Econometric Theory, **27**, 2011, 114–153.
doi:10.1017/S0266466610000149

TESTING FOR NONNESTED CONDITIONAL MOMENT RESTRICTIONS VIA CONDITIONAL EMPIRICAL LIKELIHOOD

Taisuke Otsu
Yale University

Yoon-Jae Whang
Seoul National University

We propose nonnested tests for competing conditional moment restriction models using the method of conditional empirical likelihood, recently developed by Kitamura, Tripathi, and Ahn (2004) and Zhang and Gijbels (2003). To define the test statistics, we use the implied conditional probabilities from conditional empirical likelihood, which take into account the full implications of conditional moment restrictions. We propose three types of nonnested tests: the moment-encompassing, Cox-type, and efficient score-encompassing tests. We derive the asymptotic null distributions and investigate their power properties against a sequence of local alternatives and a fixed global alternative. Our tests have distinct global power properties from some of the existing tests based on finite-dimensional unconditional moment restrictions. Simulation experiments show that our tests have reasonable finite sample properties and dominate some of the existing nonnested tests in terms of size-corrected powers.

1. INTRODUCTION

Econometric models are often written in the form of conditional moment restrictions. While researchers derive and estimate their conditional moment restriction models, those models are typically nonnested and should be evaluated by some formal tests. This paper proposes nonnested 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 implied conditional probabilities from CEL, we develop three

We would like to thank Yuichi Kitamura, a co-editor, and two anonymous referees for helpful comments. This work was financially supported by the Korea Science and Engineering Foundation. Address correspondence to Taisuke Otsu, Cowles Foundation for Research in Econonomics, Yale University, New Haven CT 06520, U.S.A; e-mail address: taisuke.otsu@yale.edu.

CEL-based nonnested tests: the moment-encompassing, Cox-type, and efficient score-encompassing tests. Compared to the existing nonnested tests, which mainly focus on testing parametric models or unconditional moment restrictions, our approach tests conditional moment restrictions, which imply an infinite number of unconditional moment restrictions. Our tests are asymptotically equivalent to some unconditional moment-based tests under the null hypothesis and a sequence of local alternatives. However, such equivalence does not hold under the alternative hypothesis, and hence the global power properties of our tests can be significantly different from those of the unconditional moment-based tests. In other words, our tests in general have nontrivial power against nonnested alternatives that may not be detected by the unconditional moment-based tests. Simulation experiments show that our tests dominate some of the existing nonnested tests in terms of size-corrected powers.

Since Cox (1961, 1962), nonnested testing for competitive statistical models has become a standard technique to evaluate the specification of a statistical model against a specific alternative model.[2] Singleton (1985), Ghysels and Hall (1990), and Smith (1992) proposed nonnested testing procedures for *unconditional* moment restriction models. Those procedures are extended by Smith (1997) and Ramalho and Smith (2002) to the generalized empirical likelihood (GEL) context.[3] Ramalho and Smith (2002) focused on the implied unconditional probabilities from the null unconditional moment restrictions and derived GEL analogs of the moment-encompassing, Cox-type, and parametric-encompassing tests. We extend the approach by Smith (1997) and Ramalho and Smith (2002) to test *conditional* moment restriction models that imply an infinite number of unconditional moment restrictions. In particular, we employ the method of CEL to obtain the implied conditional probabilities from the conditional moment restrictions and develop nonnested test statistics based on these implied conditional probabilities. Since the implied conditional probabilities contain all information from the conditional moment restrictions, we can use those probabilities to define our test statistics.

Since Owen (1988) and Qin and Lawless (1994), the method of empirical likelihood has become an attractive alternative against the conventional generalized method of moments (GMM) approach.[4] Kitamura (2001) and Newey and Smith (2004) showed desirable asymptotic properties of empirical likelihood for testing and estimating unconditional moment restriction models, respectively. To estimate conditional moment restriction models, Kitamura et al. (2004) and Zhang and Gijbels (2003) developed the method of CEL and showed that the CEL estimator is asymptotically normal and efficient. Tripathi and Kitamura (2003) proposed CEL-based consistent specification tests for conditional moment restrictions. This paper extends the CEL approach to nonnested testing problems. Compared to Tripathi and Kitamura's (2003) specification tests, our tests check the validity of the null model against some specific alternative model, and our test statistics converge at the parametric (or $\sqrt{n}$-) rate. However, as a cost of the parametric convergence rate, our tests have implicit null hypotheses, i.e., sets of distributions where the tests do not have nontrivial power. 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 two competing models have the same Kullback-Leibler information divergence (or relative entropy) from the true model. Our nonnested testing approach sets one of the competing models as the null hypothesis and checks the validity of the null model against an alternative model.

This paper is organized as follows: Section 2 introduces our basic set-up and test statistics. In Sections 3.1 and 3.2 we derive the null distributions and local power properties of the test statistics. Section 3.3 discusses the global power properties of our tests and provides a sufficient condition for the consistency of the Cox-type test. Section 3.4 compares the proposed tests with the existing unconditional moment-based 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; $||A|| = \sqrt{\text{trace}(AA')}$ 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; $I\{A\}$ is the indicator function for an event A; $\text{int}(A)$ is the interior of a set A; and $a^{(i)}$ means the ith component of a vector a.

2. SET-UP AND TEST STATISTICS

2.1. Nonnested Hypotheses

Suppose that we observe a random sample $\{x_i, z_i\}_{i=1}^n$, where $x \in \mathcal{X} \subset R^s$ and $z \in R^{d_z}$. Let $\mathcal{F}_z$ and $\mathcal{F}_x$ be the σ-algebra for z and x, respectively. We assume that $\mathcal{F}_z \nsubseteq \mathcal{F}_x$. Consider the two competing conditional moment restrictions:

$$\mathbf{H}_g : \mathrm{E}[g(z,\beta_0)|x] = 0 \quad \text{a.s.}x, \tag{1}$$

$$\mathbf{H}_h : \mathrm{E}[h(z,\gamma_0)|x] = 0 \quad \text{a.s. } x,$$

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 measurable functions up to unknown parameters $\beta_0 \in \mathcal{B} \subset R^{d_\beta}$ and $\gamma_0 \in \Gamma \subset R^{d_\gamma}$, respectively.[5] Let $\mathcal{M}_{z|x}$ be the space of all conditional measures of z given x. The spaces of conditional measures that satisfy $\mathbf{H}_g$ and $\mathbf{H}_h$ are written as

$$\mathcal{G}_{z|x} = \cup_{\beta \in \mathcal{B}} \left\{ (\mu_{z|x})_{x \in \mathcal{X}} \in \mathcal{M}_{z|x} : \int g(z,\beta)\, d\mu_{z|x} = 0 \quad \text{a.s. } x \right\}, \tag{2}$$

$$\mathcal{H}_{z|x} = \cup_{\gamma \in \Gamma} \left\{ (\mu_{z|x})_{x \in \mathcal{X}} \in \mathcal{M}_{z|x} : \int h(z,\gamma)\, d\mu_{z|x} = 0 \quad \text{a.s. } x \right\},$$

respectively. Let $(\mu^0_{z|x})_{x\in\mathcal{X}}$ be the true conditional measure of z given x. The hypotheses $\mathbf{H}_g$ and $\mathbf{H}_h$ in (1) can be written alternatively as

$$\mathbf{H}_g : (\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{G}_{z|x},$$
$$\mathbf{H}_h : (\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{H}_{z|x}.$$

It is assumed that the models $\mathbf{H}_g$ and $\mathbf{H}_h$ are nonnested, i.e.,

$$\mathcal{G}_{z|x} \subsetneq \mathcal{H}_{z|x} \quad \text{and} \quad \mathcal{H}_{z|x} \subsetneq \mathcal{G}_{z|x}. \tag{3}$$

