TESTING FOR NON-NESTED CONDITIONAL MOMENT RESTRICTIONS VIA CONDITIONAL EMPIRICAL LIKELIHOOD

By

Taisuke Otsu and Yoon-Jae Whang

September 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1533

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 Retrictions via Conditional Empirical Likelihood*

Taisuke Otsu† Yoon-Jae Whang‡

Yale University Seoul National University

September 13, 2005

Abstract

We propose non-nested tests for competing conditional moment resctriction models using a method of empirical likelihood. Our tests are based on the method of conditional empirical likelihood developed by Kitamura, Tripathi and Ahn (2004) and Zhang and Gijbels (2003). By using the conditional implied probabilities, we develop three non-nested tests: the moment encompassing, Cox-type, and effcient score encompassing tests. Compared to the existing non-nested tests which mainly focus on testing unconditional moment restrictions, our approach directly tests conditional moment restrictions which imply the infinite number of unconditional moment restrictions. We derive the null distributions and power properties of the proposed tests. Simulation experiments show that our tests have reasonable finite sample properties.

Keywords: Empirical likelihood; Non-nested tests; Encompassing tests; Cox-type tests; Conditional moment restrictions

JEL Codes: C12, C13, C14, C22

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

†Cowles Foundation for Research in Econonmics, Yale University, New Haven CT 06520, U.S.A. E-mail address: taisuke.otsu@yale.edu.

‡School of Economics, Seoul National University, Seoul 151-742, Korea. E-mail address:whang@snu.ac.kr.

1 Introduction

Empirical econometric models are often written in the forms of conditional moment restrictions. While researchers derive and estimate their conditional moment restriction models, those models are typically non-nested and to be evaluated by some formal tests. This paper proposes non-nested tests for competing conditional moment restriction models using a method of empirical likelihood. Our tests are based on the method of conditional empirical likelihood (CEL) developed by Kitamura, Tripathi and Ahn (2004) and Zhang and Gijbels (2003).[1] By using the conditional implied probabilities from CEL, we develop three CEL-based non-nested tests: the moment encompassing, Cox-type, and efficient score encompassing tests. Compared to the existing non-nested tests which mainly focus on testing parametric models or unconditional moment restrictions, our approach directly tests conditional moment restrictions which imply the infinite number of unconditional moment restrictions.

Since Cox (1961, 1962), non-nested testing for competitive statistical models has become a standard technique to evaluate specification of a statistical model against specific alternative models.[2] Singleton (1985), Ghysels and Hall (1990), and Smith (1992) proposed non-nested testing procedures for *unconditional* moment restriction models. Recently, those procedures are extended by Ramalho and Smith (2002) to the generalized empirical likelihood (GEL) context by Smith (1997) and Newey and Smith (2004). Ramalho and Smith (2002) focused on the implied unconditional probabilities from the null unconditional moment restrictions, and derived GEL analogues of the moment encompassing, Cox-type, and parametric encompassing tests. We extend the approach of Ramalho and Smith (2002) to deal with *conditional* moment restriction models, where the infinite number of unconditional moment restrictions are implied. In particular, we employ the method of CEL to obtain the conditional implied probabilities from conditional moment restrictions and derive non-nested test statistics. Since the CEL-based conditional implied probabilities contain all information from the null conditional moment restrictions, we can directly evaluate the specification of the null model against some specific alternatives.

Since Owen (1988) and Qin and Lawless (1994), the method of empirical likelihood has be-

[1] Kitamura, Tripathi, and Ahn's (2004) "smoothed" empirical likelihood and Zhang and Gijbels' (2003) "sieve" empirical likelihood are quite similar concepts. To avoid confusion, we introduce a new terminology, conditional empirical likelihood.

[2] Examples include Davidson and MacKinnon (1981), Fisher and McAleer (1981), White (1982), Gourieroux, Monfort and Trognon (1983), Loh (1985), Mizon and Richard (1986), Wooldridge (1990), Godfrey (1998), and Chen and Kuan (2002), to mention only a few. See also Gourieroux and Monfort (1994), Pesaran and Weeks (2001) and Dhaene (1997) for a review of non-nested and encompassing tests.

come an attractive alternative against the conventional generalized method of moments (GMM) approach.[3] Kitamura (2001) and Newey and Smith (2004) showed desirable properties of empirical likelihood for testing and estimating unconditional moment restriction models, respectively. To deal with conditional moment restriction models, Kitamura, Tripathi and Ahn (2004) and Zhang and Gijbels (2003) developed the method of CEL and showed that the CEL estimator is asymptotically efficient. Tripathi and Kitamura (2003) proposed CEL-based consistent specification tests for conditional moment restrictions. This paper extends the CEL approach to non-nested testing problems. Compared to Tripathi and Kitamura's (2003) specification tests, our tests check the validity of the null model against some specific alternatives, and our test statistics converge at the parametric rate, i.e., $\sqrt{n}$-rate. Kitamura (2003) employed CEL as a model selection criterion and proposed a Vuong (1989) type discrimination test for conditional moment restriction models, which tests whether some two competing models have the same distance (in terms of the Kullback-Leibler information criterion) from the true model. Our non-nested testing approach sets one of the competing models as the null hypothesis and checks the validity of the null model.

This paper is organized as follows. Section 2 introduces our basic setup and test statistics. In Section 3, we derive the asymptotic properties of the proposed non-nested tests. Section 4 reports simulation results. Section 5 concludes.

We use the following notation. The abbreviations "a.s." and "w.p.a.1" mean "almost surely" and "with probability approaching one," respectively. $\|\cdot\|$ is the Frobenius norm. A^{-}, $\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ are a g-inverse, the minimum eigenvalue, and the maximum eigenvalue of a matrix A, respectively. $I\{A\}$ 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 Setup and Test Statistics

2.1 Non-nested Hypotheses

Suppose that we observe a random sample $\{x_i, z_i\}_{i=1}^n$, where $x \in \mathcal{X} \subset R^s$ and $z \in R^{d_z}$. Consider the two competing conditional moment restrictions:

$$\begin{aligned} \mathbf{H}_g &: E[g(z,\beta_0)|x] = 0, \quad (1) \\ \mathbf{H}_h &: E[h(z,\gamma_0)|x] = 0, \end{aligned}$$

[3] See Owen (2001) for a comprehensive review of the empirical likelihood approach.

a.s. x, where $g : R^{d_z} \times \mathcal{B} \to R^{d_g}$ and $h : R^{d_z} \times \Gamma \to R^{d_h}$ are known functions, and $\beta_0 \in \mathcal{B} \subset R^{d_\beta}$ and $\gamma_0 \in \Gamma \subset R^{d_\gamma}$ are unknown parameters. These conditional moment restrictions imply the following unconditional moment restrictions:

$$\begin{aligned} \mathbf{H}_g^U &: E\left[V_g(x)\, g(z, \beta_0)\right] = 0, \\ \mathbf{H}_h^U &: E\left[V_h(x)\, h(z, \gamma_0)\right] = 0, \end{aligned} \tag{2}$$

for *any* vector of measurable functions V_g and V_h. Several papers such as Singleton (1985), Smith (1992), and Ramalho and Smith (2002) proposed non-nested tests between the unconditional moment restrictions $\mathbf{H}_g^U$ and $\mathbf{H}_h^U$ for *some* specific choices of V_g and V_h. However, if we are interested in the validity of the original conditional moment restrictions $\mathbf{H}_g$ and $\mathbf{H}_h$, the conventional non-nested tests for $\mathbf{H}_g^U$ and $\mathbf{H}_h^U$ may not be appropriate. For example, suppose that the true joint law satisfies $E\left[V_g(x)\, g(z, \beta_0)\right] = 0$ but $E[\tilde{V}_g(x)\, g(z, \beta_0)] \neq 0$ for some function $\tilde{V}_g$. Then although $\mathbf{H}_g$ is violated, the conventional non-nested tests for $\mathbf{H}_g^U$ tend to accept the null hypothesis $\mathbf{H}_g^U$. In this paper, we proposes three CEL-based non-nested tests for the conditional moment restrictions $\mathbf{H}_g$ and $\mathbf{H}_h$.

2.2 Conditional Empirical Likelihood

This subsection introduces the CEL approach. CEL is nonparametric likelihood constructed by the conditional moment restrictions in (1). Let $p_{ji}^g = \Pr\{z = z_j | x = x_i\}$ for $i, j = 1, \ldots, n$ be multinomial conditional probabilities under the null hypothesis $\mathbf{H}_g$, and $w_{ji} = \frac{K\left(\frac{x_i - x_j}{b_n}\right)}{\sum_{j=1}^n K\left(\frac{x_i - x_j}{b_n}\right)}$ be Nadaraya-Watson kernel weights, where $K : R^s \to R$ is a kernel function and b_n is a bandwidth parameter. We consider the following likelihood maximization problem using p_{ji}^g:

$$\max_{\{p_{ji}^g\}_{i,j=1}^n} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p_{ji}^g \tag{3}$$

$$\text{s.t.} \quad p_{ji}^g \geq 0, \quad \sum_{j=1}^n p_{ji}^g = 1, \quad \sum_{j=1}^n p_{ji}^g g(z_j, \beta) = 0, \quad \text{for} \quad i, j = 1, \ldots, n.$$

The conditional moment restrictions (1) are incorporated in the constraints $\sum_{j=1}^n p_{ji}^g g(z_j, \beta) = 0$. This problem can be solved by the Lagrange multiplier method. Let $\{\mu_i^g\}_{i=1}^n$ and $\{\lambda_i^g\}_{i=1}^n$ be the Lagrange multipliers. The Lagrangian is written as:

$$\mathcal{L} = \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p_{ji}^g - \sum_{i=1}^n \mu_i^g \left(\sum_{j=1}^n p_{ji}^g - 1 \right) - \sum_{i=1}^n \lambda_i^{g\prime} \left(\sum_{j=1}^n p_{ji}^g g(z_j, \beta) \right).$$

The solution (i.e., the implied conditional probability) is:

$$\hat{p}_{ji}^{g}(\beta) = \frac{w_{ji}}{1 + \lambda_i^g(\beta)' g(z_j, \beta)}, \tag{4}$$

for $i, j = 1, \ldots, n$, where $\lambda_i^g(\beta)$ satisfies:

$$\sum_{j=1}^{n} \frac{w_{ji} g(z_j, \beta)}{1 + \lambda_i^g(\beta)' g(z_j, \beta)} = 0, \tag{5}$$

for $i = 1, \ldots, n$. If we do not impose the conditional moment restriction $\sum_{j=1}^{n} p_{ji}^{g} g(z_j, \beta) = 0$ in (3), the solution of the unrestricted likelihood maximization problem is $\hat{p}_{ji}^{N} = w_{ji}$ for $i, j = 1, \ldots, n$. Using the implied conditional probabilities $\{\hat{p}_{ji}^{g}(\beta)\}_{i,j=1}^{n}$, the profile CEL function under $\mathbf{H}_g$ is defined as:

$$\ell_g(\beta) = \sum_{i=1}^{n} I_{in} \sum_{j=1}^{n} w_{ji} \log \hat{p}_{ji}^{g}(\beta) = \sum_{i=1}^{n} I_{in} \sum_{j=1}^{n} w_{ji} \log \left(\frac{w_{ji}}{1 + \lambda_i^g(\beta)' g(z_j, \beta)} \right), \tag{6}$$

where $I_{in} = I\{x_i \in \mathcal{X}_n\}$ with $\mathcal{X}_n \subset \mathcal{X}$ is a trimming term to deal with the boundary or denominator problem in the kernel estimators (see Kitamura, Tripathi and Ahn (2004, p. 1673)).

The CEL estimator is defined as $\hat{\beta}_{CEL} = \arg\max_{\beta \in \mathcal{B}} \ell_g(\beta)$. Kitamura, Tripathi and Ahn (2004) showed that $\hat{\beta}_{CEL}$ is an asymptotically normal and efficient estimator for β_0 under $\mathbf{H}_g$. In the same manner, we can define CEL $\ell_h(\gamma)$ under $\mathbf{H}_h$ and the CEL estimator $\hat{\gamma}_{CEL}$ for γ_0. Kitamura (2003) showed that if $\mathbf{H}_g$ is misspecified, $\hat{\beta}_{CEL}$ converges to the pseudo-true value β_{CEL}^{*}, that is

$$\beta_{CEL}^{*} = \arg\min_{\beta \in \mathcal{B}} E\left[\max_{\lambda^g \in R^{d_g}} E\left[\log\left(1 + \lambda^{g\prime} g(z, \beta)\right) | x \right] \right]. \tag{7}$$

The pseudo-true value γ_{CEL}^{*} for $\hat{\gamma}_{CEL}$ is defined in the same manner.

To construct our non-nested test statistics, we employ some consistent estimators $\hat{\beta}$ and $\hat{\gamma}$ for β_0 and γ_0, respectively. $\hat{\beta}$ and $\hat{\gamma}$ may be the CEL estimators or other consistent estimators such as the GMM estimators based on the unconditional moment restrictions in (2). Let β_* and γ_* be the pseudo-true values for $\hat{\beta}$ and $\hat{\gamma}$, respectively. Given $\hat{\beta}$, the implied conditional probability under $\mathbf{H}_g$ is obtained as $\{\hat{p}_{ji}^{g}(\hat{\beta})\}_{i,j=1}^{n}$ in (4). By comparing $\{\hat{p}_{ji}^{g}(\hat{\beta})\}_{i,j=1}^{n}$ and $\{\hat{p}_{ji}^{N}\}_{i,j=1}^{n}$, we derive three non-nested tests: the moment encompassing, Cox-type, and efficient score encompassing tests.

