TESTING FOR NON-NESTED CONDITIONAL MOMENT RESTRICTIONS USING UNCONDITIONAL EMPIRICAL LIKELIHOOD

By

Taisuke Otsu, Myung Hwan Seo and Yoon-Jae Whang

May 2008

COWLES FOUNDATION DISCUSSION PAPER NO. 1660

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Testing for Non-Nested Conditional Moment Restrictions using Unconditional Empirical Likelihood*

Taisuke Otsu† Yale University
Myung Hwan Seo‡ London School of Economics
Yoon-Jae Whang§ Seoul National University

May 16, 2008

Abstract

We propose non-nested hypotheses tests for conditional moment restriction models based on the method of generalized empirical likelihood (GEL). By utilizing the implied GEL probabilities from a sequence of unconditional moment restrictions that contains equivalent information of the conditional moment restrictions, we construct Kolmogorov-Smirnov and Cramér-von Mises type moment encompassing tests. Advantages of our tests over Otsu and Whang's (2007) tests are: (i) they are free from smoothing parameters, (ii) they can be applied to weakly dependent data, and (iii) they allow non-smooth moment functions. We derive the null distributions, validity of a bootstrap procedure, and local and global power properties of our tests. The simulation results show that our tests have reasonable size and power performance in finite samples.

Keywords: Empirical likelihood; Non-nested tests; Conditional moment restrictions
JEL Codes: C12, C13, C14, C22

*We would like to thank Yuichi Kitamura for helpful comments.

†Cowles Foundation and Department of Economics, Yale University, New Haven CT 06520, U.S.A. E-mail address: taisuke.otsu@yale.edu. Financial support from the National Science Foundation (SES-0720961) is gratefully acknowledged.

‡Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom. E-mail address: m.seo@lse.ac.uk

§Department of Economics, Seoul National University, Seoul 151-742, Korea. E-mail address : whang@snu.ac.kr. Thanks to Seoul National University for financial support.

1 Introduction

Since the pioneering works of Cox (1961, 1962), testing for non-nested competitive statistical models has become a standard technique to evaluate specification of a statistical model against a specific alternative model, see, e.g., MacKinnon (1983), Gourieroux and Monfort (1994) and Pesaran and Weeks (2001) for a review of non-nested testing. The purpose of this paper is to develop non-nested hypotheses tests for conditional moment restriction models which are common in economics.

Singleton (1985), Ghysels and Hall (1990), and Smith (1992) proposed non-nested testing procedures for unconditional moment restriction models. Their procedures are extended by Smith (1997) and Ramalho and Smith (2002) to the empirical likelihood context.[1] However, these procedures are not suitable to test conditional moment restriction models which imply an infinite number of unconditional moment restrictions. On the other hand, Otsu and Whang (2007) extended the empirical likelihood approach of Smith (1997) and Ramalho and Smith (2002) to test non-nested conditional moment restriction models. In particular, Otsu and Whang (2007) applied the method of conditional empirical likelihood by Kitamura, Tripathi and Ahn (2004) and Zhang and Gijbels (2003) and constructed non-nested test statistics based on the implied conditional probabilities from the conditional moment restrictions.

In this paper, we extend the results of Otsu and Whang (2007) in the following senses. First, instead of conditional empirical likelihood, we employ the ordinary unconditional empirical likelihood approach to test conditional moment restrictions. To do so, we represent conditional moment restrictions by sequences of unconditional moment restrictions indexed by real numbers. Our unconditional empirical likelihood approach has an advantage over Otsu and Whang's (2007) conditional empirical likelihood approach since the former does not require a choice of smoothing parameters, which can be arbitrary in practice. Second, our setup allows the observations to be weakly dependent so that our tests can be applied in time series applications. For example, our tests can be used to test competing asset pricing models for financial data. To the best of our knowledge, the conditional empirical likelihood approach may not be readily extended to time series contexts. Third, our setup allows non-smooth moment functions. This extension is useful, for example, to test non-nested quantile regression models. Fourth, we allow the implied probabilities to be computed by any member of generalized empirical likelihood (GEL), which includes empirical likelihood, exponential tilting, and continuously updating GMM as special cases. Since it is known that each member of GEL shows different finite sample performances, this extension might be useful to practitioners.[2]

This paper is organized as follows. Section 2 introduces our basic set-up and test statistics. Section 3 investigates the asymptotic properties of the proposed tests. Sections 3.1 derives the null distributions of the test statistics. Section 3.2 studies their local power properties. Section 3.3 discusses the global

[1] See Owen (2001) for a review of empirical likelihood.

[2] GEL is originally proposed by Smith (1997) and its higher order properties are investigated by Newey and Smith (2004).

power properties of our tests. Section 4 describes the block bootstrap procedure to compute the critical values and gives its asymptotic justification. Section 5 reports simulation results. Section 6 concludes.

We use the following notation. The abbreviations "a.s." and "w.p.a.1" mean "almost surely" and "with probability approaching one," respectively. "$\xrightarrow{p}$" and "$\Rightarrow$" mean the convergence in probability and weak convergence, respectively. $\|A\| = \sqrt{\text{trace}\,(AA')}$ is the Frobenius norm for a scalar, vector, or matrix. A^{-}, $\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ are a g-inverse, the minimum eigenvalue, and the maximum eigenvalue of a matrix A, respectively. $1(A)$ is the indicator function for an event A. $int(A)$ is the interior of a set A. $a^{(i)}$ means the i-th component of a vector a.

2 Set-up and Test Statistics

Suppose that we observe weakly dependent data $\{(X_t, Z_t) : t = 1, ...T\}$, where $X_t \in \mathcal{X} \subset \mathbb{R}^{d_x}$ and $Z_t \in \mathbb{R}^{d_z}$. Let $\mu_{X,Z}$, μ_X, μ_Z, and $\mu_{Z|X}$ denote the joint probability law of (X_t, Z_t), the marginal law of X_t, the marginal law of Z_t, and the conditional law of Z_t given X_t, respectively. Let $g : \mathbb{R}^{d_z} \times \mathcal{B} \to \mathbb{R}^{d_g}$ and $h : \mathbb{R}^{d_z} \times \Gamma \to \mathbb{R}^{d_h}$ be vectors of moment functions, where $\mathcal{B} \subset \mathbb{R}^{d_\beta}$ and $\Gamma \subset \mathbb{R}^{d_\gamma}$ are parameter spaces. Consider the two competing hypotheses written by the conditional moment restrictions:

$$\mathbf{H}_g \ : \ \int g(z, \beta_0) d\mu_{Z|X} = 0 \text{ a.s. } \mu_X \text{ for some } \beta_0 \in \mathcal{B}, \tag{1}$$

$$\mathbf{H}_h \ : \ \int h(z, \gamma_0) d\mu_{Z|X} = 0 \text{ a.s. } \mu_X \text{ for some } \gamma_0 \in \Gamma. \tag{2}$$

Except for the conditional moment restrictions, these hypotheses do not impose any parametric assumptions on the distribution forms of $\mu_{Z|X}$ and μ_X. In this sense, these hypotheses are semiparametric. It is known that the hypotheses $\mathbf{H}_g$ and $\mathbf{H}_h$ are equivalently written as (see, e.g., Billinglsey (1995, Theorem 16.10 (iii)))

$$\mathbf{H}_g \ : \ \int g(z, \beta_0) 1(x \le u) d\mu_{X,Z} = 0 \text{ for all } u \in \mathcal{X} \text{ for some } \beta_0 \in \mathcal{B}, \tag{3}$$

$$\mathbf{H}_h \ : \ \int h(z, \gamma_0) 1(x \le u) d\mu_{X,Z} = 0 \text{ for all } u \in \mathcal{X} \text{ for some } \gamma_0 \in \Gamma. \tag{4}$$

In other words, a finite number of conditional moment restrictions on the conditional law $\mu_{Z|X}$ can be equivalently represented by an infinite number of unconditional moment restrictions on the joint law $\mu_{Z,X}$. In this paper, we assume that the above hypotheses are non-nested, i.e., if $\mathbf{H}_g$ holds true, then for any $\gamma \in \Gamma$ there exists some $u \in \mathcal{X}$ such that $\int h(z, \gamma) 1(x \le u) d\mu_{X,Z} \neq 0$. Otsu and Whang (2007) proposed non-nested test statistics based on the implied conditional probabilities from the conditional moment restrictions in (1) and (2). On the other hand, this paper focuses on the sequences of the unconditional moment restrictions indexed by u in (3) and (4) and utilizes the implied unconditional probabilities to construct non-nested test statistics.

To obtain the implied unconditional probabilities from (3) and (4), we employ the GEL approach. See Smith (1997) and Newey and Smith (2004) for detailed discussions on GEL. Let $g_t(u, \beta) =$

$g(Z_t,\beta)1(X_t \le u)$ be the unconditional moment function of (3). At given u and β, the GEL function for the unconditional moment restriction (3) can be written as

$$\sup_{\lambda\in\hat{\Lambda}_T(\beta)} \sum_{t=1}^{T} \rho(\lambda' g_t(u,\beta)),$$

where $\rho : \mathcal{V} \to \mathbb{R}$ is a criterion function defined on an open interval $\mathcal{V}$ containing zero, and $\hat{\Lambda}_T(\beta) = \{\lambda \in \mathbb{R}^{d_g} : \lambda' g(Z_t,\beta) \in \mathcal{V}, t = 1,\dots,T\}$ is the support of the auxiliary parameter λ. Popular choices for ρ are $\rho(v) = \log(1-v)$ (empirical likelihood by Owen (1988) and Qin and Lawless (1994)), $\rho(v) = -e^v$ (exponential tilting by Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)), and $\rho(v) = -(1+v)^2/2$ (continuous updating GMM by Hansen, Heaton and Yaron (1996)). Suppose we have an estimator $\hat{\beta}$ for β_0 such as Dominguez and Lobato's (2004) estimator. Let $\hat{g}_t(u) = g_t(u,\hat{\beta})$,

$$\hat{\lambda}(u) = \arg\max_{\lambda\in\hat{\Lambda}_T(\hat{\beta})} \sum_{t=1}^{T} \rho(\lambda' \hat{g}_t(u)), \tag{5}$$

and $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ whenever it exists. The GEL-based implied unconditional probabilities from (3) at u and $\hat{\beta}$ are obtained as

$$p_t^g(u) = \frac{\rho_1(\hat{\lambda}(u)'\hat{g}_t(u))}{\sum_{s=1}^{T} \rho_1(\hat{\lambda}(u)'\hat{g}_s(u))},$$

for $t = 1,\dots,T$. In the same manner, we can define $\hat{\gamma}$, $h_t(u,\gamma)$, $\hat{h}_t(u)$, $\hat{\gamma}(u)$, and $p_t^h(u)$.

