Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 109–140
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models

Xiaohong Chen[a], Han Hong[b,*], Matthew Shum[c]

[a]*Department of Economics, New York University, USA*
[b]*Department of Economics, Stanford University, USA*
[c]*Department of Economics, Johns Hopkins University, USA*

Available online 8 March 2007

Abstract

We propose a nonparametric likelihood ratio testing procedure for choosing between a parametric (likelihood) model and a moment condition model when both models could be misspecified. Our procedure is based on comparing the Kullback–Leibler Information Criterion (KLIC) between the parametric model and moment condition model. We construct the KLIC for the parametric model using the difference between the parametric log likelihood and a sieve nonparametric estimate of population entropy, and obtain the KLIC for the moment model using the empirical likelihood statistic. We also consider multiple (>2) model comparison tests, when all the competing models could be misspecified, and some models are parametric while others are moment-based. We evaluate the performance of our tests in a Monte Carlo study, and apply the tests to an example from industrial organization.

JEL classification: C52

Keywords: Model selection tests; Empirical likelihood; KLIC

*Corresponding author.
E-mail addresses: xiaohong.chen@nyu.edu (X. Chen), doubleh@stanford.edu (H. Hong), mshum@jhu.edu (M. Shum).

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.01.010

1. Introduction

In many empirical applications, similar economic theoretical models imply different econometric specifications: some specify parametric (likelihood-based) models while others specify moment-based models. For a given dataset, then, applied researchers face the problem of how to select one model among the multiple competing models. For example, in demand analysis, both likelihood-based estimation methods (Gowrisankaran et al., 2003) and moment-based estimation methods (Berry et al., 1995) have been used. In empirical structural auction models, Paarsch (1992) showed that while maximum likelihood is feasible for some models, other models can be feasibly estimated by method of moments.

There has been a well developed econometric literature on model selection when all the competing models are parametric likelihood-based, and can be estimated by parametric maximum likelihood; see, for example, Vuong (1989), Sin and White (1996), Gourieroux and Monfort (1994) and the references therein. Several recent papers have also developed model selection tests when all the competing models are moment based, and can be estimated by generalized method of moments or empirical likelihood; see, for example, Hansen (1982), Kitamura (2000), Smith (1992), Rivers and Vuong (2002), Ramalho and Smith (2002) and others. In particular, using information theoretic methods,[1] Kitamura (2000) developed the nonparametric analog of Vuong's (1989) model selection tests for potentially misspecified (unconditional) moment based models, which is also extended to conditional moment restriction models in Kitamura (2002). However, to the best of our knowledge, there is no published work in the likelihood setting on model selection tests among competing parametric (likelihood-based) models and (unconditional) moment-based models. For the non-likelihood based models, Horowitz and Hardle (1994) offered a specification test between a parametric model against a semiparametric alternative.

In this paper, we formulate nonparametric likelihood ratio tests for choosing among parametric models and moment models when all these models could be misspecified. Our procedures are based on comparing the Kullback–Leibler Information Criterion (KLIC) for the parametric models and the moment condition models, and can be regarded as extensions of the likelihood ratio test of Vuong (1989) for comparing two parametric models and of the nonparametric likelihood ratio test of Kitamura (2000) for comparing two moment-based models. We construct a sample estimate of the KLIC for a parametric model using the difference between the maximized sample log-likelihood parametric function and a sieve nonparametric estimate of the population entropy[2]; we obtain a sample estimate of the KLIC for a moment-based model by the empirical likelihood method (cf. Owen, 2001; Qin and Lawless, 1994). Under mild regularity conditions, we show that both estimated KLIC are $\sqrt{n}$-consistent and asymptotically normally distributed under potential model misspecification. The difference between the estimated KLICs is then used to construct our test statistics. For comparison between two models, the proposed test is directional and applies to situations in which the two competing models

[1] cf. Kitamura and Stutzer (1997), Qin and Lawless (1994), Kitamura (1997), Imbens et al. (1998) and Kitamura and Tripathi (2001).

[2] Any choice of the sieve basis, such as spline, Fourier series, power series, neural network, Hermite polynomial, wavelet, normal mixture and many others, can be used to estimate the population entropy.

are non-nested, overlapping, or nested and whether both, one, or neither is misspecified. We derive large sample statistical properties of these tests under mild regularity conditions.

In some applications, researchers need to compare multiple econometric structural models implied by several competing economic theories, where some of the candidate models are parametric, and some are moment-based. For a multiple model comparison, we follow the approach of White (2000) by specifying a benchmark model, and require that at least one of the candidate models is not nesting nor nested by the benchmark model. The benchmark model can be either a parametric model or a moment-based model.

In this paper we use the KLIC as the model comparison criterion, which is a popular criterion in both statistics and econometrics. On the one hand, it is well known that, for parametric models, the maximum likelihood estimate minimizes the KLIC between the parametric density and the true population density even when the model is misspecified; see e.g. White (1982). On the other hand, for moment models, in which a set of moment conditions form the statistical model for approximating the population data generating process (DGP), the maximized empirical likelihood criterion function is the minimal KLIC between the set of implied densities which satisfy the moment conditions and the true population density. Therefore it seems convenient to use a KLIC criterion in comparing parametric and moment-based models, because in both cases the KLIC can be constructed from components which are outputs of the estimation procedure.

We note that KLIC is not the only choice for the discrepancy measure between different families of probability distributions defined by competing models. In some applications, alternative discrepancy measures that measure goodness-of-fit in some sense, such as the Kolmogorov–Smirnov distance, might be preferred.[3] Nevertheless, we need to stress the point that no matter which discrepancy measure is used for model selection tests, it is more desirable to use the same criterion for both the parameter estimation and the model comparison. This is because in model selection problems, all the competing models could be misspecified even under the null hypothesis that they are equally close to the true DGP according to some discrepancy measure; for this reason, it is important to apply the same criterion to define pseudo-true parameters as well as to compare the models. This is in contrast to consistent specification test problems, where the model is correctly specified under the null hypothesis. For this case, one can apply any consistent estimator for the true unknown parameters under the null and then apply other criteria to define test statistics. The crucial difference is that the true parameters only depend on the true DGP and do not depend on the criterion used to estimate them, but the pseudo-true parameters depend on both the true DGP and the discrepancy measure used to define them.[4]

The rest of the paper is organized as follows: Section 2 discusses the definition and estimation of the KLIC for parametric models and moment condition models. In Section 3 we formulate nonparametric likelihood ratio tests between a parametric model and a moment condition model when both can be misspecified. Monte Carlo evidence on the small-sample properties of this test are presented in Section 4. In Section 5, we study selection tests among multiple models where some models are parametric and others are

[3]For example, for finance and macroeconomic models, other popular model comparison criteria include the Hansen–Jaganathen distance in evaluating asset pricing models; the mean squared prediction error, the median squared prediction error, density forecasts evaluation and the conditional Kolmogorov test. See, for example, Granger (2002) and Diebold (1989) and the references contained therein.

[4]See Chen and Fan (2005) for a more detailed discussion of this matter.

moment-based. Section 6 contains an empirical illustration of our testing procedure to price search models. Section 7 briefly concludes. All technical proofs are gathered in the Appendices.

2. The KLIC for parametric and moment condition models

Throughout the paper, we assume that there is a random sample $(z_i:\ i=1,\dots,n)$ from Z, whose true but unknown probability density is $h_0(Z)$. We refer to the parametric likelihood specification as model **F**, which is given by a family of probability density functions $\{f(Z;\beta):\beta\in R^{d_\beta}\}$, where the functional form of $f(\cdot;\beta)$ is specified up to unknown finite-dimensional parameter β. We refer to the moment model as **M**, which is described by the unconditional moment restrictions $\mathrm{E}^0[m(Z;\alpha)]=0$, where the functional form of $m(\cdot;\alpha)\in R^{d_m}$ is specified up to an unknown finite-dimensional parameter $\alpha\in R^{d_\alpha}$, and the expectation $\mathrm{E}^0[\cdot]$ is taken with respect to the true DGP. For feasibility, we focus on the over-identified case where $d_m \geqslant d_\alpha$.

The KLIC defines a pseudo-distance between a family of distribution functions and the true DGP. A sample estimate of the KLIC of a parametric model can be obtained by combining the maximized sample log-likelihood parametric function with a sieve nonparametric estimate of the population entropy. A sample estimate of the KLIC for the moment condition model can be obtained by the empirical likelihood methods.

2.1. KLIC for parametric models

The KLIC from a family of densities $\{h(Z):h\in\mathscr{T}\}$ to a density $h_0(Z)$ is defined as

$$I(\{h(Z):h\in\mathscr{T}\}|h_0(\cdot))\equiv\min_{h\in\mathscr{T}}\int\left(\log\frac{h_0(z)}{h(z)}\right)h_0(z)\,\mathrm{d}z\geqslant 0,$$

where the equality is achieved at $h(\cdot)=h_0(\cdot)$ only if $h_0\in\mathscr{T}$. The family of densities $\{h(Z):h\in\mathscr{T}\}$ is called misspecified if $h_0\notin\mathscr{T}$, in which case the KLIC is strictly positive. The population KLIC for a parametric likelihood family $\{f(Z;\beta):\beta\in\mathscr{B}\}$, $\mathscr{B}$ a compact subset of R^{d_β}, is

$$I(\{f(\cdot;\beta):\beta\in\mathscr{B}\}|h_0(\cdot))=\min_{\beta\in\mathscr{B}}\int\left(\log\frac{h_0(z)}{f(z;\beta)}\right)h_0(z)\,\mathrm{d}z.$$

The parametric maximum likelihood estimator

$$\hat{\beta}=\arg\max_{\beta\in\mathscr{B}}\frac{1}{n}\sum_{t=1}^{n}\log f(z_t;\beta)$$

is well known to minimize sample versions of the population KLICs. The behavior of the parametric maximum likelihood estimator under misspecification is by now well understood (see, for example, White, 1982). Define the pseudo-true value β^* as

$$\beta^*=\arg\min_{\beta\in\mathscr{B}} I(\{f(\cdot;\beta):\beta\in\mathscr{B}\}|h_0(\cdot)). \tag{2.1}$$

Under mild regularity conditions, $\hat{\beta}$ converges to β^* at the $\sqrt{n}$-rate and is asymptotically normally distributed.

Assumption 1. Assume that the data $\{z_t\}_{t=1}^n$ are i.i.d., $\mathscr{B}$ is a compact subset of R^{d_β} with non-empty interior and the following conditions hold:

1. The solution β^* to (2.1) is unique and belongs to the interior of $\mathscr{B}$.
2. $\log f(z_t;\beta)$, is continuous at $\beta \in \mathscr{B}$ with probability one, and $\mathrm{E}^0[\sup_{\beta\in\mathscr{B}}|\log f(z_t;\beta)|] < \infty$.
3. $\log f(z_t;\beta)$ is twice continuously differentiable at $\beta \in \mathscr{N}(\beta^*)$ with probability one, where $\mathscr{N}(\beta^*)$ is a small neighborhood around β^*, and
$$\mathrm{E}^0\left[\sup_{\beta\in\mathscr{N}(\beta^*)}\left|\frac{\partial^2 \log f(z_t;\beta)}{\partial\beta\partial\beta'}\right|\right] < \infty.$$
4. A_f and Ω_f are finite and non-singular, where
$$A_f \equiv \frac{\partial^2 \mathrm{E}^0\{\log f(z_t;\beta^*)\}}{\partial\beta\partial\beta'}, \quad \Omega_f \equiv \mathrm{E}^0\left[\frac{\partial \log f(z_t,\beta^*)}{\partial\beta}\frac{\partial \log f(z_t,\beta^*)'}{\partial\beta}\right].$$

In the above, E^0 denotes the expectation with respect to $h_0(z)$, the true density function of the data z. The next theorem follows directly from Newey and McFadden (1994, Theorem 2.1, Lemma 2.4, Theorem 3.1), and is therefore stated without a proof.

Theorem 1. *Under Assumption* 1, $\sqrt{n}(\hat{\beta}-\beta^*) \xrightarrow{\mathrm{d}} \mathrm{N}(0, A_f^{-1}\Omega_f A_f^{-1})$, *and*
$$\sum_{t=1}^n(\log f(z_t;\hat{\beta}) - \log f(z_t;\beta^*)) = \mathrm{O_p}(1).$$

It is well known that the conclusion of the above theorem still holds if we replace Assumption 1 by other weaker sufficient conditions.

2.1.1. Sieve nonparametric entropy estimation

The population KLICs for parametric models can be consistently estimated by the difference between the average log-likelihood function and a nonparametric estimate of the entropy of the underlying population, given by
$$\frac{1}{n}\sum_{t=1}^n \log \hat{h}(z_t) - \frac{1}{n}\sum_{t=1}^n \log f(z_t;\hat{\beta}).$$
The first part is a nonparametric estimate of the population entropy
$$\int [\log h_0(z)]h_0(z)\,\mathrm{d}z \equiv \mathrm{E}^0\{\log h_0(Z)\}.$$

In the following, we shall present two classes of sieve maximum likelihood estimation (MLE) of the unknown true density $h_0(z)$. The first class is to use sieves to approximate the log-density function and the second class is to use sieves to approximate the square-root density function. Both classes of sieve density estimators allow for the use of a variety of sieve based functions. For example, Stone (1985) applied spline sieves, Barron and Sheu (1991) used orthogonal polynomials, splines and trigonometric sieves, and Chen and White (1999) considered neural network sieves for approximating log-densities; Chen and Shen

(1998) used trigonometric sieves, Gallant and Nychka (1987) and Coppejans and Gallant (2002) used Hermite polynomials to approximate square-root densities.

