \bar{X},$$

we have

$$(\hat{\theta}-\bar{\theta}) = \tilde{Z} - \frac{1}{2}\left(\varphi_m''(\bar{\theta})\right)^{-1/2} E_{p_{\tilde{\theta}}}\left[\left((\hat{\theta}-\bar{\theta})'U\right)^2 U\right], \tag{38}$$

then applying Theorems 2.1 and 3.1 of Portnoy (1988), in our case

$$\begin{aligned}\left|\hat{\theta}-\bar{\theta}\right| &= O_p\left(\sqrt{\frac{m}{n}}\right) \\ \left|(\hat{\theta}-\bar{\theta}) - \tilde{Z}\right| &= O_p\left(\sqrt{\frac{m}{n}}\right),\end{aligned}$$

so that from (37) and (38),

$$\begin{aligned}\frac{-2\log\bar{\Lambda}-m}{\sqrt{2m}} &= \frac{n}{\sqrt{2m}}\Big\{(\tilde{Z}'\tilde{Z}-\frac{m}{n})-\Big((\hat{\theta}-\bar{\theta})-\tilde{Z}\Big)'\Big((\hat{\theta}-\bar{\theta})-\tilde{Z}\Big) \\ &\quad +\frac{1}{6}(\hat{\theta}-\bar{\theta})'\Big((\hat{\theta}-\bar{\theta})'\varphi_m'''(\bar{\theta})(\hat{\theta}-\bar{\theta})\Big)\Big\}+O_p(\frac{m^{3/2}}{n}),\end{aligned}$$

and so (21) may be established as in Theorem 3.2, Portnoy (1988). As for (22), since $\left|\hat{\theta}-\bar{\theta}\right| = O_p\left(\sqrt{\frac{m}{n}}\right)$, then as $m^2/n \to 0$

$$\frac{n}{\sqrt{2m}}\tilde{Z}'\tilde{Z} = \frac{n}{\sqrt{2m}}\bar{Z}'\bar{Z}+o_p(1),$$

which immediately gives the limiting distribution from Theorem 2. If however, the restrictions imposed in (15) do not hold in that $f_k\left(\frac{1}{N}\sum_i \phi_k(x_i)\right)$ and $\frac{1}{N}\sum_i \phi_k(x_i)$, do not have the same probability limit,for all $k > m'$, then the second term in (17) diverges, and hence so does $\frac{-2\log\Lambda-m}{\sqrt{2m}}$. ■

Appendix B

This appendix graphs the restricted density estimator (dashed), with m' of (26) equal to 1, with the unrestricted (dotted) and the simulated (solid) for different values of ρ in the distribution of x, given by (23).