Note that the conditional moment restrictions $\mathbf{H}_g$ and $\mathbf{H}_h$ imply the following unconditional moment restrictions:

$$\mathbf{H}^U_g : \mathrm{E}\left[Q_g(x)\, g(z,\beta_0)\right] = 0, \tag{4}$$
$$\mathbf{H}^U_h : \mathrm{E}[Q_h(x)\, h(z,\gamma_0)] = 0,$$

for *any* matrices of measurable functions Q_g and Q_h, respectively. Several papers such as Singleton (1985), Smith (1992), and Ramalho and Smith (2002) proposed various nonnested tests between the unconditional moment restrictions $\mathbf{H}^U_g$ and $\mathbf{H}^U_h$ for *some* specific choices of Q_g and Q_h. However, if we are interested in the validity of the original conditional moment restriction $\mathbf{H}_g$ or $\mathbf{H}_h$, the conventional nonnested tests for $\mathbf{H}^U_g$ or $\mathbf{H}^U_h$ may not be appropriate in the following senses: First, the choice of the weight matrices $Q_g(x)$ and $Q_h(x)$ can be arbitrary if we are interested in testing (1); see also Kitamura (2006, Sec. 7) for this point. Our tests defined below, however, directly test the conditional moment restrictions (1). Second, there may be some alternative conditional moment restrictions where the existing nonnested tests for $\mathbf{H}^U_g$ or $\mathbf{H}^U_h$ do not have nontrivial power. For example, suppose that the true joint measure for (z,x) satisfies $\mathrm{E}\left[Q_g(x)\, g(z,\beta_0)\right] = 0$ but $\mathrm{E}[\tilde{Q}_g(x)\, g(z,\beta_0)] \neq 0$ for some $\tilde{Q}_g$. Then although the original null hypothesis $\mathbf{H}_g$ is violated, the existing nonnested tests based on $\mathbf{H}^U_g$ cannot reject the null hypothesis $\mathbf{H}_g$. To be more precise, consider the spaces of conditional measures that satisfy $\mathbf{H}^U_g$ and $\mathbf{H}^U_h$, i.e.,

$$\mathcal{G}^U_{z|x} = \cup_{\beta\in\mathcal{B}}\left\{(\mu_{z|x})_{x\in\mathcal{X}} \in \mathcal{M}_{z|x} : \int\int Q_g(x)\, g(z,\beta)\, d\mu_{z|x}\, d\mu_x = 0 \text{ for some } \mu_x\right\}, \tag{5}$$

$$\mathcal{H}^U_{z|x} = \cup_{\gamma\in\Gamma}\left\{(\mu_{z|x})_{x\in\mathcal{X}} \in \mathcal{M}_{z|x} : \int\int Q_h(x)\, h(z,\gamma)\, d\mu_{z|x}\, d\mu_x = 0 \text{ for some } \mu_x\right\},$$

respectively. Since $\mathbf{H}_g$ and $\mathbf{H}_h$ imply $\mathbf{H}^U_g$ and $\mathbf{H}^U_h$, respectively, we have $\mathcal{G}_{z|x} \subset \mathcal{G}^U_{z|x}$ and $\mathcal{H}_{z|x} \subset \mathcal{H}^U_{z|x}$. We will see that our nonnested test statistics have well-defined limiting distributions under $(\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{G}_{z|x}$ but they generally diverge under $(\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{G}^U_{z|x} \setminus \mathcal{G}_{z|x}$. However, nonnested test statistics based on $\mathbf{H}^U_g$ are generally (stochastically) bounded even if $(\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{G}^U_{z|x} \setminus \mathcal{G}_{z|x}$. Therefore the latter tests cannot detect those departures from $\mathbf{H}_g$. See Sections 3.3 and 3.4 for more detailed discussions.

This paper proposes three CEL-based nonnested tests for the conditional moment restrictions $\mathbf{H}_g$ against $\mathbf{H}_h$.

2.2. Conditional Empirical Likelihood

This section introduces the CEL approach. CEL is nonparametric likelihood constructed by the conditional moment restrictions in (1). Let p^g_{ji} for $i, j = 1, \ldots, n$ be multinomial conditional weights 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 maximization problem using p^g_{ji}:

$$\max_{\{p^g_{ji}\}^n_{i,j=1}} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p^g_{ji} \tag{6}$$

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

The conditional moment restriction $\mathbf{H}_g$ is incorporated in the constraints $\sum_{j=1}^n p^g_{ji} g\left(z_j, \beta\right) = 0$ for $i = 1, \ldots, n$. This problem can be solved by the Lagrange multiplier method. Let $\{\mu^g_i\}^n_{i=1}$ and $\{\lambda^g_i\}^n_{i=1}$ be the Lagrange multipliers. The Lagrangian is written as

$$\mathcal{L} = \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p^g_{ji} - \sum_{i=1}^n \mu^g_i \left(\sum_{j=1}^n p^g_{ji} - 1 \right) - \sum_{i=1}^n \lambda^{g\prime}_i \left(\sum_{j=1}^n p^g_{ji} g\left(z_j, \beta\right) \right).$$

The solution for p^g_{ji}, the implied conditional probability, is obtained as

$$\hat{p}^g_{ji}(\beta) = \frac{w_{ji}}{1 + \lambda^g_i(\beta)' g\left(z_j, \beta\right)}, \tag{7}$$

for $i, j = 1, \ldots, n$, where $\lambda^g_i(\beta)$ satisfies

$$\sum_{j=1}^n \frac{w_{ji} g\left(z_j, \beta\right)}{1 + \lambda^g_i(\beta)' g\left(z_j, \beta\right)} = 0, \tag{8}$$

for $i = 1, \ldots, n$.[6] If we do not impose the conditional moment restriction $\sum_{j=1}^n p^g_{ji} g\left(z_j, \beta\right) = 0$ in (6), the solution of the unconstrained 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 based on $\mathbf{H}_g$ is defined as

$$\ell_g(\beta) = \sum_{i=1}^{n} I_i \sum_{j=1}^{n} w_{ji} \log \hat{p}_{ji}^{g}(\beta) = \sum_{i=1}^{n} I_i \sum_{j=1}^{n} w_{ji} \log \left(\frac{w_{ji}}{1 + \lambda_i^{g}(\beta)' g(z_j, \beta)} \right), \quad \textbf{(9)}$$

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 to focus on specification testing over regions in $\mathcal{X}$ that are empirically more relevant. It also avoids the boundary problem in kernel estimation; see Tripathi and Kitamura (2003, p. 2062).

The CEL estimator is defined as $\hat{\beta}_{CEL} = \arg\max_{\beta \in \mathcal{B}} \ell_g(\beta)$. Under $\mathbf{H}_g$, $\hat{\beta}_{CEL}$ is consistent and asymptotically normal (see Kitamura et al., 2004).[7] In the same manner, we can define CEL $\ell_h(\gamma)$ based on $\mathbf{H}_h$ 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

$$\beta_{CEL}^{*} = \arg\min_{\beta \in \mathcal{B}} \mathrm{E}\left[I_i \max_{\lambda^g \in R^{d_g}} \mathrm{E}\left[\log\left(1 + \lambda^{g\prime} g(z_i, \beta)\right) | x_i \right] \right]. \quad \textbf{(10)}$$

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

To construct the nonnested test statistics, we employ some $\sqrt{n}$-consistent estimators $\hat{\beta}$ and $\hat{\gamma}$ for β_0 and γ_0, respectively. For example, $\hat{\beta}$ and $\hat{\gamma}$ may be the CEL estimators or the GMM estimators based on the unconditional moment restrictions in (4). Let β_* and γ_* be the pseudo-true values for $\hat{\beta}$ and $\hat{\gamma}$, respectively. Given $\hat{\beta}$, the implied conditional probabilities from $\mathbf{H}_g$ are obtained as $\{\hat{p}_{ji}^{g}(\hat{\beta})\}_{i,j=1}^{n}$ in (7).[8] By comparing $\hat{p}_{ji}^{g}(\hat{\beta})$ and $\hat{p}_{ji}^{N}$, we develop three nonnested tests: the moment encompassing, Cox-type, and efficient score-encompassing tests.

2.3. Test Statistics

2.3.1. Moment-Encompassing Test Statistic. We first define the CEL-based moment-encompassing test statistic. The moment-encompassing approach is employed by Ghysels and Hall (1990) and Ramalho and Smith (2002), for example. To test nonnested moment condition models, Ramalho and Smith (2002) considered a contrast of estimators for some moments based on the sample average and weighted average using the GEL implied unconditional probabilities. Here we incorporate the implied conditional probabilities to construct a contrast of estimators of moments. Consider moment indicators in the form of $\tilde{m}(x_i, z_i, \beta, \gamma) = \hat{M}(x_i, \beta, \gamma)' m(z_i, \beta, \gamma)$, where $\hat{M}(x_i, \beta, \gamma)$ is a $d_m \times d_M$ possibly random matrix of functions of $\{x_i, z_i\}_{i=1}^{n}$ and $m(z_i, \beta, \gamma)$ is a $d_m \times 1$ vector of functions of z_i. A typical choice of $m(z_i, \beta, \gamma)$ is the moment function $h(z_i, \gamma)$ for the alternative conditional moment restrictions $\mathbf{H}_h$ in (1). We assume that $\hat{M}(x_i, \hat{\beta}, \hat{\gamma})$ converges to $M(x_i, \beta_0, \gamma_*)$ uniformly on $x_i \in \mathcal{X}_*$ (Assumption 3.2(iv)). For each element of $\hat{M}(x_i, \beta, \gamma)$, we allow these cases: (i) constants or functions of (β, γ), (ii) functions of x_i or (x_i, β, γ), and (iii) weighted sums in the form of $\sum_{j=1}^{n} w_{ji} f(z_j, \beta, \gamma)$