To compute $\hat{p}_{ji}^{g}(\hat{\beta})$ in (4), we need to solve n root-finding optimizations in (5) to obtain $\lambda_i^g(\hat{\beta})$ for $i = 1, \ldots, n$. However, by using an asymptotic approximation for $\lambda_i^g(\hat{\beta})$, we can avoid the optimizations to compute $\lambda_i^g(\hat{\beta})$. Since Lemma A.4 implies that $\lambda_i^g(\hat{\beta})$ is approximated by $\tilde{\lambda}_i^g(\hat{\beta}) = \left(\sum_{j=1}^{n} w_{ji} g(z_j, \hat{\beta}) g(z_j, \hat{\beta})'\right)^{-1} \left(\sum_{j=1}^{n} w_{ji} g(z_j, \hat{\beta})\right)$, the one-step version of the implied conditional probability is obtained as[4]

$$\tilde{p}_{ji}^{g}(\hat{\beta}) = \frac{w_{ji}}{1 + \tilde{\lambda}_i^{g}(\hat{\beta})' g(z_j, \hat{\beta})}. \tag{8}$$

The non-nested test statistics based on $\hat{p}_{ji}^{g}(\hat{\beta})$ and $\tilde{p}_{ji}^{g}(\hat{\beta})$ are asymptotically equivalent.

2.3 Test Statistics

2.3.1 Moment Encompassing Test Statistic

We first define the CEL-based moment encompassing test statistic, which focuses on the multiplicative moment indicator, $\tilde{m}(x_i, z_j, \beta, \gamma) = M(x_i, \beta, \gamma)' m(z_j, \beta, \gamma)$, where $M(x_i, \beta, \gamma)$ is a $d_m \times d_M$ matrix of functions of x_i and $m(z_j, \beta, \gamma)$ is a $d_m \times 1$ vector of functions of z_j. A typical choice of $\tilde{m}(x_i, z_j, \beta, \gamma)$ is $M(x_i, \beta, \gamma) = I_{d_h}$ and $m(z_j, \beta, \gamma) = h(z_j, \gamma)$, which is based on the alternative conditional moment restrictions $\mathbf{H}_h$ in (1). We allow $M(x_i, \beta, \gamma)$ to be the form of weighted sums: $M(x_i, \beta, \gamma) = \sum_{j=1}^{n} w_{ji} M_z(x_i, z_j, \beta, \gamma)$. By using the implied conditional probability $\hat{p}_{ji}^{g}(\hat{\beta})$ and the unrestricted conditional probability $\hat{p}_{ji}^{N}$, we consider the following contrast of estimators for $E[\tilde{m}(x_i, z_i, \beta_0, \gamma_*)]$:

$$T_M = \frac{1}{n}\sum_{i=1}^{n} I_i \sum_{j=1}^{n} \hat{p}_{ji}^{g}(\hat{\beta}) \tilde{m}(x_i, z_j, \hat{\beta}, \hat{\gamma}) - \frac{1}{n}\sum_{i=1}^{n} I_i \sum_{j=1}^{n} \hat{p}_{ji}^{N} \tilde{m}(x_i, z_j, \hat{\beta}, \hat{\gamma}), \tag{9}$$

where $I_i = I\{x_i \in \mathcal{X}_*\}$ is a trimming term on a fixed subset $\mathcal{X}_* \subset \mathcal{X}$. This trimming term allows us focus to specification testing on regions in $\mathcal{X}$ which are empirically more relevant. It also let us avoid the boundary problem associated with the kernel estimators, see also Tripathi and Kitamura (2003, p.2062)[5]. If the null hypothesis $\mathbf{H}_g$ is correct, T_M converges to zero. If $\mathbf{H}_g$ is incorrect, T_M diverges in general. The moment indicator $\tilde{m}(x_i, z_j, \beta, \gamma)$ determines the direction of misspecification. Let

$$\hat{J}_i(\beta, \gamma)' = \sum_{j=1}^{n} w_{ji} m(z_j, \beta, \gamma) g(z_j, \beta)'; \; \hat{V}_i(\beta) = \sum_{j=1}^{n} w_{ji} g(z_j, \beta) g(z_j, \beta)'; \; \hat{G}_i(\beta) = \sum_{j=1}^{n} w_{ji} \partial g(z_j, \beta) / \partial \beta'.$$

The CEL-based moment encompassing test statistic for $\mathbf{H}_g$ is defined as

$$M_g = n T_M' \hat{\Phi}_M^{-} T_M, \tag{10}$$

[4] >From Lemma A.1 and Assumption 3.2 (ii), $\sum_{j=1}^{n} w_{ji} g(z_j, \hat{\beta}) g(z_j, \hat{\beta})'$ is invertible w.p.a.1.

[5] We may also allow the trimming set to be data-dependent as in Kitamura, Tripathi, and Ahn (2004) at the cost of a substantially more complicated arguments.

where

$$\hat{\Phi}_M = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i^M(\hat{\beta},\hat{\gamma})\hat{\psi}_i^M(\hat{\beta},\hat{\gamma})',$$
$$\hat{\psi}_i^M(\beta,\gamma) = -I_i M(x_i,\beta,\gamma)'\hat{J}_i(\beta,\gamma)'\hat{V}_i(\beta)^{-1}g(z_i,\beta)+\hat{H}_M(\beta,\gamma)\Delta\psi(x_i,z_i,\beta),$$
$$\hat{H}_M(\beta,\gamma) = \frac{1}{n}\sum_{i=1}^{n}I_i M(x_i,\beta,\gamma)'\hat{J}_i(\beta,\gamma)'\hat{V}_i(\beta)^{-1}\hat{G}_i(\beta).$$

Δ and $\psi(x_i,z_i,\beta)$ are defined in Assumption 3.1 (ii), which assumes the asymptotic linear form for $\hat{\beta}$:

$$n^{1/2}(\hat{\beta}-\beta_0) = -n^{-1/2}\Delta\sum_{i=1}^{n}\psi(x_i,z_i,\beta_0)+o_p(1). \tag{11}$$

The CEL-based moment encompassing test statistic for $\mathbf{H}_h$ is defined in the same manner.

2.3.2 Cox-type Test Statistic

We next define the CEL-based Cox-type test statistic, which focuses on the probability limit of the GMM-type (or Euclidean) nonparametric likelihood. Let

$$\hat{h}_i(\gamma)=\sum_{j=1}^{n}w_{ji}h(z_j,\gamma)\,;\ \hat{h}_i^g(\gamma)=\sum_{j=1}^{n}\hat{p}_{ji}^g(\hat{\beta})h(z_j,\gamma)\,;\ \hat{V}_i^h(\gamma)=\sum_{j=1}^{n}w_{ji}h(z_j,\gamma)h(z_j,\gamma)'.$$

By using $\hat{p}_{ji}^g(\hat{\beta})$ and $\hat{p}_{ji}^N=w_{ji}$, we consider the following contrast of Euclidean likelihood:[6]

$$T_C=\frac{1}{n}\sum_{i=1}^{n}I_i\hat{h}_i^g(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{h}_i^g(\hat{\gamma})-\frac{1}{n}\sum_{i=1}^{n}I_i\hat{h}_i(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{h}_i(\hat{\gamma}). \tag{12}$$

Let $\hat{J}_i^h(\beta,\gamma)'=\sum_{j=1}^{n}w_{ji}h(z_j,\gamma)g(z_j,\beta)'$. The CEL-based Cox-type test statistic for $\mathbf{H}_g$ is defined as

$$C_g=\frac{\sqrt{n}T_C}{\sqrt{\hat{\phi}_C}}, \tag{13}$$

[6]Although we may focus on the contrast of CEL for estimating γ_0:

$$\sum_{i=1}^{n}I_i\sum_{j=1}^{n}\hat{p}_{ji}^g(\hat{\beta})\log\hat{p}_{ji}^h(\hat{\gamma})-\sum_{i=1}^{n}I_i\sum_{j=1}^{n}\hat{p}_{ji}^N\log\hat{p}_{ji}^h(\hat{\gamma}),$$

the asymptotic representation of the Lagrange multiplier $\lambda_i^h(\hat{\gamma})$ in $\hat{p}_{ji}^h(\hat{\gamma})$ is less tractable under $\mathbf{H}_g$ (see Kitamura (2003)). Therefore, for its simplicity, we analyze the contrast of Euclidean likelihood.

where

$$
\begin{aligned}
\hat{\phi}_C &= \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i^C(\hat{\beta},\hat{\gamma})^2,\\
\hat{\psi}_i^C(\beta,\gamma) &= -2I_i\hat{h}_i(\gamma)'\hat{V}_i^h(\gamma)^{-1}\hat{J}_i^h(\beta,\gamma)'\hat{V}_i(\beta)^{-1}g(z_i,\beta)+\hat{H}_C(\beta,\gamma)\Delta\psi(x_i,z_i,\beta),\\
\hat{H}_C(\beta,\gamma) &= \frac{2}{n}\sum_{i=1}^{n}I_i\hat{h}_i(\gamma)'\hat{V}_i^h(\gamma)^{-1}\hat{J}_i^h(\beta,\gamma)'\hat{V}_i(\beta)^{-1}\hat{G}_i(\beta).
\end{aligned}
$$

Δ and $\psi(x_i, z_i, \beta)$ are defined in (11). The CEL-based Cox-type test statistic for $\mathbf{H}_h$ is defined in the same manner.

2.3.3 Efficient Score Encompassing Test Statistic

We finally introduce the CEL-based efficient score encompassing test statistic, which focuses on the probability limit of the asymptotic linear form of asymptotically efficient estimators for γ_0 in $\mathbf{H}_h$ (i.e., the efficient score for estimating γ_0):[7]

$$
n^{1/2}(\hat{\gamma}-\gamma_0) = -n^{-1/2}I^h(\gamma_0)^{-1}\sum_{i=1}^{n}G_i^h(\gamma_0)'V_i^h(\gamma_0)^{-1}h(z_i,\gamma_0)+o_p(1),
$$

where

$$
V_i^h(\gamma) = E\left[h(z,\gamma)h(z,\gamma)'|x_i\right];\ G_i^h(\gamma) = E\left[\partial h(z,\gamma)/\partial\gamma'|x_i\right];\ I^h(\gamma) = E\left[G_i^h(\gamma)'V_i^h(\gamma)^{-1}G_i^h(\gamma)\right].
$$

Let $\hat{G}_i^h(\gamma) = \sum_{j=1}^n w_{ji}\partial h(z_j,\gamma)/\partial\gamma'$. By using $\hat{p}_{ji}^g(\hat{\beta})$ and $\hat{p}_{ji}^N = w_{ji}$, we consider the following contrast of the efficient score:

$$
T_S = \frac{1}{n}\sum_{i=1}^{n}I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{h}_i^g(\hat{\gamma}) - \frac{1}{n}\sum_{i=1}^{n}I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{h}_i(\hat{\gamma}). \tag{14}
$$

The CEL-based efficient score encompassing test statistic is defined as

$$
S_g = nT_S'\hat{\Phi}_S^{-}T_S, \tag{15}
$$

[7] Although it requires a lengthy mathematical argument, we can consider the CEL-based parametric encompassing test statistic, which focuses on the probability limit of the CEL estimator $\hat{\gamma}_{CEL}$ for γ_0. Let

$$
\tilde{\gamma}_{CEL} = \arg\max_{\gamma\in\Gamma}\sum_{i=1}^{n}I_{in}\sum_{j=1}^{n}\hat{p}_{ji}^g(\hat{\beta}_{CEL})\log\hat{p}_{ji}^h(\gamma).
$$

Since we can expect that $\tilde{\gamma}_{CEL}$ is a consistent estimator for the pseudo-true value γ_*, the CEL-based parametric encompassing test statistic can be constructed by a quadratic form of $(\hat{\gamma}_{CEL} - \tilde{\gamma}_{CEL})$.

where

$$
\begin{aligned}
\hat{\Phi}_S &= \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_i^S(\hat{\beta},\hat{\gamma})\hat{\psi}_i^S(\hat{\beta},\hat{\gamma})', \\
\hat{\psi}_i^S(\beta,\gamma) &= -I_i\hat{G}_i^h(\gamma)'\hat{V}_i^h(\gamma)^{-1}\hat{J}_i^h(\beta,\gamma)'\hat{V}_i(\beta)^{-1}g(z_i,\beta)+\hat{H}_S(\beta,\gamma)\Delta\psi(x_i,z_i,\beta), \\
\hat{H}_S(\beta,\gamma) &= \frac{1}{n}\sum_{i=1}^{n}I_i\hat{G}_i^h(\gamma)'\hat{V}_i^h(\gamma)^{-1}\hat{J}_i^h(\beta,\gamma)'\hat{V}_i(\beta)^{-1}\hat{G}_i(\beta).
\end{aligned}
$$

The CEL-based efficient score encompassing test statistic for $\mathbf{H}_h$ is defined in the same manner.

