THE UNIVERSITY of York

Discussion Papers in Economics

No. 2000/56

Nonparametric Likelihood Ratio Tests

by

Patrick Marsh

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Nonparametric Likelihood Ratio Tests[1]

Patrick W.N. Marsh

Departments of Mathematics and Economics

University of York

Heslington, York

YO1 5DD, U.K.

tel/fax: +44 (0)1904 433084/3071

e-mail: pwnm1@york.ac.uk

October 10^{th} 2000

[1]1991 Subject Classifications: Primary-62G10, Secondary-62G20.

Key Words: exponential series density estimator, nonparametric inference, Hausman test, infinite dimension parameter space.

Thanks are due to Karim Abadir, Giovanni Forchini, Grant Hillier, Peter Phillips for useful comments and discussions, and particularly Andrew Chesher for introducing me to the problem and for advice on the implementation of the density estimator considered here. Financial support in the form of a Leverhulme Special Research Fellowship is gratefully acknowledged.

Abstract

A nonparametric likelihood ratio test is derived via the exponential series density estimator examined in Barron and Sheu (1991). Moreover, utilising the results of Chow and Teicher (1988) and Portnoy (1988) an asymptotic theory (distribution under the null and consistency under the alternative) is developed. Crucially, the null hypothesis (usually a parametric density) neither needs to be a member of the exponential family nor completely specified, leading to a nonparametric profile likelihood ratio test, which is interpreted as a Hausman (1978) test for density estimators. In a series of experiments, based on the specification of a linear regression the size and power properties of the test are examined and compared to existing tests.

Proposed Running head: Nonparametric L.R. Tests

1 Introduction

Although assessing the adequacy of a fitted parametric model for a sample generated by some unknown population distribution is perfectly feasible, see White (1982), if that adequacy is rejected we are often left with no reasonable description of the fundamental properties of that sample. For this reason a variety of non and semiparametric estimation techniques have flourished recently. Such techniques usually fall into two distinct categories; those based on nonparametric density estimation, whether by Kernel based methods (see Silverman (1978)), or based on empirical likelihood (see Owen (1988)) and those based on discriminating between two competing (incorrectly specified) parametric models (see Nishi (1988) or Kitamura (1998)).

The purpose of this paper is to provide both a new specification test based on a nonparametric density estimator introduced by Barron and Sheu (1991) and the relevant asymptotic distribution theory. Usually, within this general context (for example, see Eubank and Speigelman (1990), Härdle and Mammen (1993) or Horowitz and Härdle (1994)), a goodness-of-fit test is employed based upon the relative fit of the density estimators over the sample points. To illustrate, suppose that $p_n(x)$ is a kernel-based density estimator based on a sample of size n, generated by a population density $p(x)$, with kernel $K(v)$ and bandwidth h_n. Then tests, based on the implicit null hypothesis that $x \sim p(x)$, may be based upon the following fundamental asymptotic result; let $h_n = cn^{-1/5}$, then if $p(x)$ is twice continuously differentiable,

$$\frac{n^{2/5}\left[p_n(x) - p(x)\right] - \mu}{\sigma} \rightarrow_d N(0,1), \tag{1}$$

where

$$\mu = c^2 p''(x) \int_{-\infty}^{\infty} v^2 K(v) dv \quad ; \quad \sigma^2 = c^{-1} p(x) \int_{-\infty}^{\infty} K(v)^2 dv,$$

and the bandwidth parameter c is chosen to minimise either Asymptotic or Integrated Mean Square Error. See Silverman (1978), Stone (1980) and Horowitz (1998) for these and other related results. Application of (1) to derive nonparametric specification tests are contained in Zheng (1996) and Fan and Linton (1999).

Here, we develop analogous asymptotic results for the exponential series density estimator proposed in Barron and Sheu (1991), specifically in order to design a testing procedure for a class of specific null hypotheses against arbitrary alternatives. Suppose that the null hypothesis specifies a family of population distributions, up to some unknown parameters. Importantly these parameters may either be of interest (i.e. the distribution is fully specified) or nuisance (i.e. only a family is specified). Moreover, this null may be either fully parametric, or semiparametric in that only a sequence of moments is specified. The alternative is simply that the null is not true. Adaptation of Hausman's (1978) test, supplies the required approach. That is we compare the performance of two density estimators, both consistent under the null, but only one of which is consistent under the alternative. In the context of exponential series estimation this leads to a nonparametric likelihood ratio test.

The asymptotic theory derives from the fact that the estimation procedure is set up in such a way that consistency may be demonstrated through convergence of simple functions of the sample. This occurs because the estimation routine, see equation (6) below, equates the moments of the estimating density with those of simple functions, usually polynomials, of the sample. Consequently, hypotheses concerning the population may be formulated by the form and number of these functions required for consistency of the density estimator. For example, to test whether the sample comes from a member of the exponential family, of dimension m', say, only m' functions of the data will be required for consistency. For an arbitrary nonparametric alternative, infinitely many will be required.

Infinite dimensional inference, which in general is what we have here, has also been extensively studied. Portnoy (1988) gives asymptotic theory for a likelihood ratio test of a simple null hypothesis in an infinite dimension exponential model, while Murphy and van der Vaart (1997) consider tests on a scalar parameter in the presence of infinitely many nuisance parameters. This paper differs in that we consider the case of many interest parameters (if the null is a parametric density), and moreover these parameters are not restricted under the null. Precise specification of the kind of hypotheses considered here is contained in Section 3. The asymptotic theory presented in the paper follows from extension of Portnoy's (1988) analysis to the case where the sequence of parameters, in the infinite dimension exponential family, do not take fixed values under the null. As a consequence, the test may actually be viewed as a profile likelihood ratio, as in Murphy and van der Vaart (1997), but for more arbitrary null hypotheses.

The major practical advantage of basing such tests on the Barron and Sheu (1991) estimator is that even in the event of rejection of the null, we have a relatively simple analytic approximation for the population density. Such a feature is important if, for instance as in Chesher (1999) and Koo, Kooperberg and Park (1999), a tractable density estimator is required. Moreover, because the estimator is a member (albeit infinite) of the exponential family, it embeds many of the parametric densities we would wish to test for. However, due to the flexibility of the methods, the test also has an application as a specification test, in the spirit of Zheng (1996) and Fan and Linton (1999). Indeed, in Section 4 we apply the test to the regression hypothesis considered in those papers, in order to assess the size and power properties of the nonparametric likelihood ratio test.