or functions of such weighted sums. For brevity, we use the same notation $\hat{M}(x_i, \beta, \gamma)$ and omit other arguments such as $\{x_j\}_{j\neq i}$ and $\{z_j\}_{j=1}^n$. 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 $\mathrm{E}\left[\tilde{m}\left(x_i, z_i, \beta_0, \gamma_*\right)\right]$:

$$T_M = \frac{1}{n}\sum_{i=1}^n I_i \sum_{j=1}^n \hat{p}_{ji}^g(\hat{\beta})\tilde{m}(x_i, z_j, \hat{\beta}, \hat{\gamma}) - \frac{1}{n}\sum_{i=1}^n I_i \sum_{j=1}^n \hat{p}_{ji}^N \tilde{m}(x_i, z_j, \hat{\beta}, \hat{\gamma}), \tag{11}$$

where the first term is a nonparametric sample analog of $\mathrm{E}_x[I_i \mathrm{E}^g[\tilde{m}(x_i, z_i, \beta_0, \gamma_*)|x_i]]$ using the implied conditional probability $\hat{p}_{ji}^g(\hat{\beta})$ from $\mathbf{H}_g$, and the second term is a nonparametric sample analog of $\mathrm{E}_x[I_i \mathrm{E}[\tilde{m}\left(x_i, z_i, \beta_0, \gamma_*\right)|x_i]]$ using the Nadaraya-Watson kernel weights $\hat{p}_{ji}^N$, where $\mathrm{E}^g[\cdot|x_i]$ is the conditional expectation taken under $\mathbf{H}_g$ and $\mathrm{E}_x[\cdot]$ is the expectation for x. If the null hypothesis $\mathbf{H}_g$ is correct, these nonparametric analogs have the same probability limit, and hence we expect that T_M converges to 0. On the other hand, if $\mathbf{H}_g$ is incorrect, the two terms in (11) converge to different probability limits, and hence T_M converges to some nonzero constant vector. The moment indicator $\tilde{m}\left(x_i, z_j, \beta, \gamma\right)$ determines the direction of misspecification. Let

$$J_i(\beta, \gamma)' = \mathrm{E}[m(z_i, \beta, \gamma)\, g(z_i, \beta)'|x_i], \qquad V_i(\beta) = \mathrm{E}[g(z_i, \beta)\, g(z_i, \beta)'|x_i],$$

$$G_i(\beta) = \mathrm{E}[\partial g(z_i, \beta)/\partial \beta' | x_i].$$

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

$$M_g = n T_M' \hat{\Phi}_M^- T_M, \tag{12}$$

where $\hat{\Phi}_M^-$ is a consistent estimator for a g-inverse of $\Phi_M = \mathrm{E}[\psi_i^M(\beta_0, \gamma_*)\psi_i^M(\beta_0, \gamma_*)']$,

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

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

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

$$n^{1/2}(\hat{\beta} - \beta_0) = -n^{-1/2}\Delta \sum_{i=1}^n \psi(x_i, z_i, \beta_0) + o_p(1). \tag{13}$$

To obtain a specific form of $\hat{\Phi}_M^-$, we need to specify Δ and $\psi(x_i, z_i, \beta_0)$. Section 2.3.4 provides an example for the case of the CEL estimator. 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 based on Smith (1992, 1997) and Ramalho and Smith (2002). To construct a nonnested test statistic for moment condition models, Smith (1992) considered a contrast of estimators for the probability limit of the GMM criterion function. This idea is a natural generalization of the original nonnested test by Cox (1961, 1962) to test nonnested parametric models. Smith (1997) and Ramalho and Smith (2002) extended this approach to the GEL criterion function. We follow their approach and focus on the probability limit of the GMM-type (or Euclidean) nonparametric likelihood. Let

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

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

Note that $\hat{h}_i(\gamma), \hat{h}_i^g(\gamma)$, and $\hat{V}_i^h(\gamma)$ are nonparametric sample analogs of $\mathrm{E}\left[h(z_i, \gamma)|x_i\right]$, $\mathrm{E}^g\left[h(z_i, \gamma)|x_i\right]$, and $V_i^h(\gamma) = \mathrm{E}\left[h(z_i, \gamma) h(z_i, \gamma)' |x_i\right]$, respectively. By using $\hat{p}_{ji}^g(\hat{\beta})$ and $\hat{p}_{ji}^N = w_{ji}$, we consider the following contrast of the Euclidean likelihoods:[9]

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

We can expect that T_C converges to 0 under the null hypothesis $\mathbf{H}_g$ because both terms in T_C converge to the same probability limit $\mathrm{E}_x[I_i \mathrm{E}^g[h(z_i, \gamma_*)|x_i]' V_i^h(\gamma_*)^{-1} \mathrm{E}^g[h(z_i, \gamma_*)|x_i]]$. On the other hand, under the alternative hypothesis $\mathbf{H}_h$, T_C will converge to the probability limit $\mathrm{E}_x[I_i E^g[h(z_i, \gamma_0)|x_i]' V_i^h(\gamma_0)^{-1} \mathrm{E}^g[h(z_i, \gamma_0)|x_i]]$, which is generally nonzero if the null and alternative hypotheses are strictly nonnested; see Section 3.3 for details. Let $J_i^h(\beta, \gamma)' = \mathrm{E}[h(z_i, \gamma) g(z_i, \beta)'|x_i]$. Then the CEL-based Cox-type test statistic for $\mathbf{H}_g$ is defined as

$$C_g = \frac{\sqrt{n} T_C}{\sqrt{\hat{\phi}_C}}, \tag{15}$$

where $\hat{\phi}_C$ is a consistent estimator of $\phi_C = \mathrm{E}[\psi_i^C(\beta_0, \gamma_*)^2]$,

$$\psi_i^C(\beta, \gamma) = -2I_i \mathrm{E}[h(z_i, \gamma)|x_i]' V_i^h(\gamma)^{-1} J_i^h(\beta, \gamma)' V_i(\beta)^{-1} g(z_i, \beta) + H_C(\beta, \gamma)\Delta\psi(x_i, z_i, \beta),$$

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

Now Δ and $\psi(x_i, z_i, \beta)$ are defined in (13). To obtain a specific form of $\hat{\phi}_C$, we need to specify Δ and $\psi(x_i, z_i, \beta_0)$. Section 2.3.4 provides an example for the case of the CEL estimator. 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. We focus on the probability limit of the asymptotic linear form of an asymptotically efficient estimator for γ_0 under $\mathbf{H}_h$ (i.e., the efficient score for estimating γ_0):[10]

$$\sqrt{n}\left(\hat{\gamma}-\gamma_0\right)=-I^h\left(\gamma_0\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_i G_i^h\left(\gamma_0\right)' V_i^h\left(\gamma_0\right)^{-1}h\left(z_i,\gamma_0\right)+o_p\left(1\right),$$

where

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

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

$$T_S=\frac{1}{n}\sum_{i=1}^{n}I_i\hat{G}_i^h\left(\hat{\gamma}\right)'\hat{V}_i^h\left(\hat{\gamma}\right)^{-1}\hat{h}_i^g\left(\hat{\gamma}\right)-\frac{1}{n}\sum_{i=1}^{n}I_i\hat{G}_i^h\left(\hat{\gamma}\right)'\hat{V}_i^h\left(\hat{\gamma}\right)^{-1}\hat{h}_i\left(\hat{\gamma}\right). \quad \textbf{(16)}$$

Based on the contrast T_S, the efficient score-encompassing test can be considered as a special case of the moment-encompassing test by setting $\tilde{m}(x_i,z_j,\hat{\beta},\hat{\gamma})=\hat{G}_i^h\left(\hat{\gamma}\right)'\hat{V}_i^h\left(\hat{\gamma}\right)^{-1}h\left(z_j,\hat{\gamma}\right)$. The CEL-based efficient score-encompassing test statistic is defined as

$$S_g=nT_S'\hat{\Phi}_S^{-}T_S, \quad \textbf{(17)}$$

where $\hat{\Phi}_S^{-}$ is a consistent estimator for a g-inverse of $\Phi_S=\mathrm{E}[\psi_i^S(\beta_0,\gamma_*)\psi_i^S(\beta_0,\gamma_*)']$,

$$\psi_i^S\left(\beta,\gamma\right)=-I_i G_i^h(\gamma)'V_i^h\left(\gamma\right)^{-1}J_i^h\left(\beta,\gamma\right)'V_i\left(\beta\right)^{-1}g(z_i,\beta)$$
$$+H_S\left(\beta,\gamma\right)\Delta\psi(x_i,z_i,\beta),$$
$$H_S\left(\beta,\gamma\right)=\mathrm{E}[I_i G_i^h(\gamma)'V_i^h\left(\gamma\right)^{-1}J_i^h\left(\beta,\gamma\right)'V_i\left(\beta\right)^{-1}G_i\left(\beta\right)].$$