2.3.4 Special Case: Test Statistics with the CEL Estimator

Suppose that we use the CEL estimator $\hat{\beta}_{CEL}$ for β_0. Then from Kitamura, Tripathi and Ahn (2004, p. 1690), we can show that under certain regularity conditions, the asymptotic linear form for $\hat{\beta}_{CEL}$ is written as

$$
n^{1/2}(\hat{\beta}_{CEL}-\beta_0) = -n^{-1/2}I(\beta_0)^{-1}\sum_{i=1}^{n}G_i(\beta_0)'V_i(\beta_0)^{-1}g(z_i,\beta_0)+o_p(1),
$$

where

$$
G_i(\beta)=E\left[\partial g(z,\beta)/\partial\beta'|x_i\right];\ V_i(\beta)=E\left[g(z,\beta)g(z,\beta)'|x_i\right];\ I(\beta)=E\left[G_i(\beta)'V_i(\beta)^{-1}G_i(\beta)\right].
$$

By setting $\Delta = I(\beta_0)^{-1}$ and $\psi(x_i,z_i,\beta_0) = G_i(\beta_0)'V_i(\beta_0)^{-1}g(z_i,\beta_0)$ in (10), (13), and (15), the CEL-based non-nested test statistics are defined by the following simpler forms,

(i) the moment encompassing test statistic:

$$
M_{g,CEL} = nT_M'\hat{\Phi}_{M,CEL}^{-}T_M, \tag{16}
$$

$\hat{\Phi}_{M,CEL}$ = RSS for regression of $\hat{V}_i(\hat{\beta})^{-1/2}\hat{J}_i(\hat{\beta},\hat{\gamma})M(x_i,\hat{\beta},\hat{\gamma})$ on $\hat{V}_i(\hat{\beta})^{-1/2}\hat{G}_i(\hat{\beta})$,

(ii) the Cox-type test statistic:

$$
C_{g,CEL} = \frac{\sqrt{n}T_C}{\sqrt{\hat{\phi}_{C,CEL}}}, \tag{17}
$$

$\hat{\phi}_{C,CEL}$ = RSS for regression of $2\hat{V}_i(\hat{\beta})^{-1/2}\hat{J}_i^h(\hat{\beta},\hat{\gamma})\hat{V}_i^h(\hat{\gamma})^{-1}\hat{h}_i(\hat{\gamma})$ on $\hat{V}_i(\hat{\beta})^{-1/2}\hat{G}_i(\hat{\beta})$,

(iii) the efficient score encompassing test:

$$
S_{g,CEL} = nT_S'\hat{\Phi}_{S,CEL}^{-}T_S, \tag{18}
$$

$\hat{\Phi}_{S,CEL}$ = RSS for regression of $\hat{V}_i(\hat{\beta})^{-1/2}\hat{J}_i^h(\hat{\beta},\hat{\gamma})\hat{V}_i^h(\hat{\gamma})^{-1}\hat{G}_i^h(\hat{\gamma})$ on $\hat{V}_i(\hat{\beta})^{-1/2}\hat{G}_i(\hat{\beta})$,

where RSS denotes the residual sum of squares.

The asymptotic properties obtained in the next section hold for the above test statistics as well. The above formulae are also applicable to other semiparametric efficient estimators by Newey (1990) and Donald, Imbens and Newey (2003) for example.

3 Asymptotic Properties

3.1 Null Distributions

In this subsection, we derive the asymptotic distributions of the CEL-based non-nested test statistics under the null hypothesis $\mathbf{H}_g$. We impose the following assumptions.

Assumption 3.1

(i) $\{x_i, z_i\}_{i=1}^n$ *is an i.i.d. sample on* $\mathcal{X} \times R^{d_z}$, x *is continuously distributed with density* f, $\mathcal{X}_*$ *is compact and contained in* $int(\mathcal{X})$, *and* $\inf_{x\in\mathcal{X}_*} f(x) > 0$.

(ii) $\beta_0 \in int(\mathcal{B})$, *and* $\hat{\beta}$ *satisfies* $n^{1/2}(\hat{\beta} - \beta_0) = -n^{-1/2}\Delta \sum_{i=1}^n \psi(x_i, z_i, \beta_0) + o_p(1)$, *where* Δ *is a* $d_\beta \times d_\beta$ *non-stochastic matrix,* $E[\psi(x, z, \beta_0)] = 0$, *and* $E[||\psi(x, z, \beta_0)||^\xi] < \infty$ *for some* $\xi > 2$.

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

(iv) $K(x) = \Pi_{i=1}^s \kappa(x^{(i)})$, *where* κ *is a continuously differentiable pdf with support* $[-1, 1]$, *symmetric around the origin, and* $\inf_{x\in[-\bar{k},\bar{k}]} \kappa(x) > 0$ *for some* $\bar{k} \in (0, 1)$.

(v) $b_n = n^{-\alpha}$ *for*

$$0 < \alpha < \min\left\{\frac{1}{3s}, \frac{1}{s}\left(1 - \frac{4}{\zeta}\right), \frac{1}{s}\left(1 - \frac{4}{\zeta_m}\right), \frac{1}{s}\left(1 - \frac{2}{\zeta} - \frac{2}{\eta}\right), \frac{1}{s}\left(1 - \frac{2}{\zeta_m} - \frac{2}{\eta}\right), \frac{1}{s}\left(1 - \frac{2}{\zeta} - \frac{2}{\eta_m}\right)\right\}.$$

Assumption 3.2

(i) $E[\sup_{\beta\in\mathcal{B}} \|g(z, \beta)\|^\zeta] < \infty$ *for some* $\zeta \geq 6$.

(ii) $f(x)$ *and* $E[g(z, \beta_0) g(z, \beta_0)' |x]$ *are twice continuously differentiable on* $\mathcal{X}$, $E[\partial g(z, \beta_0) / \partial\beta' |x]$ *is continuous on* $\mathcal{X}$, $f(x)$ *and* $E[\|g(z, \beta_0)\|^\zeta |x] f(x)$ *are uniformly bounded on* $\mathcal{X}$, *and* $\inf_{x\in\mathcal{X}_*} \lambda_{\min}(E[g(z, \beta_0) g(z, \beta_0)' |x]) > 0$.

(iii) $g(z, \beta)$ *is twice continuously differentiable a.s. on a neighborhood* $\mathcal{B}_0$ *around* β_0, *for* $i = 1, \ldots, d_g$ *and* $j = 1, \ldots, d_\beta$, $\sup_{\beta\in\mathcal{B}_0} |\partial g^{(i)}(z, \beta) / \partial\beta^{(j)}| \leq d_1(z)$ *holds a.s. for a real-valued function* $d_1(z)$ *with* $E[d_1(z)^\eta] < \infty$ *for some* $\eta \geq 6$, *and for* $i = 1, \ldots, d_g$ *and* $j, k = 1, \ldots, d_\beta$, $\sup_{\beta\in\mathcal{B}_0} |\partial^2 g^{(i)}(z, \beta) / \partial\beta^{(j)}\partial\beta^{(k)}| \leq d_2(z)$ *holds a.s. for a real-valued function* $d_2(z)$ *with* $E[d_2(z)^{\eta_2}] < \infty$ *for some* $\eta_2 \geq 2$.

(iv) $\sup_{x\in\mathcal{X}_*}||M(x,\hat{\beta},\hat{\gamma}) - \bar{M}(x,\beta_0,\gamma_*)|| \xrightarrow{p} 0$, *$\bar{M}(x,\beta_0,\gamma_*)$ is uniformly bounded on $\mathcal{X}_*$, $E[\sup_{\beta\in\mathcal{B},\gamma\in\Gamma}\|m(z,\beta,\gamma)\|^{\zeta_m}] < \infty$ for some $\zeta_m \geq 6$, $m(z,\beta,\gamma)$ is continuously differentiable a.s. on a neighborhood $\mathcal{B}_0 \times \Gamma_*$ around (β_0,γ_*), and for $i = 1,\ldots,d_m$ and $j = 1,\ldots,d_\beta + d_\gamma$, $\sup_{(\beta,\gamma)\in\mathcal{B}_0\times\Gamma_*}|\partial m^{(i)}(z,\beta,\gamma)/\partial(\beta',\gamma')^{(j)}| \leq d_m(z)$ holds a.s. for a real-valued function $d_m(z)$ with $E[d_m(z)^{\eta_m}] < \infty$ for some $\eta_m \geq 6$.*

In Assumption 3.1 (i), although x should be continuous, z can be continuous, discrete, or mixed. Assumption 3.1 (ii) assumes the asymptotic linear form for $\hat{\beta}$ and implies the asymptotic normality of $\hat{\beta}$. This assumptions holds for a number of parametric and semiparametric estimators. Assumption 3.1 (iii) imposes the $\sqrt{n}$-consistency of $\hat{\gamma}$ to the pseudo-true value γ_*. Depending on the estimation method, γ_* may be different. Assumption 3.1 (iv) and (v) are conditions for the kernel function K and the bandwidth b_n. Assumption 3.1 (iv) rules out kernel functions whose orders are higher than two. Assumption 3.2 (i)-(iii) are conditions for the moment function $g(z,\beta)$, which are mainly used to derive the convergence of nonparametric components such as $\hat{V}_i(\hat{\beta})$ and $\hat{G}_i(\hat{\beta})$. Assumption 3.2 (iv) is a set of conditions for the moment indicator $\tilde{m}(x,z,\beta,\gamma)$. For the Cox-type and efficient score encompassing test statistics, we take $m(z_i,\beta,\gamma) = h(z,\gamma)$.

Let $J_i^h(\beta,\gamma)'{=}E\left[h(z,\gamma)\,g(z,\beta)'\,|x_i\right]$. The null distributions of the CEL-based non-nested test statistics are obtained as follows.

Theorem 3.1 (Null Distributions)

(i) *Suppose that Assumptions 3.1 and 3.2 hold. Then under the null hypothesis $\mathbf{H}_g$,*

$$M_g \xrightarrow{d} \chi^2_{\mathrm{rank}(\Phi_M)},$$

where Φ_M (defined below (41)) is the probability limit of $\hat{\Phi}_M$.

(ii) *Suppose that Assumptions 3.1 and 3.2 (i)-(iii) hold, and Assumption 3.2 (iv) holds for $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $M(x_i,\beta,\gamma)' = \{2\hat{h}_i(\gamma) - J_i^h(\beta,\gamma)\hat{V}_i(\beta)^{-1}\hat{g}_i(\beta)\}'\hat{V}_i^h(\gamma)^{-1}$, and $\bar{M}_i(x_i,\beta,\gamma)' = 2E[h(z,\gamma)\,|x_i]'\,V_i^h(\gamma)^{-1}$. Then under the null hypothesis $\mathbf{H}_g$,*

$$C_g \xrightarrow{d} N(0,1).$$

(iii) *Suppose that Assumptions 3.1 and 3.2 (i)-(iii) hold, and Assumption 3.2 (iv) holds for $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $M_i(x_i,\beta,\gamma)' = \hat{G}_i^h(\gamma)'\,\hat{V}_i^h(\gamma)^{-1}$, and $\bar{M}_i(x_i,\beta,\gamma)' = G_i^h(\gamma)'V_i^h(\gamma)^{-1}$. Then under the null hypothesis $\mathbf{H}_g$,*

$$S_g \xrightarrow{d} \chi^2_{\mathrm{rank}(\Phi_S)},$$

where Φ_S (defined below (43)) is the probability limit of $\hat{\Phi}_S$.

Therefore, all the non-nested test statistics follow the standard limiting distributions. Compared to the CEL-based specification test statistics by Tripathi and Kitamura (2003), our non-nested test statistics show the parametric convergence rate. For (ii) and (iii) of this theorem, the assumptions on $m(z_i, \beta, \gamma)$ and $M(x_i, \beta, \gamma)$ can be replaced with more primitive conditions, such as the conditions obtained by replacing $g(z, \beta)$, β_0, $\mathcal{B}$, and $\mathcal{B}_0$ in Assumption 3.2 (i)-(iii) with $h(z, \gamma)$, γ_*, Γ, and Γ_*, respectively.

3.2 Power Properties

This subsection studies the power properties of the CEL-based non-nested test statistics under some local alternative hypothesis. We assume that the joint distribution of (x, z) is fixed, and that there exists a nonstochastic sequence $\beta_{0n} \in \mathcal{B}$ such that

$$\mathbf{H}_{gn} : E\left[g\left(z, \beta_{0n}\right)|x\right] = n^{-1/2}\delta\left(x\right) \tag{19}$$

holds a.s. for some $\delta : \mathcal{X} \to R^{d_g}$. The null hypothesis $\mathbf{H}_g$ is satisfied if $\delta(x) = 0$[8]. We impose the following assumptions.

Assumption 3.3

(i) *$\delta(x)$ is continuous on $\mathcal{X}$, $E[\|\delta(x)\|^{\zeta}] < \infty$, $\|\beta_{0n} - \beta_0\| \to 0$ as $n \to \infty$, $\beta_0 \in int(\mathcal{B})$, and $n^{1/2}(\hat{\beta} - \beta_{0n}) = -n^{-1/2}\Delta \sum_{i=1}^{n} \psi(x_i, z_i, \beta_{0n}) + o_p(1)$, where Δ is a $d_\beta \times d_\beta$ non-stochastic matrix, $E[\psi(x, z, \beta_{0n})|x] = n^{-1/2}\delta_\psi(x)$, $\delta_\psi(x)$ is continuous on $\mathcal{X}$, and $E[\|\delta_\psi(x)\|^{\zeta}] < \infty$.*

(ii) *$f(x)$ and $E[g(z, \beta) g(z, \beta)'|x]$ are twice continuously differentiable on $\mathcal{X}$ for each $\beta \in \mathcal{B}_0$, $E[g(z, \beta) g(z, \beta)'|x]$ and $E[\partial g(z, \beta)/\partial \beta'|x]$ are continuous on $\mathcal{X} \times \mathcal{B}_0$, $f(x)$ and $\sup_{\beta \in \mathcal{B}_0} E[\|g(z, \beta)\|^{\zeta}|x] f(x)$ are uniformly bounded on $\mathcal{X}$, $\inf_{(x,\beta) \in \mathcal{X}_* \times \mathcal{B}_0} \lambda_{\min}(E[g(z, \beta) g(z, \beta)'|x]) > 0$, and $\inf_{(x,\beta) \in \mathcal{X}_* \times \mathcal{B}_0} \lambda_{\max}(E[g(z, \beta) g(z, \beta)'|x]) < \infty$.*

(iii) *$\sup_{x \in \mathcal{X}_*} ||M(x, \hat{\beta}, \hat{\gamma}) - \bar{M}(x, \beta_{0n}, \gamma_*)|| \xrightarrow{p} 0$, $\sup_{\beta \in \mathcal{B}_0} \bar{M}(x, \beta, \gamma_*)$ is uniformly bounded on $\mathcal{X}_*$, $E[\sup_{\beta \in \mathcal{B}, \gamma \in \Gamma} \|m(z, \beta, \gamma)\|^{\zeta_m}] < \infty$ for some $\zeta_m \geq 6$, $m(z, \beta, \gamma)$ is continuously differentiable a.s. on a neighborhood $\mathcal{B}_0 \times \Gamma_*$ around (β_0, γ_*), and for $i = 1, \ldots, d_m$ and*

[8] Another way to formulate the local alternatives in the spirit of Singleton (1985, p.402) would be

$$\mathbf{H}^*_{gn} : \left(1 - \frac{\eta}{\sqrt{n}}\right) E\left[g(z, \beta_0)|x\right] + \frac{\eta}{\sqrt{n}} E\left[h(z, \gamma)|x\right] = 0,$$

where $\eta \in R$ is a constant. This case can be treated similarly because $\mathbf{H}^*_{gn}$ now corresponds to $\mathbf{H}_{gn}$ with $\delta(x) = \eta\left\{E\left[g(z, \beta_0)|x\right] - E\left[h(z, \gamma)|x\right]\right\}$ and $\beta_{0n} = \beta_0$.