The plan of the paper is as follows. The next section details the density estimation procedure and it's relevant properties. Section 2.2 examines consistency of that procedure in terms of laws of large numbers (LLN) for functions of the sample, which leads to a central limit theorem (CLT) for the parameters of the estimating density, given in Section 2.3. Section 3 details the class of hypotheses of interest, derives the form of the likelihood ratio test, and provides the relevant asymptotic theory. Section 4 contains a brief numerical study of the properties of the test, in the context of specification testing in a regression model, Section 5 then concludes. Finally an Appendix contains the proofs of all theorems and tables used in the numerical analysis.

2 The Exponential Series Estimator

Most of the properties of the density estimator we consider are contained in Barron and Sheu (1991), see in particular Theorem 1 and Remarks 1 through 6. Here, though, we summarise some basic properties, but interpret, in particular, consistency in such a way as to inform the testing procedure that is eventually set up. Moreover, we present a CLT, which in conjunction with the consistency result, drives the asymptotics of the test.

2.1 Basic Properties

Before considering hypothesis tests, we suppose that we wish to estimate the density of some scalar random variable x, with distribution $P(x)$, satisfying the following assumption.

Assumption (i) Let x be defined on the bounded sample space $\Omega_X = [a, b]$, (without loss of generality $[0, 1]$) with density

$$p(x) = dP(x) : \left\{ \Omega_X \to \mathbb{R}, \int_{\Omega_X} dP(x) = 1, p(x) \geq 0 \right\},$$

and any sample $(x_1, ..., x_N)$ consists of i.i.d. copies of x.

(ii) The log-density of x satisfies

$$lp(x) = \ln [p(x)] \in W_2^r,$$

where W_2^r is the Sobolev space of functions, so that $lp^{(r-1)}(x) = \frac{d^{r-1} lp(x)}{dx^{r-1}}$ is absolutely continuous and $lp^{(r)}(x)$ is square integrable on $[0, 1]$ for all $r \geq 2$.

The density estimator is a member of the exponential family

$$p_\theta(x) = p_0(x) \exp \left\{ \sum\nolimits_{k=1}^{m} \theta_k \phi_k(x) - \varphi_m(\theta) \right\}, \tag{2}$$

where in (2) the cumulant function is defined by

$$\varphi_m(\theta) = \log \int_{\Omega_X} p_0(x) \exp \left\{ \sum\nolimits_{k=1}^{m} \theta_k \phi_k(x) \right\} dx, \tag{3}$$

where $\theta = (\theta_1, .., \theta_m)' \in \mathbb{R}^m$ 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, forming a 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 concerned with the polynomial case. Though none of the results depend on this restriction, their interpretation will.

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 (2). That is, in terms of the log-likelihood

$$\hat{\theta} : \max_{\theta} l(\theta) = \ln[p_0(x)] + \sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{m} \theta_k \phi_k(x_i) - n\varphi_m(\theta). \tag{4}$$

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 (4) some properties are immediately obtainable. First, the score:

$$l^k(\theta) = \sum\nolimits_i \phi_k(x_i) - n\varphi_m^k(\theta), \tag{5}$$

where the superscript(s) indicates, with respect to which variable(s) we are deriving. From (3),we have

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

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

$$\int \phi_k(x) p_\theta(x) dx = \frac{1}{n} \sum\nolimits_i \phi_k(x_i), \quad k = 1, ..., m. \tag{6}$$

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

$$\begin{aligned} l^{j,k}(\theta) &= -n\varphi_m^{j,k}(\theta) = \frac{d}{d\theta_k} \int \phi_j(x) p_\theta(x) dx \\ &= -n \left(\int \phi_j(x)\phi_k(x) p_\theta(x) dx - \int \phi_j(x) p_\theta(x) dx \int \phi_k(x) p_\theta(x) dx \right). \end{aligned} \tag{7}$$

In the fully parametric case, i.e. m fixed and finite, we may obtain the usual asymptotic result

$$\left(n\varphi_m^{j,k}(\theta)\right)^{1/2} l^k(\theta) \rightarrow_d N_m(0, I_m), \tag{8}$$

from which a CLT for the MLE, $\hat{\theta}$, itself may be obtained. However, nonparametric estimation requires $m \rightarrow \infty$ as $n \rightarrow \infty$, hence standard asymptotic results are not applicable. Moreover, since we will not, in general, assume $p(x)$ is an exponential density, the estimated parameters do not necessarily have any significant meaning. Consequently, in this paper, we concentrate on the case where the $\phi_k(x)$ are polynomials, where for instance, in the simplest case,

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

Considering the estimating equations (6), and defining $E_{p_\theta}[.]$ as expectations with respect to the series exponential family, we have

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

or in general $E_{p_{\hat{\theta}}}[\phi_k(x)] = \frac{1}{n}\sum_i \phi_k(x_i)$. That is (6) equates 'moments' from the family $p_\theta(x)$ with the sample moments. Alternatively, the nonparametric 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

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

so that the moments of $\phi(x)$ are themselves bounded. Consequently, for the purposes of the asymptotic analysis to follow, the properties of the density estimator will be closely related to those of the sample moments, in particular laws of large numbers and CLTs. In summary, then, letting $\phi = (\phi_1(x), .., \phi_m(x))'$, ϕ_i defined analogously, and $\bar{\phi} = n^{-1}\sum_i \phi_i$, we have the set of m estimating equations

$$\int \phi . p_{\hat{\theta}}(x) dx = \bar{\phi}. \tag{9}$$

Although for the purposes of the development of the theory we assume that x has bounded support, in practice this assumption is benign. As an example, suppose a random variable $x^* \in \mathbb{R}$ has distribution $P^1(x^*)$. Letting $F(.)$ be any distribution function defined on $\mathbb{R}$ and let $x = F(x^*)$, so that $x \in [0,1]$. Now if $F(x^*) = P^1(x^*)$, then x will be uniformly distributed, otherwise suppose that the density of x is estimated via (9), giving $p_{\hat{\theta}}(x)$, then since $x^* = F^{-1}(x)$, the density of $x^*, p^1(x^*) = dF(x^*)$, may be estimated via

$$\hat{p}^1(x^*) = p_{\hat{\theta}}(F(x^*))f(x^*).$$

Moreover, provided $F(.)$ *is* differentiable, then the asymptotic properties of the density estimators, $p_{\hat{\theta}}(x)$ and $\hat{p}^1(x^*)$ are identical. In the following section we develop these asymptotic properties, stemming from a WLLN for the (infinite) vector of sample moments $\bar{\phi}$.

2.2 Consistency

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 some other point in $\mathbb{R}^m$.

First, define the moment vector $\mu = (\mu_1, .., \mu_m)'$ for $m \to \infty$, and let C denote the hyper-plane of densities, dP, satisfying Assumption 1 and

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

Consider the population analogue of (9)

$$\int \phi . p_\theta(x) dx = \mu, \tag{10}$$

then the following Lemma, proved directly from Theorem 1 and Remark 6 of Barron and Sheu (1991).