Now Δ and $\psi(x_i,z_i,\beta)$ are defined in (13). To obtain a specific form of $\hat{\Phi}_S^{-}$, we need to specify Δ and $\psi\left(x_i,z_i,\beta_0\right)$. Section 2.3.4 provides an example for the case of the CEL estimator. 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 et al. (2004), the asymptotic linear form of $\hat{\beta}_{CEL}$ is written as

$$\sqrt{n}(\hat{\beta}_{CEL}-\beta_0)=-I\left(\beta_0\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_i G_i\left(\beta_0\right)' V_i\left(\beta_0\right)^{-1}g\left(z_i,\beta_0\right)+o_p\left(1\right),$$

where $I(\beta) = \mathrm{E}\left[I_i G_i(\beta)' V_i(\beta)^{-1} G_i(\beta)\right]$. Let

$$\hat{V}_i(\beta) = \sum_{j=1}^{n} w_{ji} g(z_j, \beta) g(z_j, \beta)', \qquad \hat{G}_i(\beta) = \sum_{j=1}^{n} w_{ji} \partial g\left(z_j, \beta\right) / \partial \beta',$$

$$\hat{J}_i(\beta, \gamma)' = \sum_{j=1}^{n} w_{ji} m\left(z_j, \beta, \gamma\right) g(z_j, \beta)', \qquad \hat{J}_i^h(\beta, \gamma)' = \sum_{j=1}^{n} w_{ji} h\left(z_j, \gamma\right) g(z_j, \beta)'.$$

By setting $\Delta = I(\beta_0)^{-1}$ and $\psi(x_i, z_i, \beta_0) = I_i G_i(\beta_0)' V_i(\beta_0)^{-1} g(z_i, \beta_0)$ in (12), (15), and (17), the CEL-based nonnested test statistics are obtained by the following simpler forms:

(i) the moment-encompassing test statistic:

$$M_{g,CEL} = n T_M' \hat{\Phi}_{M,CEL}^{-} T_M, \tag{18}$$

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

(ii) the Cox-type test statistic:

$$C_{g,CEL} = \frac{\sqrt{n} T_C}{\sqrt{\hat{\phi}_{C,CEL}}}, \tag{19}$$

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

(iii) the efficient score-encompassing test statistic:

$$S_{g,CEL} = n T_S' \hat{\Phi}_{S,CEL}^{-} T_S, \tag{20}$$

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

where RSS denotes the residual sum of squares and $\hat{\gamma}$ may be any $\sqrt{n}$-consistent estimator for the pseudo-true value γ_*.

The asymptotic properties obtained in the next section hold for the above test statistics as well. The CEL estimator $\hat{\beta}_{CEL}$ in the above formulas can be replaced with other semiparametric efficient estimators by Newey (1990) and Donald, Imbens, and Newey (2003), for example.

3. ASYMPTOTIC PROPERTIES

3.1. Null Distributions

In this section we derive the asymptotic distributions of the CEL-based nonnested test statistics under the null hypothesis $\mathbf{H}_g$. We impose some assumptions.

Assumption 3.1.

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

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

(iv) Let $K(x) = \Pi_{i=1}^s \kappa(x^{(i)})$, 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) Assume b_n satisfies $b_n \to 0$ as $n \to \infty$ and $b_n = O(n^{-\alpha})$ for some $0 < \alpha < \frac{1}{3s}$.

Assumption 3.2.

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

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

(iii) Assume $g(z, \beta)$ 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 $\mathrm{E}\left[d_1(z)^{\eta}\right] < \infty$ 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 $\mathrm{E}\left[d_2(z)^{\eta_2}\right] < \infty$ for some $\eta_2 \geq 2$.

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

(v) For the moment-encompassing test, Φ_M is nonnull and $\hat{\Phi}_M^- \xrightarrow{p} \Phi_M^-$. For the Cox-type test, ϕ_C is positive. For the efficient score-encompassing test, Φ_S is nonnull and $\hat{\Phi}_S^- \xrightarrow{p} \Phi_S^-$.

(vi) Assume $\inf_{x \in \mathcal{X}_*} \lambda_{\min}(\mathrm{E}[h(z,\gamma_*)h(z,\gamma_*)'|x]) > 0$ and $\sup_{x \in \mathcal{X}_*} \lambda_{\max}(\mathrm{E}[h(z,\gamma_*)h(z,\gamma_*)'|x]) < \infty$.

In Assumption 3.1(i), although x should be continuous, z can be continuous, discrete, or mixed.[11] Assumption 3.1(ii) assumes the asymptotic linear form for $\hat{\beta}$ that implies the asymptotic normality of $\hat{\beta}$. This assumption 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, respectively. Assumption 3.1(iv) assumes that the kernel function should be second-order. Assumption 3.1(v) implies that the bandwidth b_n can vanish arbitrarily slowly.[12] Tripathi and Kitamura (2003) and Kitamura et al. (2004) employ similar assumptions. Assumption 3.2(i)–(iii) are conditions for the moment function $g(z,\beta)$. These are mainly used to derive uniform convergence rates 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 $\tilde{m}(x,z,\beta,\gamma)$. Assumption 3.2(v) is required to obtain nondegenerate limiting distributions of the test statistics. Primitive conditions for the consistency of $\hat{\Phi}_M^-$ and $\hat{\Phi}_S^-$ can be found in Andrews (1987). Assumption 3.2(vi) guarantees that $V_i^h(\hat{\gamma})$ is invertible uniformly on $x_i \in \mathcal{X}_*$ w.p.a.1. Let $\hat{g}_i(\beta) = \sum_{j=1}^n w_{ji} g(z_j,\beta)$. The null distributions of the CEL-based nonnested test statistics are then obtained below.

THEOREM 3.1 (Null distribution).

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

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

(ii) Suppose that Assumptions 3.1 and 3.2(i)–(iii), (v), and (vi) hold. Furthermore, Assumption 3.2(iv) holds for $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $\hat{M}(x_i,\beta,\gamma)' = \{2\hat{h}_i(\gamma) - \hat{J}_i^h(\beta,\gamma)\hat{V}_i(\beta)^{-1}\hat{g}_i(\beta)\}'\hat{V}_i^h(\gamma)^{-1}$, *and* $M(x_i,\beta,\gamma)' = 2\mathrm{E}[h(z_i,\gamma)|x_i]' V_i^h(\gamma)^{-1}$.[13] *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), (v), and (vi) hold. Furthermore, Assumption 3.2(iv) holds for $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $\hat{M}(x_i,\beta,\gamma)' = \hat{G}_i^h(\gamma)'\hat{V}_i^h(\gamma)^{-1}$, *and* $M(x_i,\beta,\gamma)' = G_i^h(\gamma)'V_i^h(\gamma)^{-1}$. *Then under the null hypothesis* $\mathbf{H}_g$,

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

Therefore these nonnested test statistics follow the standard limiting distributions. Compared to Tripathi and Kitamura's (2003) CEL-based specification test statistics, our nonnested test statistics show the parametric convergence rate. Actually, the proof of Theorem 3.1 presented in the Appendix indicates that under the null hypothesis $\mathbf{H}_g$ our nonnested test statistics are asymptotically equivalent to some unconditional moment restriction test statistics. The main effort of the proof is devoted to establishing such asymptotic equivalence results. However, this asymptotic equivalence holds true only under the null and a sequence of local alternative hypotheses; see Section 3.3 for a detailed discussion. For (ii) and (iii) of this theorem, the assumptions on $m(z,\beta,\gamma)$ and $\hat{M}(x,\beta,\gamma)$ 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. Local Power

This section studies local power properties of the CEL-based nonnested tests. 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

$$\mathbf{H}_{gn} : \mathrm{E}[g(z,\beta_{0n})\,|x] = n^{-1/2}\delta_h(x) \tag{21}$$

holds a.s. for some $\delta_h : \mathcal{X} \to R^{d_g}$. The null hypothesis $\mathbf{H}_g$ is satisfied if $\delta_h(x) = 0$.[14] To obtain the local power properties, we impose additional assumptions.

Assumption 3.3.

(i) Assume $\delta_h(x)$ is continuous on $\mathcal{X}$, $\mathrm{E}[\|\delta_h(x)\|^{\zeta}] < \infty$, $\|\beta_{0n} - \beta_0\| \to 0$ as $n \to \infty$, $\beta_0 \in \mathrm{int}(\mathcal{B})$, 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 $d_\beta \times d_\beta$ nonstochastic matrix, $\mathrm{E}[\psi(x,z,\beta_{0n})\,|x] = n^{-1/2}\delta_\psi(x)$, $\delta_\psi(x)$ is continuous on $\mathcal{X}$, and $\mathrm{E}[\|\delta_\psi(x)\|^{\xi}] < \infty$ for some $\xi > 2$.