$j = 1, \dots, d_\beta + d_\gamma$, $\sup_{(\beta,\gamma)\in\mathcal{B}_0\times\Gamma_*} |\partial m^{(i)}(z,\beta,\gamma)/\partial(\beta',\gamma')^{(j)}| \leq d_m(z)$ *holds a.s. for a real-valued function* $d_m(z)$ *with* $E[d_m(z)^{\eta_m}] < \infty$ *for some* $\eta_m \geq 6$.

Assumption 3.3 (i), (ii), and (iii) are extensions of Assumptions 3.1 (ii) and 3.2 (ii) and (iv), respectively. Let $J_i(\beta,\gamma)' = E[m(z,\beta,\gamma)\, g(z,\beta)' |x_i]$, and $\chi^2_d(v)$ be the noncentral chi-squared distribution with the degree of freedom d and the noncentrality parameter v. The local power properties of the CEL-based non-nested test statistics are obtained as follows.

Theorem 3.2 (Local Power)

(i) *Suppose that Assumptions 3.1 (i) and (iii)-(v), 3.2 (i) and (iii), and 3.3 hold. Then under the local alternative hypothesis* $\mathbf{H}_{gn}$,

$$M_g \xrightarrow{d} \chi^2_{\text{rank}(\Phi_M)}\left(\mu_M' \Phi_M^- \mu_M\right),$$

where

$$\mu_M = -E\left[I_i M(x_i,\beta_0,\gamma_*)' J_i(\beta_0,\gamma_*)' V_i(\beta_0)^{-1}\delta(x_i)\right] + H_M(\beta_0,\gamma_*)\Delta E[\delta_\psi(x_i)],$$

$$H_M(\beta,\gamma) = E\left[I_i M(x_i,\beta,\gamma)' J_i(\beta,\gamma)' V_i(\beta)^{-1} G_i(\beta)\right].$$

(ii) *Suppose that Assumptions 3.1 (i) and (iii)-(v), 3.2 (i) and (iii), and 3.3 (i)-(ii) hold, and Assumption 3.3 (iii) holds for* $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $M(x_i,\beta,\gamma)' = \{2\hat{h}_i(\gamma) - J_i^h(\beta,\gamma)\hat{V}_i(\beta)^{-1}\hat{g}_i(\beta)\}'\hat{V}_i^h(\gamma)^{-1}$, *and* $\bar{M}_i(x_i,\beta,\gamma)' = 2E[h(z,\gamma)|x_i]' V_i^h(\gamma)^{-1}$. *Then under the local alternative hypothesis* $\mathbf{H}_{gn}$,

$$C_g \xrightarrow{d} N\left(\phi_C^{-1/2}\mu_C, 1\right),$$

where

$$\mu_C = -2E\left[I_i E[h(z,\gamma_*)|x_i]' V_i^h(\gamma_*)^{-1} J_i^h(\beta_0,\gamma_*)' V_i(\beta_0)^{-1}\delta(x_i)\right] + H_C(\beta_0,\gamma_*)\Delta E[\delta_\psi(x_i)],$$

$$H_C(\beta,\gamma) = 2E\left[I_i E[h(z,\gamma)|x_i]' V_i^h(\gamma)^{-1} J_i^h(\beta,\gamma)' V_i(\beta)^{-1} G_i(\beta)\right].$$

(iii) *Suppose that Assumptions 3.1 (i) and (iii)-(v), 3.2 (i) and (iii), and 3.3 (i)-(ii) hold, and Assumption 3.3 (iii) holds for* $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $M_i(x_i,\beta,\gamma)' = \hat{G}_i^h(\gamma)' \hat{V}_i^h(\gamma)^{-1}$, *and* $\bar{M}_i(x_i,\beta,\gamma)' = G_i^h(\gamma)' V_i^h(\gamma)^{-1}$. *Then under the local alternative hypothesis* $\mathbf{H}_{gn}$,

$$S_g \xrightarrow{d} \chi^2_{\text{rank}(\Phi_S)}\left(\mu_S' \Phi_S^- \mu_S\right),$$

where

$$\mu_S = -E\left[I_i G_i^h(\gamma_*)' V_i^h(\gamma_*)^{-1} J_i^h(\beta_0,\gamma_*)' V_i(\beta_0)^{-1}\delta(x_i)\right] + H_S(\beta_0,\gamma_*)\Delta E[\delta_\psi(x_i)],$$

$$H_S(\beta,\gamma) = E\left[I_i G_i^h(\gamma)' V_i^h(\gamma)^{-1} J_i^h(\beta,\gamma)' V_i(\beta)^{-1} G_i(\beta)\right].$$

Therefore, similar to the conventional non-nested tests, the local power functions are obtained from the standard noncentral distributions. While the CEL-based specification test by Tripathi and Kitamura (2003) has non-trivial power against the local alternatives with a nonparametric rate (i.e., $n^{-1/2}b_n^{-s/4}\delta(x)$), our CEL-based non-nested tests have non-trivial power against the local alternatives with the parametric rate (i.e., $n^{-1/2}\delta(x)$). For (ii) and (iii) of this theorem, we can also replace the assumptions on $m(z_i,\beta,\gamma)$ and $M(x_i,\beta,\gamma)$ with more primitive conditions, such as the conditions obtained by replacing $g(z,\beta)$, β_0, $\mathcal{B}$, and $\mathcal{B}_0$ in Assumptions 3.2 (i) and (iii) and 3.3 (ii) with $h(z,\gamma)$, γ_*, Γ, and Γ_*, respectively.

We finally derive the consistency of the CEL-based non-nested tests under the alternative hypothesis $\mathbf{H}_h$. We assume that under $\mathbf{H}_h$ the estimators $\hat{\beta}$ and $\hat{\gamma}$ converge to the pseudo-true values β_* and γ_0, respectively. Let $\mathcal{B}_*$ and Γ_0 be neighborhoods around β_* and γ_0, respectively, and

$$\lambda_*^g(x,\beta) = \arg\max_{\lambda\in R^{d_g}} E\left[\log\left(1+\lambda' g(z,\beta)\right)|x\right].$$

>From Kitamura (2003), we have $\max_{i\in I_*}||\lambda_i^g(\hat{\beta})-\lambda_*^g(x_i,\beta_*)||\xrightarrow{p}0$ under $\mathbf{H}_h$. Let

$$\begin{aligned}
J_{i*}(\beta,\gamma)' &= E\left[\left.\frac{m(z,\beta,\gamma)\,g(z,\beta)'}{1+\lambda_*^g(x_i,\beta)'\,g(z,\beta)}\right| x_i\right],\quad J_{i*}^h(\beta,\gamma)' = E\left[\left.\frac{h(z,\gamma)\,g(z,\beta)'}{1+\lambda_*^g(x_i,\beta)'\,g(z,\beta)}\right| x_i\right],\\
\hat{J}_{i*}^h(\beta,\gamma)' &= \sum_{j=1}^n w_{ji}\frac{h(z_j,\gamma)g_j(\beta)'}{1+\lambda_i^g(\beta)'g_j(\beta)}.
\end{aligned}$$

The consistency results are obtained as follows.

Theorem 3.3 (Consistency)

(i) *Suppose that for β_*, γ_0, $\mathcal{B}_*$, and Γ_0 instead of β_0, γ_*, $\mathcal{B}_0$, and Γ_*, respectively, Assumptions 3.1 and 3.2 hold. Then under the alternative hypothesis $\mathbf{H}_h$, the CEL-based moment encompassing test by M_g is consistent if $\mu_{hM}'\Phi_{hM}^-\mu_{hM}>0$, where*

$$\mu_{hM} = -E\left[I_i\bar{M}_i(x_i,\beta_*,\gamma_0)'J_{i*}(\beta_*,\gamma_0)'\,\lambda_*^g(x_i,\beta_*)\right],$$

and Φ_{hM} is the probability limit of $\hat{\Phi}_M$ under $\mathbf{H}_h$.

(ii) *Suppose that for β_*, γ_0, $\mathcal{B}_*$, and Γ_0 instead of β_0, γ_*, $\mathcal{B}_0$, and Γ_*, respectively, Assumptions 3.1 and 3.2 (i)-(iii) hold, and Assumption 3.2 (iv) holds for $m(z_i,\beta,\gamma)=h(z_i,\gamma)$, $M(x_i,\beta,\gamma)' = \left\{\sum_{j=1}^n w_{ji}\frac{2h(z_j,\gamma)}{1+\lambda_i^g(\beta)'g_j(\beta)} + \hat{J}_{i*}^h(\beta,\gamma)'\lambda_i^g(\beta)\right\}'\hat{V}_i^h(\gamma)^{-1}$, and $\bar{M}_i(x_i,\beta,\gamma)' = \left\{E\left[\left.\frac{2h(z,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)'g(z,\beta_*)}\right|x_i\right] + J_{i*}^h(\beta_*,\gamma_0)'\,\lambda_*^g(x_i,\beta_*)\right\}'V_i^h(\gamma_0)^{-1}$. Then under the alternative hypothesis $\mathbf{H}_h$, the CEL-based Cox-type test by C_g is consistent if $\mu_{hC}^2/\phi_{hC}>0$, where*

$$\begin{aligned}
\mu_{hC} &= -E[I_i\left\{E\left[\left.\frac{2h(z,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)'\,g(z,\beta_*)}\right|x_i\right] + J_{i*}^h(\beta_*,\gamma_0)'\,\lambda_*^g(x_i,\beta_*)\right\}'\\
&\quad\times V_i^h(\gamma_0)^{-1}\,J_{i*}^h(\beta_*,\gamma_0)'\,\lambda_*^g(x_i,\beta_*)],
\end{aligned}$$

and ϕ_{hC} is the probability limit of $\hat{\phi}_{hC}$ under $\mathbf{H}_h$.

(iii) *Suppose that for β_*, γ_0, $\mathcal{B}_*$, and Γ_0, instead of β_0, γ_*, $\mathcal{B}_0$, and Γ_*, respectively, Assumptions 3.1 and 3.2 (i)-(iii) hold, and Assumption 3.2 (iv) holds for $m(z_i, \beta, \gamma) = h(z_i, \gamma)$, $M_i(x_i, \beta, \gamma)' = \hat{G}_i^h(\gamma)' \hat{V}_i^h(\gamma)^{-1}$, and $\bar{M}_i(x_i, \beta, \gamma)' = G_i^h(\gamma)' V_i^h(\gamma)^{-1}$. Then under the alternative hypothesis $\mathbf{H}_h$, the CEL-based efficient score test by S_g is consistent if $\mu_{hS}' \Phi_{hS}^{-} \mu_{hS} > 0$, where*

$$\mu_{hS} = -E\left[I_i G_i^h(\gamma_0)' V_i^h(\gamma_0)^{-1} J_{i*}^h(\beta_*, \gamma_0)' \lambda_*^g(x_i, \beta_*)\right],$$

and Φ_{hS} is the probability limit of $\hat{\Phi}_S$ under $\mathbf{H}_h$.

4 Simulations

This section examines the finite sample properties of our tests against some of the existing non-nested tests using Monte-Carlo methods.

4.1 Experimental Design

We consider two simulation designs. In Design I, we consider two competing linear regression models: for $i = 1, ..., n$,

$$\begin{aligned} \mathbf{H}_g &: y_i = \beta_{01} + \beta_{02} x_{1i} + u_{gi} \\ \mathbf{H}_h &: y_i = \gamma_{01} + \gamma_{02} x_{2i} + u_{hi}, \end{aligned} \tag{20}$$

where $x_{1i} = c_0 x_{2i} + e_i$ for $c_0 \in \{1, 2\}$, $\{x_{2i}\}$ and $\{e_i\}$ are i.i.d. $N(0, 1)$, $\{u_{gi}\}$ and $\{u_{hi}\}$ are i.i.d. $N(0, 4)$, and 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 (20) correspond to the conditional moment restrictions in (1) with $g(z, \beta_0) = y - \beta_{01} - \beta_{02} x_1$ and $h(z, \gamma_0) = y - \gamma_{01} - \gamma_{02} x_2$, where $z = (y, x_1, x_2)'$ and $x = (x_1, x_2)'$.

On the other hand, in Design II, we consider the following regression models: for $i = 1, ..., n$,

$$\begin{aligned} \mathbf{H}_g &: y_i = \beta_0 x_i + u_{gi} \\ \mathbf{H}_h &: y_i = \gamma_0 x_i^3 + u_{hi}, \end{aligned} \tag{21}$$

where $\{x_i\}$, $\{u_{gi}\}$ and $\{u_{hi}\}$ are i.i.d. $N(0, 1)$ and $\beta_0 = \gamma_0 = 1$. The hypotheses (21) correspond to (1) with $g(z, \beta_0) = y - \beta_0 x$ and $h(z, \gamma_0) = y - \gamma_0 x^3$, where $z = (y, x)'$.

As benchmarks for our simulation experiments, we consider the non-nested tests of Singleton (1985, eqn. (33), p.404), labelled S, and Ramalho and Smith (2002, Simplified Cox test in Eqn.