Lemma 1 *Let $\bar{\theta}$ be a solution of (10) then*

(i) $p_{\bar{\theta}}(x)$ is the unique member of (2) in C and moreover,

(ii) the relative entropy (Kullback-Leibler divergence) of $p_{\bar{\theta}}(x)$ to $dP(x)$ is

$$\int \ln \left[\frac{p_{\bar{\theta}}(x)}{p(x)} \right] dP(x) = O_r \left(m^{-2r} \right),$$

where r is the 'smoothness' of the log-density $lp(x)$ as in Assumption 1.

(iii) suppose that $m^3/n \to 0$ as $m, n \to \infty$, then the maximum likelihood estimator in the family (2), $p_{\hat{\theta}}(x)$, given by (9) converges, in relative entropy, to $dP(x)$ according to

$$\int \ln \left[\frac{p_{\hat{\theta}}(x)}{p(x)} \right] dP(x) = O_{pr} \left(m^{-2r} + m/n \right) \tag{11}$$

Interpretation of the previous lemma is as follows; $p_{\bar{\theta}}(x)$ is the member of family (2) closest, in terms of entropy, to the underlying probability measure $dP(x)$, and has deterministic error of order m^{-2r} as $m \to \infty$. Convergence of the MLE then follows from bounds on the divergence of $p_{\hat{\theta}}(x)$ to $p_{\bar{\theta}}(x)$ and from $p_{\bar{\theta}}(x)$ to $p(x)$. An interesting corollary arises when we optimise the rate of convergence of (11), with respect to m. Doing so, yields the optimal rate of increase

$$m \propto n^{\frac{1}{2r+1}}, \tag{12}$$

giving convergence in (11) of order $O_{pr} \left(n^{\frac{-2r}{2r+1}} \right)$. Hence in the absence of any knowledge on $dP(x)$, other than it satisfies assumption 1, this implies a minimax rate of $m \propto n^{\frac{1}{5}}$, which mirrors the corresponding optimal bandwidth selection, in kernel based estimation.

For our analysis, any likelihood ratio test will be based upon a measure of distance of $\hat{\theta}$ to $\bar{\theta}$ in $\mathbb{R}^m$, but unlike in Portnoy (1988) and Murphy and van der Vaart (1997), $\bar{\theta}$ is neither known, nor is $p_{\bar{\theta}}(x)$ restricted to be the underlying probability measure. Thus any restrictions imposed on the estimation program must be placed on the estimating equations themselves, in the form of restrictions on the sample moment sequences. Details of these restrictions will be given later. Therefore, for the unrestricted case, we prove first prove a LLN for the sample moment sequence, and it is this basic result which drives the asymptotics of any hypothesis we may wish to impose.

Let $y_{k,i} = \phi_k(x_i) - \mu_k$ be the deviation of the sample moments from their (true, but unknown) expectations. Obviously, since $k = 1, ..., m$ and $m \to \infty$, strong laws for $\bar{\phi}$ are precluded. However, a strong law can be established for each individual $\phi_k(x_i)$, since it's moments are bounded, then provided m is such that $m/n \to 0$, a weak law will follow. First note the following Lemma.

Lemma 2 *Hajek-Renyi 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* $\varepsilon > 0$

$$\Pr[\max_{r<i<s} c_i|Y_1 + .. + Y_i| \geq \varepsilon] \leq \frac{1}{\varepsilon^2}\left(c_r^2 \sum\nolimits_{i=1}^{r} var[Y_i] + \sum\nolimits_{i=r+1}^{s} c_i^2 var[Y_i]\right). \quad \blacksquare \tag{13}$$

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, which is proved in Appendix A1.

Theorem 1 *Let* $|.|$ *denote the Euclidean norm in* $\mathbb{R}^m$, *and given an* $\epsilon > 0$, *then under Assumption 1,*

$$\lim_{n,m\to\infty} \Pr\left[\left|\bar{\phi} - \mu\right| > \epsilon\right] = 0, \tag{14}$$

where $\bar{\phi}$ *and* μ *are defined above.*

Theorem 1 establishes that

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

with convergence (in probability) obtained at a rate $O(\sqrt{m/n})$. Alternatively, in light of their definitions and equation (9), we have

$$\int \phi p_{\hat{\theta}}(x)dx - \int \phi p_{\bar{\theta}}(x)dx = O(\frac{m}{n}), \tag{15}$$

where $\phi = \{\phi_1(x), ..., \phi_m(x)\}'$, if $m/n \to 0$, see also Barron and Sheu (1991, eq. 2.10). Consequently, applying Jensen's inequality and by the Lebesgue dominated convergence theorem we establish

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

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

$$(\log[p_{\hat{\theta}}(x)] - \log[p_{\bar{\theta}}(x)]) = O_p(\frac{m}{n}), \tag{16}$$

that is convergence of the log-densities, alternatively the Kullback-Leibler divergence of $p_{\hat{\theta}}(x)$ from $p_{\bar{\theta}}(x)$, is obtained at this rate. Moreover, analysing the problem in this way gives a direct link between the asymptotics of the sample moments relative to their population counterparts and the MLE relative to the entropy minimising density.

2.3 A Central Limit Theorem

Our hypothesis tests are going to be based upon the stochastic differences implied by equations (11) and (16). In particular, this difference takes the form of a log-likelihood ratio. However, there are some technical aspects which need to be considered before applying any statistical procedure. First, (11) is a weak, not a strong law, convergence of the density estimator is obtained in probability only. In fact, unless $p(x)$ is itself a (finite) member of the exponential family, even asymptotically, the series log-density estimator does not converge pointwise 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) and Murphy and van der Vaart (1997) detail the analysis of estimation and testing in exponential families as the number of parameters diverges. In those papers 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 3 *(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

$$\sum\nolimits_i E\left[|R_i|^3\right] = o(s_n^3), \tag{17}$$

and also that the conditional variances are bounded, so that

$$\sum\nolimits_i E\left[\left|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right] - \sigma_i^2\right|\right] = o(s_n^2) \tag{18}$$

then

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

In order to apply Lemma 3 we need to establish some notation, note that the score vector is defined by

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

and has 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 $T_m = \left(n\varphi_m^{j,k}(x)\right)^{1/2}$, $Y_i = \phi_i - \mu$, 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 3 directly to the sum of squared elements of $\bar{Z}$, giving the following Theorem, proved in Appendix A2.