Denote the parametric space of the density $h(\cdot)$ by $\mathscr{H}$. A sieve is a sequence of approximating spaces $\mathscr{H}_n$ that is dense in $\mathscr{H}$ as $n \to \infty$, in the sense that for any $h(\cdot) \in \mathscr{H}$, there exists a sequence $\pi_n h \in \mathscr{H}_n$ such that $d(h, \pi_n h) \to 0$ as $n \to \infty$, where $d(\cdot)$ denotes a pseudo-metric.

Let h_0 be the true unknown probability density of Z on its support $\mathscr{Z}$, where $\mathscr{Z}$ is a compact set in R^{d_z} with Lipschitz continuous boundaries. For notational convenience, we use $L(h_0)$ to denote either $\log h_0$ or $\sqrt{h_0}$, and $L(h)$ to denote either $\log h$ or $\sqrt{h}$. In the following, $C^k(\mathscr{Z})$ denotes the space of k-times continuously differentiable functions on $\mathscr{Z}$, and $\frac{\partial^k L(h(z))}{\partial z^k}$ denotes the kth derivative of $L(h)$ with respect to z.

Assumption 2. Let $L(h_0) \in \mathscr{H}$:

$$\mathscr{H} = \left\{ L(h) \in C^k(\mathscr{Z}) : h(z) \geqslant c_0 > 0 \text{ for all } z \in \mathscr{Z}, \text{ and} \right.$$
$$\left. \left| \frac{\partial^k L(h(z_1))}{\partial z^k} - \frac{\partial^k L(h(z_2))}{\partial z^k} \right| \Big/ |z_1 - z_2|^\gamma \leqslant d_h \text{ for all } z_1, z_2 \in \mathscr{Z} \right\},$$

where $k = 0, 1, 2, \ldots$ and $\gamma \in (0, 1]$, and $s \equiv k + \gamma > d_z/2$.

Let $\{p_j(Z), j = 1, 2, \ldots\}$ denote a sequence of known basis functions, such as Fourier series, Hermite polynomials, splines of order $k+1$, etc., that can approximate any real-valued square-integrable function of Z arbitrarily well. Then a finite-dimensional linear sieve for $\mathscr{H}$ is

$$\mathscr{H}_n = \left\{ L(h) \in \mathscr{H} : L(h(z)) = \sum_{j=1}^{k_n} a_j p_j(z), a_j \in R \right\},$$

with $\dim(\mathscr{H}_n) = k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

An artificial neural networks (ANN) sieve is attractive when the dimension d_z is greater than or equal to 3 (see e.g. Chen and White, 1999 and the reference therein). A simple ANN sieve for $\mathscr{H}$ is

$$\mathscr{H}_n = \left\{ L(h) \in \mathscr{H} : L(h(z)) = b_0 + \sum_{j=1}^{k_n} b_j \psi(a_j' z + a_{0,j}), \sum_{j=1}^{k_n} \sum_{i=0}^{d_z} |a_{i,j}| \leqslant k_n \log n, \right.$$
$$\left. \sum_{j=0}^{k_n} |b_j| \leqslant \log(n) \right\},$$

where $\psi()$ is a logistic function, and $a_j = (a_{1,j}, \ldots, a_{d_z,j})'$.

Log-density estimation: Suppose that we want to estimate $\log h_0$, the log-density. Since $\log h_0$ is subject to the non-linear constraint $\int_{\mathscr{Z}} \exp\{\log h_0(z)\}\,\mathrm{d}z = 1$, it is more convenient to write $\log h_0 = \theta_0 - \log \int_{\mathscr{Z}} \exp\{\theta_0(z)\}\,\mathrm{d}z$, and treat θ_0 as an unknown function in some linear space Θ:

$$\theta_0 \in \Theta = \left\{ \theta \in \mathscr{H} : \mathrm{E}[\theta(Z)^2] < \infty, \int_{\mathscr{Z}} \theta(z)\,\mathrm{d}z = 0 \right\}.$$

Let Θ_n denote some sieve space for Θ:

$$\Theta_n = \left\{ \theta \in \mathscr{H}_n : \int_{\mathscr{Z}} \theta(z)\,\mathrm{d}z = 0 \right\},$$

where $\mathscr{H}_n$ can be any of the popular sieve spaces for $\mathscr{H}$. Then

$$\widehat{\theta} = \arg\max_{\theta \in \Theta_n} \frac{1}{n} \sum_{t=1}^{n} \left[\theta(z_t) - \log \int_{\mathscr{Z}} \exp \theta(z)\,\mathrm{d}z \right]$$

is a sieve maximum likelihood estimator of θ_0, and we have $\log \hat{h} = \widehat{\theta} - \log \int_{\mathscr{Z}} \exp\{\widehat{\theta}(z)\}\,\mathrm{d}z$.

Square-root density estimation: Suppose that we want to estimate the square root density $\theta_0 = \sqrt{h_0} \in \mathscr{H}$, then

$$\theta_0 \in \Theta = \left\{ \theta \in \mathscr{H} : \int_{\mathscr{Z}} [\theta(z)]^2\,\mathrm{d}z = 1 \right\}.$$

Let Θ_n denote some sieve space for Θ:

$$\Theta_n = \left\{ \theta \in \mathscr{H}_n : \int_{\mathscr{Z}} [\theta(z)]^2\,\mathrm{d}z = 1 \right\},$$

where $\mathscr{H}_n$ be any of the popular sieve spaces for $\mathscr{H}$. Then

$$\widehat{\theta} = \arg\max_{\theta \in \Theta_n} \frac{1}{n} \sum_{t=1}^{n} \log [\theta(z_t)]^2$$

is a sieve maximum likelihood estimator of θ_0, and we have $\hat{h} = [\widehat{\theta}]^2$.

Theorem 2. *Assume that the data* $\{z_t\}_{t=1}^n$ *are i.i.d. and* $Var^0[\log h_0(Z)]$ *is positive finite. Suppose* $h_0(z)$ *satisfies Assumption* 2. *Let* $\hat{h}(z)$ *be the nonparametric estimator of the density using the sieve MLE of either the log-density or the squared-root density. If* $k_n = \mathrm{O}(n^{1/(2s+d_z)})$ *for the linear sieve and* $(k_n)^{2+1/d_z} \log(k_n) = \mathrm{O}(n)$ *for ANN sieves, then*

$$\sqrt{n}\left(\frac{1}{n} \sum_{t=1}^{n} \log \hat{h}(z_t) - \mathrm{E}^0 \log h_0(Z) \right) = \sqrt{n}\left(\frac{1}{n} \sum_{t=1}^{n} \log h_0(z_t) - \mathrm{E}^0 \log h_0(Z) \right) + \mathrm{o_p}(1)$$
$$\xrightarrow{\mathrm{d}} \mathrm{N}(0, Var^0[\log h_0(Z)]).$$

Proof of Theorem 2. Under the stated conditions, by Example 3.7, Remark 3.8 and Example 3.9 in Chen and White (1999) for ANN sieve estimation of log-densities and square-root densities, and by the results in Chen and Shen (1998) for linear sieve estimation of squared root-density, we have the following two results:

$$\frac{1}{n} \sum_{t=1}^{n} (\log \hat{h}(z_t) - \log h_0(z_t)) - \mathrm{E}^0(\log \hat{h}(Z) - \log h_0(Z)) = \mathrm{o_p}(n^{-1/2})$$

and

$$\mathrm{E}^0(\log \hat{h}(Z) - \log h_0(Z)) = \mathrm{o_p}(n^{-1/2}).$$

The theorem now follows. $\square$

2.2. KLIC for moment condition models

Moment condition-based models allow researchers to focus on the main features of the econometric models without facing a high risk of misspecifying a tight parametric likelihood function. A large recent econometric literature have studied the information theoretic features of moment condition-based estimation, and developed a variety of generalized empirical likelihood estimators (see Kitamura and Stutzer, 1997; Kitamura, 1997; Imbens et al., 1998; Kitamura and Tripathi, 2001; Newey and Smith, 2001). A set of overidentified estimating equations imply a nonparametric family of probability measures that are compatible with these moment restrictions.

For each $\alpha \in \mathscr{A}$, define $\mathscr{Q}_\alpha = \{q(z) | \int m(z;\alpha)q(z)\,\mathrm{d}z = 0\}$, a set of probability densities for Z which are consistent with the moment conditions. Then we can define $\mathbf{M} = \cup_{\alpha\in\mathscr{A}}\mathscr{Q}_\alpha$ as the unconditional moment condition-based model. The model $\mathbf{M}$ is *misspecified* if the true population distribution $h_0(\cdot) \notin \mathbf{M}$.

An empirical likelihood procedure will identify an element in $\mathbf{M}$ that is closest to the true distribution h_0, in the sense of $\inf_{q(\cdot)\in\mathbf{M}} I(q(\cdot)|h_0(\cdot))$. The solution to this population KLIC problem can be characterized in two steps (see Appendix A for more details). First, for a fixed parameter vector $\alpha \in \mathscr{A}$,

$$\inf_{q(\cdot)\in\mathscr{Q}_\alpha} I(q(\cdot)|h_0(\cdot)) = I(q^*_\alpha(\cdot)|h_0(\cdot)) = \max_{\lambda\in\Lambda} \int \log(1+\lambda' m(z;\alpha))h_0(z)\,\mathrm{d}z$$
$$\equiv \int \log(1+\lambda(\alpha)' m(z;\alpha))h_0(z)\,\mathrm{d}z,$$

and the implied distribution $q^*_\alpha(\cdot)$ that minimizes the KLIC is given by

$$q^*_\alpha(Z) = \frac{h_0(Z)}{(1+\lambda(\alpha)' m(Z;\alpha))}.$$

The population empirical likelihood problem can then be stated as

$$\inf_{q(\cdot)\in\mathbf{M}} I(q(\cdot)|h_0(\cdot)) = \min_{\alpha\in\mathscr{A}} \int \log(1+\lambda(\alpha)' m(z;\alpha))h_0(z)\,\mathrm{d}z$$
$$= \min_{\alpha\in\mathscr{A}} \max_{\lambda\in\Lambda} \int \log(1+\lambda' m(z;\alpha))h_0(z)\,\mathrm{d}z$$
$$\equiv \int \log(1+\lambda^{*\prime} m(z;\alpha^*))h_0(z)\,\mathrm{d}z,$$

with $\lambda^* \equiv \lambda(\alpha^*)$. The empirical likelihood estimator optimizes the following sample analog of the population saddle-point objective function (cf. Qin and Lawless, 1994; Imbens et al., 1998):

$$(\hat{\alpha}, \hat{\lambda}) = \arg\min_{\alpha\in\mathscr{A}} \arg\max_{\lambda\in\Lambda} \hat{M}(\lambda,\alpha),$$

where

$$\hat{M}(\lambda,\alpha) \equiv \frac{1}{n}\sum_{t=1}^{n} \log(1+\lambda' m(z_t;\alpha)).$$

The implied empirical distribution subject to the moment constraints is given by

$$\hat{Q}(z) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}(z_t \leqslant z)/(1+\hat{\lambda}' m(z_t;\hat{\alpha})).$$

Under mild regularity conditions, $\hat{\alpha} \xrightarrow{p} \alpha^*$ and $\hat{\lambda} \xrightarrow{p} \lambda^*$. Both $\sqrt{n}(\hat{\alpha} - \alpha^*)$ and $\sqrt{n}(\hat{\lambda} - \lambda^*)$ are asymptotically normal (see Kitamura, 2000). These are formally stated in the following theorem. First, we list the assumptions.