(4.4), p.108), labelled SC, respectively. We compute S and SC from the following unconditional moment restrictions implied by (20) and (21): for Design I,

$$\begin{aligned}\mathbf{H}_g^U &: E\left[(1,x_{1i},x_{2i})'(y_i-\beta_{01}-\beta_{02}x_{1i})\right]=0 \\ \mathbf{H}_h^U &: E\left[(1,x_{1i},x_{2i})'(y_i-\gamma_{01}-\gamma_{02}x_{2i})\right]=0\end{aligned} \tag{22}$$

and, for Design II,

$$\begin{aligned}\mathbf{H}_g^U &: E\left[(1,x_i)'(y_i-\beta_0x_i)\right]=0 \\ \mathbf{H}_h^U &: E\left[(1,x_i^3)'\left(y_i-\gamma_0x_i^3\right)\right]=0.\end{aligned} \tag{23}$$

As another benchmark, we also consider the over-identifying test of Hansen (1982), labelled J, that tests the validity of $\mathbf{H}_g^U$ in (22) and (23) against general alternatives.

We consider two sample sizes $n \in \{100, 200\}$ and fix the number R of Monte Carlo repetitions to be 1000. Because of very long computing time required for nonlinear optimizations, we do not consider larger n and R. We use the Gaussian kernel for our CEL-based tests M_g, C_g, and S_g. For the bandwidth b_n, we consider $b_n \in [0.1, 0.2, ..., 1.0]$ in our simulations.

4.2 Simulation Results

Tables 1-3 present the rejection probabilities for the tests with nominal size of 5%. The simulation standard errors are approximately 0.007. Tables 1 and 2 give the results for Design I with $c_0 = 1$ and $c_0 = 2$, respectively. In both cases, our tests have reasonable size performance if the bandwidth is in a suitable range. The performance improves generally as n increases. The competitors J and SC also have little size distortions, though the Singleton's test S under-rejects in many cases we consider. In terms of size-corrected powers, the efficient score encompassing test S_g dominates M_g and C_g in Design I. When $c_0 = 1$, the test S which is known to have an optimality property against some local alternatives, has relatively very good (size-corrected) power performance. However, when $c_0 = 2$, the power performance of S deteriorates and is significantly dominated by that of S_g. To explain the latter phenomenon, notice that if the alternative hypothesis $\mathbf{H}_h$ in (20) is true, then the GMM estimator $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)'$ converges (in probability) to the pseudo-true value $\beta_* = (1, c_0/(1+c_0^2))'$. This implies that the sample analogue of the unconditional expectation in (22) converges

$$\frac{1}{n}\sum_{i=}^{n}\left[(1,x_{1i},x_{2i})'\left(y_i-\widehat{\beta}_1-\widehat{\beta}_2x_{1i}\right)\right] \xrightarrow{p} \left(0,0,\frac{1}{1+c_0^2}\right)'. \tag{24}$$

Therefore, since the limit in (24) degenerates to zero as c_0 increases, we can see that a test based on the sample average in (24) will have low power if c_0 is large.

Table 3 reports the simulation results for Design II. In this design, we expect that the tests based on the unconditional moments in (23) will be inconsistent. It is because, under $\mathbf{H}_h$, the estimator $\widehat{\beta}$ converges in probability to the pseudo-true value $\beta_* = 3$ and hence the sample average converges to

$$\frac{1}{n}\sum_{i=}^{n}\left[(1,x_i)'\left(y_i-\widehat{\beta}x_i\right)\right] \xrightarrow{p} E_H\left[(1,x_i)'\left(y_i-\beta_* x_i\right)\right]=(0,0)', \tag{25}$$

where E_H is the expectation taken under $\mathbf{H}_h$. This is precisely what happens to the powers of the tests J, S, and SC in Design II. On the other hand, our tests have non-trivial powers even in this case. Among the latter tests, M_g and C_g appear to have better (size-corrected) power performance than S_g in this design.

5 Conclusion

We propose three non-nested tests for competing conditional moment restriction models. Our test statistics are based on the implied conditional probabilities by conditional empirical likelihood. The proposed tests (the moment encompassing, Cox-type, and efficient score encompassing tests) follow the standard limiting distributions. Simulation results illustrate that our non-nested tests have reasonable finite sample properties and, in some cases, dominate some of the existing tests based on unconditional moment restrictions.

A Mathematical Appendix

Notation. Denote

$$\begin{aligned}
I_* &= \{i : x_i \in \mathcal{X}_*, 1 \le i \le n\}, \ c_n = \sqrt{\frac{\log n}{nb_n^s}}, \\
g_j(\beta) &= g(z_j,\beta), \ h_j(\gamma)=h(z_j,\gamma), \ m_j(\beta,\gamma)=m(z_j,\beta,\gamma), \\
M_i(\beta,\gamma) &= M(x_i,\beta,\gamma), \ K_{ji}=K\left(\frac{x_i-x_j}{b_n}\right), \ \hat{f}_i=\frac{1}{nb_n^s}\sum_{j=1}^{n}K_{ji}, \ \hat{g}_i(\beta)=\sum_{j=1}^{n}w_{ji}g_j(\beta), \\
V_i(\beta) &= E\left[g_j(\beta)g_j(\beta)'|x_i\right], \ \bar{V}_i(\beta)=E\left[\frac{1}{nb_n^s}\sum_{j=1}^{n}K_{ji}g_j(\beta)g_j(\beta)'|x_i\right], \\
J_i(\beta)' &= E[m_j(\beta,\gamma)g_j(\beta)'|x_i], \ \bar{J}_i(\beta)'=E\left[\frac{1}{nb_n^s}\sum_{j=1}^{n}K_{ji}m_j(\beta,\gamma)g_j(\beta)'|x_i\right], \\
G_i(\beta) &= E[\partial g_j(\beta)/\partial\beta'|x_i], \ \bar{G}_i(\beta)=E\left[\frac{1}{nb_n^s}\sum_{j=1}^{n}K_{ji}\partial g_j(\beta)/\partial\beta'|x_i\right].
\end{aligned}$$

A.1 Proof of Theorem 3.1

Proof of (i)

An expansion of $\hat{p}^g_{ji}(\hat{\beta})$ around $\lambda^g_i(\hat{\beta}) = 0$ yields

$$\hat{p}^g_{ji}(\hat{\beta}) = \frac{w_{ji}}{1 + \lambda^g_i(\hat{\beta})' g_j(\hat{\beta})} = w_{ji}\left(1 - \lambda^g_i(\hat{\beta})' g_j(\hat{\beta}) + r_{ji}\right), \tag{26}$$

where $r_{ji} = \frac{\lambda^g_i(\hat{\beta})' g_j(\hat{\beta}) g_j(\hat{\beta})' \lambda^g_i(\hat{\beta})}{(1+\tilde{\lambda}^{g\prime}_i g_j(\hat{\beta}))^3}$, and $\tilde{\lambda}^g_i$ is a point on the line joining $\lambda^g_i(\hat{\beta})$ and 0. Since $\hat{p}^g_{ji}(\hat{\beta}) - \hat{p}^N_{ji} = w_{ji}\left(-\lambda^g_i(\hat{\beta})' g_j(\hat{\beta}) + r_{ji}\right)$, the definition of T_M in (9) implies

$$\begin{aligned} T_M &= -\frac{1}{n}\sum_{i=1}^n I_i M_i(\hat{\beta},\hat{\gamma})' \hat{J}_i(\hat{\beta},\hat{\gamma})' \lambda^g_i(\hat{\beta}) + \frac{1}{n}\sum_{i=1}^n I_i M_i(\hat{\beta},\hat{\gamma})' \left(\sum_{j=1}^n w_{ji} r_{ji} m_j(\hat{\beta},\hat{\gamma})\right) \\ &= T^{(1)} + R^{(1)}. \end{aligned} \tag{27}$$

$R^{(1)}$ satisfies

$$\left\|R^{(1)}\right\| \le \max_{i\in I_*}\left\|M_i(\hat{\beta},\hat{\gamma})\right\| \max_{1\le j\le n}\left\|m_j(\hat{\beta},\hat{\gamma})\right\| \left(\max_{i\in I_*}\left\|\lambda^g_i(\hat{\beta})\right\|\right)^2 \left\|\frac{1}{n}\sum_{i=1}^n I_i \sum_{j=1}^n w_{ji} \frac{g_j(\hat{\beta}) g_j(\hat{\beta})'}{(1+\tilde{\lambda}^{g\prime}_i g_j(\hat{\beta}))^3}\right\|. \tag{28}$$

Assumption 3.2 (iv) implies

$$\max_{i\in I_*}\left\|M_i(\hat{\beta},\hat{\gamma})\right\| = O_p(1). \tag{29}$$

>From Assumption 3.2 (i) and (iv) and Tripathi and Kitamura (2004, Lemma C.4),

$$\max_{1\le j\le n}\left\|g_j(\hat{\beta})\right\| = o\left(n^{1/\zeta}\right), \quad \max_{1\le j\le n}\left\|m_j(\hat{\beta},\hat{\gamma})\right\| = o\left(n^{1/\zeta_m}\right). \tag{30}$$

>From Lemmas A.1 and A.4,

$$\max_{i\in I_*}\left\|\lambda^g_i(\hat{\beta})\right\| = O_p(c_n) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right). \tag{31}$$

Since (30) and (31) imply that $\max_{i\in I_*, 1\le j\le n}|\tilde{\lambda}^{g\prime}_i g_j(\hat{\beta})| = o_p(1)$, we have $\left\|\frac{1}{n}\sum_{i=1}^n I_i \sum_{j=1}^n w_{ji}\frac{g_j(\hat{\beta})g_j(\hat{\beta})'}{\left(1+\tilde{\lambda}^{g\prime}_i g_j(\hat{\beta})\right)^3}\right\| \le O_p(1)$ by Lemma A.1. Thus, from (28)-(31),

$$\left\|R^{(1)}\right\| \le O_p(1)\, o\left(n^{1/\zeta_m}\right)\left\{O_p(c_n) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right)\right\}^2 O_p(1) = o_p\left(n^{-1/2}\right), \tag{32}$$

where the equality follows from $\alpha < \frac{1}{s}\left(1 - \frac{4}{\zeta_m}\right)$ and $\frac{1}{\zeta_m} + \frac{2}{\eta} \le \frac{1}{2}$. >From (27) and Lemma A.4,

$$\begin{aligned} T_M &= -\frac{1}{n}\sum_{i=1}^n I_i M_i(\hat{\beta},\hat{\gamma})' \hat{J}_i(\hat{\beta},\hat{\gamma})' \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta}) - \frac{1}{n}\sum_{i=1}^n I_i M_i(\hat{\beta},\hat{\gamma})' \hat{J}_i(\hat{\beta},\hat{\gamma})' r^g_i + o_p\left(n^{-1/2}\right) \\ &= T^{(2)} + R^{(2)} + o_p\left(n^{-1/2}\right). \end{aligned} \tag{33}$$

>From (29) and Lemmas A.2 and A.4, $R^{(2)}$ satisfies

$$
\begin{aligned}
\left\|R^{(2)}\right\| &\leq \max_{i \in I_*}\left\|M_i(\hat{\beta},\hat{\gamma})\right\| \max_{i \in I_*}\|r_i^g\| \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \hat{J}_i(\hat{\beta},\hat{\gamma})\right\| \\
&= O_p(1)\, o_p\left(n^{1/\zeta}\right)\left\{O_p\left(c_n^2\right) + o_p\left(n^{-1+\frac{2}{\eta}}\right)\right\} O_p(1) = o_p\left(n^{-1/2}\right), \qquad (34)
\end{aligned}
$$

where the last equality follows from $\alpha < \frac{1}{s}\left(1 - \frac{4}{\zeta}\right)$ and $\frac{1}{\zeta} + \frac{2}{\eta} \leq \frac{1}{2}$. Thus, from (33),

$$
\begin{aligned}
T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i M_i(\hat{\beta},\hat{\gamma})' \hat{J}_i(\hat{\beta},\hat{\gamma})' \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta}) + o_p\left(n^{-1/2}\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1} \hat{g}_i(\hat{\beta}) + R^{(3)} + o_p\left(n^{-1/2}\right). \qquad (35)
\end{aligned}
$$

$R^{(3)}$ is implicitly defined and satisfies

$$
\begin{aligned}
||R^{(3)}|| &\leq \left\|\frac{1}{n}\sum_{i=1}^{n} I_i\{M_i(\hat{\beta},\hat{\gamma}) - \bar{M}_i(\beta_0,\gamma_*)\}' \hat{J}_i(\hat{\beta},\hat{\gamma})' \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta})\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)'\{\hat{J}_i(\hat{\beta},\hat{\gamma}) - \hat{J}_i(\beta_0,\gamma_*)\}' \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta})\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)'\{\hat{V}_i(\hat{\beta})^{-1} - \hat{V}_i(\beta_0)^{-1}\} \hat{g}_i(\hat{\beta})\right\| \\
&= ||R_a^{(3)}|| + ||R_b^{(3)}|| + ||R_c^{(3)}||.
\end{aligned}
$$

>From Assumption 3.2 (iv) and a similar argument to derive (40) shown below, we have $||R_a^{(3)}|| = o_p\left(n^{-1/2}\right)$. Assumption 3.2 (iv) and Lemmas A.1, A.2, and A.4 yield

$$
\begin{aligned}
||R_b^{(3)}|| &\leq \max_{i \in I_*} ||\bar{M}_i(\beta_0,\gamma_*)|| \max_{i \in I_*} ||\hat{J}_i(\hat{\beta},\hat{\gamma}) - \hat{J}_i(\beta_0,\gamma_*)|| \max_{i \in I_*} ||\hat{V}_i(\hat{\beta})^{-1}|| \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \hat{g}_i(\hat{\beta})\right\| \\
&= O_p(1)\left\{o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta_m}+\frac{1}{\eta}}\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta}+\frac{1}{\eta_m}}\right)\right\} O_p(1)\left\{O_p(c_n) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right)\right\} = o_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the last equality follows from $\frac{1}{\zeta_m} + \frac{2}{\eta} \leq \frac{1}{2}$, $\frac{1}{\zeta} + \frac{1}{\eta_m} + \frac{1}{\eta} \leq \frac{1}{2}$, and Assumption 3.1 (v). Similarly, Assumption 3.2 (iv) and Lemmas A.1, A.2, and A.4 imply that $||R_b^{(3)}|| = o_p\left(n^{-1/2}\right)$.