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

(iii) Assume $\sup_{x\in\mathcal{X}_*}||\hat{M}(x,\hat{\beta},\hat{\gamma}) - M(x,\beta_{0n},\gamma_*)|| \xrightarrow{p} 0$, $\sup_{\beta\in\mathcal{B}_0}M(x,\beta,\gamma_*)$ is uniformly bounded on $\mathcal{X}_*$, $\mathrm{E}[\sup_{\beta\in\mathcal{B},\gamma\in\Gamma}\|m(z,\beta,\gamma)\|^{\zeta_m}] < \infty$ 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 $j = 1,\ldots,d_\beta + d_\gamma$, $\sup_{(\beta,\gamma)\in\mathcal{B}_0\times\Gamma_*}|\partial m^{(i)}(z,\beta,\gamma)/\partial\left(\beta',\gamma'\right)^{(j)}| \leq d_m(z)$ holds a.s. for a real-valued function $d_m(z)$ with $\mathrm{E}\left[d_m(z)^{\eta_m}\right] < \infty$ 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 $\chi_d^2(\upsilon)$ be the noncentral chi-square distribution with the degree of freedom d and the noncentrality parameter υ. The local power properties of the CEL-based nonnested test statistics are obtained below.

THEOREM 3.2 (Local power).

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

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

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

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

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

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

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

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

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

For (ii) and (iii) of Theorem 3.2, we can replace the assumptions on $m(z,\beta,\gamma)$ and $\hat{M}(x,\beta,\gamma)$ 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. Similar to the existing nonnested tests, the local power functions are obtained from the standard noncentral distributions. While the CEL-based specification test by Tripathi and Kitamura (2003) has nontrivial power against local alternatives with a nonparametric rate (i.e., $n^{-1/2}b_n^{-s/4}\delta_h(x)$), our CEL-based nonnested tests have nontrivial power against local alternatives with the parametric rate (i.e., $n^{-1/2}\delta_h(x)$). However, at the cost of the parametric rate, our nonnested tests require additional conditions to guarantee that the noncentrality parameters μ_M, μ_C, and μ_S are nonzero.

The proof of Theorem 3.2 implies that under the local alternative hypothesis $\mathbf{H}_{gn}$ the test statistics M_g, C_g, and S_g are asymptotically equivalent to some unconditional moment restriction test statistics. Thus we can apply the results of Singleton (1985) and Ramalho and Smith (2002) to analyze the local power optimality. We can show that the nonnested tests defined by M_g, C_g, and S_g have asymptotic local optimal power against local alternatives (or choices of $\delta_h(x)$ and $\delta_\psi(x)$) such that $\mu_M = \Phi_M a$, $\mu_C = \phi_C a$, and $\mu_S = \Phi_S a$, respectively, for some nonzero vector or constant a.

3.3. Global Power

We now analyze the global power properties of the CEL-based nonnested tests under the alternative hypothesis $\mathbf{H}_h$. Assume that under $\mathbf{H}_h$ the estimators $\hat{\beta}$ and $\hat{\gamma}$ converge in probability to the pseudo-true values β_* and γ_0, respectively. Define

$$\lambda_*^g(x,\beta) = \underset{\lambda \in R^{d_g}}{\arg\max} \, \mathrm{E}\left[\log\left(1+\lambda' g(z,\beta)\right)|x\right], \tag{22}$$

which is interpreted as the pseudo-true value of the Lagrange multiplier $\lambda_i^g(\beta)$. From Kitamura (2003), we can show that $\max_{i\in\{i:x_i\in\mathcal{X}_*,1\le i\le n\}}||\lambda_i^g(\hat{\beta}) - \lambda_*^g(x_i, \beta_*)|| \xrightarrow{p} 0$ under $\mathbf{H}_h$. Note that under $\mathbf{H}_h$, $\lambda_*^g(x,\beta_*)$ is generally nonzero. Let

$$J_{i*}(\beta,\gamma)' = \mathrm{E}\left[\left.\frac{m(z_i,\beta,\gamma)\, g(z_i,\beta)'}{1+\lambda_*^g(x_i,\beta)' g(z_i,\beta)}\right| x_i\right],$$

$$J_{i*}^h(\beta,\gamma)' = \mathrm{E}\left[\left.\frac{h(z_i,\gamma)\, g(z_i,\beta)'}{1+\lambda_*^g(x_i,\beta)' g(z_i,\beta)}\right| x_i\right].$$

Let $\mathcal{B}_*$ and Γ_0 be neighborhoods around β_* and γ_0, respectively. The global power properties are obtained below.

THEOREM 3.3 (Global power).

(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(i)–(iv) hold. Furthermore, assume that the probability limit of $\hat{\Phi}_M$ under $\mathbf{H}_h$ (denoted Φ_{hM}) is nonnull. 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

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

(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) and (vi) hold. Furthermore, assume that the probability limit of $\hat{\phi}_C$ under $\mathbf{H}_h$ (denoted ϕ_{hC}) is positive, and Assumption 3.2(iv) holds for $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $\hat{M}(x_i,\beta,\gamma) =$

$\sum_{j=1}^{n} w_{ji} \frac{h(z_j,\gamma)}{1+\lambda_i^g(\beta)' g(z_j,\beta)}$, *and* $M(x_i,\beta,\gamma) = \mathrm{E}\left[\frac{h(z_i,\gamma)}{1+\lambda_*^g(x_i,\beta)' g(z_i,\beta)} \middle| x_i \right]$. *Then under the alternative hypothesis* $\mathbf{H}_h$, *the CEL-based Cox-type test by* C_g *is consistent if* $\mu_{hC}/\sqrt{\phi_{hC}} \neq 0$, *where*

$$\mu_{hC} = \mathrm{E}\left[I_i \mathrm{E}\left[\frac{h(z_i,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)' g(z_i,\beta_*)} \middle| x_i \right]' V_i^h(\gamma_0)^{-1} \right.$$

$$\left. \times \mathrm{E}\left[\frac{h(z_i,\gamma_0)}{1+\lambda_*^g(x_i,\beta_*)' g(z_i,\beta_*)} \middle| x_i \right] \right].$$

(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) and (vi) hold. Furthermore, assume that the probability limit of* $\hat{\Phi}_S$ *under* $\mathbf{H}_h$ *(denoted* Φ_{hS}*) is nonnull, and Assumption 3.2(iv) holds for* $m(z_i,\beta,\gamma) = h(z_i,\gamma)$, $\hat{M}(x_i,\beta,\gamma)' = \hat{G}_i^h(\gamma)' \hat{V}_i^h(\gamma)^{-1}$, *and* $M(x_i,\beta,\gamma)' = G_i^h(\gamma)' V_i^h(\gamma)^{-1}$. *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*

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

The noncentrality parameters μ_{hM}, μ_{hC}, and μ_{hS} depend on $\lambda_*^g(x_i,\beta_*)$, the probability limit of the Lagrange multiplier $\lambda_i^g(\hat{\beta})$. Since $\lambda_i^g(\hat{\beta})$ does not converge to 0 under $\mathbf{H}_h$ in general, the asymptotic relation $\lambda_i^g(\hat{\beta}) = \hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) + o_p(1)$ (see Lemma A.4) no longer holds under $\mathbf{H}_h$. Thus it is generally difficult to obtain an explicit form (or approximation) for those noncentrality parameters in terms of the moment function $g(z_i,\beta_*)$ instead of $\lambda_*^g(x_i,\beta_*)$.

The implications of Theorem 3.3 are as follows: First, consider the moment-encompassing and efficient score-encompassing tests. Under the alternative hypothesis $\mathbf{H}_h$, these two tests have nontrivial asymptotic global power as far as $\mu_{hM}' \Phi_{hM}^- \mu_{hM} \neq 0$ and $\mu_{hS}' \Phi_{hS}^- \mu_{hS} \neq 0$. However, even when $\lambda_*^g(x_i,\beta_*)$ is nonzero it is possible that the marginal measure for x_i satisfies $\mu_{hM} = 0$ or $\mu_{hS} = 0$, which in turn yields 0 noncentrality parameters, i.e., $\mu_{hM}' \Phi_{hM}^- \mu_{hM} = 0$ or $\mu_{hS}' \Phi_{hS}^- \mu_{hS} = 0$. This drawback, called the implicit null hypothesis, is common in the nonnested hypotheses and encompassing testing literature. Using the notation of Section 2.1, this inconsistency problem can be interpreted as the discrepancy between $\mathcal{H}_{z|x}$ (the set of conditional measures satisfying $\mathbf{H}_h$) and $\mathcal{H}_{z|x}^M = \left\{ (\mu_{z|x})_{x\in\mathcal{X}} \in \mathcal{M}_{z|x} : \mu_{hM}' \Phi_{hM}^- \mu_{hM} = 0 \text{ for some } \mu_x \right\}$ or $\mathcal{H}_{z|x}^S = \left\{ (\mu_{z|x})_{x\in\mathcal{X}} \in \mathcal{M}_{z|x} : \mu_{hS}' \Phi_{hS}^- \mu_{hS} = 0 \text{ for some } \mu_x \right\}$.