Theorem 2 *Let $C_n = n^2 \bar{Z}'\bar{Z} - nm$, then under Assumption 1*
a) C_n is a martingale and
b) if $m \propto n^{\frac{1}{2r+1}}$ as $n \to \infty$

$$\frac{C_n}{n\sqrt{2m}} = \frac{n\bar{Z}'\bar{Z} - m}{\sqrt{2m}} \to_d N(0,1). \quad \blacksquare \tag{19}$$

The theorem 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 (6). We may also pick out two special cases for consideration, in order to generate some usable asymptotics, since the asymptotics in (19) depend upon the unknown r. Letting $m = \alpha n^{\frac{1}{2r+1}}$ gives, in the absence of knowledge of r, (i.e. we assume $r = 2$)

$$\frac{n^{\frac{11}{10}}\bar{Z}'\bar{Z} - \alpha n^{\frac{1}{10}}}{\sqrt{2\alpha}} \to_d N(0,1), \tag{20}$$

or alternatively if $lp(x)$ is analytic

$$\frac{n\bar{Z}'\bar{Z} - \alpha}{\sqrt{2\alpha}} \to_d N(0,1). \tag{21}$$

These alternative asymptotic representations are important because, in practice r will not be known, however, for any experiment we may construct to analyse the density estimator, $lp(x)$ will be analytic.

In the following section, we will relate the quantity C_n of Theorem 2 to the log-likelihood difference in (16), and hence derive an asymptotic distribution for the likelihood ratio test for the validity of certain restrictions placed on the estimating equations.

3 Non-Parametric Likelihood Ratios

In the previous section, 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 (2) and moreover that the hypothesis under consideration was simple, i.e.

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

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. Suppose that $p(x)$ is specified only in that the number of estimation equations in (6) 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

$$\begin{aligned} n^{-1}\sum \phi_k(x_i) &= \int \phi_k(x)p_{\theta^*}(x)dx = \bar{\phi}_k, \quad k = 1, ..., m', \\ \lim_{m\to\infty} f_j\left(\bar{\phi}_1, ..., \bar{\phi}_{m'}\right) &= \int \phi_j(x)p_{\theta^*}(x)dx, \quad j \geq m' + 1, \end{aligned} \tag{22}$$

where m' is fixed and finite, and the $\{f_k(.)\}_{k=1}^m$ are known functions of the first m' sample moments. Estimation programme (22) 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). Notice that the analysis of Portnoy (1988) is only applicable when we both assume that the whole set of moments, $\{\mu_k\}$ is known under the null hypothesis and that $p(x)$ *is* a member of (2), in which case in which case we obtain $\theta^* = \bar{\theta}$, of Lemma 1. Formally then, the specific hypothesis we test is

$$\begin{aligned} H_0 &: \mu_j = f_j(\mu_1, .., \mu_{m'}) \quad vs. \\ H_1 &: \mu_j \neq f_j(\mu_1, .., \mu_{m'}) \quad \text{for } j \geq m' + 1. \end{aligned} \tag{23}$$

To summarise, we have unconstrained estimation, as in (6), which gives the parameter $\hat{\theta}$, while constrained estimation, as in (22), gives θ^*, which in the fully parametric case equals the 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 (2). Finally, we have the 'true' density $p(x)$.

For a given sample, we wish to test (23), via

$$-2 \log \Lambda = -2 \log \left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)} \right], \tag{24}$$

which forms a profile log-likelihood ratio test. Notice also that under either formulation of the null hypothesis (22) both $p_{\hat{\theta}}(x)$, the 'unrestricted' density estimator and $p_{\theta^*}(x)$ the 'restricted' estimator are consistent estimators for $p(x)$. However, under the implicit alternative that the restrictions do not hold, only $p_{\hat{\theta}}(x)$ is consistent. Hence the procedure we propose is essentially a Hausman (1978) test, applied to density estimation.

To continue, we decompose the likelihood ratio,

$$\begin{aligned} \log \left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)} \right] &= \log \left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)} \right] + \log \left[\frac{p_{\theta^*}(x)}{p_{\bar{\theta}}(x)} \right] \\ &= \log \left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)} \right] + \log \left[\frac{p_{\theta^*}(x)}{p(x)} \right] + \log \left[\frac{p(x)}{p_{\bar{\theta}}(x)} \right], \end{aligned} \tag{25}$$

and examine convergence in the last two terms in (25). Supposing the restrictions in (22) are valid, then from Theorem 1, we have

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

because the event $|\bar{\phi} - \mu| > \varepsilon$ may be decomposed into a finite union, rather than the infinite (41). Hence, applying Jensen's inequality, we have

$$\log\left[\frac{p_{\theta^*}(x)}{p(x)}\right] = O_p(n^{-1/2}), \tag{26}$$

if H_0 is true. Moreover, from Lemma 1, if $m \propto n^{\frac{1}{2r+1}}$,

$$E_{dP(x)}\left[\log\left[\frac{p_{\theta^*}(x)}{p_{\hat{\theta}}(x)}\right] - \log\left[\frac{p_{\bar{\theta}}(x)}{p_{\hat{\theta}}(x)}\right]\right] = O_p(n^{-\frac{2r}{2r+1}}),$$

and hence

$$-2\log\Lambda = -2\log\bar{\Lambda} + O_p\left(n^{-\frac{2r}{2r+1}}\right), \tag{27}$$

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,

$$H_0' : \theta = \bar{\theta} \quad vs. \quad H_1' : \theta \neq \bar{\theta}, \tag{28}$$

in the family of exponential densities (2).

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 before, 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 (27), it is asymptotically equivalent, to order $O(n^{-\frac{2r}{2r+1}})$, to the criterion in Portnoy (1988). As a consequence, the asymptotic distribution of $-2\log\bar{\Lambda}$ is given by the following Theorem, proved in Appendix A3.

Theorem 3 *Let $-2\log\bar{\Lambda}$ be the likelihood ratio test for the simple hypothesis (28), then if $m \propto n^{\frac{1}{2r+1}}$ as $n \to \infty$*

$$\frac{-2\log\bar{\Lambda} - m}{\sqrt{2m}} = \frac{n\tilde{Z}'\tilde{Z} - m}{\sqrt{2m}} + o_p(1), \tag{29}$$

where $\tilde{Z} = \left(\varphi''_m(\bar{\theta})\right)^{-1/2}(\phi - \mu)$ *and, as a consequence, if the restrictions implied by (22) are true, then*

$$\frac{-2\log\Lambda - m}{\sqrt{2m}} \rightarrow_d N(0,1), \tag{30}$$

where $\Lambda = \log\left[\frac{p_{\theta*}(x)}{p_{\hat{\theta}}(x)}\right]$, *otherwise the criterion diverges.*

Notice that the effect of not assuming that the true $p(x)$ is a fully specified member of (2) is a slower rate of convergence of the criterion to it's asymptotic distribution, than in Portnoy (1988). However, what we gain is the ability to test whether the sample comes from a family of distributions, rather than a specific member of that family. Notice also that for specific values of r, i.e. for $r = 2$ and analytic functions, the asymptotic results follow from (20) and (21) respectively.