Thus, from (35),

$$\begin{aligned}
T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1}\hat{g}_i(\hat{\beta}) + o_p\left(n^{-1/2}\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1}\{\hat{g}_i(\beta_0) + \hat{G}_i(\tilde{\beta})(\hat{\beta}-\beta_0)\} + o_p\left(n^{-1/2}\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1}\hat{g}_i(\beta_0) + \hat{H}_M(\beta_0,\gamma_*)\Delta\frac{1}{n}\sum_{i=1}^{n}\psi(x_i,z_i,\beta_0) \\
&\quad + R^{(4)} + o_p\left(n^{-1/2}\right) \\
&= T_{Ma} + T_{Mb} + R^{(4)} + o_p\left(n^{-1/2}\right), \qquad (36)
\end{aligned}$$

where the second equality follows from an expansion of $\hat{g}_i(\hat{\beta})$ around $\hat{\beta}=\beta_0$, and $\tilde{\beta}$ is a point on the line joining $\hat{\beta}$ and β_0. $R^{(4)}$ is implicitly defined and satisfies

$$\begin{aligned}
\left\|R^{(4)}\right\| &\leq \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1}\{\hat{G}_i(\tilde{\beta}) - \hat{G}_i(\beta_0)\}\right\| ||\hat{\beta}-\beta_0|| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1}\hat{G}_i(\beta_0)\right\| o_p\left(n^{-1/2}\right) \\
&\leq \max_{i\in I_*}||\bar{M}_i(\beta_0,\gamma_*)|| \max_{i\in I_*}||\hat{J}_i(\beta_0,\gamma_*)|| \max_{i\in I_*}||\hat{V}_i(\beta_0)^{-1}|| \left\|\frac{1}{n}\sum_{i=1}^{n} I_i\{\hat{G}_i(\tilde{\beta}) - \hat{G}_i(\beta_0)\}\right\| ||\hat{\beta}-\beta_0|| \\
&\quad + \max_{i\in I_*}||\bar{M}_i(\beta_0,\gamma_*)|| \max_{i\in I_*}||\hat{J}_i(\beta_0,\gamma_*)|| \max_{i\in I_*}||\hat{V}_i(\beta_0)^{-1}|| \max_{i\in I_*}||\hat{G}_i(\beta_0)|| o_p\left(n^{-1/2}\right) \\
&= o_p\left(n^{-1+\frac{1}{\eta_2}}\right) + o_p\left(n^{-1/2}\right) = o_p\left(n^{-1/2}\right),
\end{aligned}$$

where the equality follows from Assumption 3.2 (iv) and Lemmas A.1, A.2, and A.3. Thus, from (36), we have $T_M = T_{Ma} + T_{Mb} + o_p\left(n^{-1/2}\right)$. T_{Ma} is written as

$$\begin{aligned}
T_{Ma} &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_i E[\hat{f}_i|x_i]^{-1}\bar{M}_i(\beta_0,\gamma^*)' \bar{J}_i(\beta_0,\gamma_*)' \bar{V}_i(\beta_0)^{-1}\frac{1}{nb_n^s}K_{ji}g_j(\beta_0) + R_a^{(5)} \\
&= \bar{T}_{Ma} + R_a^{(5)}, \qquad (37)
\end{aligned}$$

where $R_a^{(5)}$ is implicitly defined and satisfies

$$\begin{aligned}
\left\|R_a^{(5)}\right\| &\leq \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_0,\gamma_*)' \left\{\hat{J}_i(\beta_0,\gamma_*) - E[\hat{f}_i|x_i]^{-1}\bar{J}_i(\beta_0,\gamma_*)\right\}' \hat{V}_i(\beta_0)^{-1}\hat{g}_i(\beta_0)\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i E[\hat{f}_i|x_i]^{-1}\bar{M}_i(\beta_0,\gamma_*)' \bar{J}_i(\beta_0,\gamma_*)' \left\{\hat{V}_i(\beta_0)^{-1} - E[\hat{f}_i|x_i]\bar{V}_i(\beta_0)^{-1}\right\}\hat{g}_i(\beta_0)\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_i\left\{\hat{f}_i^{-1} - E[\hat{f}_i|x_i]^{-1}\right\}\bar{M}_i(\beta_0,\gamma_*)' \bar{J}_i(\beta_0,\gamma_*)' \bar{V}_i(\beta_0)^{-1}\frac{1}{nb_n^s}K_{ji}g_j(\beta_0)\right\| \\
&= ||R_{aa}^{(5)}|| + ||R_{ab}^{(5)}|| + ||R_{ac}^{(5)}||.
\end{aligned}$$

>From Assumption 3.2 (iv), Lemmas A.1 and A.2, and Tripathi and Kitamura (2004, Lemma C.1), we have $||R_{aa}^{(5)}|| \leq O_p\left(c_n^2\right) = o_p(n^{-1/2})$ from $\alpha < \frac{1}{3s}$. Similarly, we have $||R_{ab}^{(5)}|| \leq O_p\left(c_n^2\right) = o_p(n^{-1/2})$. Moreover, Assumption 3.2 (iv), Lemmas A.1 and A.2, and Tripathi and Kitamura (2004, eqn. (C.1)) imply $||R_{ac}^{(5)}|| \leq O_p\left(c_n^2\right) = o_p(n^{-1/2})$. Thus, from (37), we have $T_{Ma} = \bar{T}_{Ma} + o_p\left(n^{-1/2}\right)$. By applying the U-statistic arguments of Kitamura, Tripathi and Ahn (2004, pp.1696-1698) and Powell, Stock and Stoker (1989, Lemma 3.1), we have the asymptotic linear forms for $\bar{T}_{Ma}$:

$$n^{1/2}\bar{T}_{Ma} = -n^{-1/2}\sum_{i=1}^{n} I_i\bar{M}_i\left(\beta_0,\gamma_*\right)' J_i\left(\beta_0,\gamma_*\right)' V_i\left(\beta_0\right)^{-1} g_i(\beta_0) + o_p\left(1\right). \tag{38}$$

>From Lemmas A.1, A.2, and A.3, and a weak law of large numbers, we can show that $\hat{H}_M\left(\beta_0,\gamma_*\right) \xrightarrow{p} E[I_i\bar{M}_i\left(\beta_0,\gamma_*\right)' J_i\left(\beta_0,\gamma_*\right)' V_i\left(\beta_0\right)^{-1} G_i\left(\beta_0\right)] = H_M\left(\beta_0,\gamma_*\right)$. Therefore, $\bar{T}_{Mb}$ satisfies

$$n^{1/2}T_{Mb} = n^{-1/2}\sum_{i=1}^{n} H_M\left(\beta_0,\gamma_*\right)\Delta\psi(x_i,z_i,\beta_0) + o_p(1). \tag{39}$$

From (36), (38), and (39), a central limit theorem yields

$$\begin{aligned} n^{1/2}T_M &= n^{1/2}\bar{T}_{Ma} + n^{1/2}T_{Mb} + o_p\left(1\right) = n^{-1/2}\sum_{i=1}^{n}\psi_i^M\left(\beta_0,\gamma_*\right) + o_p\left(1\right) \\ &\xrightarrow{d} N\left(0,\Phi_M\right), \end{aligned} \tag{40}$$

where

$$\psi_i^M\left(\beta,\gamma\right) = -I_i\bar{M}_i\left(\beta,\gamma\right)' J_i\left(\beta,\gamma\right)' V_i\left(\beta\right)^{-1} g(z_i,\beta) + H_M\left(\beta,\gamma\right)\Delta\psi(x_i,z_i,\beta), \tag{41}$$

and $\Phi_M = E\left[\psi_i^M\left(\beta_0,\gamma_*\right)\psi_i^M\left(\beta_0,\gamma_*\right)'\right]$. >From Lemmas A.1, A.2, and A.3, we can show that $\hat{\Phi}_M \xrightarrow{p} \Phi_M$. Therefore, we have

$$M_g = nT_M'\hat{\Phi}_M^{-}T_M \xrightarrow{d} \chi^2_{\mathrm{rank}(\Phi_M)}.$$

■

Proof of (ii)

>From (26) and Lemma A.4, T_C in (12) is written as

$$\begin{aligned} T_C &= \frac{1}{n}\sum_{i=1}^{n} I_i\left\{\sum_{j=1}^{n}(\hat{p}_{ji}^g(\hat{\beta}) + \hat{p}_{ji}^N)h(z_j,\hat{\gamma})\right\}' \hat{V}_i^h(\hat{\gamma})^{-1}\left\{\sum_{j=1}^{n}(\hat{p}_{ji}^g(\hat{\beta}) - \hat{p}_{ji}^N)h(z_j,\hat{\gamma})\right\} \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_i\left\{\sum_{j=1}^{n}(2w_{ji} - w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta}))h(z_j,\hat{\gamma})\right\}' \hat{V}_i^h(\hat{\gamma})^{-1}\left\{\sum_{j=1}^{n}(w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta}))h(z_j,\hat{\gamma})\right\} + R^{(1c)}, \end{aligned}$$

where $R^{(1c)}$ is implicitly defined. From a similar argument to derive (32), $R^{(1c)}$ satisfies

$$\begin{aligned}
\left\|R^{(1c)}\right\| &\leq \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \left\{\sum_{j=1}^{n}(2w_{ji} - w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta}))h(z_j,\hat{\gamma})\right\}' \hat{V}_i^h(\hat{\gamma})^{-1}\left\{\sum_{j=1}^{n} w_{ji}r_{ji}h(z_j,\hat{\gamma})\right\}\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \left\{\sum_{j=1}^{n} w_{ji}r_{ji}h\left(z_j,\hat{\gamma}\right)\right\}' \hat{V}_i^h\left(\hat{\gamma}\right)^{-1}\left\{\sum_{j=1}^{n}\left\{w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta})\right\}h\left(z_j,\hat{\gamma}\right)\right\}\right\| \\
&\quad + \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \left\{\sum_{j=1}^{n} w_{ji}r_{ji}h\left(z_j,\hat{\gamma}\right)\right\}' \hat{V}_i^h\left(\hat{\gamma}\right)^{-1}\left\{\sum_{j=1}^{n} w_{ji}r_{ji}h\left(z_j,\hat{\gamma}\right)\right\}\right\| \\
&\leq o\left(n^{1/\zeta_m}\right)\left\{O_p\left(c_n\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right)\right\}^2 + o\left(n^{1/\zeta_m}\right)\left\{O_p\left(c_n\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right)\right\}^3 \\
&\quad + o\left(n^{2/\zeta_m}\right)\left\{O_p\left(c_n\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right)\right\}^4 \\
&= o_p\left(n^{-1/2}\right).
\end{aligned}$$

Thus, from Lemma A.4, we have

$$\begin{aligned}
T_C &= -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{\sum_{j=1}^{n}(2w_{ji} - w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta}))h(z_j,\hat{\gamma})\right\}' \hat{V}_i^h(\hat{\gamma})^{-1}\left\{\sum_{j=1}^{n}(w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta}))h(z_j,\hat{\gamma})\right\} \\
&\quad + o_p\left(n^{-1/2}\right) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{2\hat{h}_i(\hat{\gamma}) - J_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta})\right\}' \hat{V}_i^h(\hat{\gamma})^{-1}\left\{\hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta})\right\} \\
&\quad + R^{(2c)} + o_p\left(n^{-1/2}\right),
\end{aligned}$$

where $R^{(2c)}$ is implicitly defined. A similar argument to show (34) yields that $\left\|R^{(2c)}\right\| = o_p\left(n^{-1/2}\right)$. By setting

$$\begin{aligned}
M_i(x_i,\beta,\gamma)' &= \{2\hat{h}_i(\gamma) - J_i^h(\beta,\gamma)'\hat{V}_i(\beta)^{-1}\hat{g}_i(\beta)\}'\hat{V}_i^h\left(\gamma\right)^{-1}, \\
\bar{M}_i(x_i,\beta,\gamma)' &= 2E\left[h\left(z,\gamma\right)|x_i\right]' V_i^h\left(\gamma\right)^{-1}, \\
m(z_j,\beta,\gamma) &= h(z_j,\gamma),
\end{aligned}$$

we can apply the same argument as the proof of Theorem 3.1 (i). Thus,

$$\begin{aligned}
n^{1/2}T_C &= n^{-1/2}\sum_{i=1}^{n}\psi_i^C\left(\beta_0,\gamma_*\right) + o_p\left(1\right) \\
&\xrightarrow{d} N\left(0,\phi_C\right),
\end{aligned}$$

where

$$\psi_i^C\left(\beta,\gamma\right) = -I_i\bar{M}_i(x_i,\beta,\gamma)'J_i^h\left(\beta,\gamma\right)' V_i\left(\beta\right)^{-1} g(z_i,\beta) + H_C\left(\beta,\gamma\right)\Delta\psi(x_i,z_i,\beta), \qquad (42)$$

$\phi_C = E[\psi_i^C(\beta_0,\gamma_*)^2]$, and $H_C(\beta,\gamma) = E[I_i\bar{M}_i(\beta,\gamma)' J_i^h(\beta,\gamma)' V_i(\beta)^{-1} G_i(\beta)]$. From Lemmas A.1, A.2, and A.3, we can show that $\hat{\phi}_C \xrightarrow{p} \phi_C$. Therefore, we have

$$C_g = \frac{\sqrt{n}T_C}{\sqrt{\hat{\phi}_C}} \xrightarrow{d} N(0,1).$$

■

Proof of (iii)