Assumption 3. Assuming that the data $\{z_t\}_{t=1}^n$ are i.i.d., and:

1. $\mathscr{A} \subset R^{d_\alpha}$, $\lambda \in \Lambda \subset R^{d_m}$. Both $\mathscr{A}$ and Λ are compact, and
$$\sup_{\alpha \in \mathscr{A}, z \in \mathscr{Z}} |m(z;\alpha)| < \infty \quad \text{and} \quad \inf_{\alpha \in \mathscr{A}, \lambda \in \Lambda, z \in \mathscr{Z}} \{1 + \lambda' m(z;\alpha)\} > 0.$$
2. $\mathrm{E}^0\{m(z_t;\alpha)m(z_t;\alpha)'\}$ is positive definite uniformly over $\alpha \in \mathscr{A}$.
3. $M(\lambda, \alpha) \equiv \mathrm{E}^0\{\log(1 + \lambda' m(z_t;\alpha))\}$ is continuously differentiable over Λ and $\mathscr{A}$.
4. The problem of $\min_{\alpha \in \mathscr{A}} \max_{\lambda \in \Lambda} M(\lambda, \alpha)$ has a unique saddle point solution (α^*, λ^*) with $\alpha^* \in \mathrm{int}(\mathscr{A})$ and $\lambda^* \in \mathrm{int}(\Lambda)$.
5. $D = \frac{\partial^2 M(\lambda,\alpha)}{\partial \lambda \partial \alpha'}|_{(\alpha^*,\lambda^*)}$ is of full column rank.
6. $V = Var^0\left(\frac{m(z_t;\alpha^*)}{1+\lambda^{*\prime} m(z_t;\alpha^*)}\right) = \mathrm{E}^0 \frac{m(z_t;\alpha^*)m(z_t;\alpha^*)'}{(1+\lambda^{*\prime} m(z_t;\alpha^*))^2}$ is finite and positive definite.
7. $\sup_{\alpha \in \mathscr{A}} \sup_{\lambda \in \Lambda} |\hat{M}(\lambda,\alpha) - M(\lambda,\alpha)| = \mathrm{o_p}(1)$.
8. Stochastic equi-continuity for sample moment conditions. For any $\delta_n \to 0$,
$$\sup_{|\alpha-\alpha^*| \leqslant \delta_n} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{m(z_t;\alpha)}{1+\lambda^{*\prime} m(z_t;\alpha)} - \frac{m(z_t;\alpha^*)}{1+\lambda^{*\prime} m(z_t;\alpha^*)} - \mathrm{E}^0 \frac{m(z_t;\alpha)}{1+\lambda^{*\prime} m(z_t;\alpha)} \right) \right| = \mathrm{o_p}(1).$$
9. Stochastic equi-continuity for KLIC saddle-point objective function. For any $\delta_n \to 0$,
$$\sup_{|\alpha-\alpha^*| \leqslant \delta_n, |\lambda-\lambda^*| \leqslant \delta_n} \sqrt{n} |\hat{M}(\lambda,\alpha) - \hat{M}(\lambda^*,\alpha^*) - [M(\lambda,\alpha) - M(\lambda^*,\alpha^*)]| = \mathrm{o_p}(1).$$

Remark 1. The uniform boundedness condition (1) on $m(z_t;\alpha)$ is doubtless very strong. We assume it now to avoid the technical difficulty that the non-negativity condition that $1 + \lambda' m(z_t;\alpha) > 0$ may reduce Λ to a singleton (zero) when $m(z_t;\alpha)$ is unbounded. Although Newey and Smith (2001) is able to relax this condition for correctly specified moment models, we suspect that attempts to relax this assumption for misspecified models may be technically difficult. In practice, whether or not this condition is satisfied depends on both z and the functional form of $m(z,\alpha)$. The actual specification of m does offer the user some flexibility in terms of the assumptions they would make in practice.

Remark 2. The interior point condition (4) of (λ^*, α^*) requires the moment condition model not to be too misspecified. The interior point condition implies that

$$\mathrm{E}^0\{m(z_t;\alpha^*)/(1 + \lambda^{*\prime} m(z_t;\alpha^*))\} = 0,$$

which rules out cases where $m(z;\alpha) > 0$ with probability one for all $\alpha \in \mathscr{A}$. Essentially, we require that the set $\mathbf{M}$ be non-empty.

Theorem 3. *Under Assumption* 3, $\sqrt{n}(\hat{\alpha}-\alpha^*)=O_p(1)$, $\sqrt{n}(\hat{\lambda}-\lambda^*)=O_p(1)$, *and*

$$\sum_{t=1}^{n}(\log(1+\hat{\lambda}'m(z_t;\hat{\alpha}))-\log(1+\lambda^{*\prime}m(z_t;\alpha^*)))=O_p(1).$$

3. Nonparametric selection test between two models

Given the KLIC measures for the parametric and moment condition-based models, their difference indicates which model is closer to the true DGP. Define

$$LR_\infty(\beta^*,h_0;\lambda^*,\alpha^*)=E^0\log h_0(Z)-E^0\log f(Z;\beta^*)-E^0\log(1+\lambda^{*\prime}m(Z;\alpha^*)),$$

and consider the following sets of null and alternative hypotheses:

$$H_0: LR_\infty(\beta^*,h_0;\lambda^*,\alpha^*)=0,$$

meaning that model **F** and model **M** are equivalent, against

$$H_g: LR_\infty(\beta^*,h_0;\lambda^*,\alpha^*)<0,$$

meaning that model **F** is better than model **M**, or

$$H_m: LR_\infty(\beta^*,h_0;\lambda^*,\alpha^*)>0,$$

meaning that model **M** is better than model **F**, in the sense of providing a better (smaller) KLIC to the data.

Hence a nonparametric quasi-likelihood ratio test statistic between a parametric model and a moment condition model can be formulated as the sample analog of these population quantities

$$LR_n(\hat{\beta},\hat{h};\hat{\lambda},\hat{\alpha})=\frac{1}{n}\sum_{t=1}^{n}\log\hat{h}(z_t)-\frac{1}{n}\sum_{t=1}^{n}\log f(z_t;\hat{\beta})-\frac{1}{n}\sum_{t=1}^{n}\log(1+\hat{\lambda}'m(z_t,\hat{\alpha})).$$

For example, a statistically significantly negative value of $LR_n(\hat{\beta},\hat{h};\hat{\lambda},\hat{\alpha})$ indicates that the model **F** is likely to be better than model **M**.

Remark 3. In some applications, the parametric likelihood model and the moment restriction model may have different numbers of unknown parameters. To take this into account, we follow the approach in Sin and White (1996) by adopting a general penalization of model complexity. Let $Pen(d_\beta,n)$ denote a penalization term such that $Pen(d_\beta,n)$ increases with d_β, decreases with n, and $Pen(d_\beta,n)/n\to 0$. Then the penalized LR statistic is

$$\begin{aligned}PLR_n(\hat{\beta},\hat{h};\hat{\lambda},\hat{\alpha})&=LR_n(\hat{\beta},\hat{h};\hat{\lambda},\hat{\alpha})+\frac{Pen(d_\beta,n)-Pen(d_\alpha,n)}{n}\\&=LR_n(\hat{\beta},\hat{h};\hat{\lambda},\hat{\alpha})+o_p(1).\end{aligned}$$

We note that $Pen(d_\beta,n)=d_\beta$ corresponds to AIC, and $Pen(d_\beta,n)=0.5d_\beta\log n$ corresponds to SIC criterion.

3.1. Large sample properties of the test statistics

The limit distribution of the test statistic $\mathrm{LR}_n(\hat{\beta}, \widehat{h}; \hat{\lambda}, \hat{\alpha})$ is an immediate consequence of Theorems 1–3. In the following we denote

$$v_g(z_t) = (\log h_0(z_t) - \mathrm{E}^0 \log h_0(z_t)) - (\log f(z_t; \beta^*) - \mathrm{E}^0 \log f(z_t; \beta^*)) \\ - (\log(1 + \lambda^{*\prime} m(z_t; \alpha^*)) - \mathrm{E}^0 \log(1 + \lambda^{*\prime} m(z_t; \alpha^*))).$$

Theorem 4. *Under conditions for Theorems* 1–3,

$$\sqrt{n}(\mathrm{LR}_n(\hat{\beta}, \widehat{h}; \hat{\lambda}, \hat{\alpha}) - \mathrm{LR}_\infty(\beta^*, h_0; \lambda^*, \alpha^*)) \\ \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_g^2) \quad \textit{for } \sigma_g^2 = \mathrm{E}^0[v_g(z_t)^2].$$

Suppose further $\frac{Pen(d_\beta, n) - Pen(d_\alpha, n)}{\sqrt{n}} = \mathrm{o_p}(1)$, *then*

$$\sqrt{n}(\mathrm{PLR}_n(\hat{\beta}, \widehat{h}; \hat{\lambda}, \hat{\alpha}) - \mathrm{LR}_\infty(\beta^*, h_0; \lambda^*, \alpha^*)) \to^{\mathrm{d}} \mathrm{N}(0, \sigma_g^2).$$

This result follows directly from the fact that due to Theorems 1–3,

$$\sqrt{n}(\mathrm{LR}_n(\hat{\beta}, \widehat{h}; \hat{\lambda}, \hat{\alpha}) - \mathrm{LR}_\infty(\beta^*, h_0; \lambda^*, \alpha^*)) = \frac{1}{\sqrt{n}} \sum_{t=1}^n v_g(z_t) + \mathrm{o_p}(1).$$

The asymptotic distribution described in Theorem 4 can be used to test the null hypothesis of H_0 when the limiting variance is not degenerate. Degeneracy occurs if and only if $v_g(Z) = 0$ with Z-probability one, which occurs when

$$\mathrm{H}_E : f(z; \beta^*) \equiv \frac{h_0(z)}{(1 + \lambda^{*\prime} m(z; \alpha^*))} \quad \text{for almost all } z,$$

namely that the probability measure induced by the parametric family coincides with the probability measure induced by the moment functions. When this happens, the two models are observationally equivalent, even if they can be both misspecified.

A consistent estimate of σ_g^2 can be formed by $\widehat{\sigma}_g^2 = \frac{1}{n} \sum_{t=1}^n [\hat{v}_g(z_t)]^2$, where

$$\hat{v}_g(z_t) = \left(\log \widehat{h}(z_t) - \frac{1}{n} \sum_{j=1}^n \log \widehat{h}(z_j) \right) - \left(\log f(z_t; \widehat{\beta}) - \frac{1}{n} \sum_{j=1}^n \log f(z_j; \widehat{\beta}) \right) \\ - \left(\log(1 + \hat{\lambda}' m(z_t, \hat{\alpha})) - \frac{1}{n} \sum_{j=1}^n \log(1 + \hat{\lambda}' m(z_j, \hat{\alpha})) \right).$$

Lemma 1. *Under conditions for Theorems* 1–3, $\widehat{\sigma}_g^2 \xrightarrow{\mathrm{p}} \sigma_g^2$.

Similar to Vuong (1989), the asymptotic distribution of this likelihood-ratio-like statistic is normal. This differs from the familiar case where the competing models are nested, and the unrestricted model is correctly specified. For this case, the leading term in the Taylor expansion of the likelihood ratio statistic is the second-order term, leading to a limiting χ^2 distribution. On the other hand, in the non-nested case, the difference between the sample log-likelihood functions of the competing models determines the asymptotic distribution, leading to a limiting normal distribution via a central limit theorem argument. For further discussion, see Vuong (1989, pp. 313–314).

3.2. Testing the degeneracy condition

The degeneracy hypothesis H_E can be tested. For comparing between two parametric models, Vuong (1989) used the variance estimate $\widehat{\sigma}_g^2$ as the test statistic and obtained its distribution in a manner related to a second order expansion of the statistic $\mathrm{LR}_n(\hat{\beta}, \hat{h}; \hat{\lambda}, \hat{\alpha})$. While a similar strategy can be used here, the statistical properties are more complicated than the Vuong (1989) test. Therefore we consider more direct tests of the equality between the two implied distributions. We consider both density- and CDF-based test statistics for the degeneracy hypothesis.

Density-based degeneracy test statistic: The degeneracy hypothesis H_E implies that the pseudo-density functions should be equal to each other: $h_0(z) = f(z;\beta^*)(1 + \lambda^{*\prime} m(z;\alpha^*))$. Therefore a test statistic can be based on an empirical analog of a norm of the difference between these two densities. For example, a von Mises type statistic can be formulated as

$$V_n = \int [f(z;\hat{\beta})(1 + \hat{\lambda}' m(z,\hat{\alpha})) - \widetilde{h}(z)]^2 \, \mathrm{d}z, \tag{3.2}$$

where $\widetilde{h}(z)$ is a nonparametric density estimate using either kernel- or sieve-based methods. The results of Fan (1994) can be used to describe the asymptotic distribution of V_n when $\widetilde{h}(z)$ is a kernel density estimate.[5] A drawback of the density-based degeneracy test is that it converges at the slower nonparametric rate, therefore eroding power of the sequential testing procedure. Appendix C gives some more details.

CDF-based degeneracy test statistics: Next we focus on CDF-based tests for the degeneracy hypotheses H_E using the implied empirical distribution functions. These tests converge faster than the density-based statistics, and therefore are more powerful than those based on density estimates. However, the faster convergence rate also implies that the sampling variation introduced by using the estimated $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\alpha}$ in computing the test statistic needs to be taken into account for valid large sample inferences. Note that the degeneracy hypothesis can alternatively be stated as

$$\mathrm{H}_E : \int^Z h(z)\,\mathrm{d}z \equiv \int^Z f(z;\beta^*)(1 + \lambda^{*\prime} m(z;\alpha^*))\,\mathrm{d}z.$$

Thus a consistent test can be based on some notation of functional distance between the empirical distribution and its implied parametric form:

$$\rho\left(\frac{1}{n}\sum_{t=1}^{n} 1(z_t \leqslant Z) - \int^Z f(z;\hat{\beta})(1 + \hat{\lambda}' m(z;\hat{\alpha}))\,\mathrm{d}z\right),$$

where $\rho(\cdot)$ is a pseudo-metric that is bounded and continuous with respect to the L_1 norm. These include both Kolmogorov–Smirnov type sup-norm statistic

$$\rho_1 = \sup_Z \left| \frac{1}{n}\sum_{i=1}^{n} 1(z_i \leqslant Z) - \int^Z f(z,\hat{\beta})(1 + \hat{\lambda}' m(z,\hat{\alpha}))\,\mathrm{d}z \right|, \tag{3.3}$$

[5] If the variance v_g is estimated using kernel density methods, then Vuong's variance test will be closely related to the kernel based von-Mises statistic. We thank Yuichi Kitamura for pointing this out to us. Kitamura (2002) uses sample splitting as an alternative to the degeneracy test.

and Cramer–von Mises type integrated square statistic

$$\rho_2 = \int \left(\frac{1}{n} \sum_{i=1}^{n} 1(z_i \leqslant Z) - \int^{Z} f(z, \hat{\beta})(1 + \hat{\lambda}' m(u, \hat{\alpha}))\, \mathrm{d}u \right)^2 w(Z)\, \mathrm{d}Z, \tag{3.4}$$

where $w(Z)$ is a weighting function. Appendix D gives more details about the asymptotic distributions of the empirical CDF based tests of nondegeneracy.