We consider a testing problem for the null hypothesis $\mathbf{H}_g$ against the non-nested alternative hypothesis $\mathbf{H}_h$. To construct test statistics, we adopt the moment encompassing approach (see, e.g., Ramalho and Smith (2002)). Our test statistics are based on the following contrast:

$$M_T(u) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t(u) - \sum_{t=1}^{T} p_t^g(u)\hat{h}_t(u)\right),$$

for $u \in \mathcal{X}$. Note that the first and second terms are sample analogs of the population moment $E[h(Z_t,\gamma)1(X_t \le u)]$ evaluated under the uniform weight (i.e., $1/T$) and the implied probabilities $\{p_t^g(u) : t = 1,\dots,T\}$. Under the null hypothesis $\mathbf{H}_g$, both analogs are consistent for $E[h(Z_t,\gamma)1(X_t \le u)]$ and thus we can expect that the contrast $M_T(u)$ will not diverge. On the other hand, if the null hypothesis $\mathbf{H}_g$ does not hold, the second term $\sum_{t=1}^{T} p_t^g(u)\hat{h}_t(u)$ is typically inconsistent for estimating $E[h(Z_t,\gamma)1(X_t \le u)]$ and the contrast $M_T(u)$ will diverge in general. Therefore, based on the sequence of the contrasts $\{M_T(u) : u \in \mathcal{X}\}$, we propose the Kolmogorov-Smirnov- and Cramér-von Mises-type test statistics

$$\begin{aligned} M_{KS} &= \sup_{u\in\mathcal{X}} \|M_T(u)\|, \\ M_{CM} &= \int \|M_T(u)\|^2 \, d\hat{\mu}_X, \end{aligned}$$

where $\hat{\mu}_X$ is the empirical measure of $\{X_t : t = 1,\dots,T\}$. The test statistics for testing the null hypothesis $\mathbf{H}_h$ against the alternative hypothesis $\mathbf{H}_g$ can be constructed in the same manner. Note that although the non-nested test statistics of Otsu and Whang (2007) require to choose smoothing parameters to compute the conditional implied probabilities, our test statistics are free from those smoothing parameters.

3 Asymptotic Properties

3.1 Null Distributions

This subsection derives the asymptotic null distributions of our non-nested test statistics. We introduce some notation. Denote

$$\begin{aligned}\Omega(u,\beta) &= E\left[g_t(u,\beta)g_t(u,\beta)'\right], \quad \Omega(u)=\Omega(u,\beta_0),\\ \Omega_h(u,\beta,\gamma) &= E\left[h_t(u,\gamma)\,g_t(u,\beta)'\right], \quad \Omega_h(u)=\Omega_h(u,\beta_0,\gamma_*),\\ D(u,\beta) &= \partial E\left[g_t(u,\beta)\right]/\partial\beta', \quad D(u)=D(u,\beta_0).\end{aligned}$$

Let $g_t(u)=g_t(u,\beta_0)$ and

$$\nu_T(u,\beta)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(g_t(u,\beta)-E[g_t(u,\beta)]\right)-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(g_t(u)-E[g_t(u)]\right) \tag{6}$$

be the empirical process evaluated at $g_t(u,\beta)-g_t(u)$. Let γ_* be the probability limit of $\hat{\gamma}$ under $\mathbf{H}_g$. To obtain the null distributions, we impose the following assumptions.

Assumption 3.1 *$\rho:\mathcal{V}\to\mathbb{R}$ is twice continuously differentiable and concave on an open interval $\mathcal{V}$ containing zero, and $\rho_j(0)=-1$ for $j=1$ and 2.*

Assumption 3.2 *$\{(X_t,Z_t):t=1,2,\ldots\}$ is a strictly stationary β-mixing sequence on $\mathcal{X}\times\mathbb{R}^{d_z}$ whose mixing coefficient is of order $O\left(n^{-b}\right)$ for some $b>r/(r-1)$ with some $r>1$, $\mathcal{X}\subset\mathbb{R}^{d_x}$ is compact, and there exists a constant $C>0$ such that $\Pr\{X_t^{(i)}\in[a,b]\}\le C(b-a)$ for any $a\le b\in\mathcal{X}$ and $i=1,...,d_x$.*

Assumption 3.3

(i) *$\mathcal{B}$ is compact and $\beta_0\in int(\mathcal{B})$.*

(ii) *$\nu_T(u,\beta)$ satisfies*

$$\sup_{\beta_1,\beta_2\in\{\beta_1,\beta_2\in\mathcal{B}:|\beta_1-\beta_2|<\delta_T\},u\in\mathcal{X}}\left\|\nu_T(u,\beta_1)-\nu_T(u,\beta_2)\right\|\xrightarrow{p}0,$$

for any sequence $\delta_T\to 0$.

(iii) *There exists a neighborhood $\mathcal{B}_0$ around β_0 such that $E[\sup_{\beta\in\mathcal{B}_0}\|g(Z_t,\beta)\|^{\alpha}]<\infty$ for some $\alpha>4$ and $\sup_{\beta\in\mathcal{B}_0,u\in\mathcal{X}}|D(u,\beta)|<\infty$, $D(u,\beta)$ is continuous at $\beta=\beta_0$ uniformly in $u\in\mathcal{X}$, and $D(u)$ is full column rank for all $u\in\mathcal{X}$.*

(iv) $\inf_{u\in\mathcal{X}} \lambda_{\min}(\Omega(u)) > 0$*, and* $\Omega(u,\beta)$ *is continuous at* $\beta = \beta_0$ *uniformly in* $u \in \mathcal{X}$*.*

Assumption 3.4

(i) $\sup_{\gamma\in\Gamma_*,u\in\mathcal{X}} \left\| \frac{1}{T}\sum_{t=1}^{T} h_t(u,\gamma) - E[h_t(u,\gamma)] \right\| \xrightarrow{p} 0$ *and*
$\sup_{\gamma\in\Gamma_*,\beta\in\mathcal{B},u\in\mathcal{X}} \left\| \frac{1}{T}\sum_{t=1}^{T} h_t(u,\gamma) g_t(u,\beta)' - E\left[h_t(u,\gamma) g_t(u,\beta)'\right] \right\| \xrightarrow{p} 0.$

(ii) *There exists a neighborhood* Γ_* *around* γ_* *such that* $E[\sup_{\gamma\in\Gamma_*} \|h(Z_t,\gamma)\|^{\alpha}] < \infty$ *for some* $\alpha > 4$*.*

(iii) $\Omega_h(u)$ *is finite and is full column rank for all* $u \in \mathcal{X}$*.*

Assumption 3.5

(i) $\sqrt{T}(\hat{\beta} - \beta_0) = -T^{-1/2}\Delta_0 \sum_{t=1}^{T} \psi(X_t, Z_t, \beta_0) + o_p(1)$*, where* Δ_0 *is a non-stochastic* $d_\beta \times d_\beta$ *matrix,* $E[\psi(X_t, Z_t, \beta_0)] = 0$ *and* $E\left[\|\psi(X_t, Z_t, \beta_0)\|^{\xi}\right] < \infty$ *for some* $\xi > 2$*.*

(ii) $\|\hat{\gamma} - \gamma_*\| = O_p(T^{-1/2})$*.*

Assumption 3.1 is on the GEL criterion function. Popular criterion functions such as empirical likelihood, exponential tilting, and continuous updating GMM satisfy this assumption. Assumption 3.2 is on the data $\{(X_t, Z_t) : t = 1, 2, \ldots\}$. Compared to Otsu and Whang (2007) who assume iid data, we allow dependent data. Thus, for example, our method can be applied to test competing asset pricing models based on financial time series data. The second condition in this assumption is used to establish a stochastic equicontinuity and is satisfied when the density of X_t is bounded. If the support of X_t is finite, this condition is irrelevant and the limit process $M(u)$ in Theorem 3.1 below becomes a multivariate Normal. Assumption 3.3 contains the conditions for the moment function $g(Z_t, \beta)$. Assumption 3.3 (i) is standard. Assumption 3.3 (ii) is a stochastic equicontinuity condition on the empirical process $\nu_T(u, \beta)$. It is important to note that the moment function $g(Z_t, \beta)$ need not to be smooth in β. Note that

$$\|g(Z_t, \beta_1) 1(X_t \le u) - g(Z_t, \beta_2) 1(X_t \le u)\| \le \|g(Z_t, \beta_1) - g(Z_t, \beta_2)\|$$

for any $\beta_1, \beta_2 \in \mathcal{B}$. So, Assumption 3.3 (ii) is satisfied if g is in a P-Donsker class, which includes for example Lipschitz continuous functions and indicator functions (see, e.g., Andrews (1993, Section 4) for a detail). Therefore, for example, our setup allows quantile regression models, where $g(Z_t, \beta) = 1(Y_t - X_t'\beta \le 0) - q$ and $Z_t = (Y_t, X_t')'$ for some $q \in (0, 1)$. In contrast, Otsu and Whang (2007) do not allow non-smooth moment functions. Assumption 3.3 (iii) and (iv) are boundedness and rank conditions for moments related to $g(Z_t, \beta)$ that are required to apply limit theorems. Assumption 3.4 lists the conditions for the alternative moment function $h(Z_t, \gamma)$. Assumption 3.4 (i) contains uniform laws of large numbers for $h_t(u, \gamma)$ and $h_t(u, \gamma) g_t(u, \beta)'$. See, e.g., Andrews (1987) and Pötscher and Prucha (1989) for more primitive conditions on this assumption. Our setup also allows non-smoothness on $h_t(u,\gamma)$. Assumption 3.5 is on the estimators $\hat{\beta}$ and $\hat{\gamma}$. This assumption is satisfied by many $T^{1/2}-$ consistent parametric and semiparametric estimators in the literature, e.g., maximum likelihood estimator, generalized method of moment estimators and the estimators of Donald, Imbens and Newey (2003), Kitamura, Tripathi and Ahn (2004), and Dominguez and Lobato (2004). Although Assumption 3.5 (i) implies the asymptotic normality of $\hat{\beta}$, it does not require the estimator $\hat{\beta}$ to be asymptotically efficient.

To derive the asymptotic null distribution, we need some additional notation. Let $\psi_t = \psi(X_t, Z_t, \beta_0)$ and define a mean zero Gaussian process $(G(u)', \Psi')'$ with the covariance kernel

$$V(u_1,u_2) = \lim_{T\to\infty} \operatorname{cov}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} g_t(u_1) \\ \psi_t \end{pmatrix}, \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} g_t(u_2) \\ \psi_t \end{pmatrix}' \right).$$

Then $F(u) = \Omega(u)^{-1}(G(u) - D(u)\Delta_0\Psi)$ is a mean zero Gaussian process with the covariance kernel

$$V_F(u_1,u_2) = \Omega(u_1)^{-1}[I_{d_g} : -D(u_1)\Delta_0]V(u_1,u_2)\left[I_{d_g} : -D(u_2)\Delta_0\right]'\Omega(u_2)^{-1}.$$

Based on the above assumptions and notation, we obtain the null distribution of our test statistics.

Theorem 3.1 (Null distributions) *Suppose that Assumptions 3.1-3.5 hold. Then under the null hypothesis* $\mathbf{H}_g$,

$$M_T(u) \Rightarrow M(u),$$

with $M(u) = \Omega_h(u)F(u)$, *and*

$$M_{KS} \Longrightarrow \sup_{u\in\mathcal{X}} \|M(u)\|,$$
$$M_{CM} \Longrightarrow \int \|M(u)\|^2 d\mu_X.$$

The limit process $\|M(u)\|^2$ is a chi-square process as the distribution of $\|M(u)\|^2$ for a fixed u is chi-square with degree of freedom d_h. The asymptotic distributions of the test statistics M_{KS} and M_{CM} cannot be tabulated in general as they depend on several unknown components such as the covariance kernel $V(u_1,u_2)$. Therefore, we consider a bootstrap procedure to conduct valid inference in Section 4.