>From (26) and Lemma A.4, we have

$$\begin{aligned} T_S &= \frac{1}{n}\sum_{i=1}^n I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\{\hat{p}_{ji}^g(\hat{\beta}) - \hat{p}_{ji}^N\}h_j(\hat{\gamma}) \\ &= -\frac{1}{n}\sum_{i=1}^n I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\{w_{ji}\lambda_i^g(\hat{\beta})'g_j(\hat{\beta})\}h_j(\hat{\gamma}) + R^{(1s)} \\ &= -\frac{1}{n}\sum_{i=1}^n I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\{\hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta})\} + R^{(1s)} + R^{(2s)}, \end{aligned}$$

where $R^{(1s)}$ and $R^{(2s)}$ are implicitly defined. Similar arguments to derive (32) and (34) yield $\|R^{(1s)}\| = o_p(n^{-1/2})$ and $\|R^{(2s)}\| = o_p(n^{-1/2})$, respectively. By setting

$$\begin{aligned} M_i(x_i,\beta,\gamma)' &= \hat{G}_i^h(\gamma)'\hat{V}_i^h(\gamma)^{-1}, \\ \bar{M}_i(x_i,\beta,\gamma)' &= G_i^h(\gamma)'V_i^h(\gamma)^{-1}, \\ m(z_j,\beta,\gamma) &= h(z_j,\gamma), \end{aligned}$$

we can apply the same argument as the proof of Theorem 3.1 (i). Thus,

$$\begin{aligned} n^{1/2}T_S &= n^{-1/2}\sum_{i=1}^n \psi_i^S(\beta_0,\gamma_*) + o_p(1) \\ &\xrightarrow{d} N(0,\Phi_S), \end{aligned}$$

where

$$\psi_i^S(\beta,\gamma) = -I_i\bar{M}_i(x_i,\beta,\gamma)'J_i^h(\beta,\gamma)'V_i(\beta)^{-1}g(z_i,\beta) + H_S(\beta,\gamma)\Delta\psi(x_i,z_i,\beta), \qquad (43)$$

$\Phi_S = E[\psi_i^S(\beta_0,\gamma_*)\psi_i^S(\beta_0,\gamma_*)']$, and $H_S(\beta,\gamma) = E[I_i\bar{M}_i(\beta,\gamma)' J_i^h(\beta,\gamma)' V_i(\beta)^{-1} G_i(\beta)]$. From Lemmas A.1, A.2, and A.3, we can show that $\hat{\Phi}_S \xrightarrow{p} \Phi_S$. Therefore, we have

$$S_g = nT_S'\hat{\Phi}_S^- T_S \xrightarrow{d} \chi^2_{\mathrm{rank}(\Phi_S)}.$$

■

A.2 Proof of Theorem 3.2

Proof of (i)

Assume that n is large enough so that $\hat{\beta} \in \mathcal{B}_0$ and $\beta_{0n} \in \mathcal{B}_0$. Note that Lemmas A.1-A.3 remain valid when β_0 is replaced by β_{0n}. Thus, from the proof of Tripathi and Kitamura (2003, Lemma B.1),

$$I_i \lambda_i^g(\hat{\beta}) = I_i \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta}) + I_i \tilde{r}_i^g,$$

where $\|\tilde{r}_i^g\| = o_p(n^{1/\zeta}) \left\{ \left(\max_{i \in I_*} \left\| \sum_{j=1}^n w_{ji} g_j(\beta_{0n}) \right\| \right)^2 + ||\hat{\beta} - \beta_{0n}||^2 \sum_{j=1}^n w_{ji} d_1(z_j)^2 \right\}$, and the $o_p(n^{1/\zeta})$ term does not depend on $i \in I_*$. From the continuity of $\delta(x)$ and $f(x)$, and the compactness of $\mathcal{X}_*$, an adapted version of Tripathi and Kitamura (2003, Lemma C.1) yields $\max_{i \in I_*} \left\| \sum_{j=1}^n w_{ji} g_j(\beta_{0n}) \right\| = O_p(c_n)$. Thus, Lemma A.4 also remains valid when β_0 is replaced by β_{0n}. Since the adapted versions of Lemmas A.1-A.4 are valid, we can proceed as in the proof of Theorem 3.1 (i) by replacing β_0 with β_{0n}. Therefore, under $\mathbf{H}_{gn}$,

$$\begin{aligned}
n^{1/2} T_M &= n^{-1/2} \sum_{i=1}^n \psi_i^M(\beta_{0n}, \gamma_*) + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \left\{ \psi_i^M(\beta_{0n}, \gamma_*) - E[\psi_i^M(\beta_{0n}, \gamma_*)] \right\} \\
&\quad + \{ -E\left[I_i \bar{M}_i(\beta_{0n}, \gamma_*)' J_i(\beta_{0n}, \gamma_*)' V_i(\beta_{0n})^{-1} E[g(z_i, \beta_{0n}) | x_i] \right] \\
&\quad + E\left[H_M(\beta_{0n}, \gamma_*) \Delta E[\psi(x_i, z_i, \beta_{0n}) | x_i] \right] \} + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \left\{ \psi_i^M(\beta_{0n}, \gamma_*) - E[\psi_i^M(\beta_{0n}, \gamma_*)] \right\} + \mu_M + o_p(1) \\
&\xrightarrow{d} N(\mu_M, \Phi_M).
\end{aligned}$$

$>$From adapted versions of Lemmas A.1-A.3, we can show that $\hat{\Phi}_M \xrightarrow{p} \Phi_M$ under $\mathbf{H}_{gn}$. Therefore, the conclusion is obtained. ■

Proof of (ii)

A similar argument to the proof of Theorem 3.2 (i) yields that under $\mathbf{H}_{gn}$,

$$
\begin{aligned}
n^{1/2}T_C &= n^{-1/2}\sum_{i=1}^{n}\psi_i^C\left(\beta_{0n},\gamma_*\right)+o_p\left(1\right)\\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\psi_i^C\left(\beta_{0n},\gamma_*\right)-E[\psi_i^C\left(\beta_{0n},\gamma_*\right)]\right\}\\
&\quad +\{-2E[I_iE\left[h\left(z,\gamma_*\right)|x_i\right]'V_i^h\left(\gamma_*\right)^{-1}J_i^h\left(\beta_{0n},\gamma_*\right)'V_i\left(\beta_{0n}\right)^{-1}E\left[g(z_i,\beta_{0n})|x_i\right]]\\
&\quad +E\left[H_C\left(\beta_{0n},\gamma_*\right)\Delta E\left[\psi(x_i,z_i,\beta_{0n})|x_i\right]\right]\}+o_p\left(1\right)\\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\psi_i^C\left(\beta_{0n},\gamma_*\right)-E[\psi_i^C\left(\beta_{0n},\gamma_*\right)]\right\}+\mu_C+o_p\left(1\right)\\
&\quad \xrightarrow{d} N\left(\mu_C,\phi_C\right).
\end{aligned}
$$

>From adapted versions of Lemmas A.1-A.3, we can show that $\hat{\phi}_C \xrightarrow{p} \phi_C$ under $\mathbf{H}_{gn}$. Therefore, the conclusion is obtained. ■

Proof of (iii)

A similar argument to the proof of Theorem 3.2 (i) yields that under $\mathbf{H}_{gn}$,

$$
\begin{aligned}
n^{1/2}T_S &= n^{-1/2}\sum_{i=1}^{n}\psi_i^S\left(\beta_{0n},\gamma_*\right)+o_p\left(1\right)\\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\psi_i^S\left(\beta_{0n},\gamma_*\right)-E[\psi_i^S\left(\beta_{0n},\gamma_*\right)]\right\}\\
&\quad \{-E[I_iG_i^h\left(\gamma_*\right)'V_i^h\left(\gamma_*\right)^{-1}J_i^h\left(\beta_{0n},\gamma_*\right)'V_i\left(\beta_{0n}\right)^{-1}E\left[g(z_i,\beta_{0n})|x_i\right]]\\
&\quad +E\left[H_S\left(\beta_{0n},\gamma_*\right)\Delta E\left[\psi(x_i,z_i,\beta_{0n})|x_i\right]\right]\}+o_p\left(1\right)\\
&= n^{-1/2}\sum_{i=1}^{n}\left\{\psi_i^S\left(\beta_{0n},\gamma_*\right)-E[\psi_i^S\left(\beta_{0n},\gamma_*\right)]\right\}+\mu_S+o_p\left(1\right)\\
&\quad \xrightarrow{d} N\left(\mu_S,\Phi_S\right).
\end{aligned}
$$

>From adapted versions of Lemmas A.1-A.3, we can show that $\hat{\Phi}_S \xrightarrow{p} \Phi_S$ under $\mathbf{H}_{gn}$. Therefore, the conclusion is obtained. ■

A.3 Proof of Theorem 3.3

Proof of (i)

Let $\tilde{J}_i(\beta,\gamma)' = \sum_{j=1}^{n} w_{ji} \frac{m(z_j,\beta,\gamma) g_j(\beta)'}{1+\lambda_i^g(\beta)' g_j(\beta)}$. By the definition of $\hat{p}_{ji}^g(\beta)$ in (4) and T_M in (9),

$$\begin{aligned}
T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i M_i(\hat{\beta},\hat{\gamma})' \tilde{J}_i(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_*,\gamma_0)' \tilde{J}_i(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) + o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_*,\gamma_0)' \tilde{J}_i(\hat{\beta},\hat{\gamma})' \lambda_*^g(x_i,\beta_*) + o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \bar{M}_i(\beta_*,\gamma_0)' J_{i*}(\beta_*,\gamma_0)' \lambda_*^g(x_i,\beta_*) + o_p(1) \\
&= \mu_{hM} + o_p(1),
\end{aligned}$$

under $\mathbf{H}_h$, where the second equality follows from Assumption 3.2 (iv), the third equality follows from $\max_{i\in I_*} ||\lambda_i^g(\hat{\beta}) - \lambda_*^g(x_i,\beta_*)|| \xrightarrow{p} 0$, and fourth equality follows by applying similar arguments as Lemma A.2 and Newey (1994, Lemma B.3). Therefore, we have $M_g/n \xrightarrow{p} \mu_{hM}' \Phi_{hM}^{-} \mu_{hM}$ under $\mathbf{H}_h$, and the conclusion is obtained. ■

Proof of (ii)

By the definition of $\hat{p}_{ji}^g(\beta)$ in (4) and T_C in (12),

$$\begin{aligned}
T_C &= -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} w_{ji} \frac{2h(z_j,\hat{\gamma})}{1+\lambda_i^g(\hat{\beta})' g_j(\hat{\beta})} + \hat{J}_{i*}^h(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) \right\}' \hat{V}_i^h(\hat{\gamma})^{-1} \hat{J}_{i*}^h(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{ E\left[\frac{2h(z,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)' g(z,\beta_*)} \middle| x_i \right] + J_{i*}^h(\beta_*,\gamma_0)' \lambda_*^g(x_i,\beta_*) \right\}' \\
&\quad \times V_i^h(\gamma_0)^{-1} \hat{J}_{i*}^h(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) + o_p(1) \\
&= -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{ E\left[\frac{2h(z,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)' g(z,\beta_*)} \middle| x_i \right] + J_{i*}^h(\beta_*,\gamma_0)' \lambda_*^g(x_i,\beta_*) \right\}' \\
&\quad \times V_i^h(\gamma_0)^{-1} J_{i*}^h(\beta_*,\gamma_0)' \lambda_*^g(x_i,\beta_*) + o_p(1) \\
&= \mu_{hC} + o_p(1),
\end{aligned}$$

under $\mathbf{H}_h$, where the second equality follows from Assumption 3.2 (iv), and the third equality follows from $\max_{i\in I_*} ||\lambda_i^g(\hat{\beta}) - \lambda_*^g(x_i,\beta_*)|| \xrightarrow{p} 0$ and similar arguments as Lemma A.2 and Newey (1994, Lemma B.3). Therefore, we have $C_g/\sqrt{n} \xrightarrow{p} \mu_{hC}/\sqrt{\phi_{hC}}$ under $\mathbf{H}_h$, and the conclusion is obtained. ■

Proof of (iii)

By the definition of $\hat{p}^g_{ji}(\beta)$ in (4) and T_S in (14),

$$\begin{aligned} T_S &= -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{G}_i(\hat{\gamma})' \hat{V}_i^h(\hat{\gamma})^{-1} \hat{J}^h_{i*}(\hat{\beta},\hat{\gamma})'\lambda_i^g(\hat{\beta}) \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_i G_i^h(\gamma_0)' V_i^h(\gamma_0)^{-1} \hat{J}^h_{i*}(\hat{\beta},\hat{\gamma})'\lambda_i^g(\hat{\beta}) + o_p(1) \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_i G_i^h(\gamma_0)' V_i^h(\gamma_0)^{-1} J^h_{i*}(\beta_*,\gamma_0)'\lambda_*^g(x_i,\beta_*) + o_p(1) \\ &= \mu_{hS} + o_p(1), \end{aligned}$$

under $\mathbf{H}_h$, where the second equality follows from Assumption 3.2 (iv), and the third equality follows from $\max_{i\in I_*}||\lambda_i^g(\hat{\beta}) - \lambda_*^g(x_i,\beta_*)|| \xrightarrow{p} 0$ and similar arguments to Lemma A.2 and Newey (1994, Lemma B.3). Therefore, we have $S_g/n \xrightarrow{p} \mu_{hS}'\Phi_{hS}^{-}\mu_{hS}$ under $\mathbf{H}_h$, and the conclusion is obtained. ■

A.4 Auxiliary Lemmas

Lemma A.1 *Suppose that Assumptions 3.1 (i), (ii), and (iv) and 3.2 (i)-(iii) hold. If* $\log n/n^{1-4/\zeta}b_n^s \to 0$*, then*

$$\sup_{x_i\in\mathcal{X}_*}\left\|\hat{V}_i(\hat{\beta}) - \hat{V}_i(\beta_0)\right\| = o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta}+\frac{1}{\eta}}\right), \quad \sup_{x_i\in\mathcal{X}_*}\left\|\hat{V}_i(\hat{\beta})^{-1} - \hat{V}_i(\beta_0)^{-1}\right\| = o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta}+\frac{1}{\eta}}\right),$$

$$\sup_{x_i\in\mathcal{X}_*}\left\|\hat{V}_i(\beta_0) - E[\hat{f}_i|x_i]^{-1}\bar{V}_i(\beta_0)\right\| = O_p(c_n), \quad \sup_{x_i\in\mathcal{X}_*}\left\|\hat{V}_i(\beta_0)^{-1} - E[\hat{f}_i|x_i]\bar{V}_i(\beta_0)^{-1}\right\| = O_p(c_n).$$