In the following section a brief experiment is considered in order to illustrate the implementation of the density estimation technique, and the numerical properties of the likelihood ratio test.

4 An Application

4.1 Implementation

Details of the numerical properties of the density estimator (2) itself are contained in the original article by Barron and Sheu (1991) and in applications due to Chesher (1999) and Koo, Kooperberg and Park (1999). In this section, in order to compare with existing results, we will apply the nonparametric likelihood ratio test (24) as a specification test in the following regression model as analysed in Bierens (1990), Zheng (1996) and Fan and Linton (1999);

$$E[y_i|z_i] = g(z_i, \beta) \quad ; \quad i = 1, .., n, \tag{31}$$

where z_i and β are $k \times 1$ vectors of covariates and coefficients, respectively, and $g(.,.)$ is a possibly nonlinear function. As in the latter papers we will assume that the null hypothesis specifies a Gaussian linear regression model, viz.

$$H_0 : y_i = z_i'\beta + \varepsilon_i \quad ; \quad \varepsilon_i \sim NID(0, 1). \tag{32}$$

According to classical hypothesis testing theory (Cox and Hinkley (1974) and Hillier (1987)), let $y = (y_1, .., y_n)'$, $Z = (z_1, ..., z_n)'$, define $M_Z = I_n - Z(Z'Z)^{-1}Z'$ then the $(n-k)$ vector $v = C'y$, where C is the singular value decomposition of M_Z (i.e. $CC' = M_Z$ and $C'C = I_{n-k}$) characterises both the class of similar (having size independent of the nuisance parameter β) and invariant (with respect to affine group action on y) tests of H_0. Indeed, the majority of the proposed tests for specific alternatives (e.g. omitted variables, functional form, dependence in the error structure, non-normality etc.) are functions of the data only through y, and generally take the form

$$test = v'A(H_1)v,$$

where $A(H_1)$ is a matrix chosen for each alternative H_1. However, although one may deduce an 'optimal' (however optimality is defined) test for a given alternative, there is no guarantee that this test will even be unbiased or consistent against other alternatives, let alone 'optimal'. To apply our test in this context note that under H_0,

$$v \sim N(0, I_{n-k}),$$

and so if we let $x_i^* = \Phi(v_i)$, where v_i is the i^{th} element of v and $\Phi(.)$ is the standard normal cumulative distribution function, then under H_0, x_i^* is uniform on $[0, 1]$. Moreover, under any alternative whatsoever, x_i^* has some other distribution on $[0, 1]$, consequently so long as the null and alternative distributions of x_i^* may be consistently estimated by (2), then the nonparametric likelihood ratio test will be consistent.

As it turns out, direct nonparametric estimation of a uniform density proves numerically troublesome, consequently we let $x_i = (x_i^*)^2$ and apply the procedure described in Section 2.1 to estimate the density of x_i. Practical implementation of the procedure is as follows: By definition the MLE satisfies (4), which we rewrite as

$$\begin{aligned} l(\theta) - \ln[p_0(x)] &= \sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{m} \theta_k(\phi_k(x_i) - \bar{\phi}_k) \\ &\quad - n \log \int_0^1 \exp\left\{ \sum\nolimits_{k=1}^{m} \theta_k(\phi_k(x_i) - \bar{\phi}_k) \right\} dx, \end{aligned} \tag{33}$$

and since, at the MLE, the contribution of the first term of (33) is zero, $\hat{\theta}$ formally minimises,

$$\lim_{m,n,J \to \infty} R_u(\theta) = \frac{1}{J+1} \sum_{j=1}^{J} \exp\left\{ \sum\nolimits_{k=1}^{m} \theta_k(\phi_k(\beta_j) - \bar{\phi}_k) \right\} \tag{34}$$

subject to $m/n \to 0$, and where $\beta_j = (j-1)/J$. Reasonable density estimators are then found by setting m and J to large positive integers.

Equation (34) delivers the unconstrained density estimator for x, given an independent sample $(x_1, .., x_n)$. Following the analysis outlined in the previous section, under either of the null hypotheses, we estimate the density using the restrictions as in (22). As above, in this case, the restricted MLE, minimises

$$\lim_{m,n,J \to \infty} R_r(\theta) = \frac{1}{J+1} \sum_{j=1}^{J} \exp\left\{ \sum\nolimits_{k=1}^{m'} \theta_k(\phi_k(\beta_j)) - \bar{\phi}_k) + \sum\nolimits_{k=m'}^{m} \theta_k(\phi_k(\beta_j) - f_k(\bar{\phi})) \right\}, \tag{35}$$

where $f(.)$ is a known function of $(\bar{\phi}_1, .., \bar{\phi}_{m'})$. Again, the values $J = 150$ and $m = 7$ were used, while the value m' depends upon the particular null hypothesis under consideration.

For the null hypothesis (32), since $\sqrt{x_i}$ is uniform, then under H_0 the moment sequence of x_i is given by

$$E[x_i^k] = \frac{2}{3k+2},$$

for all k. Consequently, setting $m' = 1$, the restrictions to be imposed are of the form

$$f_j(\bar{\phi}) = \bar{\phi}_1 \times \frac{5}{3j+2}, \text{ for } j = 2, .., m,$$

and $\bar{\phi}_1 = \sum_{i=1}^{n-k} x_i/n$ is the only estimated moment. Consequently, letting $\hat{\theta}$ be the unrestricted MLE (which minimises (34)) and θ^* be the restricted MLE (which minimises (34)) the nonparametric likelihood ratio test is then

$$LR = \frac{-2\log\Lambda - m}{\sqrt{2m}}, \tag{36}$$

where $\Lambda = \log[p_{\theta^*}(x)/p_{\hat{\theta}}(x)]$, and we reject (32) for small values of LR.

4.2 Numerical Properties

In this subsection, results from a Monte Carlo study are used to evaluate the numerical properties of the test (36) for the null hypothesis specified in (32). All numerical calculations were performed in *Mathematica* (see Wolfram (1999)). Following Zheng (1996), we let $k = 3$ and ζ_{1i} and ζ_{2i} be independent drawings from the standard normal distribution, and so the three regressors included in (32) are $z_{1i} = 1$, $z_{2i} = \zeta_{1i}$ and $z_{3i} = (\zeta_{1i} + \zeta_{2i})/\sqrt{2}$, and so formally the hypothesis to be tested is

$$H_0 : y_i = \beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i} + \varepsilon_i \quad ; \quad \varepsilon_i \sim NID(0,1). \tag{37}$$

Values for ζ_{1i} and ζ_{2i} for $i = 1, ..., 400$, were generated once and for all.