3.2 Local Power

We now evaluate the local power properties of the proposed non-nested test statistics. Consider a sequence of local alternatives that converge to the null hypothesis at $T^{-1/2}$ rate, that is,

$$\mathbf{H}_{gT} : \int g(z,\beta_0)d\mu^T_{Z|X} = T^{-1/2}\pi(X), \quad \beta_0 \in \mathcal{B} \text{ (a.s. } \mu_X),$$

where $\mu^T_{Z|X}$ denote the conditional law of Z_t given X_t under the local alternatives $\mathbf{H}_{gT}$, and $\pi : \mathcal{X} \to \mathbb{R}^{d_g}$ is a non-zero function satisfying $\int \|\pi(x)\| \, d\mu_X < \infty$. Let $\Pi(u) = E[\pi(X_t) 1(X_t \leq u)]$. Similar to (3), the local alternative hypothesis $\mathbf{H}_{gT}$ can be equivalently written as

$$\mathbf{H}_{gT} : \int g(z, \beta_0) 1(x \leq u) \, d\mu^T_{X,Z} = T^{-1/2}\Pi(u), \quad \beta_0 \in \mathcal{B} \text{ for all } u \in \mathcal{X}. \tag{7}$$

Let

$$\varrho_T(u, \beta) = \int g(z, \beta) 1(x \leq u) \, d\mu^T_{X,Z}$$

compared to $\varrho(u, \beta) = \int g(z, \beta) 1(x \leq u) \, d\mu_{X,Z}$, which is the limit law of $\mu^T_{X,Z}$ as $T \to \infty$. Let μ^T_Z be the marginal law of Z_t under the local alternatives $\mathbf{H}_{gT}$. This local alternative also demands a change in the limit of $\hat{\beta}$. The relevant change and some strengthening of the previous assumptions are collected here:

Assumption 3.6

(i) $\sqrt{T}(\hat{\beta} - \beta_0) = -T^{-1/2}\Delta_0 \sum_{t=1}^T \psi(X_t, Z_t, \beta_0) + o_p(1)$, *where* Δ_0 *is a non-stochastic* $d_\beta \times d_\beta$ *matrix,* $T^{1/2} \int\int \psi(x, z, \beta_0) d\mu^T_{Z|X} d\mu_X \to \eta_0 < \infty$, $\sup_{T \geq 1} \iint ||\psi(x, z, \beta_0)||^\xi d\mu^T_{Z|X} d\mu_X < \infty$ *for some* $\xi > 2$, *and* $\iint \psi(x, z, \beta_0)\psi(x, z, \beta_0)' d\left(\mu^T_{Z|X} - \mu_{Z|X}\right) d\mu_X \to 0$.

(ii) $\sup_{T \geq 1} \int \sup_{\beta \in \mathcal{B}_0} \|g(z, \beta)\|^\alpha \, d\mu^T_Z < \infty$ *for some* $\alpha > 4$, *and* $\varrho_T(u, \beta) = \varrho(u, \beta) + T^{-1/2}\Pi(u, \beta)$, *where* $\Pi(u, \beta)$ *is continuous at* $\beta = \beta_0$ *uniformly in* $u \in \mathcal{X}$ *and* $\Pi(u, \beta_0) = \Pi(u)$. *Furthermore, Assumption 3.3 (ii) holds under* $\mu^T_{Z,X}$.

(iii) $\sup_{T \geq 1} \int \sup_{\gamma \in \Gamma_*} \|h(z, \gamma)\|^\alpha \, d\mu^T_Z < \infty$, *for some* $\alpha > 4$ *and* $\sup_{\gamma \in \Gamma_*, u \in \mathcal{X}} \left\| \frac{1}{T} \sum_{t=1}^T h_t(u, \gamma) - \int h_t(u, \gamma) \, d\mu^T_{Z,X} \right\| = o_p(1)$.

(iv) $\iint [g(z, \beta_0)g(z, \beta_0)' 1(x \leq u)] \, d\left(\mu^T_{Z|X} - \mu_{Z|X}\right) d\mu_X \to 0$ *uniformly in* $u \in \mathcal{X}$ *and* $\iint [g(z, \beta_0)h(z, \gamma)' 1(x \leq u)] \, d\left(\mu^T_{Z|X} - \mu_{Z|X}\right) d\mu_X \to 0$ *uniformly in* $u \in \mathcal{X}$ *and in* $\gamma \in \Gamma_*$.

Theorem 3.2 (Local power) *Suppose Assumptions 3.1-3.4, and 3.6 hold. Then under the local alternative hypothesis* $\mathbf{H}_{gT}$,

$$M_{KS} \implies \sup_{u \in \mathcal{X}} \|M(u) + \zeta(u)\|,$$
$$M_{CM} \implies \int \|M(u) + \zeta(u)\|^2 \, d\mu_X.$$

where $\zeta(u) = \Omega_h(u) [\Pi(u) - D(u)\Delta_0\eta_0]$.

Theorem 2 implies that our tests have non-trivial power against a sequence of $T^{-1/2}$ local alternatives and asymptotic local powers of the tests M_{KS} and M_{CM} are given by $\Pr\left\{\sup_{u \in \mathcal{X}} \|M(u) + \zeta(u)\| > c^{KS}_\alpha\right\}$ and $\Pr\left\{\int \|M(u) + \zeta(u)\|^2 \, d\mu_X > c^{CM}_\alpha\right\}$, where c^{KS}_α and c^{CM}_α denote the $(1-\alpha)$-th quantile of the asymptotic null distributions of M_{KS} and M_{CM} given in Theorem 3.1, respectively.

3.3 Global Power

This subsection investigates the global power properties of the proposed non-nested tests under the fixed (or non-local) alternative hypothesis $\mathbf{H}_h$. Since our non-nested test statistics are constructed against a specific alternative hypothesis $\mathbf{H}_h$, the global power analysis is crucial to assess the validity of the proposed test statistics. Let β_* be the probability limit of the estimator $\hat{\beta}$ under the true measure satisfying $\mathbf{H}_h$. Since the conditional moment restriction in $\mathbf{H}_g$ is misspecified, $\hat{\beta}$ does not converge to β_0 in general. Similarly, let $\lambda_*(u)$ be the probability limit of $\hat{\lambda}(u)$ under $\mathbf{H}_h$. Note that $\lambda_*(u)$ depends on the choice of the GEL criterion function ρ. We impose the following conditions.

Assumption 3.7

(i) $\hat{\beta} \xrightarrow{p} \beta_*$, $\hat{\gamma} \xrightarrow{p} \gamma_0$, *and* $\sup_{u \in \mathcal{X}} \left\| \hat{\lambda}(u) - \lambda_*(u) \right\| \xrightarrow{p} 0$.

(ii) *There exist neighborhoods* $\mathcal{B}_*$, Γ_0, *and* $\Lambda_*(u)$ *for* $u \in \mathcal{X}$ *around* β_*, γ_0, *and* $\lambda_*(u)$, *respectively, such that*

$$\sup_{\gamma \in \Gamma_0} \left\| \frac{1}{T} \sum_{t=1}^{T} h_t(u, \gamma) - E[h_t(u, \gamma)] \right\| \xrightarrow{p} 0,$$

$$\sup_{\lambda \in \Lambda_*(u), \beta \in \mathcal{B}_*} \left\| \frac{1}{T} \sum_{t=1}^{T} \rho_1\left(\lambda' g_t(u, \beta)\right) - E\left[\rho_1\left(\lambda' g_t(u, \beta)\right)\right] \right\| \xrightarrow{p} 0,$$

$$\sup_{\lambda \in \Lambda_*(u), \beta \in \mathcal{B}_*, \gamma \in \Gamma_0} \left\| \frac{1}{T} \sum_{t=1}^{T} \rho_1\left(\lambda' g_t(u, \beta)\right) h_t(u, \gamma) - E\left[\rho_1\left(\lambda' g_t(u, \beta)\right) h_t(u, \gamma)\right] \right\| \xrightarrow{p} 0,$$

uniformly in $u \in \mathcal{X}$.

(iii) $E[h_t(u, \gamma)]$, $E\left[\rho_1\left(\lambda' g_t(u, \beta)\right)\right]$, *and* $E\left[\rho_1\left(\lambda' g_t(u, \beta)\right) h_t(u, \gamma)\right]$ *are continuous at* γ_0, $(\lambda_*(u), \beta_*)$, *and* $(\lambda_*(u), \beta_*, \gamma_0)$ *uniformly in* $u \in \mathcal{X}$, *respectively.*

Assumption 3.7 (i) is on the consistency of $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\lambda}(u)$ to β_*, γ_0, and $\lambda_*(u)$, respectively. We do not need $\sqrt{T}$-consistency of the estimators for the global power analysis. Assumption 3.7 (ii) contains uniform laws of large numbers for $h_t(u, \gamma)$, $\rho_1\left(\lambda' g_t(u, \beta)\right)$, and $\rho_1\left(\lambda' g_t(u, \beta)\right) h_t(u, \gamma)$. Since $\hat{\lambda}(u)$ does not converge to zero under the alternative hypothesis $\mathbf{H}_h$, we need to directly control the behaviors of $\rho_1\left(\lambda' g_t(u, \beta)\right)$ with respect to λ (see, Hong, Preston and Shum (2003) and Chen, Hong and Shum (2007) for similar arguments). Assumption 3.7 (iii) is on the continuity of the probability limits in Assumption 3.7 (ii). Assumption 3.7 (i) and (ii) are relatively higher level. More primitive assumptions can be found by applying similar arguments to Lemma B.1 and Hong, Preston and Shum (2003, Lemma 1). The consistency results for the tests based on M_{KS} and M_{CM} are obtained as follows.

Theorem 3.3 (Consistency) *Suppose that Assumption 3.7 holds. Then under the alternative hypothesis* $\mathbf{H}_h$,

$$\sup_{u \in \mathcal{X}} \left\| \frac{M_T(u)}{\sqrt{T}} - \mu_*(u) \right\| \xrightarrow{p} 0,$$

where

$$\mu_*(u) = -\frac{E\left[\rho_1(\lambda_*(u)'g_t(u,\beta_*))h_t(u,\gamma_0)\right]}{E[\rho_1(\lambda_*(u)'g_t(u,\beta_*))]}.$$

Therefore, if there exists $u_* \in \mathcal{X}$ such that $\|\mu_*(u_*)\| \neq 0$, then the non-nested tests based on M_{KS} and M_{CM} are consistent against $\mathbf{H}_h$.

This theorem says that in order to guarantee the consistency of our non-nested tests, we need to check whether the noncentrality parameter $\mu_*(u)$ has at least one point u_* in $\mathcal{X}$ satisfying $\|\mu_*(u_*)\| \neq 0$. Since the noncentrality parameter $\mu_*(u)$ depends on the first-order derivative of the GEL criterion function ρ_1 and moment functions g and h, it is difficult to find a general condition to guarantee the consistency. Thus, we hereafter consider some specific examples.