3.3. Nested, non-nested and overlapping models

Since $\mathrm{H}_E \subset \mathrm{H}_0$, if we fail to reject H_E, then we are done. On the other hand, if we reject H_E, H_0 may still hold, in which case Theorem 4 provides the basis for a directional test for H_0 vs H_f and H_m. Using the classification method of Vuong (1989), the relation between the parametric model and the moment model can be nested, non-nested or overlapping.

The parametric model and the moment-based model are strictly non-nested if and only if $\mathbf{F} \cap \mathbf{M} = \emptyset$. Formally this holds if and only if

$$\forall \beta \in \mathscr{B}, \ \forall \alpha \in \mathscr{A}, \quad \int m(z;\alpha) f(z;\beta)\, \mathrm{d}z \neq 0.$$

Since model **F** and model **M** have no common implied probability measure, the degeneracy hypothesis H_E cannot hold for this case. Hence Theorem 4 can be applied.

Models **M** and **F** are overlapping if and only if (i) $\mathbf{M} \cap \mathbf{F} \neq \emptyset$; (ii) $\mathbf{M} \not\subset \mathbf{F}$ and $\mathbf{F} \not\subset \mathbf{M}$. This holds if and only if $\mathscr{B} = \mathscr{B}_1 \cup \mathscr{B}_2$ such that

$$\forall \beta \in \mathscr{B}_1, \ \exists \alpha \in \mathscr{A}, \quad \int m(z;\alpha) f(z;\beta)\, \mathrm{d}z = 0,$$

$$\forall \beta \in \mathscr{B}_2, \ \forall \alpha \in \mathscr{A}, \quad \int m(z;\alpha) f(z;\beta)\, \mathrm{d}z \neq 0,$$

$$\exists P \in \mathbf{M} \ \text{such that} \ P \notin \mathbf{F}.$$

Since $\mathbf{F} \cap \mathbf{M} \neq \emptyset$, the observational equivalence relation may or may not hold. Thus a general sequential procedure consists in testing first whether H_E holds. In the second step, if H_E is not rejected, we can conclude that **F** and **M** cannot be discriminated given the data. On the other hand, if H_E is rejected, H_0 may still hold, and we can proceed by using Theorem 4 to test for H_0. The significance level for the sequential testing procedure is bounded from above by those of each of the two steps.

The two models can also be nested. It is probably more common for the moment-based model to nest the fully parametric model than for the fully parametric model to nest the moment-based model. Model **M** nests model **F** if and only if $\mathbf{F} \subset \mathbf{M}$, namely

$$\forall \beta \in \mathscr{B}, \ \exists \alpha \in \mathscr{A} \quad \text{such that} \int m(z;\alpha) f(z;\beta)\, \mathrm{d}z = 0.$$

For nested models, the pseudo-true measure $h_0(z)/(1 + \lambda^{*\prime} m(z;\alpha^*))$ may not necessarily equal $f(z;\beta^*)$ since it is may not belong to $\{f(\cdot;\beta) : \beta \in \mathscr{B}\}$. However, H_E and H_0 are equivalent statements for nested models. A proof of this result would follow the same reasoning as Lemma 7.1 in Vuong (1989). Given this equivalence, we can use the degeneracy test to test for H_0.

4. Monte Carlo simulations

In this section we report the results from a small Monte Carlo simulation exercise to illustrate the finite sample properties of the above testing procedure. We specify the true DGP to be a standard normal distribution. The data in the experiments are drawn from this distribution. We consider two unconditional competing models, one parametric and one based on moment conditions, such that neither of them nests the true normal distribution (so that both are misspecified).

The unconditional moment model is specified to be the first six moments calculated from the double exponential distribution density, which takes the form of $g(z;\alpha) = \frac{1}{2\alpha_1}\exp(-\frac{|(z-\alpha_2)|}{\alpha_1})$. In other words, $m(z;\alpha) = (m_k(z;\alpha),\ k = 1,\ldots,6)$ has six components, for $k = 1, 2, 3$:

$$m_{2k}(z;\alpha) = |z-\alpha_2|^{2k} - (2k)!\alpha_1^{2k},$$
$$m_{2k+1}(z;\alpha) = |z-\alpha_2|^{2k+1}.$$

The KLIC will be considered a function of the location parameter α_2, while the scale parameter α_1 is a nuisance parameter that is preset to a certain value of $\frac{1}{6}$ as described below.

The unconditional parametric model $f(z;\beta)$ is specified to be a double Weibull distribution:

$$f(z;\beta) = \frac{\beta_2}{2\beta_1}|z-\beta_3|^{\beta_2-1}\exp\left(-\frac{|z-\beta_3|^{\beta_2}}{\beta_1}\right), \quad -\infty < z < \infty.$$

The shape parameters β_1 and β_2 are considered nuisance parameters that are prefixed. The KLIC will be considered a function of the location parameter β_3. Analytic derivations show that for all fixed values of α_1, β_1 and β_2, the KLICs are minimized at $\alpha_2 = 0$ and $\beta_3 = 0$, respectively.

4.1. Final-sample size

The first set of experiments studies the size of the proposed test. In order to do this, we impose the null by calibrating the parameters $\alpha_1, \beta_1, \beta_2$ such that the two competing models are equi-distant (in terms of the KLIC) from the true normal distribution. Hence, in this design, although both models are misspecified, the null that both models fit equally well (in terms of KLIC) holds, enabling us to consider the empirical rejection probability for small samples.

Given that the KLICs are minimized at $\alpha_2 = 0$ and $\beta_3 = 0$, we calibrate the parameters $\alpha_1, \beta_1, \beta_2$ in the following way. First of all, we fix $\alpha_1 = \frac{1}{6}$ and find, via Monte Carlo simulations (with 10 million simulation draws) to obtain the KLIC between the moment model $m(z;\alpha) = \{m_k(z;\alpha),\ k = 1,\ldots,6\}$ and the standard normal distribution. Through this numerical procedure we also found the vector of Lagrange multiplier associated with these moment conditions, which are

$$\lambda = (-0.1377,\ 75.3769,\ 3.4499,\ 11.4009,\ -1.9334,\ 58.6049)'.$$

Next, we fix $\beta_3 = 0$ and one of the shape parameters of the double Weibull distribution $\beta_2 = 3$. Then, using Monte Carlo simulations and a numerical root finder, we found the value of β_1 that equates the KLIC between the double Weibull distribution with ($\beta_2 = 3, \beta_3 = 0$) and the true DGP of standard normal distribution with the previously found KLIC between the moment model $m(z;\alpha)$ and the true DGP of standard normal distribution. The value of β_1 that equates the two KLICs turns out to be $\beta_1 = 0.2554$. For these values of $\alpha_1, \beta_1, \beta_2$, the KLICs for the two competing models are identical, and equal to 3.9945.

To illustrate, Fig. 1 shows the KLICs as functions of α_2 and β_3, using the prespecified values of the nuisance parameters α_1, β_1 and β_2. It also presents a mesh plot of the difference in KLICs as a function of α_2 and β_3. Fig. 1 clearly shows that the KLICs are minimized at $\beta_3 = 0$ and $\alpha_2 = 0$, respectively.

In running the size simulations, we generate data sets of sizes 100 and 200, respectively, and repeat the experiments 10,000 times. In each experiment, the parameter estimates $\hat{\alpha}_2$ and $\hat{\beta}_3$ are obtained through empirical likelihood methods and maximum likelihood numerically, and the test statistics are computed. The nonparametric entropy estimate was obtained using cardinal B-spline wavelet functions.

Table 1 tabulates the empirical rejection rates of these 10,000 experiments. It is clear from the table that the empirical rejection rates are close to the ones predicted by the asymptotic theory. The accuracy of the asymptotic theory increases with sample size.

4.2. Finite-sample power properties

In the second set of Monte Carlo experiments, we examine the finite sample power properties of the testing procedure. The setups of the experiments are identical to the first one, except that we calibrate β_1 at different values so that the minimized population KLICs for the two competing models are no longer identical. In Figs. 2–4 we report the rejection frequencies under the alternatives where the differences in the KLICs are non-zero. In these calculations we vary the sample sizes from 100 to 2000.

Fig. 2 corresponds to the case where $\beta_1 = 0.26$ and the difference in the KLICs is -0.41. This alternative is very close to the null hypothesis. We can see that for smaller sample sizes there is a sizable distortion to the power curve. The rejection frequencies start to be a monotonically increasing function of the sample size for larger size samples. But for the 1% test, the distortion to the power function is still visible even at sample size 2000.

Fig. 3 corresponds to the case where $\beta_1 = 0.30$ and the difference in the KLICs is -0.95. This alternative is further away from the null distribution. Obviously the results improve over those in Fig. 2. There is still a visible distortion to the power functions for small size samples, but such distortions vanish as soon as the sample size becomes moderately large, even for the 1% size test.

Fig. 4 corresponds to the case where $\beta_1 = 0.40$ and the difference in the KLICs is -1.39. This alternative is quite far from the null distribution. At these alternative parameter values the distortion of the power function completely vanishes. The rejection rates increase monotonically as the sample size increases. In particular, the power curve for the 1% test is very similar to the power curve for the other two tests.

5. Multiple model comparison

In this section, we extend our procedure to handle the comparison of multiple (>2) models, following the idea of White (2000). In this case, all the candidate models are

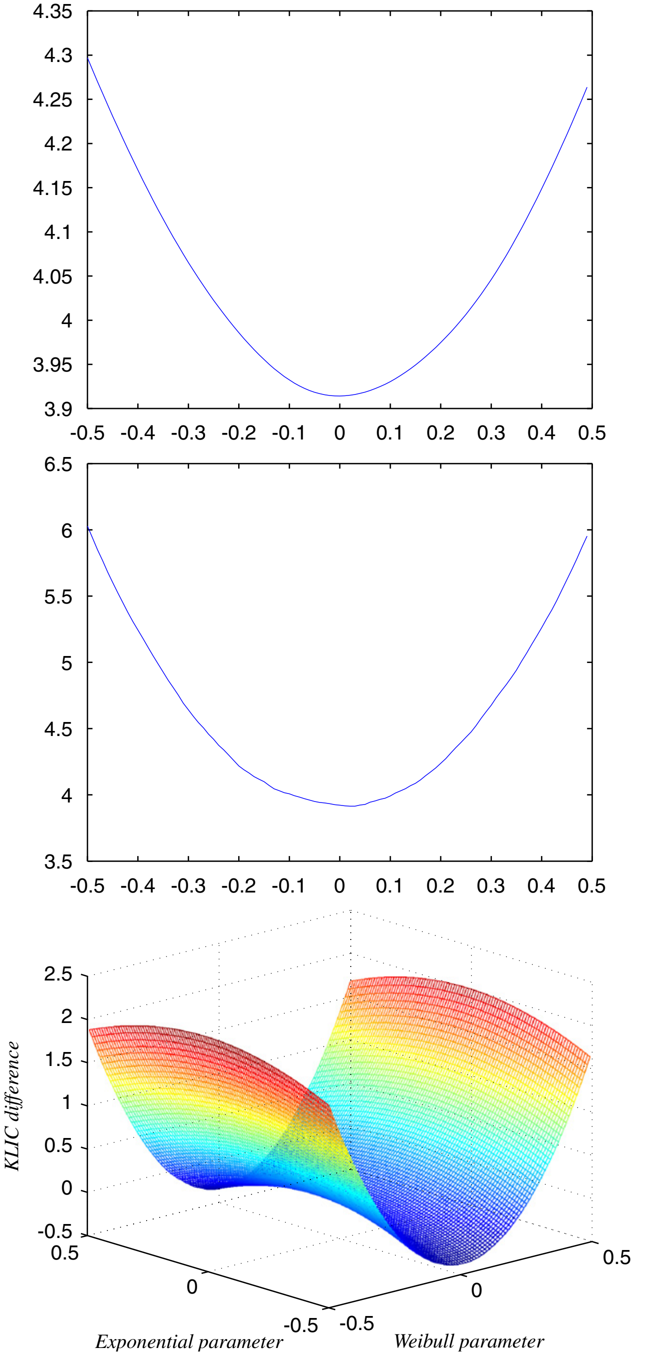

Fig. 1. KLICs as functions of the Exponential and Weibull parameters.

Table 1
Monte Carlo results

Sample size	$z_{1\%}$ (%)	$z_{5\%}$ (%)	$z_{10\%}$ (%)
$n = 100$	3.40	7.87	12.79
$n = 200$	1.52	5.41	11.89

Empirical rejection probabilities under null test sizes for the corresponding critical values.