A number of relevant alternative hypotheses were considered, involving both the mean function and the error term, viz.

$$\begin{aligned}
H_1 &: \quad y_i = \beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i} + \frac{a}{10} z_{2i} z_{3i} + \varepsilon_i \\
H_2 &: \quad y_i = (\beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i})^{1-a/9} + \varepsilon_i \\
H_3 &: \quad y_i = \beta_1 + \beta_2 z_{2i} + \beta_3 z_{3i} + u_i, \quad \text{where} \\
u_i &\sim \begin{cases} N(0,1) & \text{with probability } 1 - a/10 \\ U[-2,2] & \text{with probability } a/10, \end{cases}
\end{aligned} \tag{38}$$

and we consider values of $a = 1, 2, .., 10$ (for H_2 the real root is used) and various sample sizes, $n = 50, 100, 200, 400$. Notice that the null (37) is embedded in each of the alternatives via $a = 0$. In particular the alternatives considered by Zheng (1996) are recovered from H_1 with $a = 1$ and H_2 with $a = 6$.

The remaining issue is the choice of the dimension of the series density estimator, i.e. m. Since in all cases the log-density of x_i is analytic we may choose m to grow arbitrarily slowly with n, c.f. (12). In fact, we set $m = 4$ for $n = 50$, and increased m by 1 as n doubled. Allowing m to grow either more rapidly or more slowly had an insignificant impact upon the outcomes of the experiments, and so details are not reported.

In the first experiment the null distribution of the test LR, (36), was simulated with 20,000 replications and for sample sizes of $n = 50, 100, 200$ and 400. Table 1 below reports the nominal (based on the limiting standard normal) versus the simulated rejection frequencies of LR, for each sample size. Furthermore, Table 2, contained in the Appendix reports the nominal versus simulated quantiles for the null distribution of LR, again for each sample size.

Table 1 : Nominal vs. True Size of the

LR test (36) under H_0 based on 20,000 replications.

	$n = 50$	$n = 100$	$n = 200$	$n = 400$
1%	.004	.006	.007	.008
5%	.034	.039	.043	.046
10%	.078	.084	.092	.101

By way of comparison, the kernel based specification test of Zheng (1996) seems to offer less finite sample accuracy under the null hypothesis. Not necessarily in terms of absolute accuracy, but certianly in terms of the monotonicity of accuracy as the sample size increases, although this may be explained by the relatively small number of replications, and hence high standard error in that study. Fan and Linton (1999) present results for a higher-order correction to a Kernel based specification tests. Comparing Table 1 above with their Tables 1 and 2 would suggest that the likelihood based approach offers reasonable accuracy even compared to their higher-order correction.

Three further experiments were conducted, in order to assess the power properties of LR (36), given each of the alternatives. Specifically the test was simulated under each of the alternatives given in (38) for sample sizes of $n = 100, 200$ and 400, and for each value of $a = 1, 2, .., 10$, based on 5,000 replications. For the purposes of these experiments we fixed $\beta_1 = \beta_2 = \beta_3 = 1$ (again for comparison with Zheng (1996)). The power was simulated using three significance levels (1%, 5% and 10%) and the critical values used were those obtained in the null simulation study detailed above. Tables 3 through 5 contained in the Appendix contain the simulated rejection frequencies under each of the alternatives listed in (38) for each value of a and for each sample size. Again, where comparisons can be made with the results of Zheng (1996, Tables 2 and 3), the power of the LR test, (36), compares favourably with that of the kernel based test.

Before concluding, some important qualitative distinctions between the likelihood based approach of this paper and previous kernel based approaches need to be examined. Although here, the rate at which m increases with n has an insignificant impact upon the results, the same is not true for the choice of bandwidth, $h = cn^{1/5}$ for kernel density estimators. Indeed Fan and Linton (1999) find that larger values of c tend to degrade the accuracy of the nominal size as an approximation to the 'true' size, under the null. However, Zheng (1996) finds that larger values of c increase the power of the test. This implied trade-off does not appear to be present in the approach of this paper.

5 Conclusions

The goal of this paper has been to develop some asymptotic theory for a test based upon the exponential series density estimator proposed by Barron and Sheu (1991). By exploiting both the properties of the estimator itself and the fundamental idea of Hausman (1978) a nonparametric likelihood ratio may be constructed, which may be seen as an alternative to other nonparametric approaches, such as empirical likelihood or kernel based methods.

As in the analysis of Portnoy (1988) and Murphy and van der Vaart (1997) the asymptotic theory for the test is nonstandard, involving convergence of infinite dimensional seres, and hence convergence is slow in comparison with the classical case. However, the asymptotics involved with the test are clearly mirrored with those tests derived from kernel density estimators (e.g. see Silverman (1978) and Horowitz (1998)).

Since it is based upon a nonparametric density estimator, the flexibility of the test is impressive. However, when applied to a particular example, previously studied in the literature, we find favourable numerical properties, as compared to more established techniques. Equally, the test benefits from three beneficial properties. First, the intuition behind the test comes from Hausman's (1978) principle, secondly the test is relatively simple to implement and finally, even when the test is rejected, one is still left with an analytic approximation to the density of the statistic of interest.

Appendix A: Proofs

A1: Proof of Theorem 1.

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

$$\Pr[\max_{r\leq i\leq n} |\bar{y}_{i,k}| > \epsilon_k] \leq \frac{1}{\epsilon_k^2}\left(\frac{1}{r^2}\sum\nolimits_{i=1}^{r} var[y_{k,i}] + \sum\nolimits_{i=r+1}^{n} i^{-2} var[y_{k,i}]\right). \quad (39)$$

Since x is a bounded random variable, 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

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

Substituting into (39), 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 SLLN for $\phi_k(x_i)$, namely

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

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

$$\Pr[\left|\frac{1}{n}\sum\nolimits_{i} \phi_k(x_i) - \mu_k\right| > \epsilon_k] = O(n^{-1}). \quad (40)$$

Now consider the m terms in the estimating equations (6), $\left\{\frac{1}{n}\sum_i \phi_k(x_i)\right\}_{k=1}^{m}$ 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 (40),

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

Let the m vector $A_m = (a_1, .., a_m)$, so that the probability statement in (14) 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

$$\lim_{m\to\infty} A = \lim_{m\to\infty} \bigcup\nolimits_{k=1}^{m} a_k. \tag{41}$$

Since,

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

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

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

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

Appendix A2: 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 (6), and define for any positive integer $n^* \le n$, $\bar{Z}_{(n^*)} = (n^*)^{-1} \sum_i^{n^*} Z_i$. 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 $E[.] = E_{dP(x)}[.]$. We assume that the number of estimating equations grows according to its optimal rate, i.e. $m \propto n^{\frac{1}{2r+1}}$, so that all orders of magnitude may be written solely in terms of powers of n. Asymptotics are driven by the usual triangular array, so that $\dim \bar{Z}_{(n^*)} = \dim \bar{Z}_{(n)}$ for all $n^* < n$, although notation will be suppressed.