First, in order to compare popular members of GEL, such as the continuous updating GMM and empirical likelihood, we specify the GEL criterion function by the Cressie and Read (1984) divergence family:

$$\rho(v) = -\frac{(1+\phi v)^{(\phi+1)/\phi}}{\phi+1},$$

for some constant ϕ. This family includes the continuous updating GMM ($\phi = 1$), empirical likelihood ($\phi \to -1$), and exponential tilting ($\phi \to 0$) as special cases. From $\rho_1(v) = -(1+\phi v)^{1/\phi}$, the noncentrality parameter $\mu_*(u)$ is written as

$$\mu_*(u) = -\frac{E\left[(1+\phi\lambda_*(u)'g_t(u,\beta_*))^{1/\phi}h_t(u,\gamma_0)\right]}{E[(1+\phi\lambda_*(u)'g_t(u,\beta_*))^{1/\phi}]}.$$

Note that since $\mu_*(u)$ depends on $\lambda_*(u)$, the probability limit of the Lagrange multiplier $\hat{\lambda}(u)$ under $\mathbf{H}_h$, it is not easy to find an intuitive condition for the consistency. To proceed furthermore, we focus on the case of the continuous updating GMM, i.e., the case of $\phi = 1$. In this case, the Lagrange multiplier $\hat{\lambda}(u)$ defined in (5) has an explicit solution, that is

$$\hat{\lambda}(u) = \hat{\Omega}(u,\hat{\beta})^{-1}\frac{1}{T}\sum_{t=1}^{T} g_t(u,\hat{\beta}),$$

where $\hat{\Omega}(u,\beta) = \frac{1}{T}\sum_{t=1}^{T}\left(g_t(u,\beta) - \frac{1}{T}\sum_{t=1}^{T} g_t(u,\beta)\right)\left(g_t(u,\beta) - \frac{1}{T}\sum_{t=1}^{T} g_t(u,\beta)\right)'$. Under certain regularity conditions the probability limit of $\hat{\lambda}(u)$ under $\mathbf{H}_h$ can be written as

$$\lambda_*(u) = \Omega_*(u,\beta_*)^{-1}E[g_t(u,\beta_*)],$$

where $\Omega_*(u,\beta_*) = E[g_t(u,\beta_*)g_t(u,\beta_*)'] - E[g_t(u,\beta_*)]E[g_t(u,\beta_*)]'$. Therefore, in this case, the noncentrality parameter $\mu_*(u)$ is

$$\begin{aligned}\mu_*(u) &= -\frac{E\left[(1+E[g_t(u,\beta_*)]'\Omega_*(u,\beta_*)^{-1}g_t(u,\beta_*))h_t(u,\gamma_0)\right]}{1+E[g_t(u,\beta_*)]'\Omega_*(u,\beta_*)^{-1}E[g_t(u,\beta_*)]} \\ &\underset{\text{under } \mathbf{H}_h}{=} -\frac{E\left[\left\{E[g_t(u,\beta_*)]'\Omega_*(u,\beta_*)^{-1}g_t(u,\beta_*)\right\}h_t(u,\gamma_0)\right]}{1+E[g_t(u,\beta_*)]'\Omega_*(u,\beta_*)^{-1}E[g_t(u,\beta_*)]}\end{aligned}$$

Since the hypotheses $\mathbf{H}_g$ and $\mathbf{H}_h$ are non-nested, there exists a non-empty set $U_h = \{u \in \mathcal{X} : E[g_t(u, \beta_*)] \neq 0 \text{ under } \mathbf{H}_h\}$. Then as far as there exists $u_* \in U_h$ and $j = 1, \ldots, d_h$ such that $\Omega_* (u_*, \beta_*)$ is positive definite and

$$E[g_t(u_*, \beta_*) h_t^{(j)}(u_*, \gamma_0)] \neq 0,$$

we can guarantee the consistency of the non-nested tests based on M_{KS} and M_{CM}. Intuitively, the above condition requires that the alternative moment function $h_t(u, \gamma_0)$ must have some correlation or prediction power with the null moment function $g_t(u, \beta_*)$. This finding is summarized in the following corollary.

Corollary 3.1 (Continuous updating GMM) *If there exists $u_* \in \mathcal{X}$ such that*

$$\begin{aligned} E[g_t(u_*, \beta_*)] &\neq 0, \quad \Omega_* (u_*, \beta_*) \text{ is positive definite,} \\ E[g_t(u_*, \beta_*) h_t^{(j)}(u_*, \gamma_0)] &\neq 0 \quad \text{for some } j = 1, \ldots, d_h, \end{aligned} \tag{8}$$

then under Assumption 3.7 with $\lambda_*(u) = \Omega_* (u, \beta_*)^{-1} E[g_t(u, \beta_*)]$, the non-nested tests based on M_{KS} and M_{CM} using the continuous updating GMM criterion function are consistent against $\mathbf{H}_h$.

Next, to explore the conditions in (8), we consider the non-nested (possibly nonlinear) regression models:

$$\begin{aligned} \mathbf{H}_g^{reg} &: Y_t = G(X_t, \beta_0) + e_t, \quad E[e_t | X_t] = 0 \text{ (a.s. } \mu_X), \\ \mathbf{H}_h^{reg} &: Y_t = H(X_t, \gamma_0) + v_t, \quad E[v_t | X_t] = 0 \text{ (a.s. } \mu_X). \end{aligned}$$

Since $\mathbf{H}_g^{reg}$ and $\mathbf{H}_h^{reg}$ are non-nested, the first condition in (8) is satisfied, i.e., for some $u_* \in \mathcal{X}$,

$$E[g_t(u_*, \beta_*)] = E[(H(X_t, \gamma_0) - G(X_t, \beta_*)) 1 (X_t \leq u_*)] \neq 0. \tag{9}$$

For the second and third conditions in (8), note that

$$\begin{aligned} \Omega_* (u, \beta_*) &= E\left[(Y_t - G(X_t, \beta_*))^2 1 (X_t \leq u)\right] - E[(Y_t - G(X_t, \beta_*)) 1 (X_t \leq u)]^2 \\ &= E\left[E\left[v_t^2 | X_t\right] 1 (X_t \leq u)\right] + Var\left(\{H(X_t, \gamma_0) - G(X_t, \beta_*)\} 1 (X_t \leq u)\right), \end{aligned}$$

$$\begin{aligned} E[g_t(u, \beta_*) h_t(u, \gamma_0)] &= E[(Y_t - G(X_t, \beta_*)) (Y_t - H(X_t, \gamma_0)) 1 (X_t \leq u)] \\ &= E\left[E\left[v_t^2 | X_t\right] 1 (X_t \leq u)\right]. \end{aligned}$$

Therefore, the second and third conditions in (8) are satisfied under mild assumptions: $E\left[v_t^2 | X_t\right] > 0$ (a.s. μ_X) and $Var\left((H(X_t, \gamma_0) - G(X_t, \beta_*)) 1 (X_t \leq u_*)\right) > 0$ for $u_* \in \mathcal{X}$ satisfying (9). We summarize this result in the following corollary.

Corollary 3.2 (Regression models) *If* $Var\left(\left(H(X_t,\gamma_0)-G(X_t,\beta_*)\right)1\left(X_t\le u_*\right)\right)>0$ *for* $u_*\in\mathcal{X}$ *satisfying*
$E[(H(X_t,\gamma_0)-G(X_t,\beta_*))\,1\,(X_t\le u^*)]\neq 0$ *and* $E\left[v_t^2|X_t\right]>0$ *a.s.* μ_X*, then under Assumption 3.7 with* $\lambda_*(u)=\Omega_*\left(u,\beta_*\right)^{-1}E[g_t(u,\beta_*)]$*, the non-nested tests for* $\mathbf{H}_g^{reg}$ *against* $\mathbf{H}_h^{reg}$ *based on* M_{KS} *and* M_{CM} *using the continuous updating GMM criterion function are consistent against* $\mathbf{H}_h^{reg}$*.*

Finally, to obtain further insights on the condition of the consistency, we apply the theory of information geometry or ϕ-divergence by Csiszár (1975, 1995). Here we assume that the sample $\{(X_t,Z_t):t=1,\ldots,T\}$ is iid. Observe that the noncentrality parameter $\mu_*(u)$ is written as

$$\begin{aligned}\mu_*(u) &= -\int h_t(u,\gamma_0)\frac{\rho_1(\lambda_*(u)'g_t(u,\beta_*))}{E[\rho_1(\lambda_*(u)'g_t(u,\beta_*))]}d\mu_{X,Z}\\ &= -\int h_t(u,\gamma_0)d\mu_{X,Z}^{(u)*},\end{aligned}$$

where $\frac{d\mu_{X,Z}^{(u)*}}{d\mu_{X,Z}}=\frac{\rho_1(\lambda_*(u)'g_t(u,\beta_*))}{E[\rho_1(\lambda_*(u)'g_t(u,\beta_*))]}$. Let $\mathcal{G}(u,\beta)$ be a set of joint measures for (X_t,Z_t) satisfying $E[g_t(u,\beta)]=0$, i.e.,

$$\mathcal{G}(u,\beta)=\left\{\mu\in\mathcal{M}_{X,Z}:\int g_t(u,\beta)d\mu=0\right\},$$

where $\mathcal{M}_{X,Z}$ is the set of all joint measures for (X_t,Z_t). Based on Csiszár (1995), the probability measure $\mu_{X,Z}^{(u)*}$ can be considered as the best approximation of the true joint measure $\mu_{X,Z}$ to the set $\mathcal{G}(u,\beta_*)$ by using some information divergence for probability measures. In theory of information geometry, this measure $\mu_{X,Z}^{(u)*}$ is called the ϕ- projection of $\mu_{X,Z}$ to $\mathcal{G}(u,\beta_*)$. For example, if the GEL criterion function ρ is concave and the moment function $g_t(u,\beta_*)$ is bounded, then the ϕ-projection $\mu_{X,Z}^{(u)*}$ always exists and satisfies $\mu_{X,Z}^{(u)*}\in\mathcal{G}(u,\beta_*)$ for each $u\in\mathcal{X}$ (see, Csiszár (1995)). Based on the above notation, the null hypothesis $\mathbf{H}_g$ is written as

$$\mathbf{H}_g:\mu_{X,Z}\in\mathcal{G}=\cup_{\beta\in\mathcal{B}}\cap_{u\in\mathcal{X}}\mathcal{G}(u,\beta).$$

Similarly, the alternative hypothesis $\mathbf{H}_h$ is

$$\mathbf{H}_h:\mu_{X,Z}\in\mathcal{H}=\cup_{\gamma\in\Gamma}\cap_{u\in\mathcal{X}}\mathcal{H}(u,\gamma),$$

where

$$\mathcal{H}(u,\gamma)=\left\{\mu\in\mathcal{M}_{X,Z}:\int h_t(u,\gamma)d\mu=0\right\}.$$

To guarantee $\mu_*(u)=-\int h_t(u,\gamma_0)d\mu_{X,Z}^{(u)*}\neq 0$ at some $u\in\mathcal{X}$, it is sufficient to check that $\mu_{X,Z}^{(u)*}\notin\mathcal{H}$ for some $u\in\mathcal{X}$. For example, this condition is satisfied when $\mu_{X,Z}^{(u)*}\in\mathcal{G}(u,\beta_*)$ for all $u\in\mathcal{X}$ and $\mathcal{G}(u,\beta_*)\cap\mathcal{H}$ is empty for some $u\in\mathcal{X}$. The latter condition says that the set $\mathcal{G}(u,\beta_*)$ created by the null moment function $g_t(u,\beta_*)$ should not intersect with the set of the alternative hypothesis $\mathcal{H}$. We obtain the following result.

Corollary 3.3 (Information geometry) *Suppose that Assumption 3.7 holds with iid data* $\{(X_t,Z_t):t=1,\ldots,T\}$ *and* $g\left(Z_t,\beta_*\right)$ *is bounded. If there exists* $u_*\in\mathcal{X}$ *such that* $\mathcal{G}(u_*,\beta_*)\cap\mathcal{H}$ *is empty, then the non-nested tests based on* M_{KS} *and* M_{CM} *are consistent against* $\mathbf{H}_h$*.*

4 Block Bootstrap

To obtain critical values for the non-nested tests based on M_{KS} and M_{CM}, we apply the moving block bootstrap (MBB) by Künsch (1989), which accommodates general dependent data. Let $W_t = (X_t, Z_t)$, b be a chosen block length, and k be the smallest integer such that $bk \geq T$. Then, define blocks $B_t = (W_t, W_{t+1}, ..., W_{t+b-1})$ for $t = 1, ..., T-b+1$ and sample k blocks independently with replacement (denoted as $B_t^*, t = 1, ..., k$). Connect those blocks end-to-end and delete the observations once the sample size reaches T to get the bootstrap sample $\{W_t^* : t = 1, \ldots, T\}$ (All the quantities corresponding to the conditional bootstrap probability measure P^* are supplied with an asterisk*). Once we obtain a bootstrap sample, we define $\hat{\beta}^*, \hat{\gamma}^*$, and the implied probabilities $p_t^{g^*}(u)$ in the same manner as $\hat{\beta}, \hat{\gamma}$ and $p_t^g(u)$. Then, the bootstrap test statistics are constructed based on

$$M_T^*(u) = \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t^*(u) - \sum_{t=1}^{T}p_t^{g^*}(u)\,\hat{h}_t^*(u)\right) - M_T(u),$$

where $\hat{h}_t^*(u) = h(Z_t^*, \hat{\gamma}^*)\,1(X_t^* \leq u)$. Construct $M_{KS}^* = \sup_{u\in\mathcal{X}}\|M_T^*(u)\|$ and $M_{CM}^* = \int \|M_T^*(u)\|\,d\hat{\mu}_X$. Now, we repeat this a large number of times and obtain the bootstrap p-values by calculating the proportion of the bootstrap statistics that are larger than the original statistics respectively.