Proof. See the proof of Tripathi and Kitamura (2003, Lemma C.2). ■

Lemma A.2 *Suppose that Assumptions 3.1 (i)-(iv) and 3.2 hold. If* $\log n/n^{1-4/\min\{\zeta,\zeta_m\}}b_n^s \to 0$*, then*

$$\sup_{x_i\in\mathcal{X}_*}\left\|\hat{J}_i(\hat{\beta},\hat{\gamma}) - \hat{J}_i(\beta_0,\gamma_*)\right\| = o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta_m}+\frac{1}{\eta}}\right) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\zeta}+\frac{1}{\eta_m}}\right),$$

$$\sup_{x_i\in\mathcal{X}_*}\left\|\hat{J}_i(\beta_0,\gamma_*) - E[\hat{f}_i|x_i]^{-1}\bar{J}_i(\beta_0,\gamma_*)\right\| = O_p(c_n).$$

Proof. (First part) An expansion of $\hat{J}_i(\hat{\beta},\hat{\gamma})'$ around $(\hat{\beta},\hat{\gamma}) = (\beta_0,\gamma_*)$ and Assumption 3.2

(iii) and (iv) yield

$$
\begin{aligned}
& \sup_{x_i \in \mathcal{X}_*} \left\| \hat{J}_i(\hat{\beta}, \hat{\gamma})' - \hat{J}_i(\beta_0, \gamma_*)' \right\| \\
= & \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} \left(m_j\left(\beta_0, \gamma_*\right) + \frac{\partial m_j(\tilde{\beta}, \tilde{\gamma})}{\partial(\beta', \gamma')} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_* \end{pmatrix} \right) \left(g_j(\beta_0) + \frac{\partial g_j(\tilde{\beta})}{\partial \beta'} (\hat{\beta} - \beta_0) \right)' \right. \\
& \left. - \sum_{j=1}^{n} w_{ji} m_j\left(\beta_0, \gamma_*\right) g_j(\beta_0)' \right\| \\
\leq & \ ||\hat{\beta} - \beta_0|| \max_{1 \leq j \leq n} \left\| m_j\left(\beta_0, \gamma_*\right) \right\| \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} d_1\left(z_j\right) \right\| + \left\| \begin{matrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_* \end{matrix} \right\| \max_{1 \leq j \leq n} \left\| g_j(\beta_0) \right\| \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} d_m\left(z_j\right) \right\| \\
& + ||\hat{\beta} - \beta_0|| \left\| \begin{matrix} \hat{\beta} - \beta_0 \\ \hat{\gamma} - \gamma_* \end{matrix} \right\| \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} d_1\left(z_j\right) d_m\left(z_j\right) \right\| \\
= & \ R_a^J + R_b^J + R_c^J,
\end{aligned}
$$

where $(\tilde{\beta}, \tilde{\gamma})$ is a point on the line joining $(\hat{\beta}, \hat{\gamma})$ and (β_0, γ_*). From (30), Assumption 3.1 (ii) and (iii), and Tripathi and Kitamura (2003, Lemma C.6), we have

$$
R_a^J = o_p\left(n^{-\frac{1}{2} + \frac{1}{\zeta_m} + \frac{1}{\eta}}\right), \quad R_b^J = o_p\left(n^{-\frac{1}{2} + \frac{1}{\zeta} + \frac{1}{\eta_m}}\right), \quad R_c^J = o_p\left(n^{-1 + \max\{2/\eta, 2/\eta_m\}}\right).
$$

Since $\eta, \eta_m \geq 6$, R_c^J is negligible. Therefore, the first part is obtained.

(Second part) The second part is obtained from the proof of Newey (1994, Lemma B.3). ■

Lemma A.3 *Suppose that Assumptions 3.1 (i), (ii), and (iv) and 3.2 (i)-(iii) hold. If* $\log n / n^{1-2/\eta} b_n^s \to 0$*, then*

$$
\sup_{x_i \in \mathcal{X}_*} \left\| \hat{G}_i(\hat{\beta}) - \hat{G}_i\left(\beta_0\right) \right\| = o_p\left(n^{-\frac{1}{2} + \frac{1}{\eta_2}}\right),
$$

$$
\sup_{x_i \in \mathcal{X}_*} \left\| \hat{G}_i\left(\beta_0\right) - E[\hat{f}_i | x_i]^{-1} \bar{G}_i\left(\beta_0\right) \right\| = O_p\left(c_n\right).
$$

Proof. (First part) An expansion of $\partial g_j^{(k)}(\hat{\beta}) / \partial \beta^{(\ell)}$ around $\hat{\beta} = \beta_0$ and Assumption 3.2 (iii) yield

$$
\begin{aligned}
& \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} \frac{\partial g_j^{(k)}(\hat{\beta})}{\partial \beta^{(\ell)}} - \sum_{j=1}^{n} w_{ji} \frac{\partial g_j^{(k)}\left(\beta_0\right)}{\partial \beta^{(\ell)}} \right\| \leq \sup_{x_i \in \mathcal{X}_*} \left\| \sum_{j=1}^{n} w_{ji} d_2\left(z_j\right) \right\| \left\| \hat{\beta} - \beta_0 \right\| \\
= & \ o\left(n^{1/\eta_2}\right) O_p\left(n^{-1/2}\right),
\end{aligned}
$$

where the equality follows from Assumption 3.1 (ii) and Tripathi and Kitamura (2003, Lemma C.6). Therefore, the first part is obtained.

(Second part) The second part is obtained from the proof of Newey (1994, Lemma B.3). ■

Lemma A.4 *Suppose that Assumptions 3.1 (i), (ii), and (iv) and 3.2 (i)-(iii) hold. If* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}\left(1 - \frac{4}{\zeta}\right)$, *then under* $\mathbf{H}_g$

$$\max_{i \in I_*} ||\hat{g}_i(\hat{\beta})|| = O_p(c_n) + o_p\left(n^{-\frac{1}{2}+\frac{1}{\eta}}\right),$$

and

$$I_i \lambda_i^g(\hat{\beta}) = I_i \hat{V}_i(\hat{\beta})^{-1} \hat{g}_i(\hat{\beta}) + I_i r_i^g,$$

where

$$\max_{i \in I_*} \|r_i^g\| = o_p\left(n^{1/\zeta}\right)\left\{O_p\left(c_n^2\right) + o_p\left(n^{-1+\frac{2}{\eta}}\right)\right\}.$$

Proof. See the proof of Tripathi and Kitamura (2003, Lemma A.1). Note that Assumptions 3.1 (i), (ii), and (iv) and 3.2 (i)-(iii) imply Tripathi and Kitamura (2003, Assumptions 3.1-3.7).■

References

[1] Chen, Y. and C. Kuan (2002): "The pseudo-true score encompassing test for non-nested hypotheses," *Journal of Econometrics*, 106, 271-295.

[2] 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.

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

[4] Davidson, R. and J. MacKinnon (1981): "Several tests for model specification in the presence of alternative hypothesis," *Econometrica*, 49, 781-793.

[5] Dhaene, G. (1997): *Encompassing: Formulation, Properties and Testing*, Springer.

[6] Donald, S. G., Imbens, G. W. and W. K. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55-93.

[7] Fisher, G. and M. McAleer (1981): "Alternative procedures and associated tests of significance for non-nested hypotheses," *Journal of Econometrics*, 16, 103-119.

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

[9] Godfrey, L. G. (1998): "Tests of non-nested regression models: some results on small sample behaviour and the bootstrap," *Journal of Econometrics*, 84, 59-74.

[10] 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.

[11] Gourieroux, C., A. Monfort and A. Trognon (1983): "Testing nested or non-nested hypotheses," *Journal of Econometrics*, 21, 83-115.

[12] Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029-1054.

[13] Kitamura, Y. (2001): "Asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, 69, 1661-1672.

[14] Kitamura, Y. (2003): "A likelihood-based approach to the analysis of a class of nested and non-nested models," manuscript.

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

[16] Loh, W. (1985): "A new method for testing separate families of hypotheses," *Journal of the American Statistical Association*, 80, 362-368.

[17] Mizon, G. and J. Richard (1986): "The encompassing principle and its application to testing non-nested hypotheses," *Econometrica*, 54, 657-678.

[18] Newey, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, 58, 809-837.

[19] Newey, W. K. (1994): "Kernel estimation of partial means and a general variance estimator," *Econometric Theory*, 10, 233-253.

[20] Newey,W. K. and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

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

[22] Owen, A. B. (2001): *Empirical Likelihood*, Chapman and Hall.

[23] 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.

[24] Powell, J. L., Stock, J. L. and T. M. Stoker (1989): "Semiparametric estimation of index coefficients," *Econometrica*, 57, 1403-1430.

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

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

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

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

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

[30] Tripathi, G. and Y. Kitamura (2003): "Testing conditional moment restrictions," *Annals of Statistics*, 31, 2059-2095.

[31] Vuong, Q. H. (1989): "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, 57, 307-333.

[32] White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1-26.

[33] Wooldridge, J. (1990): "An encompassing approach to conditional mean tests with application to testing nonnested hypotheses," *Journal of Econometrics*, 45, 331-350.

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

Table 1. Estimated Sizes and Powers of the tests with nominal size of 5%[9]

(Design I, $c_0 = 1$)

		$n = 100$			$n = 200$		
Test	b_n	Size	A-P	S-P	Size	A-P	S-P
	0.7	.170	.778	.528	.135	.936	.878
M_g	0.8	.100	.777	.678	.090	.947	.923
	0.9	.064	.775	.749	.060	.966	.961
	1.0	.046	.781	.796	.029	.960	.969
	0.7	.070	.500	.399	.038	.600	.703
C_g	0.8	.030	.389	.581	.023	.462	.848
	0.9	.010	.281	.684	.007	.343	.889
	1.0	.005	.202	.726	.001	.211	.899
	0.7	.329	.970	.823	.174	.989	.978
S_g	0.8	.244	.968	.905	.110	.996	.992
	0.9	.164	.982	.945	.070	.997	.995
	1.0	.123	.989	.971	.045	.999	.999
J		.041	.926	.934	.052	.999	.998
S		.008	.911	.972	.007	.997	1.00
SC		.055	.935	.934	.054	.999	.999

[9] Tests M_g,C_g, and S_g refer to the moment encompassing, Cox-type, and efficient score encompassing tests, repectively. Also, tests J, S, and SC refer to Hansen's (1982) overidentifying test, Singleton's (1985) test, and Ramalho and Smith's (2002) simplified Cox test, respectively. A-P and S-P denote Actual Power and Size-Corrected Power, respectively.

Table 2. Estimated Sizes and Powers of the tests with nominal size of 5%[10]

(Design I, $c_0 = 2$)

		$n = 100$			$n = 200$		
Test	b_n	Size	A-P	S-P	Size	A-P	S-P
	0.7	.176	.537	.262	.138	.752	.517
M_g	0.8	.104	.500	.357	.084	.745	.644
	0.9	.071	.460	.415	.057	.732	.711
	1.0	.039	.442	.473	.038	.716	.748
	0.7	.064	.272	.221	.036	.244	.327
C_g	0.8	.029	.165	.309	.021	.147	.467
	0.9	.013	.095	.390	.008	.076	.584
	1.0	.003	.046	.403	.001	.036	.601
	0.7	.325	.953	.807	.175	.986	.971
S_g	0.8	.230	.957	.876	.117	.987	.981
	0.9	.164	.965	.908	.071	.988	.985
	1.0	.126	.958	.931	.039	.992	.994
J		.044	.563	.572	.056	.868	.865
S		.021	.554	.666	.023	.863	.906
SC		.055	.589	.582	.053	.878	.876

[10] Tests M_g,C_g, and S_g refer to the moment encompassing, Cox-type, and efficient score encompassing tests, repectively. Also, tests J, S, and SC refer to Hansen's (1982) overidentifying test, Singleton's (1985) test, and Ramalho and Smith's (2002) simplified Cox test, respectively. A-P and S-P denote Actual Power and Size-Corrected Power, respectively.

Table 3. Estimated Sizes and Powers of the tests with nominal size of 5%[11]
(Design II)

		$n=100$			$n=200$		
Test	b_n	Size	A-P	S-P	Size	A-P	S-P
	0.1	.062	.624	.502	.043	.635	.696
M_g	0.2	.018	.604	.913	.015	.608	.959
	0.3	.009	.538	.967	.008	.568	.984
	0.4	.007	.452	.984	.004	.471	.981
	0.1	.164	.685	.428	.112	.670	.454
C_g	0.2	.061	.660	.639	.040	.675	.675
	0.3	.029	.664	.803	.027	.680	.883
	0.4	.018	.644	.897	.017	.707	.948
	0.1	.095	.292	.140	.078	.334	.234
S_g	0.2	.053	.356	.339	.040	.414	.486
	0.3	.034	.412	.589	.027	.427	.729
	0.4	.020	.433	.791	.017	.489	.837
J		.048	.027	.027	.053	.040	.034
S		.011	.021	.158	.009	.031	.172
SC		.008	.075	.174	.004	.070	.165

[11] Tests M_g,C_g, and S_g refer to the moment encompassing, Cox-type, and efficient score encompassing tests, repectively. Also, tests J, S, and SC refer to Hansen's (1982) overidentifying test, Singleton's (1985) test, and Ramalho and Smith's (2002) simplified Cox test, respectively. A-P and S-P denote Actual Power and Size-Corrected Power, respectively.