Consider the difference

$$\begin{aligned} R_n &= C_n - C_{n-1} = n^2 \bar{Z}'\bar{Z} - nm - (n-1)^2 \bar{Z}'_{(n-1)} \bar{Z}_{(n-1)} - (n-1)m \\ &= 2(n-1) Z'_n \bar{Z}_{(n-1)} + \left(Z'_n Z_n - m\right), \end{aligned} \tag{42}$$

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 (42), we have

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

As far as the limiting distribution is concerned, in order to apply Lemma 2 we merely have to check that conditions (17) and (18) 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 $E\left[Tr[(Z_iZ_i')^2]\right] = O(m^2) = O(n^{\frac{2}{2r+1}})$, we have

$$\begin{aligned} s_n^2 &= 4\sum\nolimits_i (i-1)^2 Tr\left[E[\bar{Z}_{(i-1)}\bar{Z}_{(i-1)}']I_m\right] + O(n^{\frac{2r+3}{2r+1}}) \qquad (43) \\ &= 2n(n-1)m + O(n^{\frac{2r+3}{2r+1}}) \\ &= 2n^{\frac{4r+3}{2r+1}}\left(1 + O(1)\right), \end{aligned}$$

which establishes the rate of divergence of the right hand side of (17).

As for the left hand side of (17), from (42), we have

$$\begin{aligned} \sum\nolimits_i E\left[|R_i|^3\right] &= \sum\nolimits_i E\left[\left|2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right|^3\right] \\ &= \sum\nolimits_i E\left[\left\{\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^6\right\}^{1/2}\right]. \qquad (44) \end{aligned}$$

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

$$\left(E\left[|D|\right]\right)^2 \leq E\left[|D|^2\right]. \tag{45}$$

Applying inequality (45) to (44) we find

$$\begin{aligned}
\sum_i E\left[|R_i|^3\right] &\leq \sum_i \left(E\left[\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^6\right]\right)^{1/2} \\
&\leq \sum_i \left\|\left(2(i-1)Z_i'\bar{Z}_{(i-1)} + (Z_i'Z_i - m)\right)^3\right\|_2,
\end{aligned} \tag{46}$$

in terms of the L_2 norm. Thus applying the Minkowski inequality to (46)

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

which on account of $E\left[(Z_i'Z_i)^6\right] = O(m^6) = O(n^{\frac{6}{2r+1}})$, gives, for some positive constant b_2,

$$\sum_i E\left[|R_i|^3\right] \leq 8\sum_i \left(E\left[\left(2(i-1)Z_i'\bar{Z}_{(i-1)}\right)^6\right]\right)^{1/2} + b_2 n^{\frac{2r+4}{2r+1}}.$$

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 positive constant b_3

$$\begin{aligned}
\sum_i E\left[|R_i|^3\right] &\leq b_3 n^{\frac{10r+9}{2(2r+1)}} \left(1 + O\left(n^{-\frac{r}{2r+1}}\right)\right) \\
&= o(s_n^{-3}),
\end{aligned}$$

so that the Lindeberg condition holds.

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

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

Again Proposition (A.3) of Portnoy (1988) is applicable to the remaining expectation, which yields, for some positive constant b_4

$$\begin{aligned}\sum_i E\left[\left|E\left[R_i^2 \,|\, \mathcal{F}_{i-1}\right]-\sigma_i^2\right|\right] &\leq b_4 n^{\frac{4r+3}{2r+1}}\left(1+O\left(n^{-\frac{r}{2r+1}}\right)\right) \\ &= o(s_n^{-2}),\end{aligned}$$

so that (18) holds, and finally noting, from (43), we get the asymptotic equivalence

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

the theorem is established. ■

A3: 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,

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

Since,

$$-2\log\bar{\Lambda}=2n\left\{(\hat{\theta}-\bar{\theta})'\bar{X}-\left(\varphi_m(\hat{\theta})-\varphi_m(\bar{\theta})\right)\right\} \tag{47}$$

and on account of

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

we have

$$(\hat{\theta} - \bar{\theta}) = \tilde{Z} - \frac{1}{2}\left(\varphi''_m(\bar{\theta})\right)^{-1/2} E_{p_{\tilde{\theta}}}\left[\left((\hat{\theta} - \bar{\theta})'U\right)^2 U\right], \tag{48}$$

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

$$\begin{aligned} \left|\hat{\theta} - \bar{\theta}\right| &= O_p\left(n^{-\frac{r}{2r+1}}\right) \\ \left|(\hat{\theta} - \bar{\theta}) - \tilde{Z}\right| &= O_p\left(n^{-\frac{r}{2r+1}}\right), \end{aligned}$$

so that from (47) and (48),

$$\begin{aligned} \frac{-2\log\bar{\Lambda} - m}{\sqrt{2m}} &= \frac{n}{\sqrt{2m}}\Big\{(\tilde{Z}'\tilde{Z} - \frac{m}{n}) - \left((\hat{\theta} - \bar{\theta}) - \tilde{Z}\right)'\left((\hat{\theta} - \bar{\theta}) - \tilde{Z}\right) \\ &\quad + \frac{1}{6}(\hat{\theta} - \bar{\theta})'\left((\hat{\theta} - \bar{\theta})'\varphi'''_m(\bar{\theta})(\hat{\theta} - \bar{\theta})\right)\Big\} + O_p(n^{-\frac{4r+1}{2(2r+1)}}), \end{aligned}$$

and so (29) may be established as in Theorem 3.2, Portnoy (1988). As for (30), since $\left|\hat{\theta} - \bar{\theta}\right| = O_p\left(n^{-\frac{r}{2r+1}}\right)$, then

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

which immediately gives the limiting distribution from Theorem 2. If however, the restrictions imposed in (22) do not hold in that $f_j\left(\bar{\phi}_1, .., \bar{\phi}_{m'}\right)$ and $\frac{1}{n}\sum_i \phi_j(x_i)$, do not have the same probability limit, for all $j > m'$, then the second term in (25) diverges, and hence so does $\frac{-2\log\Lambda - m}{\sqrt{2m}}$. ■