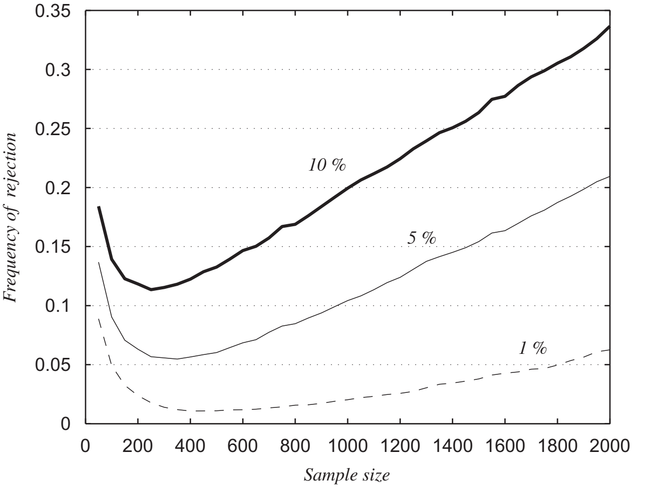

Fig. 2. Rejection probabilities under alternatives: $\beta_3 = 0.26$, $KLIC = -0.41$.

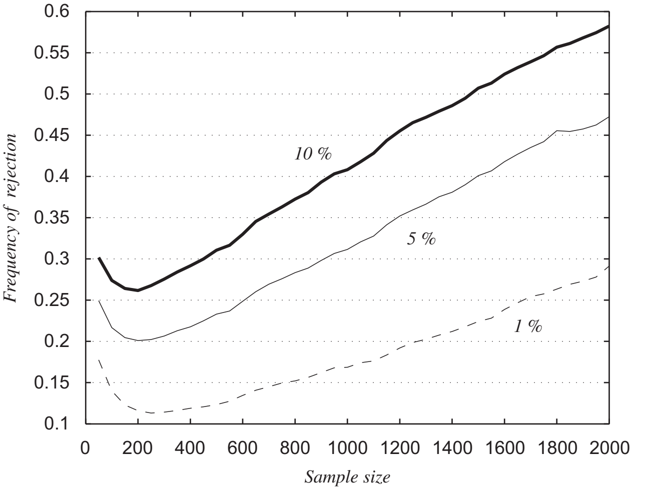

Fig. 3. Rejection probabilities under alternatives: $\beta_3 = 0.30$, $KLIC = -0.95$.

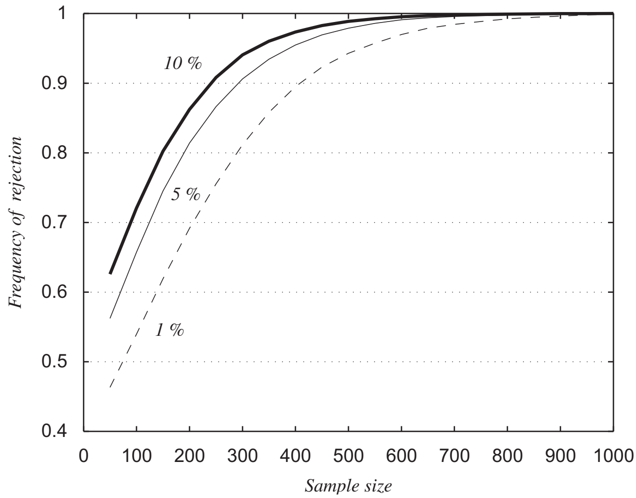

Fig. 4. Rejection probabilities under alternatives: $\beta_3 = 0.40$, $KLIC = -1.39$.

compared with a benchmark model. If no candidate model is closer to the true model than the benchmark model in terms of KLIC, the benchmark model is chosen; otherwise, the candidate model that is closest to the true model in terms of KLIC will be selected.

In the following we denote I_0 as the KLIC associated with the benchmark model. When the benchmark model is a parametric model $\{f_0(z;\beta_0) : \beta_0 \in \mathscr{B}_0\}$, then

$$\begin{aligned} I_0 = I(\beta_0^*|h_0) &= \mathrm{E}^0\{\log h_0(Z) - \log f_0(Z;\beta_0^*)\} \\ &= \min_{\beta_0\in\mathscr{B}_0} \mathrm{E}^0\{\log h_0(Z) - \log f_0(Z;\beta_0)\}. \end{aligned}$$

When the benchmark model is a moment-based model $\{m_0(\cdot;\alpha_0) : \alpha_0 \in \mathscr{A}_0\}$, then

$$\begin{aligned} I_0 = I(\lambda_0^*,\alpha_0^*|h_0) &= \mathrm{E}^0\{\log(1+\lambda_0^{*\prime} m_0(Z;\alpha_0^*))\} \\ &= \min_{\alpha_0\in\mathscr{A}_0} \max_{\lambda_0\in\Lambda_0} \mathrm{E}^0\{\log(1+\lambda_0' m_0(Z;\alpha_0))\}. \end{aligned}$$

Let $\{f_i(z;\beta_i)\}_{i=1}^{M_f}$ be the candidate parametric models and $\{m_i(z;\alpha_i)\}_{i=1}^{M_m}$ be the candidate moment-based models. For $i = 1,\ldots,M_f$ define

$$I(\beta_i^*|h_0) = \mathrm{E}^0\{\log h_0(Z) - \log f_i(Z;\beta_i^*)\} = \min_{\beta_i\in\mathscr{B}_i} \mathrm{E}^0\{\log h_0(Z) - \log f_i(Z;\beta_i)\},$$

and for $j = 1,\ldots,M_m$ define

$$\begin{aligned} I(\lambda_j^*,\alpha_j^*|h_0) &= \mathrm{E}^0\{\log(1+\lambda_j^{*\prime} m_j(Z;\alpha_j^*))\} \\ &= \min_{\alpha_j\in\mathscr{A}_j} \max_{\lambda_j\in\Lambda_j} \mathrm{E}^0\{\log(1+\lambda_j' m_j(Z;\alpha_j))\}. \end{aligned}$$

We are interested in testing whether the best candidate model outperforms the benchmark parametric model. Hence, the null hypothesis is

$$\mathrm{H}_0^M : I_0 \leqslant \min\left\{ \min_{i=1,\ldots,M_f} I(\beta_i^*|h_0), \min_{j=1,\ldots,M_m} I(\lambda_j^*,\alpha_j^*|h_0) \right\},$$

or equivalently

$$\mathrm{H}_0^M : \max\left\{ \max_{i=1,\ldots,M_f} \{I_0 - I(\beta_i^*|h_0)\}, \max_{j=1,\ldots,M_m} \{I_0 - I(\lambda_j^*, \alpha_j^*|h_0)\} \right\} \leqslant 0,$$

meaning that no candidate model is closer to the true model than the benchmark model in terms of KLIC, and the alternative hypothesis is

$$\mathrm{H}_1^M : I_0 > \min\left\{ \min_{i=1,\ldots,M_f} I(\beta_i^*|h_0), \min_{j=1,\ldots,M_m} I(\lambda_j^*, \alpha_j^*|h_0) \right\},$$

or equivalently

$$\mathrm{H}_1^M : \max\left\{ \max_{i=1,\ldots,M_f} \{I_0 - I(\beta_i^*|h_0)\}, \max_{j=1,\ldots,M_m} \{I_0 - I(\lambda_j^*, \alpha_j^*|h_0)\} \right\} > 0,$$

meaning that there exists a candidate model that is closer to the true model than the benchmark model in terms of KLIC.

For $i = 1, \ldots, M_f$, let

$$\widehat{\beta}_i = \arg\max_{\beta_i \in \mathscr{B}_i} \frac{1}{n} \sum_{t=1}^{n} \log f_i(z_t; \beta_i) \quad \text{and} \quad \widehat{I}(\widehat{\beta}_i|\widehat{h}) = \frac{1}{n} \sum_{t=1}^{n} \{\log \widehat{h}(z_t) - \log f_i(z_t; \widehat{\beta}_i)\},$$

and for $j = 1, \ldots, M_m$ let

$$(\widehat{\alpha}_j, \widehat{\lambda}_j) = \arg\min_{\alpha_j \in \mathscr{A}_j} \arg\max_{\lambda_j \in \Lambda_j} \frac{1}{n} \sum_{t=1}^{n} \log(1 + \lambda_j' m_j(z_t; \alpha_j)),$$

$$\widehat{I}(\widehat{\lambda}_j, \widehat{\alpha}_j|\widehat{h}) = \frac{1}{n} \sum_{t=1}^{n} \log(1 + \widehat{\lambda}_j' m_j(z_t; \widehat{\alpha}_j)).$$

Our test will be based on the following nonparametric likelihood ratio statistic vector of length $M = M_f + M_m$:

$$(\widehat{I}_0 - \widehat{I}(\widehat{\beta}_1|\widehat{h}), \ldots, \widehat{I}_0 - \widehat{I}(\widehat{\beta}_{M_f}|\widehat{h}); \widehat{I}_0 - \widehat{I}(\widehat{\lambda}_1, \widehat{\alpha}_1|\widehat{h}), \ldots, \widehat{I}_0 - \widehat{I}(\widehat{\lambda}_{M_m}, \widehat{\alpha}_{M_m}|\widehat{h}))'.$$

We consider several cases.

Case 1: If the benchmark is a parametric model $I_0 = I(\beta_0^*|h_0)$, we denote

$$\Omega = \Omega_{f_0} = (\sigma_{ik})_{i,k=1}^M,$$

where

1(i) for $i, k = 1, \ldots, M_f$,

$$\sigma_{ik} = Cov^0\left[\log \frac{f_i(Z; \beta_i^*)}{f_0(Z; \beta_0^*)}, \log \frac{f_k(Z; \beta_k^*)}{f_0(Z; \beta_0^*)}\right],$$

1(ii) for $i = 1, \ldots, M_f$ and $k = M_f + k'$, $k' = 1, \ldots, M_m$,

$$\sigma_{ik} = Cov^0\left[\log \frac{f_i(Z; \beta_i^*)}{f_0(Z; \beta_0^*)}, \log \frac{h_0(Z)}{f_0(Z; \beta_0^*)} - \log(1 + \lambda_{k'}^{*\prime} m_{k'}(Z; \alpha_{k'}^*))\right],$$

1(iii) for $i = M_f + i'$, $i' = 1, \ldots, M_m$ and $k = 1, \ldots, M_f$,

$$\sigma_{ik} = Cov^0\left[\log \frac{h_0(Z)}{f_0(Z;\beta_0^*)} - \log(1 + \lambda_{i'}^{*\prime} m_{i'}(Z;\alpha_{i'}^*)), \log \frac{f_k(Z;\beta_k^*)}{f_0(Z;\beta_0^*)}\right],$$

1(iv) for $i = M_f + i'$, $i' = 1, \ldots, M_m$ and $k = M_f + k'$, $k' = 1, \ldots, M_m$,

$$\sigma_{ik} = Cov^0\left[\log \frac{h_0(Z)}{f_0(Z;\beta_0^*)} - \log(1 + \lambda_{i'}^{*\prime} m_{i'}(Z;\alpha_{i'}^*)), \log \frac{h_0(Z)}{f_0(Z;\beta_0^*)} - \log(1 + \lambda_{k'}^{*\prime} m_{k'}(Z;\alpha_{k'}^*))\right].$$

Case 2: If the benchmark is a moment-based model $I_0 = I(\lambda_0^*, \alpha_0^* | h_0)$, we have

$$\Omega = \Omega_{m_0} = (\sigma_{ik})_{i,k=1}^M,$$

where

2(i) for $i, k = 1, \ldots, M_f$,

$$\sigma_{ik} = Cov^0\left[\log(1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)) - \log \frac{h_0(Z)}{f_i(Z;\beta_i^*)}, \log(1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)) - \log \frac{h_0(Z)}{f_k(Z;\beta_k^*)}\right],$$

2(ii) for $i = 1, \ldots, M_f$ and $k = M_f + k'$, $k' = 1, \ldots, M_m$,

$$\sigma_{ik} = Cov^0\left[\log(1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)) - \log \frac{h_0(Z)}{f_i(Z;\beta_i^*)}, \log \frac{1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)}{1 + \lambda_{k'}^{*\prime} m_{k'}(Z;\alpha_{k'}^*)}\right],$$

2(iii) for $i = M_f + i'$, $i' = 1, \ldots, M_m$ and $k = 1, \ldots, M_f$,

$$\sigma_{ik} = Cov^0\left[\log \frac{1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)}{1 + \lambda_{i'}^{*\prime} m_{i'}(Z;\alpha_{i'}^*)}, \log(1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)) - \log \frac{h_0(Z)}{f_k(z;\beta_k^*)}\right],$$

2(iv) for $i = M_f + i'$, $i' = 1, \ldots, M_m$ and $k = M_f + k'$, $k' = 1, \ldots, M_m$,

$$\sigma_{ik} = Cov^0\left[\log \frac{1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)}{1 + \lambda_{i'}^{*\prime} m_{i'}(Z;\alpha_{i'}^*)}, \log \frac{1 + \lambda_0^{*\prime} m_0(Z;\alpha_0^*)}{1 + \lambda_{k'}^{*\prime} m_{k'}(Z;\alpha_{k'}^*)}\right].$$