Next we discuss the global power property of the Cox-type test. Since ϕ_{hC} is finite under very mild conditions, we focus on the conditions for $\mu_{hC} \neq 0$. If $V_i^h(\gamma_0) = \mathrm{E}\left[h(z_i,\gamma_0) h(z_i,\gamma_0)' | x_i \right]$ is positive definite (a.s. x_i) under $\mathbf{H}_h$, a sufficient condition for $\mu_{hC} \neq 0$ is

$$\mathrm{E}\left[\frac{h(z,\gamma_0)}{1+\lambda_*^g(x,\beta_*)' g(z,\beta_*)} \middle| x \right] = \int \frac{h(z,\gamma_0)}{1+\lambda_*^g(x,\beta_*)' g(z,\beta_*)} d\mu_{z|x}^0$$
$$= \int h(z,\gamma_0)\, dP_{z|x}^* \neq 0, \tag{23}$$

for some subset of $\mathcal{X}$, where the conditional measure $(P_{z|x}^*)_{x\in\mathcal{X}}$ is defined by

$$\frac{dP_{z|x}^*}{d\mu_{z|x}^0} = \frac{1}{1+\lambda_*^g(x,\beta_*)' g(z,\beta_*)}.$$

Suppose that β_* is the pseudo-true value of the CEL estimator $\hat{\beta}_{CEL}$ and the support of $g(z,\beta)$ is bounded a.s. for all $\beta \in \mathcal{B}$. Then Kitamura (2003) showed that $(P_{z|x}^*)_{x\in\mathcal{X}}$ becomes the best approximation of the true conditional measure $\mu_{z|x}^0$ to the space of conditional measures $\mathcal{G}_{z|x}$ satisfying $\mathbf{H}_g$ by the conditional relative entropy, and that it satisfies $(P_{z|x}^*)_{x\in\mathcal{X}} \in \mathcal{G}_{z|x}$.[15] For simplicity, assume that $\mathbf{H}_g$ and $\mathbf{H}_h$ are *strictly* nonnested, i.e., $\mathcal{G}_{z|x} \cap \mathcal{H}_{z|x} = \phi$.[16] Then $(P_{z|x}^*)_{x\in\mathcal{X}} \in \mathcal{G}_{z|x}$ implies $(P_{z|x}^*)_{x\in\mathcal{X}} \notin \mathcal{H}_{z|x}$ and hence condition (23) holds. Thus the Cox-type test is always consistent in this case. This result is summarized below.

COROLLARY 3.1 (A sufficient condition for the consistency of the Cox-type test). *Suppose that $\mathcal{G}_{z|x} \cap \mathcal{H}_{z|x} = \phi$ and the same assumptions as Theorem 3.3(ii) hold for the CEL estimator $\hat{\beta}_{CEL}$. Furthermore, assume that (i) $\phi_{hC} < \infty$; (ii) $V_i^h(\gamma_0)$ is positive definite (a.s. x_i); and (iii) the support of $g(z,\beta)$ is bounded a.s. for all $\beta \in \mathcal{B}$. Then the Cox test is consistent against $\mathbf{H}_h$.*

Note that this corollary does not require somewhat artificial assumptions such as $\mu_{hM}'\Phi_{hM}^{-}\mu_{hM} \neq 0$ and $\mu_{hS}'\Phi_{hS}^{-}\mu_{hS} \neq 0$ in the moment-encompassing and efficient score-encompassing tests, respectively. Although the bounded support assumption for $g(z,\beta)$ can be restrictive in some contexts, this assumption is very easy to check.[17] Another important requirement is that we must use the CEL estimator $\hat{\beta}_{CEL}$ to obtain the above corollary. If we employ a different estimator, its pseudo-true value β_* may differ from that of the CEL estimator, and the result of Kitamura (2003) is not applicable.[18]

3.4. Comparison with Unconditional Moment-Based Tests

This section compares the proposed (*conditional* moment-based) nonnested tests with the *unconditional* moment-based tests. Under the null hypothesis $\mathbf{H}_g$: $(\mu_{z|x}^0)_{x\in\mathcal{X}} \in \mathcal{G}_{z|x}$, the statistics T_M, T_C, and T_S can be respectively written as (see (A.10), (A.16), and (A.18) in the Appendix)

$$T_M = -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}(x_i,\hat{\beta},\hat{\gamma})'\hat{J}_i(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) + o_p(n^{-1/2}),$$

$$T_C = -\frac{1}{n}\sum_{i=1}^{n} I_i \left\{2\hat{h}_i(\hat{\gamma}) - \hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta})\right\}'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{J}_i^h(\hat{\beta},\hat{\gamma})' \times \hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) + o_p(n^{-1/2}), \tag{24}$$

$$T_S = -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) + o_p(n^{-1/2}).$$

Based on the above relationships, we may consider testing for the following unconditional moment restrictions:

$$\mathbf{H}_g^U : \mathrm{E}[I_i Q_a(x_i,\beta_0,\gamma_*)\, g(z_i,\beta_0)] = 0, \qquad a = M, C, S, \tag{25}$$

where

$$Q_M(x_i,\beta,\gamma) = M(x_i,\beta,\gamma)'J_i(\beta,\gamma)'V_i(\beta)^{-1},$$

$$Q_C(x_i,\beta,\gamma) = \left\{2\mathrm{E}[h(z_i,\gamma)|x] - J_i^h(\beta,\gamma)'V_i(\beta)^{-1}\mathrm{E}[g(z_i,\beta)|x]\right\}' \times V_i^h(\gamma)^{-1}J_i^h(\beta,\gamma)'V_i(\beta)^{-1},$$

$$Q_S(x_i,\beta,\gamma) = G_i^h(\gamma)'V_i^h(\gamma)^{-1}J_i^h(\beta,\gamma)'V_i(\beta)^{-1}.$$

From Smith (1997), these unconditional moment restrictions can be tested by using the sample analogs $T_a^U = n^{-1}\sum_{i=1}^{n}\hat{Q}_a(x_i,\hat{\beta},\hat{\gamma})g(z_i,\hat{\beta})$ for $a = M$, C, and S, where $\hat{Q}_a(x_i,\beta,\gamma)$ is a nonparametric estimator for $Q_a(x,\beta,\gamma)$. We can show that under the original null hypothesis $\mathbf{H}_g : (\mu^0_{z|x})_{x\in\mathcal{X}} \in \mathcal{G}_{z|x}$ and the local alternative hypothesis $\mathbf{H}_{gn}$, T_a and T_a^U are asymptotically equivalent, i.e.,

$$n^{1/2}T_a = n^{1/2}T_a^U + o_p(1), \qquad a = M, C, S \quad \text{under } \mathbf{H}_g \text{ and } \mathbf{H}_{gn}.$$

In other words, we can construct unconditional moment restrictions whose test statistics have the same asymptotic properties under the null and and local alternative hypotheses as our nonnested test statistics. However, an inspection of the proofs in the Appendix shows that such an asymptotic equivalence generally does not hold under $\mathbf{H}_h$ because the remainder terms (in Taylor expansions to get (A.10), (A.16), and (A.18)) generally do not vanish under $\mathbf{H}_h$. Thus, under $\mathbf{H}_h$, T_a and T_a^U have different noncentrality parameters, i.e.,

$$T_a \xrightarrow{p} \mu_{ha},$$

$$T_a^U \xrightarrow{p} \mathrm{E}[Q_a(x_i,\beta_*,\gamma_0)\, g(z_i,\beta_*)],$$

where μ_{ha} is defined in Theorem 3.3. In this sense, our tests cannot be regarded as special cases of unconditional moment tests. Section 4 compares those global power properties in finite samples.

We do not claim that our tests are more powerful than unconditional moment-based tests against all alternative conditional moment restrictions. Indeed, our tests can be useful complements to the unconditional moment-based tests. To see more specifically, denote

$$\mathcal{G}_{z|x}^{a} = \cup_{\beta \in \mathcal{B}} \left\{ (\mu_{z|x})_{x \in \mathcal{X}} \in \mathcal{M}_{z|x} : \int \int Q_a(x, \beta, \gamma_*) g(z, \beta) \, d\mu_{z|x} \, d\mu_x = 0 \text{ for some } \mu_x \right\}.$$

Note that $\mathcal{G}_{z|x} \subset \mathcal{G}_{z|x}^{a}$. Suppose $(\mu_{z|x}^{0})_{x \in \mathcal{X}} \in \mathcal{G}_{z|x}^{a} \setminus \mathcal{G}_{z|x}$, i.e., the original null hypothesis $\mathbf{H}_g$ is violated but $\mathbf{H}_g^U$ holds. Then the test based on T_a^U is inconsistent since $\mathrm{E}[Q_a(x_i, \beta_*, \gamma_0) g(z_i, \beta_*)] = 0$, whereas our test based on T_a can be consistent as far as $\mu_{ha} \neq 0$. We also analyze this situation in simulation studies in the next section.