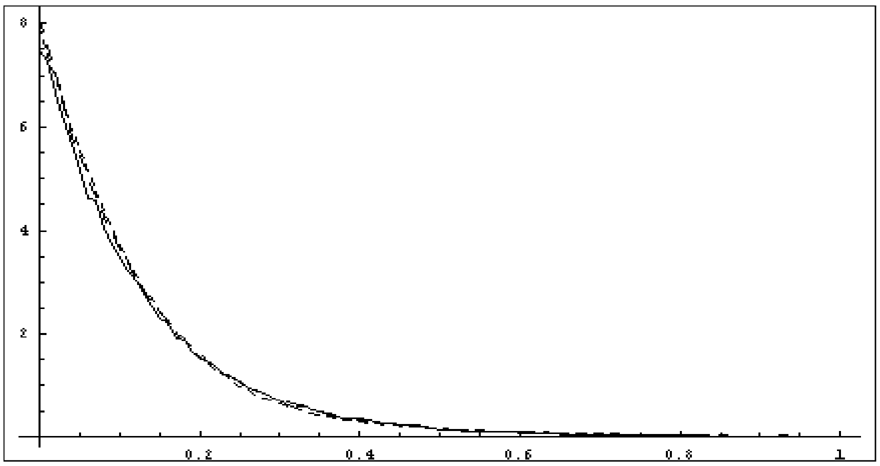

Fig 1: $\rho = 0$

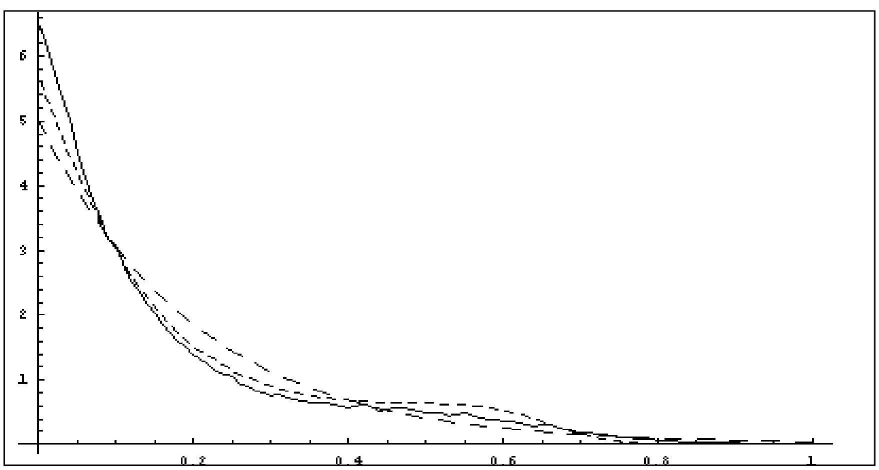

Fig 2: $\rho = 0.15$

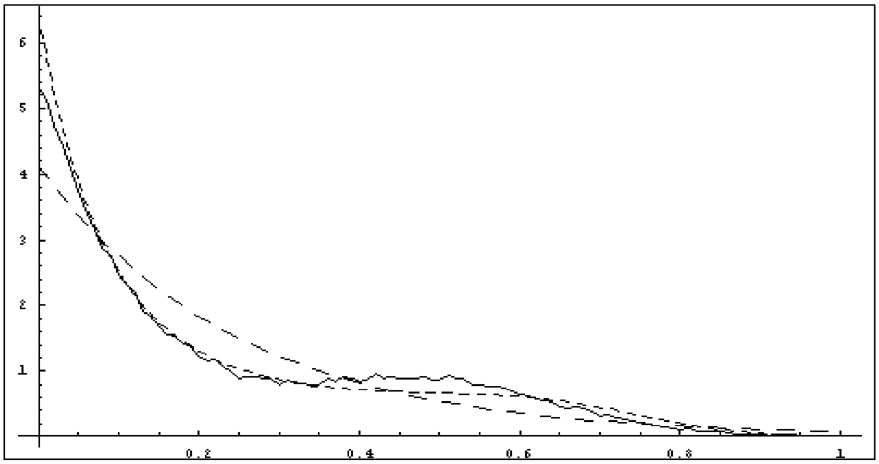

Fig 3: $\rho = 0.3$

Appendix C

This appendix graphs the restricted density estimator (dashed), with m' of (26) equal to 2, with the unrestricted (dotted) and the simulated (solid) for different values of ρ in the distribution of x, given by (23).

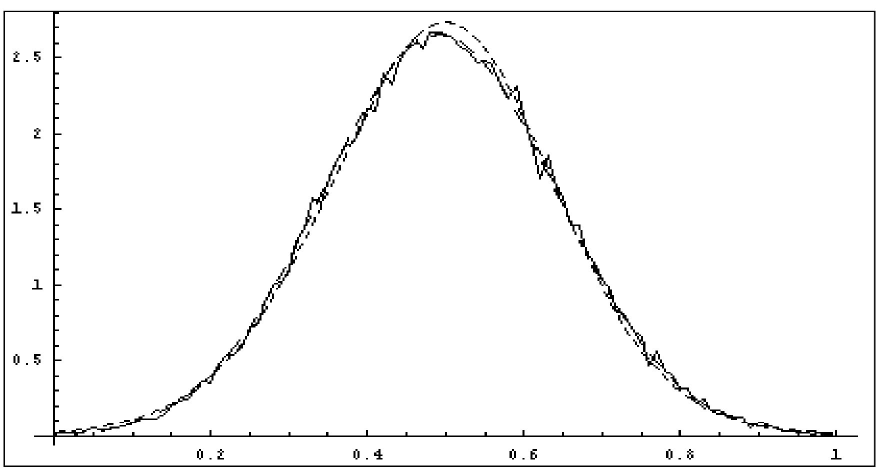

Fig 4: $\rho = 1$

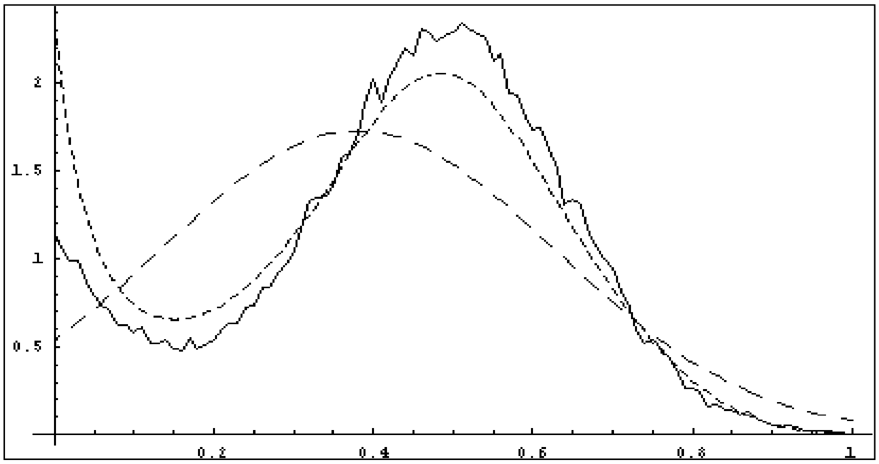

Fig 5: $\rho = 0.85$

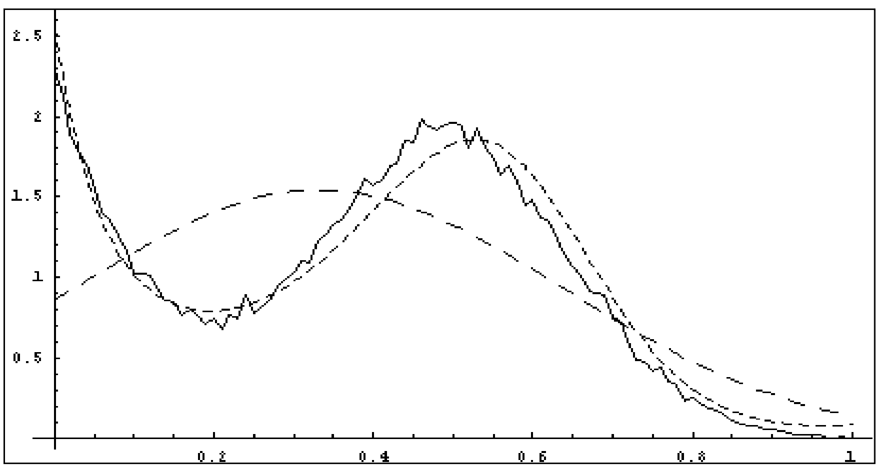

Fig 6: $\rho = 0.7$