Corollary 5. *Suppose that the parametric model* $i = 1, 2, \ldots, M_f$ *satisfies conditions of Theorems* 1 and 2, *and moment model* $j = 1, \ldots, M_m$ *satisfies conditions of Theorem* 3. *Suppose that the benchmark satisfies conditions of Theorems* 1 and 2 (*if it is parametric*) *or conditions of Theorem* 3 (*if it is moment-based*). *Suppose that* Ω *is positive semi-definite and* *its largest eigenvalue is positive. Then*

$$\sqrt{n}\begin{bmatrix} \widehat{I}_0-\widehat{I}(\widehat{\beta}_1|\widehat{h})-[I_0-I(\beta_1^*|h_0)] \\ \vdots \\ \widehat{I}_0-\widehat{I}(\widehat{\beta}_{M_f}|\widehat{h})-[I_0-I(\beta_{M_f}^*|h_0)] \\ \widehat{I}_0-\widehat{I}(\widehat{\lambda}_1,\widehat{\alpha}_1|\widehat{h})-[I_0-I(\lambda_1^*,\alpha_1^*|h_0)] \\ \vdots \\ \widehat{I}_0-\widehat{I}(\widehat{\lambda}_{M_m},\widehat{\alpha}_{M_m}|\widehat{h})-[I_0-I(\lambda_{M_m}^*,\alpha_{M_m}^*|h_0)] \end{bmatrix} \longrightarrow^{\mathrm{d}} \begin{bmatrix} Z_1 \\ \vdots \\ Z_{M_f} \\ Z_{M_f+1} \\ \vdots \\ Z_M \end{bmatrix},$$

where $(Z_1,\dots,Z_M)'\sim\mathcal{N}(0,\Omega)$. *Further* $\Omega=\Omega_{f_0}$ *if the benchmark is parametric* $I_0=I(\beta_0^*|h_0)$, *and* $\Omega=\Omega_{m_0}$ *if the benchmark is moment-based* $I_0=I(\lambda_0^*,\alpha_0^*|h_0)$.

The proof of this proposition directly follows from those of Theorems 1–3; hence, we omit it.

White's (2000) reality check test is under the Least Favorable Configuration (LFC) (i.e. $I_0=I(\beta_i^*|h_0)=I(\lambda_j^*,\alpha_j^*|h_0)$ for all $i=1,\dots,M_f$ and $j=1,\dots,M_m$),

$$\begin{aligned} T_n^{\mathrm{W}} &\equiv \max\left[\max_{i=1,\dots,M_f} n^{1/2}\{\widehat{I}_0-\widehat{I}(\widehat{\beta}_i|\widehat{h})\},\ \max_{j=1,\dots,M_m} n^{1/2}\{\widehat{I}_0-\widehat{I}(\widehat{\lambda}_j,\widehat{\alpha}_j|\widehat{h})\}\right] \\ &\longrightarrow^{\mathrm{d}} \max_{j=1,\dots,M} Z_j \quad \text{under } \mathrm{H}_0^M \text{ and LFC.} \end{aligned}$$

Since Hansen (2003) shows via simulation that the power of White's test could be poor, we follow Hansen's (2003) suggestion and consider a modified test

$$\begin{aligned} T_n^{\mathrm{H}} &\equiv \max\left[\max_{i=1,\dots,M_f} n^{1/2}\{\widehat{I}_0-\widehat{I}(\widehat{\beta}_i|\widehat{h})\},\ \max_{j=1,\dots,M_m} n^{1/2}\{\widehat{I}_0-\widehat{I}(\widehat{\lambda}_j,\widehat{\alpha}_j|\widehat{h})\},0\right] \\ &\to^{\mathrm{d}} \max\left[\max_{i=1,\dots,M_0} Z_i,0\right] \quad \text{under } \mathrm{H}_0^M, \end{aligned}$$

where M_0 is the set of competing models that are binding under H_0^M. Since the limiting distribution of both T_n^{W} and T_n^{H} are complicated, we can use bootstrap critical values to implement these tests.

Step 1: Draw a bth independent random bootstrap sample $\{z_t^b\}_{t=1}^n$ of size n from the raw data $\{z_t\}_{t=1}^n$ with replacement, and compute the bth bootstrap estimates $\widehat{I}_0^b$, $\widehat{I}(\widehat{\beta}_i^b|\widehat{h}^b)$, $\widehat{I}(\widehat{\lambda}_j^b,\widehat{\alpha}_j^b|\widehat{h}^b)$ for $i=1,\dots,M_f$, $j=1,\dots,M_m$.

Step 2: Compute the bth bootstrap centered terms.

For White's test: $\widehat{I}_i^{b,\mathrm{W}c}\equiv\widehat{I}_0^b-\widehat{I}(\widehat{\beta}_i^{\,b}|\widehat{h}^b)-[\widehat{I}_0-\widehat{I}(\widehat{\beta}_i|\widehat{h})]$ for $i=1,\dots,M_f$, and $\widehat{I}_j^{b,\mathrm{W}c}\equiv\widehat{I}_0^b-\widehat{I}(\widehat{\lambda}_j^{\,b},\widehat{\alpha}_j^b|\widehat{h}^b)-[\widehat{I}_0-\widehat{I}(\widehat{\lambda}_j,\widehat{\alpha}_j|\widehat{h})]$ for $j=1,\dots,M_m$.

For Hansen's test: $\widehat{I}_i^{b,\mathrm{H}c}\equiv\widehat{I}_0^b-\widehat{I}(\widehat{\beta}_i^b|\widehat{h}^b)-[\widehat{I}_0-\widehat{I}(\widehat{\beta}_i|\widehat{h})]1\{\widehat{I}_0-\widehat{I}(\widehat{\beta}_i|\widehat{h})\geqslant-a_n\}$ for $i=1,\dots,M_f$, and $\widehat{I}_j^{b,\mathrm{H}c}\equiv\widehat{I}_0^b-\widehat{I}(\widehat{\lambda}_j^b,\widehat{\alpha}_j^b|\widehat{h}^b)-[\widehat{I}_0-\widehat{I}(\widehat{\lambda}_j,\widehat{\alpha}_j|\widehat{h})]1\{\widehat{I}_0-\widehat{I}(\widehat{\lambda}_j,\widehat{\alpha}_j|\widehat{h})\geqslant-a_n\}$ for $j=1,\dots,M_m$. a_n is a sequence of constants where $a_n\to0$ and $a_nn^{1/2}\to\infty$. For example, we can take $a_n=\log n/\sqrt{n}$.

Step 3: Compute the bth bootstrap estimates of the tests:

$$T_n^{\mathrm{W},b} \equiv \max\left[\max_{i=1,\ldots,M_f} n^{1/2}\widehat{I}_i^{b,\mathrm{W}c}, \max_{j=1,\ldots,M_m} n^{1/2}\widehat{I}_j^{b,\mathrm{W}c}\right],$$

$$T_n^{\mathrm{H},b} \equiv \max\left[\max_{i=1,\ldots,M_f} n^{1/2}\widehat{I}_i^{b,\mathrm{H}c}, \max_{j=1,\ldots,M_m} n^{1/2}\widehat{I}_j^{b,\mathrm{H}c}, 0\right].$$

Step 4: Repeat Steps 1–3 for $b = 1, \ldots, B$ with $B = 500$ (say). Compute the bootstrap estimates of the *p-value*:

$$\widehat{p}_{\mathrm{W}} \equiv \frac{1}{B}\sum_{b=1}^{B} 1\{T_n^{\mathrm{W},b} > T_n^{\mathrm{W}}\},$$

$$\widehat{p}_{\mathrm{H}} \equiv \frac{1}{B}\sum_{b=1}^{B} 1\{T_n^{\mathrm{H},b} > T_n^{\mathrm{H}}\}.$$

The White (2000) test rejects the null when $\widehat{p}_{\mathrm{W}}$ is small and does not reject when $\widehat{p}_{\mathrm{W}}$ is big. The Hansen (2003) test rejects the null if $\widehat{p}_{\mathrm{H}}$ is small and does not reject if $\widehat{p}_{\mathrm{H}}$ is big.

6. Empirical illustration: sequential vs. non-sequential search models

To illustrate the use of our testing procedure, we present an example from industrial organization of testing between sequential and non-sequential models of consumer search models, using price data obtained from internet websites. Since the search models we compare are relatively stylized, and omit important details of online shopping behavior, we emphasize that the results in this section are presented more as an illustration of the testing procedure described above, rather than a detailed analysis of search behavior in online markets.

As is discussed in Hong and Shum (2001), the parameters characterizing consumers' optimal search and firms' optimal pricing strategies in an equilibrium pricing model with *non-sequential* consumer search behavior can be identified nonparametrically, in the sense that the population moment restrictions implied by the equilibrium model are enough to identify these parameters, without additional functional form restrictions on the shape of consumers' search costs. On the other hand, the parameters of an equilibrium pricing model with sequential consumer search behavior cannot be identified without functional form assumptions on the distribution of consumer search costs. Therefore, as an illustration of our proposed test, we test between several parameterizations of the sequential search model, and the moment-based non-sequential search model.

For both models, we assume that there are a continuum of firms and consumers, and interpret the equilibrium price distribution F_p as the symmetric equilibrium mixed strategy employed by all firms. We assume that consumers have inelastic demand for a single unit of the goods. Consumers incur a search cost c to receive a single price quote, and we assume that search costs are i.i.d. drawn across consumers from a distribution F_c. Let $\underline{p}$ and $\bar{p}$ denote, respectively, the lower and upper bound of the support of F_p, and r denote the common unit production cost of each firm. Since all firms produce homogeneous products, only search frictions (arising from consumers' imperfect information about stores' prices) generate price dispersion in this market.

Before proceeding to the test, we briefly describe the estimating procedures for each model in turn. Details on the derivations of these equations are given in Hong and Shum (2001).

6.1. Non-sequential search model

Consumers who search non-sequentially commit to buying from the lowest-priced store after obtaining a random sample of n prices. A consumer with search cost c chooses the number of stores n to canvass to minimize her total expected cost, which is the sum of her search costs as well as the price she expects to pay for the product:

$$n^*(c) \equiv \arg\min_{n>1} \mathscr{C}(n;c) \equiv c*(n-1) + \int_{\underline{p}}^{\bar{p}} np(1-F(p))^{n-1} f(p)\,\mathrm{d}p.$$

Let $\hat{F}_p$ denote the empirical distribution of the observed prices. Define:

$\tilde{q}_1$: the equilibrium proportion of consumers who search only one store,

$\tilde{q}_2$: the equilibrium proportion of consumers who search two stores,

$\tilde{q}_3$: the equilibrium proportion of consumers who search three stores, etc.

In the mixed strategy equilibrium, the following indifference condition must hold for every price $p \in [\underline{p}, \bar{p})$:

$$(\bar{p}-r)\tilde{q}_1 = (p-r)\left[\sum_{k=1}^{\infty} \tilde{q}_k k(1-F(p))^{k-1}\right].$$

Rearranging, we get that

$$F_p^{-1}(\tau) = r + \frac{(\bar{p}-r)\cdot\tilde{q}_1}{[\sum_{k=1}^{K-1} \tilde{q}_k k(1-\tau)]},$$

for any quantile $\tau \in [0,1]$ and where $K(<N)$ denotes the maximal number of stores at which consumers will search. From this equation, we can a derive set of $M(\geqslant K)$ moment conditions

$$\mathrm{E}\left\{\mathbf{1}\left(p_t \leqslant r + \frac{(\bar{p}-r)\tilde{q}_1}{[\sum_{k=1}^{K} \tilde{q}_k k(1-\tau_m)^{k-1}]}\right) - \tau_m\right\} = 0, \quad m = 1,\ldots,M, \;\; M \geqslant K,$$

which we can use to estimate the unknowns $\{r, \tilde{q}_1, \ldots, \tilde{q}_{K-1}\}$.

The sample analogs (with a dataset consisting of $T+1$ prices, with $p_{T+1} = \bar{p}$) are

$$\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\left(p_t \leqslant r + \frac{(\bar{p}-r)\tilde{q}_1}{[\sum_{k=1}^{K} \tilde{q}_k k(1-\tau_m)^{k-1}]}\right) - \tau_m\right] = 0. \tag{6.5}$$

6.2. Sequential search model

Consumers who search sequentially follow a "reservation price" policy whereby they obtain additional price quotes until they find one which is lower than their (optimally chosen) reservation price.

Table 2
Summary statistics on online prices for two different products

Product	# Obs	List	Mean	St. dev.	Median	$\underline{p}$	$\bar{p}$
Including shipping and handling costs							
Palm Pilot Vx	18		238.97	38.96	219.45	190.00	310.62
Billingsley: *Probability and Measure*	18	99.95	95.23	6.14	98.90	83.58	100.87

Price data for all products downloaded from pricescan.com: February 5, 2002.