Finally, it should be noted that the unconditional counterpart of the Cox-type statistic (i.e., T_C^U) does not yield the consistency result as in Corollary 3.1 under the same assumptions.

4. SIMULATIONS

This section examines the finite sample properties of our tests against some of the existing nonnested tests using Monte Carlo simulations.

4.1. Experimental Designs

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

$$\mathbf{H}_g : y_i = \beta_{01} + \beta_{02} x_{1i} + u_{gi} \tag{26}$$

$$\mathbf{H}_h : y_i = \gamma_{01} + \gamma_{02} x_{2i} + u_{hi},$$

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 (26) 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 $i = 1, \ldots, n$,

$$\mathbf{H}_g : y_i = \beta_0 x_i + u_{gi} \tag{27}$$

$$\mathbf{H}_h : y_i = \gamma_0 x_i^3 + u_{hi},$$

where $\{x_i\}$, $\{u_{gi}\}$, and $\{u_{hi}\}$ are i.i.d. $N(0,1)$ and $\beta_0 = \gamma_0 = 1$. The hypotheses (27) correspond to (1) with $g(z,\beta_0) = y - \beta_0 x$ and $h(z,\gamma_0) = y - \gamma_0 x^3$, where $z = (y,x)'$.

To calculate our test statistics, we use the CEL estimators $\hat{\beta}_{CEL}$ and $\hat{\gamma}_{CEL}$ to estimate β_0 and γ_0, respectively. The moment-encompassing, Cox-type, and efficient score-encompassing test statistics used in our simulations are then defined by (18), (19), and (20), respectively, with $I_i = 1$ (i.e., no trimming), $\hat{M}(x,\beta,\gamma) = 1$, and $m(z,\beta,\gamma) = h(z,\gamma)$.

To compare our tests with those based on unconditional moments, we consider the following unconditional versions of our test statistics using the first-order expansions (24), under the null hypothesis $\mathbf{H}_g$:

$$M_g^U = nT_M^{U\prime}\hat{\Phi}_{M,CEL}^{-}T_M^U, \tag{28}$$

$$C_g^U = n\left(T_C^U\right)^2/\hat{\phi}_{C,CEL}, \tag{29}$$

$$S_g^U = nT_S^{U\prime}\hat{\Phi}_{S,CEL}^{-}T_S^U, \tag{30}$$

where

$$T_M^U = \frac{1}{n}\sum_{i=1}^n I_i\hat{M}(x_i,\hat{\beta},\hat{\gamma})'\hat{J}_i(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}g(z_i,\hat{\beta}), \tag{31}$$

$$T_C^U = \frac{1}{n}\sum_{i=1}^n I_i\{2\hat{h}_i(\hat{\gamma}) - \hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta})\}'\hat{V}_i^h\left(\hat{\gamma}\right)^{-1}\hat{J}_i^h(\hat{\beta},\hat{\gamma})' \times\hat{V}_i(\hat{\beta})^{-1}g(z_i,\hat{\beta}), \tag{32}$$

$$T_S^U = \frac{1}{n}\sum_{i=1}^n I_i\hat{G}_i^h(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}\hat{J}_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}g(z_i,\hat{\beta}), \tag{33}$$

$\hat{\beta} = \hat{\beta}_{CEL}$, $\hat{\gamma} = \hat{\gamma}_{CEL}$, and $\hat{\Phi}_{M,CEL}$, $\hat{\phi}_{C,CEL}$, and $\hat{\Phi}_{S,CEL}$ are defined in (18)–(20), respectively. Note that the test statistics (28)–(30) can be viewed as the ones for testing the unconditional moment restrictions in (25). Under the null hypothesis $\mathbf{H}_g$, the test statistics (28)–(30) have the chi-square limiting distributions. Under the alternative hypothesis $\mathbf{H}_h$, however, they are not asymptotically equivalent to our tests and hence are expected to have different power performances from ours.

As other benchmarks for our simulation experiments, we consider the nonnested tests of Singleton (1985, eqn. (33), p. 404), labeled S, and Ramalho and Smith (2002, Simplified Cox test in eqn. (4.4), p. 108), labeled SC, respectively. We compute S and SC from the following unconditional moment restrictions implied by (26) and (27): for Design I,

$$\mathbf{H}_g^U : \mathrm{E}\left[(1, x_{1i}, x_{2i})'(y_i - \beta_{01} - \beta_{02}x_{1i})\right] = 0, \tag{34}$$

$$\mathbf{H}_h^U : \mathrm{E}\left[(1, x_{1i}, x_{2i})'(y_i - \gamma_{01} - \gamma_{02}x_{2i})\right] = 0,$$

and, for Design II,

$$\mathbf{H}_g^U : \mathrm{E}\left[(1, x_i)'(y_i - \beta_0 x_i)\right] = 0, \tag{35}$$

$$\mathbf{H}_h^U : \mathrm{E}\left[(1, x_i^3)'\left(y_i - \gamma_0 x_i^3\right)\right] = 0.$$

Furthermore, we consider the overidentifying restriction test of Hansen (1982), labeled J, that tests the validity of $\mathbf{H}_g^U$ in (34) and (35) against general alternatives. For tests S, SC, and J, the parameters β_0 and γ_0 are estimated by the GMM.

We consider two sample sizes $n \in \{100, 200\}$ and fix the number R of Monte Carlo repetitions to be 1,000. We use the Gaussian kernel for our CEL-based tests M_g, C_g, and S_g (and their unconditional moment versions M_g^U, C_g^U, and S_g^U). For the bandwidth b_n, we consider $b_n \in \{0.1, 0.2, \ldots, 1.5\}$ 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 error is 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 M_g, C_g, and S_g have reasonable size performances if the bandwidth is in a suitable range and the performance improves generally as n increases. The size performances of the unconditional moment tests M_g^U, C_g^U, and S_g^U are similar but appear to be less sensitive to the choice of the bandwidth b_n. On the other hand, the competitors J and SC have little size distortion, though the Singleton's test S underrejects in many cases we consider.

In terms of size-corrected powers, the efficient score-encompassing test S_g dominates the other tests in Design I. When $c_0 = 1$, the test S, which is known to have an optimality property against some local alternatives, also has very good (size-corrected) power performance in Design I. However, when $c_0 = 2$, the power performance of S deteriorates and is significantly dominated by that of S_g. On the other hand, the powers of the unconditional moment tests M_g^U, C_g^U, and S_g^U are quite different from those of our tests M_g, C_g, and S_g and are generally lower than the latter. For the former unconditional moment-based tests, powers depend sensitively on the choice of the weighting matrix $Q_a(x, \beta, \gamma)$ and c_0. In particular, M_g^U appears to be inconsistent against $\mathbf{H}_h$.

To explain these findings intuitively, consider an estimator $\hat{\beta}$ of $\beta_0 = (\beta_{01}, \beta_{02})'$ that converges to the pseudo-true value $\beta_* = (1, c_0/(1 + c_0^2))'$ under the alternative hypothesis $\mathbf{H}_h$ in (26).[19] This implies that the sample analog of the unconditional

TABLE 1. Design I, $c_0 = 1$: Estimated sizes and powers of tests with nominal size of 5%

Test	b_n	$n = 100$ Size	A-P	S-P	$n = 200$ Size	A-P	S-P
	0.7	.170 (.001)	.778 (.011)	.528 (.118)	.135 (.001)	.936 (.024)	.878 (.174)
	0.8	.100 (.002)	.777 (.025)	.678 (.143)	.090 (.001)	.947 (.046)	.923 (.208)
M_g (M_g^U)	0.9	.064 (.004)	.775 (.043)	.749 (.156)	.060 (.003)	.966 (.077)	.961 (.262)
	1.0	.046 (.006)	.781 (.060)	.796 (.173)	.029 (.004)	.960 (.107)	.969 (.297)
	1.1	.031 (.009)	.747 (.080)	.791 (.197)	.026 (.004)	.958 (.141)	.971 (.338)
	0.7	.078 (.028)	.529 (.112)	.399 (.137)	.040 (.035)	.644 (.651)	.703 (.651)
	0.8	.033 (.022)	.417 (.268)	.581 (.314)	.024 (.029)	.513 (.830)	.848 (.830)
C_g (C_g^U)	0.9	.011 (.021)	.306 (.414)	.684 (.475)	.008 (.025)	.373 (.907)	.895 (.907)
	1.0	.006 (.026)	.223 (.533)	.735 (.580)	.001 (.030)	.251 (.954)	.904 (.954)
	1.1	.001 (.034)	.125 (.641)	.748 (.693)	.000 (.034)	.145 (.969)	.920 (.969)
	0.7	.230 (.176)	.949 (.915)	.823 (.684)	.096 (.120)	.986 (.992)	.978 (.978)
	0.8	.150 (.179)	.959 (.937)	.905 (.706)	.057 (.118)	.993 (.996)	.992 (.983)
S_g (S_g^U)	0.9	.101 (.198)	.971 (.954)	.945 (.746)	.032 (.133)	.993 (.996)	.995 (.986)
	1.0	.079 (.208)	.976 (.960)	.971 (.783)	.017 (.149)	.996 (.998)	.999 (.982)
	1.1	.066 (.219)	.976 (.965)	.973 (.790)	.010 (.157)	.996 (.997)	.997 (.985)
J		.041	.926	.934	.052	.999	.998
S		.008	.911	.972	.007	.997	1.00
SC		.055	.935	.934	.054	.999	.999