Appendix B: Tables

Table 2 : Asymptotic ($N(0,1)$) and empirical quantiles of the LR test (36), under H_0 (37) based on 20,000 replications.

q	$N(0,1)$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
1	-1.282	-1.163	-1.221	-1.236	-1.267
2	-0.842	-0.826	-0.836	-0.848	-0.846
3	-0.524	-0.551	-0.552	-0.541	-0.537
4	-0.253	-0.340	-0.333	-0.291	-0.274
5	0	-0.117	-0.093	-0.031	-0.020
6	0.253	0.121	0.135	0.224	0.245
7	0.524	0.389	0.414	0.491	0.518
8	0.842	0.730	0.781	0.817	0.834
9	1.282	1.239	1.319	1.305	1.297

Table 3 : Simulated power of the LR test (36) under H_1 of (38) based on 5,000 replications.

	n	=	100	n	=	200	n	=	400
a	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	0.01	0.05	0.10	0.01	0.07	0.12	0.02	0.06	0.10
2	0.02	0.06	0.11	0.02	0.11	0.18	0.04	0.11	0.18
3	0.03	0.10	0.18	0.07	0.20	0.32	0.09	0.22	0.34
4	0.06	0.17	0.27	0.15	0.36	0.49	0.23	0.46	0.61
5	0.11	0.29	0.44	0.36	0.66	0.76	0.56	0.76	0.85
6	0.21	0.44	0.59	0.64	0.87	0.92	0.86	0.95	0.98
7	0.35	0.63	0.75	0.86	0.96	0.99	0.99	1.00	1.00
8	0.54	0.80	0.88	0.96	1.00	1.00	1.00	1.00	1.00
9	0.75	0.93	0.96	0.99	1.00	1.00	1.00	1.00	1.00
10	0.91	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Table 4 : Simulated power of the LR test (36) under H_2 of (38) based on 5,000 replications.

	n	$=$	100	n	$=$	200	n	$=$	400
a	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	0.01	0.06	0.11	0.01	0.05	0.12	0.01	0.07	0.13
2	0.02	0.08	0.13	0.02	0.07	0.15	0.04	0.13	0.23
3	0.03	0.12	0.22	0.03	0.13	0.23	0.09	0.22	0.33
4	0.04	0.15	0.26	0.05	0.16	0.27	0.18	0.36	0.47
5	0.05	0.20	0.33	0.07	0.22	0.35	0.25	0.48	0.61
6	0.08	0.27	0.41	0.11	0.32	0.45	0.40	0.62	0.74
7	0.10	0.29	0.43	0.17	0.41	0.53	0.54	0.77	0.86
8	0.17	0.41	0.53	0.26	0.54	0.67	0.76	0.91	0.95
9	0.24	0.49	0.64	0.42	0.68	0.79	0.89	0.97	0.99
10	0.34	0.63	0.76	0.66	0.86	0.93	0.98	1.00	1.00

Table 5 : Simulated power of the LR test (36) under H_3 of (38) based on 5,000 replications.

	n	=	100	n	=	200	n	=	400
a	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	0.01	0.07	0.13	0.01	0.07	0.14	0.02	0.10	0.20
2	0.02	0.10	0.20	0.03	0.12	0.21	0.05	0.16	0.25
3	0.03	0.12	0.22	0.04	0.15	0.25	0.11	0.27	0.38
4	0.04	0.15	0.26	0.07	0.24	0.35	0.18	0.38	0.54
5	0.05	0.18	0.31	0.09	0.30	0.44	0.25	0.49	0.63
6	0.07	0.25	0.39	0.14	0.37	0.51	0.36	0.61	0.73
7	0.10	0.29	0.45	0.18	0.42	0.56	0.48	0.69	0.81
8	0.12	0.33	0.48	0.25	0.52	0.68	0.59	0.82	0.90
9	0.13	0.38	0.55	0.30	0.60	0.73	0.70	0.87	0.94
10	0.19	0.45	0.65	0.39	0.69	0.80	0.82	0.95	0.98

6 Bibliography

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

Bierens, H. J., 1990, A consistent conditional moment test of functional form, Econometrica, 58, 1443–1458.

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

Chesher, A., 1999, Structural estimation in the presence of covariate measurement error, mimeo, University of Bristol, U.K.

Cox, D.R. and D.V. Hinkley, 1974, Theoretical Statistics, (Chapman and Hall, London).

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

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

Eubank, R.L. and C.H. Speigelman, 1990, Testing the goodness of fit of a linear model via nonparametric regression techniques, Journal of the American Statistical Association, 85, 387-392.

Fan, Y. and O. Linton, 1997, Some higher order theory for a consistent nonparametric model specification test, mimeo, London School of Economics, U.K.

Härdle, W. and E. Mammen, 1993, Comparing nonparametric versus parametric regression fits, Annals of Statistics, 21, 1926-1947.

Hausman, J. A., 1978, Specification tests in econometrics, Econometrica, 46, 1251–1271.

Hillier, G.H., 1987, Classes of similar regions and their power properties for some econometric testing problems, Econometric Theory, 3, 1-44.

Hong, Y. and H. White, 1995, Consistent specification testing via nonparametric series regression, Econometrica, 63, 1133-1159.

Horowitz, J.L., 1998, Semiparametric methods in econometrics, Lecture notes in statistics (Springer-Verlag, New York).

Horowitz, J.L. and W. Härdle, W., 1994, Testing a parametric model against a semiparametric alternative, Econometric Theory, 10, 821-848.

Kitamura, Y, 1998, Comparing misspecified dynamic econometric models using nonparametric likelihood, mimeo, University of Minnesota.

Koo, J-Y., Kooperberg, C. and J. Park, 1999, Logspline density estimation under censoring and truncation, Scandinavian Journal of Statistics, 26, 87–105.

Murphy, S.A. and A.W. Van der Vaart, 1997, Semiparametric likelihood ratio inference, Annals of Statistics, 25, 1471-1509.

Nishi, R. 1988, Maximum likelihood principle and model selection when the true model is unspecified, Journal of Multivariate Analysis, 27, 392-403.

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

Rao, C.R., 1973, Linear statistical inference and it's applications, (Wiley, New York).

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.

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).

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica, 50, 1-25.

Wolfram, S., 1999, The Mathematica book, 3^{rd} ed., (Cambridge University Press).

Zheng, J.X., 1996, A consistent test of functional form via nonparametric estimation techniques, Journal of Econometrics, 75, 263–289