As discussed in Hong and Shum (2001), the sequential search model is not nonparametrically identified given only data on prices. Hence, we consider parametric MLE of this model, assuming that the search cost distribution $F_c(\cdot;\theta)$ is parameterized by a (finite-dimensional) vector θ.[6] The likelihood function for each observed price is the equilibrium price density, which is given by (see Hong and Shum, 2001 for details)[7]

$$f_p(s;\theta) = -\frac{2(\underline{p}-r)}{(s-r)^3 * f_c(c(1-\frac{\bar{p}-r}{s-r};\theta);\theta)} - \frac{(\underline{p}-r)^2 f'_c(c(1-\frac{\underline{p}-r}{s-r};\theta))}{(s-r)^4 * [f_c(c(1-\frac{\underline{p}-r}{s-r};\theta);\theta)]^3}.$$

In the above equations, $c(\tau;\theta) \equiv F_c^{-1}(\tau;\theta)$, the inverse CDF for the search cost distribution. Given θ, the auxiliary parameter r can be determined by the restriction that $F_p(\bar{p}) = 1$ so that r must satisfy:

$$1 = F_p(\bar{p}) = \frac{(\underline{p}-r)}{(\bar{p}-r)^2 * f_c(c(1-\frac{\underline{p}-r}{\bar{p}-r});\theta)}.$$

The likelihood function for the whole sample of prices, then, is just $L(\theta,r) = \prod_{t=1}^{T} f_p(p_t;\theta)$.

6.3. Results

To illustrate the use of our tests, we present the test results for a sample of online prices collected for two products: the *Palm Pilot Vx*, and the statistics textbook *Probability and Measure*, by P. Billingsley. These prices were gathered from the `pricescan.com` website on February 5, 2002, with shipping and handling costs confirmed by visiting the individual websites. The summary statistics for these price data are given in Table 2.

The test results are reported in Table 3. We considered three models: (1) the non-sequential search models, estimated via the moment condition (6.5); (ii) a parametric sequential search model, with a Gamma search cost distribution; (iii) a parametric sequential search model with a log-normal search cost distribution. We only report results for the White test statistics; the Hansen test results were qualitatively similar, and we do not report them for convenience. We performed the tests taking each of the three models, in turn, as the benchmark model. The nonparametric sieve estimation of the entropy was done using a cardinal B-spline wavelet basis.

[6] $\underline{p}$ and $\bar{p}$ can be (super-consistently) estimated from the data. See Donald and Paarsch (1993) for a discussion of MLE when a subset of the parameters can be super-consistently estimated.

[7] In deriving this equation, we are implicitly assuming that the flow of consumers visiting any store is slow enough so that, from the firms' point of view, the consumer population is identical over time. This assumption is made to avoid difficult theoretical issues; see Hong and Shum (2001) for a more extended discussion.

Table 3
Results from multiple model comparison tests

Benchmark model for test	T_n^{w} (White test statistic)	Bootstrap p-value[a]
Palm Pilot Vx		
Model 1	0.397	0.953
Model 2	0.640	0.162
Model 3	−0.397	0.847
Billingsley textbook		
Model 1	0.034	0.386
Model 2	0.640	0.205
Model 3	−0.034	0.988

Statistic	Full-sample statistic	Mean from bootstrap samples	St. dev. from bootstrap samples	Number of bootstrap samples
Palm Pilot Vx				
$KLIC_1$	0.193	0.182	0.086	148
$KLIC_2$	0.956	0.055	1.215	
$KLIC_3$	0.316	0.085	2.423	
NP entropy est.	2.182	1.163	0.383	
Billingsley textbook				
$KLIC_1$	0.713	0.733	0.281	83
$KLIC_2$	1.319	1.839	1.606	
$KLIC_3$	0.679	1.916	2.658	
NP entropy est.	2.545	2.801	0.224	

Model 1: Non-sequential search.
Model 2: Parametric sequential search model, with gamma search cost distribution.
Model 3: Parametric sequential search model, with log-normal search cost distribution.
All prices include shipping and handling charges.
[a] $Prob(T_n^{\mathrm{w},b} \geqslant T_n^{\mathrm{w}})$, where $T_n^{\mathrm{w},b}$ denotes bootstrapped value for test statistic T_n^{w}, and $Prob(\cdots)$ denotes the empirical frequency from bootstrap re-samples.

The test results are qualitatively similar for both products. For the Palm Pilot Vx, the bootstrapped p-values for the test statistics using Models 1 and 3 as the benchmark models are large (at 0.953 and 0.847, respectively) which appear to imply that both the non-sequential and the parametric models assuming a log-normal search cost distribution perform equally well in fitting the data. The p-value for the test statistics when Model 2 is used as the benchmark model is substantially lower (at 0.162), suggesting the inferiority of the parametric model assuming a Gamma search cost distribution.

For the Billingsley text, the lowest p-value is again obtained when Model 2 is used as the benchmark model. These results indicate perhaps that it is difficult to distinguish (on the basis of fit) between the non-sequential search models, and certain parameterizations of the sequential search model. This may be due in part to the small sample size used in this example (we were able to obtain only 18 prices for each product).

For completeness, Table 3 also includes descriptive statistics for the KLIC and entropy estimates for the three models, across bootstrap resamples. Not surprising, given the test results, we see that the KLICs for Models 1 and 3 are noticeably smaller than the KLIC for Model 2. Furthermore, the KLICs, as well as the nonparametric entropy estimates, are quite stable across bootstrap resamples.

While this application is rather specialized, there are other potentially interesting industrial organization applications of our testing procedure. For example, Bresnahan (1981) and Berry et al. (1995) have presented competing equilibrium demand and supply models for the automobile industry. Bresnahan (1981) considers a vertical differentiation context, and estimates model parameters using MLE assuming a parametric distribution for measurement errors in price. Berry et al. (1995) consider a more general model with both vertical and horizontal differentiation, and present a moment-based GMM estimation approach. Testing between these two specifications could be done using our approach.

7. Conclusion

In this paper we first develop a nonparametric likelihood ratio model selection test between two competing models where one of the models is parametric likelihood, and the other is moment-based model. Our tests extend the likelihood-ratio model selection tests presented previously in Vuong (1989) for two parametric likelihood models and Kitamura (2000) for two moment-based models to the scenario where one model is parametric and the other is moment-based. We then extend our procedure to handle the comparison of multiple (>2) models where some candidate models are parametric and other candidate models are moment-based. We present a short Monte Carlo simulation study for our procedure, and also present an industrial organization example of our tests. Using data on observed online prices for an electronic product, we examine whether the prices are generated from an equilibrium search model where consumers follow sequential or non-sequential search strategies.

For notational clarity we have focused on the case with continuous observations in the paper. The extension to a mixture of discrete and continuous regressors is straightforward but involves more tedious notations. In particular, we only need to modify the nonparametric estimation of the entropy of the true data generating process by partitioning the sample based on the value of the discrete regressors.

We believe that these model selection testing situations arise often in practice, especially given that both parametric likelihood specification and moment condition specification have become popular in structural econometric modeling. There are potential applications to, *inter alia*, discrete-choice differentiated product demand models in industrial organization, sample selection models in labor economics, as well as dynamic models in macroeconomics and financial economics.

We also note that our model selection test can be used in conjunction with other model specification tests and selection procedures. While our test has taken the view that most models are likely to be misspecified, a variety of specification tests can be used to detect whether a model is statistically correctly specified. Lakatos suggested that one should always choose more detailed theories until they can be falsified.[8] Given that a parametric model typically offers a more detailed description of the economic phenomenon than a collection of moments, one may naturally choose the parametric model when both are correctly specified.

Therefore, a practitioner might wish to first perform model specification tests on both a parametric model and a moment condition model. If both models are statistically refuted

[8]We thank a referee for pointing this out to us.

by the data, then our model selection test can be used to choose the one with a closer K-L measure.

Acknowledgments

We thank the associate editor, three referees, J. Geweke, J. Horowitz, Y. Kitamura, O. Linton, N. Meddahi, S. Ng, H. Paarsch, R. Porter and seminar participants in Carnegie-Mellon, U. Iowa, the 2003 Econometric Society Winter Meetings in Washington DC, SITE 2003, and the 2004 Semiparametrics conference in Rio for helpful comments. We also thank D. Chen and in particular D. Nekipelov for excellent research assistance in the simulation study. All three authors gratefully acknowledge supports from the National Science Foundation, and Hong also acknowledges support from the Sloan Foundation.

Appendix A. Population empirical likelihood function

We restrict Q to the set of probability measures with continuously differentiable density functions. We find the solution for the constrained optimization problem:

$$\min_{q(\cdot)} I(q(\cdot)|h_0(\cdot)) = \int \log \frac{h_0(Z)}{q(Z)} h_0(Z)\,\mathrm{d}Z$$
$$\text{s.t.} \quad \int m(Z;\alpha)q(Z)\,\mathrm{d}Z = 0 \quad \text{and} \quad \int q(Z)\,\mathrm{d}Z = 1.$$

The Lagrangian for this constrained optimization problem is

$$\mathscr{L}_{q(\cdot),\bar{\lambda},\bar{\beta}} = \int \log \frac{h_0(Z)}{q(Z)} h_0(Z)\,\mathrm{d}Z + \bar{\lambda}' \int m(Z;\alpha)q(Z)\,\mathrm{d}Z + \bar{\beta}\left(\int q(Z)\,\mathrm{d}Z - 1\right).$$

Using calculus of variation to concentrate out $q(\cdot)$ given λ, β, we find that

$$q^*(\cdot|\bar{\lambda},\bar{\beta}) = \arg\min_{q(\cdot)} L_{q(\cdot),\bar{\lambda},\bar{\beta}} = \frac{h_0(Z)}{\bar{\beta} + \bar{\lambda}' m(Z;\alpha)}.$$

Define $\lambda = \bar{\lambda}/\bar{\beta}$. Using the constraint that $q(\cdot)$ integrates to 1, we can solve for $\bar{\beta}$:

$$\bar{\beta} = \int \frac{h_0(Z)}{1 + \lambda' m(Z;\alpha)}\,\mathrm{d}Z \quad \text{and} \quad q(Z) = \frac{h_0(Z)}{1 + \lambda' m(Z;\alpha)} \Big/ \int \frac{h_0(Z)}{1 + \lambda' m(Z;\alpha)}\,\mathrm{d}Z.$$

Concentrating out $q(\cdot)$ and $\bar{\beta}$ given λ leads to

$$I(q^*(\cdot)|h(\cdot)) = \mathscr{L} = \int \log(1 + \lambda' m(Z;\alpha))h_0(Z)\,\mathrm{d}Z + \log \int \frac{h_0(Z)}{1 + \lambda' m(Z;\alpha)}\,\mathrm{d}Z.$$

The moment constraint $\int \frac{m(Z;\alpha)}{1+\lambda' m(Z;\alpha)} h_0(Z)\,\mathrm{d}Z = 0$ implies that $\int \frac{1}{1+\lambda' m(Z;\alpha)} h_0(Z)\,\mathrm{d}Z = 1$. Hence

$$I(q^*(\cdot)|h_0(\cdot)) = \max_{\lambda} \int \log(1 + \lambda' m(Z;\alpha))h_0(Z)\,\mathrm{d}Z.$$

Appendix B. Proof of Theorem 3

Consistency and asymptotic normality for generalized empirical likelihood estimators can be found in Kitamura and Stutzer (1997), Christoffersen et al. (2001), Chernozhukov and Hansen (2001) and Newey and Smith (2001). In particular, we follow closely the arguments of Christoffersen et al. (2001), which apply to misspecified models. We review their main steps.

B.1. Consistency

Step 1: Condition 2 implies that $\lambda(\alpha)$ is continuous in α, where

$$\lambda(\alpha) = \arg\max_{\lambda} \mathrm{E}^0 \log(1 + \lambda' m(z_i;\alpha)).$$

Step 2: This combined with conditions 5 and 7 and the saddle-point property shows that for L defined as $\mathrm{E}^0 \log(1 + \lambda^{*\prime} m(z_i;\alpha^*))$, and for all $\delta > 0$, there exists $\eta > 0$ such that

$$P\left(\frac{1}{n}\sum_{i=1}^{n} \log(1 + \hat{\lambda}(\alpha)' m(z_i;\alpha)) < L + \eta\right) \leqslant P\left(\frac{1}{n}\sum_{i=1}^{n} \log(1 + \lambda(\alpha)' m(z_i;\alpha)) < L + \eta\right) \to 0.$$

Step 3: Using convexity lemmas, $\hat{\lambda}(\alpha^*) \xrightarrow{\mathrm{p}} \lambda^*$, and

$$P\left(\frac{1}{n}\sum_{i=1}^{n} \log(1 + \hat{\lambda}(\alpha^*)' m(z_i;\alpha^*)) > L + \eta\right) \to 0.$$

Combining steps 2 and 3 gives $\hat{\alpha} \xrightarrow{\mathrm{p}} \alpha^*$. Using convexity lemma again, since

$$\frac{1}{n}\sum_{i=1}^{n} \log(1 + \lambda' m(z_i;\hat{\alpha})) \xrightarrow{\mathrm{p}} E^0\left[\log(1 + \lambda' m(Z;\alpha^*))\right]$$

pointwise and therefore uniform in λ,

$$\hat{\lambda} = \hat{\lambda}(\hat{\alpha}) \xrightarrow{\mathrm{p}} \lambda^* = \arg\max_{\lambda} \mathrm{E}^0 \log(1 + \lambda' m(z_i,\alpha^*)).$$

B.2. $\sqrt{n}$ Convergence rate

Step 1: Conditions 4 and 6 imply $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{m(z_i;\alpha^*)}{1+\lambda^{*\prime} m(z_i;\alpha^*)} = \mathrm{O_p}(1)$ and

$$\sum_{i=1}^{n} \log(1 + \hat{\lambda}(\alpha^*)' m(z_i;\alpha^*)) = \mathrm{O_p}(1).$$

Step 2: Using the saddle-point property, denote $\hat{M}(\lambda,\alpha) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \lambda' m(z_i;\alpha))$. Then for all t,

$$n\hat{M}\left(\lambda^* + \frac{t}{\sqrt{n}}, \hat{\alpha}\right) \leqslant n\hat{M}(\hat{\lambda},\hat{\alpha}) \leqslant n\hat{M}(\hat{\lambda}(\alpha^*),\alpha^*) = \mathrm{O_p}(1).$$

A Taylor expansion of the left-hand side up to second order then shows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{m(z_i;\hat{\alpha})}{1 + \lambda^{*\prime} m(z_i;\hat{\alpha})} = \mathrm{O_p}(1).$$

Step 3: Using condition 8,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(z_i;\hat{\alpha})}{1+\lambda^{*\prime}m(z_i;\hat{\alpha})}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(z_i;\alpha^*)}{1+\lambda^{*\prime}m(z_i;\alpha^*)}+\sqrt{n}\mathrm{E}\,\frac{m(z_i;\hat{\alpha})}{1+\lambda^{*\prime}m(z_i;\hat{\alpha})}+\mathrm{o_p}(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(z_i;\alpha^*)}{1+\lambda^{*\prime}m(z_i;\alpha^*)}+(D+\mathrm{o_p}(1))'\sqrt{n}(\hat{\alpha}-\alpha^*)+\mathrm{o_p}(1).$$

Then by condition 5, $\sqrt{n}(\hat{\alpha}-\alpha^*)=\mathrm{O_p}(1)$.