We show that the MBB of M_{KS} (and M_{CM}) is asymptotically valid and has the same local power as the asymptotic test. This is done by showing that M_{KS}^* converges weakly to $\sup_{u\in\mathcal{X}}\|M(u)\|$ *in* P under both the null and the local alternatives. See Andrews (1997) for more discussion. The centering of the bootstrap statistic $M_T^*(u)$ eliminates the bias term $\zeta(u)$ asymptotically, which is present in the local asymptotic limit of the sample statistic $M_T(u)$.

Theorem 4.1 (Validity of bootstrap) *Suppose that Assumptions 3.1-3.5 hold. Assume further that the bootstrap consistently estimates the asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}^* - \hat{\gamma} = O_p(T^{-1/2})$. Let the mixing coefficient in Assumption 3.3 (i) is of order $O(n^{-q})$ for some $q > \alpha/(\alpha-2)$ and the block length $b = O(n^{\varepsilon})$ for some $0 < \varepsilon < (\alpha-2)/2(\alpha-1)$. Then under $\mathbf{H}_g$,*

$$\begin{aligned} M_{KS}^* &\Longrightarrow^* \sup_{u\in\mathcal{X}}\|M(u)\| \text{ in } P, \\ M_{CM}^* &\Longrightarrow^* \int \|M(u)\|^2 d\mu_X \text{ in } P. \end{aligned}$$

These also hold true under $\mathbf{H}_{gT}$ when Assumption 3.5 is replaced by Assumption 3.6.

The result draws on the bootstrap central limit theorem of Radulović (1996) for empirical processes of stationary β-mixing processes for the VC-subgraph classes of functions. $\{g_t(u)\}_{u\in\mathcal{X}}$ is a VC-subgraph class of functions by e.g. Lemma 2.6.18 of Van der Vaart and Wellner (1996). We may employ Bühlmann (1995) as his bracketing condition is satisfied as shown in Lemma B.2. However, his moment and mixing conditions are more restrictive than Radulović's while the block length condition for the MBB is more general. As the optimal length of blocks is at the order of $n^{1/3}$, our condition is not restrictive for $\alpha > 4$.

It is a corollary of this theorem and Theorem 3.3 that the test based on the MBB is consistent since M_{KS}^{*} is $O_p(1)$ under the alternative $\mathbf{H}_h$. The proof is straightforward after replacing β_0 with the pseudo true value β_* and thus omitted.

5 Monte Carlo Experiments

In this section, we investigate the finite sample performance of the tests M_{KS} and M_{CM} using Monte Carlo experiments. We consider two simulation designs. In Design I, we consider binary choice models for independent observations.[3] The logit model is the null model and the Gumbel and Burr models are the non-nested alternative models compared. The Gumbel and Burr models assume asymmetric errors, while the logit model assumes symmetric errors. These models are defined by the following conditional probabilities : For $t = 1, ..., T$,

$$\mathbf{H}_g : \Pr\{Y_t = 1|X_t\} = \frac{\exp(X_t\beta_0)}{1+\exp(X_t\beta_0)} : \text{ Logit} \tag{10}$$

$$\mathbf{H}_h \;:\; \Pr\{Y_t = 1|X_t\} = 1 - \exp(-\exp(X_t\gamma_0)) \;:\; \text{Gumbel} \tag{11}$$

$$\mathbf{H}_h \;:\; \Pr\{Y_t = 1|X_t\} = \left(\frac{\exp(X_t\gamma_0)}{1+\exp(X_t\gamma_0)}\right)^{\tau},\ \tau > 0 : \text{ Burr}(\tau) \tag{12}$$

where $\{X_t\}$ is drawn independently from the standard normal distribution and the true parameters are given by $\beta_0 = \gamma_0 = 1$. The Burr model has negative skewness for $\tau < 1$ and positive skewness for $\tau > 1$ and reduces to the logit model when $\tau = 1$.[4] We consider $\tau \in \{1/3, 2/3, 3/2, 3\}$. Note that the hypotheses (10) and (11) or (12) correspond to the conditional moment restrictions in (1) - (2) with $g(Z,\beta_0) = Y - \exp(X\beta_0)/(1+\exp(X\beta_0))$ and $h(Z,\gamma_0) = Y - 1 + \exp(-\exp(X\gamma_0))$ or $h(Z,\gamma_0) = Y - [\exp(X\gamma_0)/(1+\exp(X\gamma_0))]^{\tau}$, where $Z = (Y, X)'$.

In Design II, we consider non-nested linear quantile regression models for dependent observations: for $t = 1, \ldots, T$,

$$\mathbf{H}_g \;:\; Y_t = \beta_{01} + \beta_{02}X_{1t} + u_{gt} \tag{13}$$

$$\mathbf{H}_h \;:\; Y_t = \gamma_{01} + \gamma_{02}X_{2t} + u_{ht}, \tag{14}$$

where $X_{1t} = 0.5X_{2t} + e_{1t}$, $X_{2t} = 0.5X_{2,t-1} + e_{2t}$ $(X_{20} = 0)$, $\{e_{1t}\}$ and $\{e_{2t}\}$ are iid $N(0,1)$, $\{u_{gt}\}$ and $\{u_{ht}\}$ are iid $\log(N(0,1)) - \exp\left(\Phi^{-1}(q)\right)$ so that $\Pr\{u_{gt} \le 0\} = \Pr\{u_{ht} \le 0\} = q$ for $0 < q < 1$. The true parameters are given by $\beta_0 = (\beta_{01}, \beta_{02})' = (1,1)'$ and $\gamma_0 = (\gamma_{01}, \gamma_{02})' = (1,1)'$. Note that the hypotheses (13) - (14) correspond to the conditional moment restrictions in (1) - (2) with

[3] Pesaran and Pesarn (1993), Weeks (1996), and Santos Silva (2001) also consider non-nested tests for binary choice model. In contrast to our tests, however, their tests are based on (finite-dimensional) unconditional moment restrictions.

[4] However, in our context, the Burr model is not nested with the logit model since τ is fixed a priori.

$g(Z, \beta_0) = 1(Y \leq \beta_{01} + \beta_{02} X_1) - q$ and $h(Z, \gamma_0) = 1(Y \leq \gamma_{01} + \gamma_{02} X_2) - q$, where $Z = (Y, X_1, X_2)'$ and $X = (X_1, X_2)'$.

We estimate the true parameters by maximum likelihood estimators for Design I and by linear quantile regression estimators of Koenker and Bassett (1978) for Design II. We compare 6 different types of tests: M_{KS} and M_{CM} with the GEL implied probabilities given by empirical likelihood ($\phi = -1$), exponential tilting ($\phi = 0$) and continuous updating GMM ($\phi = 1$), i.e., M_{KS}^{el}, M_{CM}^{el}, M_{KS}^{et}, M_{CM}^{et}, M_{KS}^{cu}, M_{CM}^{cu}. In computing the supremum and integral in the test statistics, we took a maximum and sum over an equally spaced grids of size 20 on the range of empirical distributions. In computing the test statistics using bootstrap samples, we used the same grid of points as we used in the original test statistics.[5] When the observations are independent (Design I), we set the block length b for bootstrap to be unity, while when the observations are dependent (Design II), we consider several different values of b in a wide range of integers to see how sensitively the finite sample performance depends on the choice of b.

We consider two sample sizes $n \in \{100, 200\}$ and quantile probabilities $q \in \{0.50, 0.75, 0.90\}$. We fix the number of Monte Carlo repetitions to be $1,000$ and restrict the number of bootstrap repetitions to be 100 due to high computation cost.

5.1 Simulation Results

Tables 1-4 present the rejection probabilities for the tests with nominal size of 5%. The simulation standard error is approximately 0.0069.

Table 1 shows the size performance of the tests for Design I. The tests have reasonable size performance even under the small sample size ($n = 100$) when the implied probabilities are computed by empirical likelihood and exponential tilting, but tend to over-reject the null hypothesis when the implied probabilities are computed by the continuous updating GMM. However, the size distortions appear to vanish as n increases.

Table 2 gives the rejection probabilities when the alternative model of Design I is true. It is remarkable that the tests have non-trivial power against all of the alternative models we considered even with small samples. Also, as we expected, the rejection frequencies increase as we move further away from the null model, that is as we have more asymmetry in the Burr model.

Tables 3 and 4 report the simulation results for Design II. Table 3 shows that, for all quantiles q we considered, our tests have reasonable finite sample size performance when the block length b for the bootstrap procedure is in a suitable range. On the other hand, Table 4 shows that the tests have non-negligible small sample power but the latter tend to decrease as we have more extreme quantiles, i.e. as we move from $q = 0.5$ to $q = 0.90$, which is not very surprising because at extreme quantiles we have less observations to distinguish the alternative model from the null model.

[5] We experimented with a variety of such grids, but found that our simulation results are not sensitive to the choice of grids.

To sum up, the overall impression is that our tests work reasonably well in samples above 100. Among the implied probabilities, the empirical likelihood and exponential tilting work slightly better under the null hypothesis, while the continuous updating GMM implied probability works better under the alternative hypothesis. Among the Kolmogorov-Smirnov type and Cramér-von Mises type tests, the former works slightly better under the null hypothesis and the latter works better under the alternative hypothesis.

6 Conclusion

In this paper, we use the method of generalized empirical likelihood to propose tests of non-nested hypotheses of models that are specified solely in terms of conditional moment restrictions. In particular, we propose moment encompassing tests using the implied probabilities from the conditional moment restrictions that contain all information from the null model. The tests have advantages over some of the existing tests in the sense that they do not depend on smoothing parameters and allow weakly dependent data and that the criterion functions are allowed to be non-smooth. Extensions to strongly dependent or nonstationarity data, panel data, and the moment functions with infinite dimensional parameters would be interesting future topics.

A Proofs

A word on notation. As we use the mean value expansions repeatedly with respect to λ, β, or γ, we use $\vec{\lambda}$, $\vec{\beta}$, or $\vec{\gamma}$ as a generic mean value each case.

A.1 Proof of Theorem 3.1

Recall that $M_T(u) = -\sqrt{T}\sum_{t=1}^{T}\left(p_t^g(u) - \frac{1}{T}\right)\hat{h}_t(u)$ and note that a mean value expansion of $p_t^g(u)$ around $\hat{\lambda} = 0$ yields

$$\begin{aligned} & p_t^g(u) - \frac{1}{T} \\ = \; & \frac{\hat{\lambda}(u)'}{T}\left(\frac{\hat{g}_t(u)\rho_2\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)}{\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\hat{g}_s(u)\right)} - \frac{\rho_1\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)\frac{1}{T}\sum_{s=1}^{T}\hat{g}_s(u)\rho_2\left(\vec{\lambda}(u)'\hat{g}_s(u)\right)}{\left(\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\hat{g}_s(u)\right)\right)^2}\right). \end{aligned}$$

Lemma B.2 shows that $\sqrt{T}\hat{\lambda}(u) \Longrightarrow F(u)$ and thus $\sup_{u\in\mathcal{X}}\sqrt{T}\hat{\lambda}(u) = O_p(1)$. Thus, to obtain the convergence of $M_T(u)$, it remains to show that

$$\max_u\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)\rho_1\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)\frac{\frac{1}{T}\sum_{t=1}^{T}\hat{g}_t(u)\rho_2\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)}{\left(\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\hat{g}_s(u)\right)\right)^2}\right\| = o_p(1), \tag{A.1}$$

and that

$$\frac{\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)\hat{g}_t(u)'\rho_2\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)}{\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\hat{g}_s(u)\right)} \xrightarrow{p} \Omega_h(u), \tag{A.2}$$

uniformly in $u \in \mathcal{X}$. Arguments for $(A.1)$ and $(A.2)$ are similar. First, $\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\hat{g}_s(u)\right) \xrightarrow{p} -1$ uniformly in $u \in \mathcal{X}$ since $\max_{t,u\in\mathcal{X}}\left|\hat{\lambda}'\hat{g}_t(u)\right| = o_p(1)$ due to $(A.5)$. By the same reason

$$\sup_{u\in\mathcal{X}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{g}_t(u)\rho_2\left(\vec{\lambda}(u)'\hat{g}_t(u)\right)\right\| = \sup_{u\in\mathcal{X}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{g}_t(u)\right\|,$$

which is $O_p\left(T^{-1/2}\right)$ due to $(A.8)$. Similarly, $\sup_{u\in\mathcal{X}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)\right\| = O_p(1)$ by Assumption 3.4 and $\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)\hat{g}_t(u)' \xrightarrow{p} \Omega_h(u)$ by the argument similar to the proof of $(A.4)$, which involves element by element applications of a mean value expansion and a uniform law of large numbers (e.g. Lemma 2.4 of Newey and McFadden, 1994). This establishes the weak convergence of $M_T(u)$, which implies the null distributions of M_{KS} and M_{CM}. ■

A.2 Proof of Theorem 3.2

The proof is similar to that of Theorem 3.1. Lemma B.1 holds true under $\mathbf{H}_{gT}$ and Assumption 3.6. Some arguments in Lemma B.2 need to modified. In particular, $\varrho(u, \hat{\beta})$ in the equation $(A.9)$ now becomes $\varrho_T\left(u,\hat{\beta}\right)$. Then, by Assumption 3.6,

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varrho_T\left(u,\hat{\beta}\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varrho\left(u,\hat{\beta}\right)+\Pi\left(u,\hat{\beta}\right) \\ &= D\left(u,\tilde{\beta}\right)\sqrt{T}\left(\hat{\beta}-\beta_0\right)+\Pi\left(u,\hat{\beta}\right),\end{aligned}$$

and by the continuity of D and Π and by Assumption 3.6 again, we obtain the desired result following the steps of the proof of Theorem 3.1 and thus omitted. ■

A.3 Proof of Theorem 3.3

Observe that for each $u \in \mathcal{X}$,

$$\begin{aligned}\frac{M_T(u)}{\sqrt{T}} &= \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)-\sum_{t=1}^{T}p_t^g(u)\,\hat{h}_t(u) \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)-\frac{\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\hat{\lambda}(u)'\,\hat{g}_t(u)\right)\hat{h}_t(u)}{\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\hat{\lambda}(u)'\,\hat{g}_s(u)\right)}.\end{aligned}$$

From Assumption 3.7 and the continuous mapping theorem, we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(u)\xrightarrow{p}E\left[h(Z_t,\gamma_0)1(X_t\le u)\right]=0,$$

$$\frac{1}{T}\sum_{t=1}^{T}\rho_1\left(\hat{\lambda}(u)'\,\hat{g}_t(u)\right)\hat{h}_t(u)\xrightarrow{p}E\left[\rho_1(\lambda_*(u)'\,g(Z_t,\beta_*)\,h_t(Z_t,\gamma_0)\,1(X_t\le u))\right],$$

$$\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\hat{\lambda}(u)'\,\hat{g}_s(u)\right)\xrightarrow{p}E\left[\rho_1(\lambda_*(u)'\,g(Z_t,\beta_*)\,1(X_t\le u))\right],$$

uniformly in $u \in \mathcal{X}$. Therefore, the continuous mapping theorem yields the conclusion. ■

A.4 Proof of Theorem 4.1

The same reasoning as the proof of Theorem 3.1 applies once we prove Lemma B.3 and B.4. In particular, a mean value expansion of $p_t^{g^*}(u)$ around $\hat{\lambda}^*=0$ yields

$$\begin{aligned}&p_t^{g^*}(u)-\frac{1}{T} \\ &= \frac{\hat{\lambda}^*(u)'}{T}\left(\frac{\hat{g}_t^*(u)\,\rho_2\left(\vec{\lambda}(u)'\,\hat{g}_t^*(u)\right)}{\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\,\hat{g}_s^*(u)\right)}-\frac{\rho_1\left(\vec{\lambda}(u)'\,\hat{g}_t^*(u)\right)\frac{1}{T}\sum_{s=1}^{T}\hat{g}_s^*(u)\,\rho_2\left(\vec{\lambda}(u)'\,\hat{g}_s^*(u)\right)}{\left(\frac{1}{T}\sum_{s=1}^{T}\rho_1\left(\vec{\lambda}(u)'\,\hat{g}_s^*(u)\right)\right)^2}\right).\end{aligned}$$

Therefore, the centering by $M_T(u)$ allows the application of Lemma B.4 and the remaining steps are identical to the proof of Theorem 3.1 and thus omitted. ■

B Lemmas

Lemma B.1 *Suppose that Assumptions 3.1-3.5 hold. Then under* $\mathbf{H}_g$,

$$\sup_{u\in\mathcal{X}}\left\|\widehat{\lambda}(u)\right\| = O_p(T^{-1/2}).$$

Proof: By a standard argument (see Owen (1990, proof of Theorem 1)), it suffices to verify that

$$\sup_{u\in\mathcal{X}}\left\|T^{-1}\sum_{t=1}^{T}\hat{g}_t(u)\right\| = O_p(T^{-1/2}), \tag{A.3}$$

$$P\left\{\inf_{(\xi,u)\in S\times\mathcal{X}}\xi'\left(T^{-1}\sum_{t=1}^{T}\hat{g}_t(u)\hat{g}_t(u)'\right)\xi\geq d_0\right\} \rightarrow 1 \text{ for some constant } d_0>0, \tag{A.4}$$

$$\max_{1\leq t\leq T}\sup_{u\in\mathcal{X}}\|\hat{g}_t(u)\| = o_p(T^{1/2}), \tag{A.5}$$

where $S=\{\xi\in\mathbb{R}^{d_g}:\|\xi\|=1\}$.

It is easy to see that (A.3) is a direct consequence of $(A.7)$ and $(A.10)$ in Lemma B.2. The uniform law of large numbers and the continuity of $\Omega(u,\beta)$ at $\beta=\beta_0$ together with the consistency of $\hat{\beta}$ yield that

$$\sup_{u\in\mathcal{X}}\left\|T^{-1}\sum_{t=1}^{T}\hat{g}_t(u)\hat{g}_t(u)'-\Omega(u)\right\|=o_p(1).$$

This and Assumption 3.3(iv) imply (A.4) since $\inf_u|f(u)|\geq-\sup_u|f(u)-g(u)|+\inf_u|g(u)|$ for arbitrary functions f and g.

Finally, note for (A.5) that for some $\varepsilon>0$

$$\max_{1\leq t\leq T}\sup_{u\in\mathcal{X}}\|\hat{g}_t(u)\|\leq\max_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}_0}\|g(Z_t,\beta)\|+\max_{1\leq t\leq T}\sup_{u\in\mathcal{X}}\|\hat{g}_t(u)\|\,1\left(\left\|\hat{\beta}-\beta\right\|>\varepsilon\right).$$

Then, by Assumption 3.5 the last term is $o_p(1)$ and by 3.3(ii) and the Markov inequality the first term on the right hand side of the inequality is $o_p(T^{1/2})$. This completes the proof. ■

Lemma B.2 *Suppose that Assumptions 3.1-3.5 hold. Then under* $\mathbf{H}_g$,

$$\sqrt{T}\hat{\lambda}(u)\Rightarrow\Omega(u)^{-1}(G(u)-D(u)\Delta_0\Psi).$$

Proof: Consider an empirical process

$$G_T(u)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[g_t(u)-E[g_t(u)]]$$

Also define

$$\Psi_T=\frac{k_1}{T}\sum_{t=1}^{T}\psi(X_t,Z_t,\beta_0)-E\psi(X_t,Z_t,\beta_0).$$

We show the following two

$$\sup_{u\in\mathcal{X}}\left\|T^{1/2}\widehat{\lambda}(u)-\Omega(u)^{-1}\left[G_T(u)-D(u)\Delta_0T^{1/2}\Psi_T\right]\right\|=o_p(1), \tag{A.6}$$

and

$$\begin{pmatrix} G_T(\cdot) \\ T^{1/2}\Psi_T \end{pmatrix} \Rightarrow \begin{pmatrix} G(\cdot) \\ \Psi \end{pmatrix}. \tag{A.7}$$

We first show (A.6). By expanding the first order condition for (5) at $\lambda=0$, we have

$$0=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{g}_t(u)-\left[\frac{1}{T}\sum_{t=1}^{T}\rho_2\left(\vec{\lambda}'\hat{g}_t(u)\right)\hat{g}_t(u)\hat{g}_t(u)'\right]T^{1/2}\hat{\lambda}(u). \tag{A.8}$$

Recalling the notation $\varrho(u,\beta)=E\left[g\left(Z_t,\beta\right)1\left(X_t\leq u\right)\right]$, and by the fact that $\varrho(u,\beta_0)=0$ under the null hypothesis and by the mean value theorem we may write that for $\tilde{\beta}$ between β_0 and $\hat{\beta}$

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{g}_t(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varrho\left(u,\hat{\beta}\right)+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{g}_t\left(u\right)-\varrho\left(u,\hat{\beta}\right)\right) \\ &= D\left(u,\tilde{\beta}\right)\sqrt{T}\left(\hat{\beta}-\beta_0\right)+G_T\left(u\right) \\ &\quad+\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{g}_t\left(u\right)-\varrho\left(u,\hat{\beta}\right)\right)-G_T\left(u\right)\right).\end{aligned} \tag{A.9}$$

However, it follows from the conditions on g in Assumption 3.3 that $D\left(u,\tilde{\beta}\right)\to D\left(u\right)$ uniformly in $u\in\mathcal{X}$ and that

$$\sup_{u\in\mathcal{X}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{g}_t\left(u\right)-\hat{g}\left(u\right)\right)-G_T\left(u\right)\right\|=o_p\left(1\right),$$

given that $\hat{\beta}$ is consistent. Thus, by Assumption 3.5,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{g}_t(u)=G_T(u)-D(u)\Delta_0T^{1/2}\Psi_T+o_p(1)\text{ uniformly in }u\in\mathcal{X}. \tag{A.10}$$

Turning to $\frac{1}{T}\sum_{t=1}^{T}\rho_2\left(\vec{\lambda}'\hat{g}_t(u)\right)\hat{g}_t(u)\hat{g}_t(u)'$, we note that $\sup_{u\in\mathcal{X}}\left|\vec{\lambda}'\hat{g}_t\left(u\right)\right|=o_p\left(1\right)$ by $(A.5)$ and Lemma B.1. Since ρ_2 is continuous and $\rho_2\left(0\right)=-1$, the uniform law of large numbers yields that

$$\frac{1}{T}\sum_{t=1}^{T}\rho_2\left(\vec{\lambda}'\hat{g}_t(u)\right)\hat{g}_t(u)\hat{g}_t(u)'\xrightarrow{p}\Omega\left(u\right),$$

uniformly in $u\in\mathcal{X}$. This and (A.10) yield (A.6).