Step 4: A Taylor expansion of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(z_i;\hat{\alpha})}{1+\hat{\lambda}'m(z_i;\hat{\alpha})}=0$ in $\hat{\lambda}$ around λ^* shows that

$$\sqrt{n}(\hat{\lambda}-\lambda^*)=(V+\mathrm{o_p}(1))^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(z_i;\hat{\alpha})}{1+\lambda^{*\prime}m(z_i;\hat{\alpha})}+\mathrm{o_p}(1)=\mathrm{O_p}(1).$$

Empirical likelihood objective functions: using conditions 3 and 9,

$$\sqrt{n}(\hat{M}(\hat{\lambda},\hat{\alpha})-\hat{M}(\lambda^*,\alpha^*))=\sqrt{n}(M(\hat{\lambda},\hat{\alpha})-M(\lambda^*,\alpha^*))+\mathrm{o_p}(1)$$
$$=\left(\frac{\partial}{\partial\lambda}M(\lambda^*,\alpha^*)+\mathrm{o_p}(1)\right)\sqrt{n}(\hat{\lambda}-\lambda^*)+\left(\frac{\partial}{\partial\alpha}M(\lambda^*,\alpha^*)+\mathrm{o_p}(1)\right)\sqrt{n}(\hat{\alpha}-\alpha^*)+\mathrm{o_p}(1).$$

Appendix C. Kernel based test of nondegeneracy

The limit distribution of V_n, the kernel based nondegeneracy test (defined in Eq. (3.2) in the main text) can be derived following the results of Fan (1994), who shows that the convergence rate and the limit distribution of the test statistics depends on whether the data is under-smoothed, over-smoothed or optimally-smoothed. In particular, Fan (1994) showed that the effect of the first step estimation of $\hat{\beta}$, $\hat{\lambda}$ and $\hat{\alpha}$ needs to be taken into account in the cases of over-smoothing and optimal-smoothing. On the other hand, in the undersmoothing case, preliminary parametric estimation has no effect on the limit variance of the test statistics. In the following we restate the result for the under-smoothing case from case (c2) of Corollary 2.4 of Fan (1994):

Proposition 1. *Let $K(\cdot)$ be an mth order kernel and let b be the bandwidth used in the kernel estimation. If $nb^{d_z/2+2m}\longrightarrow 0$, then*

$$nb^{d_z/2}\left(V_n-\frac{1}{nb^{d_z}}\int K(u)^2\,\mathrm{d}u\right)$$
$$\xrightarrow{\mathrm{d}}\mathrm{N}\left(0,\frac{1}{4}\left[\int h^2(Z)\,\mathrm{d}Z\right]\left\{\int\left[\int K(u)K(u+v)\,\mathrm{d}u\right]^2\mathrm{d}v\right\}\right).$$

Appendix D. Distribution function-based test of nondegeneracy

We first introduce some notations to describe the limiting distribution of the CDF-based test statistics for degeneracy. Let

$$J_\beta(Z)=\int^{Z}\frac{\partial}{\partial\beta}f(z,\beta^*)(1+\lambda^{*\prime}m(z,\alpha^*))\,\mathrm{d}z,$$

$$J_{\lambda,\alpha}(Z) = \begin{cases} \int^Z f(z,\beta^*) m(z,\alpha^*)\,\mathrm{d}z, \\ \int^Z f(z,\beta^*)(1+\lambda^{*\prime}\frac{\partial}{\partial\alpha} m(z,\alpha^*))\,\mathrm{d}z. \end{cases}$$

Then

$$\sqrt{n}\left(\int^Z f(z,\hat\beta)(1+\hat\lambda' m(z,\hat\alpha))\,\mathrm{d}z - \int^Z f(z,\beta^*)(1+\lambda^{*\prime} m(z,\alpha^*))\,\mathrm{d}z\right)$$
$$= J_\beta(Z)\sqrt{n}(\hat\beta-\beta^*) + J_{\lambda,\alpha}(x)\sqrt{n}(\hat\lambda-\lambda^*, \hat\alpha-\alpha^*) + \mathrm{o_p}(1).$$

Using the linear representation of the parametric estimates, we can write

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n 1(z_i \leqslant Z) - \int^Z f(z,\hat\beta)(1+\hat\lambda' m(z,\hat\alpha))\,\mathrm{d}z - h(Z) - \int^Z f(z,\beta^*)(1+\lambda^{*\prime} m(z,\alpha^*))\,\mathrm{d}z\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1(z_i \leqslant Z) - h(Z) - J_\beta(Z) A_f^{-1} \frac{\partial}{\partial\beta}\log f(z_i;\beta^*)\right.$$
$$\left. - J_{\lambda,\alpha}(x) A_m^{-1} \begin{pmatrix} \frac{m(z_i;\alpha^*)}{1+\lambda^{*\prime} m(z_i;\alpha^*)} \\ \frac{\frac{\partial}{\partial\alpha} m(z_i;\hat\alpha)\lambda^*}{1+\lambda^{*\prime} m(z_i;\alpha^*)} \end{pmatrix}\right) + \mathrm{o_p}(1).$$

In the above we have used the notation

$$A_m = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

where

$$S_{11} = -\mathrm{E}\,\frac{m(Z;\alpha^*)m(Z;\alpha^*)'}{(1+\lambda^{*\prime} m(Z;\alpha^*))^2},$$
$$S_{12} = \mathrm{E}\,\frac{\frac{\partial}{\partial\alpha} m(Z;\alpha^*)}{1+\lambda^{*\prime} m(;\alpha^*)} - \mathrm{E}\,\frac{m(Z;\alpha^*)}{(1+\lambda^{*\prime} m(Z;\alpha^*))^2}\lambda^{*\prime}\frac{\partial}{\partial\lambda} m(Z;\alpha^*),$$
$$S_{21} = \mathrm{E}\,\frac{\frac{\partial}{\partial\alpha} m(Z;\alpha^*)}{1+\lambda' m(Z;\alpha^*)} - \mathrm{E}\,\frac{\frac{\partial}{\partial\alpha} m(Z;\alpha^*)\lambda^* m(Z;\alpha^*)}{(1+\lambda^{*\prime} m(Z;\alpha^*))^2},$$
$$S_{22} = \mathrm{E}\,\frac{\lambda^{*\prime}\frac{\partial^2}{\partial\alpha\partial\alpha'} m(Z;\alpha^*)}{1+\lambda' m(Z;\alpha^*)} - \mathrm{E}\,\frac{\frac{\partial}{\partial\alpha} m(Z;\alpha^*)\lambda^*\lambda^{*\prime}\frac{\partial}{\partial\alpha} m(Z;\alpha^*)}{(1+\lambda^{*\prime} m(Z;\alpha))^2}.$$

This linear representation then converges weakly as $n \to \infty$ to the Gaussian process $\mathscr{G}(\cdot)$ with covariance function $G(u)G(v) = \mathrm{E}v(z_i,u)v(z_i,v)$, where $v(z_i,u)$ is

$$\left(1(z_i \leqslant u) - h(u) - J_\beta(u) A_f^{-1}\frac{\partial}{\partial\beta}\log f(z_i;\beta^*) - J_{\lambda,\alpha}(u) A_m^{-1} \begin{pmatrix} \frac{m(z_i;\alpha^*)}{1+\lambda^{*\prime} m(z_i;\alpha^*)} \\ \frac{\frac{\partial}{\partial\alpha} m(z_i;\hat\alpha)\hat\lambda}{1+\hat\lambda' m(z_i;\hat\alpha)} \end{pmatrix}\right).$$

Given the derivations above, it is imperative to characterize the limit distribution for ρ_1 and ρ_2, the two CDF-based nondegeneracy test statistics introduced in Eqs. (3.3) and (3.4) in the main text:

Proposition 2. *Under* H_0:

$$\sqrt{n}\rho_1 \xrightarrow{\mathrm{d}} \sup_Z \mathscr{G}(Z),$$

$$n\rho_2 \xrightarrow{\mathrm{d}} \int \mathscr{G}(u)^2 w(u)\,\mathrm{d}u.$$

References

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

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841–890.

Bresnahan, T., 1981. Departures from marginal cost pricing in the American automobile industry. Journal of Econometrics 17, 201–227.

Chen, X., Fan, Y., 2005. A model selection test for bivariate failure-time data. Working paper. New York University and Vanderbilt University.

Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66, 289–314.

Chen, X., White, H., 1999. Improved rates and asymptotic normality for nonparametric neural network estimators. IEEE Transactions on Information Theory 45, 682–691.

Chernozhukov, V., Hansen, C., 2001. An IV model of quantile treatment effects. MIT Department of Economics Working Paper.

Christoffersen, P., Hahn, J., Inoue, A., 2001. Testing, comparing and combining value at risk measures. Journal of Empirical Finance 8, 325–342.

Coppejans, M., Gallant, A.R., 2002. Cross-validated SNP density estimates. Journal of Econometrics 110, 27–65.

Diebold, F.X., 1989. Forecast combination and encompassing: reconciling two divergent literature. International Journal of Forecasting 5, 589–592.

Donald, S., Paarsch, H., 1993. Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. International Economic Review 34, 121–148.

Fan, Y., 1994. Testing the goodness of fit of a parametric density function by kernel method. Econometric Theory 10, 316–356.

Gallant, A.R., Nychka, D.W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Gourieroux, G., Monfort, A., 1994. Testing non-nested hypotheses. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam.

Gowrisankaran, G., Geweke, J., Town, R., 2003. Bayesian inference for hospital quality in a selection model. Econometrica 71 (4), 1215–1239.

Granger, C., 2002. Time series concept for conditional distributions. Manuscript, UCSD.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, P.R., 2003. A test for superior predictive ability. Manuscript, Stanford University.

Hong, H., Shum, M., 2001. Estimating Search Costs Using Equilibrium Models. Mimeo, Princeton University.

Horowitz, J., Hardle, W., 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821–848.

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

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Annals of Statistics 25 (5), 2084–2102.

Kitamura, Y., 2000. Comparing misspecified dynamic econometric models using nonparametric likelihood. Department of Economics, University of Wisconsin.

Kitamura, Y., 2002. Econometric comparison of conditional models. Manuscript, University of Pennsylvania.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Kitamura, Y., Tripathi, G., 2001. Empirical likelihood-based inference in conditional moment restriction models. Department of Economics, University of Wisconsin.

Newey, W., McFadden, D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2113–2241.

Newey, W., Smith, R., 2001. Higher order properties of GMM and generalized empirical likelihood estimators. MIT and University of Bristol.

Owen, A., 2001. Empirical Likelihood. Chapman & Hall, CRC Press, London, Boca Raton.

Paarsch, H., 1992. Deciding between the common and private value paradigms in empirical models of auctions. Journal of Econometrics 51, 191–215.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Ramalho, J.J., Smith, R.J., 2002. Generalized empirical likelihood non-nested tests. Journal of Econometrics, pp. 1–28.

Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. The Econometrics Journal 5, 1–39.

Sin, C., White, H., 1996. Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics 71, 207–225.

Smith, R.J., 1992. Non-nested tests for competing models estimated by generalized method of moments. Econometrica 60 (4), 973–980.

Stone, C., 1985. Additive regression and other nonparametric models. Annals of Statistics 13, 689–705.

Vuong, Q., 1989. Likelihood-ratio tests for model selection and non-nested hypotheses. Econometrica, pp. 307–333.

White, H., 1982. Maximum likelihood estimation for misspecified models. Econometrica 50, 1–25.

White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.