We now show (A.7). It is sufficient to show the weak convergence of $G_T\left(u\right)$. As finite dimensional distribution is straightforward, we establish the stochastic equicontinuity of the empirical process $G_T\left(u\right)$. We show that it satisfies the L^2-continuity condition in Andrews (1993), which in turn satisfies the entropy condition (2.16) in Doukhan et al. (1995). However, the continuity follows because for any $u \in \mathcal{X}$ and $\varepsilon > 0$, and for some $C > 0$ and $\grave{\eta}$ s.t. $1/\eta + 1/\grave{\eta} = 1$,

$$\begin{aligned} & E\left[\sup_{u_1 \in \mathcal{X}: |u_1 - u| < \varepsilon} \|g_t(u) - g_t(u_1)\|^2\right] \\ \leq & \left(E\|g(Z_t, \beta_0)\|^{2+\eta}\right)^{1/\eta} P\{|X_t - u_1| < \varepsilon\}^{1/\grave{\eta}} \\ \leq & C\varepsilon^{1/\grave{\eta}}, \end{aligned}$$

by Assumption 3.2 and 3.3. ■

Lemma B.3 *Suppose that the assumptions in Theorem 4.1 hold. Then,*

$$\sup_{u \in \mathcal{X}} \left\|\widehat{\lambda}^*(u) - \hat{\lambda}(u)\right\| = O_p^*(T^{-1/2}) \text{ in } P$$

Proof: As in Lemma B.1, it suffices to verify that

$$\sup_{u \in \mathcal{X}} \left\|T^{-1}\sum_{t=1}^{T} \hat{g}_t^*(u)\rho_1\left(\hat{\lambda}'\hat{g}_t^*(u)\right)\right\| = O_p^*(T^{-1/2}) \text{ in P}, \tag{A.11}$$

$$P^*\left\{\inf_{(\xi,u) \in S \times \mathcal{X}} \xi'\left(T^{-1}\sum_{t=1}^{T} \hat{g}_t^*(u)\hat{g}_t^*(u)'\right)\xi \geq d_0\right\} \to 1 \text{ in P}, \text{ for a constant } d_0 > 0, \tag{A.12}$$

$$\max_{1 \leq t \leq T} \sup_{u \in \mathcal{X}} \|\hat{g}_t^*(u)\| = o_p^*(T^{1/2}) \text{ in P}, \tag{A.13}$$

where $\hat{g}_t^*(u) = g\left(Z_t^*, \hat{\beta}^*\right) 1\{X_t^* \leq u\}$ and S is defined in Lemma B.1. First, we show (A.13). It holds because for some $\varepsilon > 0$

$$\begin{aligned} \max_{1 \leq t \leq T} \sup_{u \in \mathcal{X}} \|\hat{g}_t(u)\| & \leq \max_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}_0} \|g(Z_t, \beta)\| + \max_{1 \leq t \leq T} \sup_{u \in \mathcal{X}} \|\hat{g}_t(u)\| 1\left\{\left|\hat{\beta} - \beta\right| > \varepsilon\right\} \\ & = o_p\left(T^{1/2}\right) + o_p(1), \end{aligned}$$

by Assumption 3.3, the Markov inequality and the assumption that $\hat{\beta}^* - \beta_0 = o_p(1)$. To see this, note that it follows from the Markov inequality that for any $\varepsilon > 0$ and $p > 0$

$$P\left\{\max_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}_0} \|g(Z_t^*, \beta)\| > \sqrt{T}\varepsilon\right\} \leq \frac{E\left[\sum_{t=1}^{T} \sup_{\beta \in \mathcal{B}_0} \|g(Z_t^*, \beta)\|^p\right]}{T^{p/2}\varepsilon^p}.$$

Assuming $bk = T$ for simplicity, we calculate

$$\begin{aligned} E\left[\sum_{t=1}^{T} \sup_{\beta \in \mathcal{B}_0} \|g(Z_t^*, \beta)\|^p\right] & = \left[k\sum_{i=0}^{b-1} \frac{1}{T - b + 1} \sum_{t=1}^{T-b+1} E \sup_{\beta \in \mathcal{B}_0} \|g(Z_{t+b}, \beta)\|^p\right] \\ & = TE \sup_{\beta \in \mathcal{B}_0} \|g(Z_{t+b}, \beta)\|^p, \end{aligned}$$

which yields that

$$\max_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}_0} \|g(Z_t^*, \beta)\| = o_p(T^{1/2})$$

since $E\sup_{\beta\in\mathcal{B}_0}\|g(Z_{t+b},\beta)\|^p<\infty$ for some $p>2$.

Provided $(A.13)$ and the proof of Lemma B.4, the same steps showing $(A.3)$ and $(A.4)$ in the proof of Lemma B.1 yield $(A.11)$ and $(A.12)$. ■

Lemma B.4 *Suppose that the assumptions in Theorem 4.1 hold. Then,*

$$\sqrt{T}\left(\hat{\lambda}^*(u)-\hat{\lambda}(u)\right)\Rightarrow^*\Omega(u)^{-1}(G(u)-D(u)\Delta_0\Psi)\ \ in\ P.$$

Proof: We show that $(A.6)$ and $(A.7)$ in Lemma B.2 hold under both the null and the local alternative. For $(A.6)$, expand the first order condition for (5) at $\hat{\lambda}^*=\hat{\lambda}$ to obtain

$$0=A_T+B_T\sqrt{T}\left(\hat{\lambda}^*(u)-\hat{\lambda}(u)\right),$$

where

$$A_T=\frac{1}{\sqrt{T}}\sum\hat{g}_t^*(u)\rho_1\left(\hat{\lambda}(u)'\hat{g}_t^*(u)\right),$$

and

$$B_T=\frac{1}{T}\sum_{t=1}^T\rho_2\left(\vec{\lambda}^*(u)'\hat{g}_t^*(u)\right)\hat{g}_t^*(u)\hat{g}_t^*(u)'.$$

The analysis of A_T and B_T are similar as the empirical CLT implies the empirical LLN in probability and thus we focus on A_T. The mean value expansion of A_T around $\hat{\lambda}=0$ yields

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum\hat{g}_t^*(u)\rho_1\left(\hat{\lambda}'\hat{g}_t^*(u)\right)\\ =\ &-\frac{1}{\sqrt{T}}\sum\hat{g}_t^*(u)+\frac{1}{T}\sum\rho_2\left(\vec{\lambda}'\hat{g}_t^*(u)\right)\hat{g}_t^*(u)\hat{g}_t^*(u)'\sqrt{T}\hat{\lambda}\\ =\ &-\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\hat{g}_t^*(u)-\hat{g}_t(u)\right)+o_p(1),\end{aligned}$$

uniformly in $u\in\mathcal{X}$, where the last equality follows from $(A.8)$ using Theorem 1 of Radulović (1996), which together with Lemma B.3 and $(A.13)$ yields

$$\frac{1}{T}\sum\rho_2\left(\vec{\lambda}'\hat{g}_t^*(u)\right)\hat{g}_t^*(u)\hat{g}_t^*(u)'\xrightarrow{p^*}\Omega(u)\ \ in\ \mathrm{P},$$

uniformly in $u\in\mathcal{X}$.

Let

$$\nu_T^*(u,\beta)=\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(g_t^*(u,\beta)-E^*g_t^*(u,\beta)\right)-\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(g_t^*(u)-E^*g_t^*(u)\right)$$

and recall (6) for $\nu_T(u,\beta)$. Since

$$\begin{aligned}E^*T^{-1}\sum_{t=1}^T g_t^*(u)&=b^{-1}\sum_{i=0}^{b-1}(T-b-1)^{-1}\sum_{t=1}^{T-b-1}g_t(u)\\ &=T^{-1}\sum_{t=1}^{T-b-1}g_t(u)+O_p\left(\frac{b}{T-b-1}\right),\end{aligned}$$

we may write

$$
\begin{aligned}
& -\frac{1}{\sqrt{T}} \sum \left(\hat{g}_t^*(u) - \hat{g}_t(u)\right) \\
= \; & \nu_T^*\left(u, \hat{\beta}^*\right) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(g_t^*(u) - E^* g_t^*(u)\right) \\
& + \nu_T\left(u, \hat{\beta}^*\right) - \nu_T\left(u, \hat{\beta}\right) + g\left(u, \hat{\beta}^*\right) - g\left(u, \hat{\beta}\right) + o_p(1),
\end{aligned}
$$

uniformly in $u \in \mathcal{X}$. Due Assumption 3.3, both $\nu_T\left(u, \hat{\beta}^*\right)$, and $\nu_T\left(u, \hat{\beta}\right)$ are $o_p(1)$ uniformly in $u \in \mathcal{X}$ since $\hat{\beta}^*$ and $\hat{\beta}$ converge in probability to β_0. Similarly,

$$
\nu_T^*\left(u, \hat{\beta}^*\right) = o_p(1),
$$

uniformly in $u \in \mathcal{X}$ by Theorem 1 of Radulović (1996). And it follows from the mean value theorem and continuity of $D(u, \beta)$ in β that

$$
\begin{aligned}
g\left(u, \hat{\beta}^*\right) - g\left(u, \hat{\beta}\right) &= D\left(u, \tilde{\beta}^*\right)\left(\hat{\beta}^* - \beta_0\right) - D\left(\tilde{\beta}\right)\left(\hat{\beta} - \beta_0\right) \\
&= D(u)\left(\hat{\beta}^* - \hat{\beta}\right) + o_p(1),
\end{aligned}
$$

uniformly in $u \in \mathcal{X}$. Putting these together, we get

$$
\frac{1}{\sqrt{T}} \sum \hat{g}_t^*(u) \rho_1\left(\hat{\lambda}' \hat{g}_t^*(u)\right) = -\frac{1}{\sqrt{T}} \sum \left[g_t^*(u) - E^* g_t^*(u)\right] - D(u)\sqrt{T}\left(\hat{\beta}^* - \hat{\beta}\right).
$$

Finally, it follows from Theorem 1 of Radulović (1996) that

$$
\frac{1}{\sqrt{T}} \sum \left[g_t^*(u) - E^* g_t^*(u)\right] \Longrightarrow^* G(u) \;\; in \;\; P.
$$

Note that these convergences hold under both the null and local alternative. As the bootstrap statistic $\sqrt{T}\left(\hat{\beta}^* - \hat{\beta}\right)$ estimate the asymptotic distribution of $\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ consistently under the null and the local alternative, the lemma is proved. ■

C Tables

Table 1. Size performance of the tests with nominal size 0.05 (Design I)

		M^{el}_{KS}	M^{el}_{CM}	M^{et}_{KS}	M^{et}_{CM}	M^{cu}_{KS}	M^{cu}_{CM}
	Gumbel	.069	.059	.070	.059	.092	.122
	$Burr(\frac{1}{3})$	.056	.058	.058	.059	.092	.127
$n = 100$	$Burr(\frac{2}{3})$	.068	.061	.069	.060	.092	.124
	$Burr(\frac{3}{2})$	.071	.059	.070	.061	.091	.124
	$Burr(3)$	.065	.059	.065	.064	.090	.119
	Gumbel	.055	.048	.054	.047	.072	.100
	$Burr(\frac{1}{3})$	.052	.047	.053	.046	.071	.097
$n = 200$	$Burr(\frac{2}{3})$	.055	.047	.056	.049	.072	.101
	$Burr(\frac{3}{2})$	.055	.048	.055	.047	.071	.102
	$Burr(3)$	.053	.049	.051	.044	.071	.104

Table 2. Power performance of the tests with nominal size 0.05 (Design I)

		M^{el}_{KS}	M^{el}_{CM}	M^{et}_{KS}	M^{et}_{CM}	M^{cu}_{KS}	M^{cu}_{CM}
	Gumbel	.782	.776	.784	.777	.838	.872
	$Burr(\frac{1}{3})$	1.000	1.000	1.000	1.000	1.000	1.000
$n = 100$	$Burr(\frac{2}{3})$	.717	.700	.719	.704	.770	.825
	$Burr(\frac{3}{2})$	.768	.762	.769	.764	.836	.877
	$Burr(3)$	1.000	1.000	1.000	1.000	1.000	1.000
	Gumbel	.977	.978	.977	.978	.985	.988
	$Burr(\frac{1}{3})$	1.000	1.000	1.000	1.000	1.000	1.000
$n = 200$	$Burr(\frac{2}{3})$	.937	.930	.937	.930	.953	.970
	$Burr(\frac{3}{2})$	.972	.971	.973	.971	.982	.985
	$Burr(3)$	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. Size performance of the tests with nominal size 0.05 (Design II)

(A) $q = 0.50$

		M_{KS}^{el}	M_{CM}^{el}	M_{KS}^{et}	M_{CM}^{et}	M_{KS}^{cu}	M_{CM}^{cu}
	$b = 40$	.011	.007	.012	.007	.010	.008
$n = 100$	$b = 50$	.042	.023	.047	.024	.051	.030
	$b = 60$	.053	.036	.061	.039	.067	.043
	$b = 70$	.121	.094	.125	.096	.131	.107
	$b = 70$	.021	.012	.024	.011	.028	.010
$n = 200$	$b = 80$	.032	.019	.030	.016	.029	.017
	$b = 90$	.045	.032	.048	.035	.054	.037
	$b = 100$	.087	.061	.097	.067	.104	.077

(B) $q = 0.75$

		M_{KS}^{el}	M_{CM}^{el}	M_{KS}^{et}	M_{CM}^{et}	M_{KS}^{cu}	M_{CM}^{cu}
	$b = 40$	.007	.003	.010	.002	.008	.003
$n = 100$	$b = 50$	.037	.022	.040	.021	.044	.025
	$b = 60$	.051	.028	.054	.032	.060	.037
	$b = 70$	.107	.064	.119	.065	.128	.079
	$b = 80$	.019	.017	.019	.017	.019	.018
$n = 200$	$b = 90$	.035	.031	.034	.031	.040	.033
	$b = 100$	.072	.052	.080	.052	.088	.058
	$b = 110$	.084	.066	.092	.064	.096	.071

(C) $q = 0.90$

		M_{KS}^{el}	M_{CM}^{el}	M_{KS}^{et}	M_{CM}^{et}	M_{KS}^{cu}	M_{CM}^{cu}
	$b = 40$	.004	.002	.007	.002	.008	.002
$n = 100$	$b = 50$	.014	.014	.022	.014	.027	.014
	$b = 60$	.026	.015	.037	.016	.037	.021
	$b = 70$	.065	.042	.080	.043	.080	.046
	$b = 90$	.023	.009	.027	.010	.033	.014
$n = 200$	$b = 100$	.049	.029	.051	.030	.062	.039
	$b = 110$	.053	.032	.059	.032	.063	.040
	$b = 120$	.067	.038	.072	.038	.080	.039

Table 4. Power performance of the tests with nominal size 0.05 (Design II)

(A) $q = 0.50$

		M^{el}_{KS}	M^{el}_{CM}	M^{et}_{KS}	M^{et}_{CM}	M^{cu}_{KS}	M^{cu}_{CM}
	$b = 40$	.860	.912	.892	.909	.939	.916
$n = 100$	$b = 50$	.923	.951	.945	.964	.972	.965
	$b = 60$	.937	.961	.957	.968	.982	.972
	$b = 70$	.962	.980	.978	.980	.988	.986
	$b = 70$	.997	1.000	1.000	1.000	1.000	1.000
$n = 200$	$b = 80$	.995	1.000	.998	1.000	1.000	1.000
	$b = 90$	.997	1.000	.999	1.000	1.000	1.000
	$b = 100$	.996	1.000	1.000	1.000	1.000	1.000

(B) $q = 0.75$

		M^{el}_{KS}	M^{el}_{CM}	M^{et}_{KS}	M^{et}_{CM}	M^{cu}_{KS}	M^{cu}_{CM}
	$b = 40$	.513	.560	.529	.564	.638	.638
$n = 100$	$b = 50$	.710	.737	.724	.736	.784	.786
	$b = 60$	.745	.776	.761	.777	.812	.808
	$b = 70$	.819	.843	.830	.846	.878	.872
	$b = 80$	.935	.968	.936	.967	.957	.978
$n = 200$	$b = 90$	.957	.977	.963	.980	.972	.989
	$b = 100$	.981	.991	.983	.990	.985	.993
	$b = 110$	.978	.991	.983	.990	.989	.994

(C) $q = 0.90$

		M^{el}_{KS}	M^{el}_{CM}	M^{et}_{KS}	M^{et}_{CM}	M^{cu}_{KS}	M^{cu}_{CM}
	$b = 40$	.065	.027	.067	.027	.105	.052
$n = 100$	$b = 50$	.164	.101	.167	.101	.212	.137
	$b = 60$	.183	.120	.192	.122	.241	.155
	$b = 70$	.318	.226	.328	.228	.368	.256
	$b = 90$	.358	.327	.367	.333	.426	.372
$n = 200$	$b = 100$	.463	.457	.473	.460	.539	.514
	$b = 110$	.462	.456	.470	.461	.542	.509
	$b = 120$	.516	.486	.522	.490	.576	.526

References

[1] ANDREWS, D. W. K. (1987): "Consistency in nonlinear econometric models: a generic uniform law of large numbers," *Econometrica,* 55, 1465-1471.

[2] ANDREWS, D. W. K. (1993): "An introduction to econometric applications of empirical process theory for dependent random variables," *Econometric Reviews*, 12, 183-216.

[3] ANDREWS, D. W. K. (1997): "A conditional Kolmogorov test," *Econometrica,* 65, 1097-1128.

[4] BILLINGSLEY, P. (1995): "*Probability and Measure*," Wiley.

[5] BÜHLMANN, P. (1995): "The blockwise bootstrap for general empirical processes of stationary sequences," *Stochastic Processes and their Applications*, 58, 247-265.

[6] CHEN, X., H. HONG AND M. SHUM (2007): "Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models," *Journal of Econometrics*, 141, 109-140.

[7] COX, D. R. (1961): "Tests of separate families of hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, 105-123, University of California Press.

[8] COX, D. R. (1962): "Further results on tests of separate families of hypotheses," *Journal of the Royal Statistical Society,* B, 24, 406-424.

[9] CRESSIE, N. AND T. R. C. READ (1984): "Multinomial goodness-of-fit tests," *Journal of the Royal Statistical Society*, B, 46, 440-464.

[10] CSISZÁR, I. (1975): "I-divergence geometry of probability distributions and minimization problems," *Annals of Probability*, 3, 146-158.

[11] CSISZÁR, I. (1995): "Generalized projections for non-negative functions," *Acta Mathematica Hungarica*, 68, 161-185.

[12] DOMINGUEZ M. A. AND I. N. LOBATO (2004): "Consistent estimation of models defined by conditional moment restrictions," *Econometrica*, 72, 1601-1615.

[13] DONALD, S. G., G. W. IMBENS AND W. K. NEWEY (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55-93.

[14] DOUKHAN, P., P. MASSART AND E. RIO (1995): "Invariance principles for absolutely regular empirical processes," *Annals de l'institut Henri Poincaré Probabilité et Statistiques*, 31, 393-427.

[15] Ghysels, E. and A. Hall (1990): "Testing nonnested Euler conditions with quadrature-based methods of approximation," *Journal of Econometrics*, 46, 273-308.

[16] Gourieroux, C. and A. Monfort (1994): "Testing non-nested hypotheses," in: R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, vol. IV, 2583-2637, Elsevier, Amsterdam.

[17] Hansen, L. P., J. Heaton and A. Yaron (1996): "Finite-sample properties of some alternative GMM estimators," *Journal of Business and Economic Statistics,* 14, 262-280.

[18] Hong, H., B. Preston and M. Shum (2003): "Generalized empirical likelihood-based model selection criteria for moment condition models," *Econometric Theory,* 19, 923-943.

[19] Imbens, G. W., R. H. Spady and P. Johnson (1998): "Information theoretic approaches to inference in moment condition models" *Econometrica,* 66, 333-357.

[20] Kitamura, Y. and M. Stutzer (1997): "An information-theoretic alternative to generalized method of moments estimation," *Econometrica,* 65, 861-874.

[21] Kitamura, Y., G. Tripathi and H. Ahn (2004): "Empirical likelihood-based inference in conditional moment restriction models," *Econometrica,* 72, 1667-1714.

[22] Koenker, R. and G. Bassett (1978): "Regression quantiles," *Econometrica,* 46, 33-50.

[23] Künsch, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, 17, 1217-1241.

[24] MacKinnon, J. G. (1983): "Model specification tests against non-nested alternatives," *Econometric Reviews,* 2, 85-110.

[25] Newey, W. K. and D. L. McFadden (1994): "Large sample estimation and hypothesis testing," in: R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, vol. IV, 2111-2245, Elsevier, Amsterdam.

[26] Newey, W. K. and R. J. Smith (2004): "Higher order properties of GMM and generalized empirical likelihood estimators," *Econometrica*, 72, 219-255.

[27] Otsu, T. and Y. -J. Whang (2007): "Testing for Non-nested Conditional Moment Restrictions via Conditional Empirical Likelihood," forthcoming in *Econometric Theory.*

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

[29] Owen, A. (1990): "Empirical likelihood ratio confidence regions," *Annals of Statistics*, 18, 90-120.

[30] Owen, A. (2001): "*Empirical Likelihood*," CRC Press.

[31] Pesaran, M. and B. Peseran (1993): "A simulation approach to the problem of computing Cox's statistic for testing non-nested models," *Journal of Econometrics*, 57, 377-392.

[32] Pesaran, M. and M. Weeks (2001): "Non-nested hypothesis testing: an overview," in B. Baltagi, ed., *A Companion to Econometric Theory*, Ch. 13, 279-309, Blackwell Publishers, Oxford.

[33] Pötscher, B. M. and I. R. Prucha (1989): "A uniform law of large numbers for dependent and heterogeneous data processes," *Econometrica,* 57, 675-684.

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

[35] Radulović, D. (1996): "The bootstrap for empirical processes based on stationary observations," *Stochastic Processes and their Applications*, 65, 259-279.

[36] Ramalho, J. J. S. and R. J. Smith (2002): "Generalized empirical likelihood non-nested tests," *Journal of Econometrics,* 107, 99-125.

[37] Santos Silva, J. M. C. (2001): :A score test for non-nested hypotheses with applications to discrete data models," Journal of Applied Econometrics, 16, 577-597.

[38] Singleton, K. J. (1985): "Testing specifications of economic agents' intertemporal optimum problems in the presence of alternative models," *Journal of Econometrics*, 30, 391-413.

[39] Smith, R. J. (1992): "Non-nested tests for competing models estimated by generalized method of moments," *Econometrica,* 60, 973-980.

[40] Smith, R. J. (1997): "Alternative semi-parametric likelihood approaches to generalized method of moments estimation," *Economic Journal*, 107, 503-519.

[41] van der Vaart, A. W. and J. A. Wellner (1996): *Weak convergence and empirical processes*, Springer, New York.

[42] Weeks, M. (1996): "Testing the binomial and multinomial choice models using Cox's non-nested test," *Journal of the American Statistical Association*, 105, 519-530.

[43] Zhang, J. and I. Gijbels (2003): "Sieve empirical likelihood and extensions of the generalized least squares," *Scandinavian Journal of Statistics*, 30, 1-24.