Notes: Tests M_g, C_g, and S_g refer to the moment-encompassing, Cox-type, and efficient score-encompassing tests, repectively. Also, M_g^U, C_g^U, and S_g^U refer to the unconditional versions of tests M_g, C_g, and S_g, respectively. 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.

expectation in (34) converges to

$$\frac{1}{n}\sum_{i=1}^{n}\left[(1, x_{1i}, x_{2i})'(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{1i})\right] \xrightarrow{p} \left(0, 0, \frac{1}{1+c_0^2}\right)'. \quad \textbf{(36)}$$

Then the sample averages in (31)–(33) converge to

$$T_M^U \xrightarrow{p} 0, \qquad T_C^U \xrightarrow{p} -\frac{1}{4(1+c_0^2)^2}, \qquad T_S^U \xrightarrow{p} \left(0, -\frac{1}{4(1+c_0^2)}\right)'. \quad \textbf{(37)}$$

Therefore, as c_0 increases, the limits in (36) and (37) degenerate to 0, and the tests based on T_C^U and T_S^U have lower power. Also, the result in (37) confirms our simulation finding that M_g^U is inconsistent against $\mathbf{H}_h$.

Table 3 reports the simulation results for Design II. As in Design I, all of the tests considered have reasonable size performances and M_g^U, C_g^U, and S_g^U behave quite differently from our tests in terms of powers. In this design, we

TABLE 2. Design I, $c_0 = 2$: Estimated sizes and powers of tests with nominal size of 5%

		$n=100$			$n=200$		
Test	b_n	Size	A-P	S-P	Size	A-P	S-P
	0.7	.176 (.001)	.537 (.000)	.262 (.018)	.138 (.001)	.752 (.000))	.517 (.022)
	0.8	.104 (.001)	.500 (.001)	.357 (.022)	.084 (.002)	.745 (.000)	.644 (.033)
M_g (M_g^U)	0.9	.071 (.003)	.460 (.005)	.415 (.025)	.057 (.004)	.732 (.006)	.711 (.047)
	1.0	.039 (.003)	.442 (.008)	.473 (.031)	.038 (.007)	.716 (.010)	.748 (.065)
	1.1	.028 (.008)	.407 (.014)	.476 (.039)	.035 (.008)	.694 (.016)	.740 (.088)
	0.7	.069 (.021)	.293 (.008)	.221 (.028)	.039 (.030)	.264 (.054)	.327 (.071)
	0.8	.034 (.022)	.186 (.010)	.309 (.029)	.022 (.030)	.167 (.160)	.467 (.186)
C_g (C_g^U)	0.9	.016 (.023)	.111 (.043)	.388 (.058)	.008 (.033)	.086 (.292)	.585 (.320)
	1.0	.003 (.032)	.055 (.085)	.408 (.109)	.001 (.039)	.038 (.440)	.625 (.455)
	1.1	.002 (.035)	.023 (.145)	.433 (.167)	.000 (.044)	.015 (.529)	.639 (.537)
	0.7	.234 (.192)	.930 (.804)	.807 (.565)	.108 (.136)	.982 (.923)	.971 (.834)
	0.8	.147 (.194)	.939 (.834)	.876 (.592)	.060 (.137)	.983 (.936)	.981 (.841)
S_g (S_g^U)	0.9	.097 (.197)	.948 (.867)	.908 (.652)	.030 (.147)	.984 (.949)	.985 (.867)
	1.0	.069 (.205)	.941 (.891)	.931 (.680)	.019 (.156)	.985 (.951)	.994 (.882)
	1.1	.057 (.223)	.942 (.907)	.938 (.713)	.012 (.164)	.987 (.959)	.993 (.889)
J		.044	.563	.572	.056	.868	.865
S		.021	.554	.666	.023	.863	.906
SC		.055	.589	.582	.053	.878	.876

Notes: Tests M_g, C_g, and S_g refer to the moment-encompassing, Cox-type, and efficient score-encompassing tests, repectively. Also, M_g^U, C_g^U, and S_g^U refer to the unconditional versions of tests M_g, C_g, and S_g, respectively. 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.

expect that the tests based on the unconditional moments in (35) and T_M^U will be inconsistent. Consider an estimator $\hat{\beta}$ of β_0 that converges to the pseudo-true value $\beta_* = 3$ under the alternative hypothesis $\mathbf{H}_h$ in (27). This condition is satisfied for the GMM estimator. Then the sample analog of (35) and T_M^U converge to

$$\frac{1}{n}\sum_{i=}^{n}\left[(1,x_i)'(y_i-\hat{\beta}x_i)\right] \xrightarrow{p} \mathrm{E}^h\left[(1,x_i)'(y_i-\beta_* x_i)\right] = (0,0)', \tag{38}$$

$$T_M^U \xrightarrow{p} 4\times \mathrm{E}^h\left[\frac{r(x_i)}{r(x_i)^2+4}\right] = 0, \tag{39}$$

using the fact that $x_i \sim N(0,1)$ and $r(x) = x^3 - 3x$ is an odd function, where E^h is the expectation taken under $\mathbf{H}_h$. On the other hand, C_g^U and S_g^U are expected to

TABLE 3. Design II: Estimated sizes and powers of tests with nominal size of 5%

Test	b_n	$n = 100$ Size	$n = 100$ A-P	$n = 100$ S-P	$n = 200$ Size	$n = 200$ A-P	$n = 200$ S-P
	0.1	.062 (.051)	.624 (.133)	.502 (.133)	.043 (.053)	.635 (.141)	.696 (.140)
	0.2	.018 (.047)	.604 (.137)	.913 (.144)	.015 (.057)	.608 (.159)	.959 (.153)
M_g (M_g^U)	0.3	.009 (.046)	.538 (.142)	.967 (.148)	.008 (.055)	.568 (.165)	.984 (.155)
	0.4	.007 (.047)	.452 (.134)	.984 (.138)	.004 (.055)	.471 (.154)	.981 (.150)
	0.5	.004 (.047)	.348 (.127)	.977 (.131)	.005 (.054)	.336 (.147)	.985 (.142)
	0.1	.185 (.125)	.701 (.067)	.428 (.046)	.129 (.100)	.698 (.787)	.454 (.763)
	0.2	.070 (.031)	.674 (.576)	.639 (.599)	.055 (.035)	.700 (.971)	.675 (.976)
C_g (C_g^U)	0.3	.030 (.026)	.685 (.802)	.774 (.835)	.022 (.028)	.715 (.993)	.829 (.993)
	0.4	.019 (.027)	.680 (.880)	.835 (.898)	.014 (.031)	.733 (.996)	.878 (.999)
	0.5	.012 (.030)	.662 (.911)	.845 (.928)	.008 (.030)	.727 (.999)	.879 (.999)
	0.1	.095 (.064)	.292 (.027)	.140 (.021)	.078 (.061)	.334 (.038)	.234 (.030)
	0.2	.053 (.067)	.356 (.057)	.339 (.042)	.040 (.046)	.414 (.118)	.486 (.121)
S_g (S_g^U)	0.3	.034 (.067)	.412 (.136)	.589 (.106)	.027 (.050)	.427 (.259)	.729 (.259)
	0.4	.020 (.064)	.433 (.253)	.791 (.232)	.017 (.057)	.489 (.417)	.837 (.393)
	0.5	.013 (.071)	.467 (.392)	.871 (.348)	.009 (.064)	.522 (.640)	.901 (.631)
J		.048	.027	.027	.053	.040	.034
S		.011	.021	.158	.009	.031	.172
SC		.008	.075	.174	.004	.070	.165

Notes: Tests M_g, C_g, and S_g refer to the moment encompassing, Cox-type, and efficient score encompassing tests, repectively. Also, M_g^U, C_g^U, and S_g^U refer to the unconditional versions of tests M_g, C_g, and S_g, respectively. 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.

be consistent since

$$T_C^U \xrightarrow{p} -4 \times \mathrm{E}^h \left[\left(\frac{r(x_i)}{r(x_i)^2 + 4} \right)^2 \right] \neq 0, \qquad \textbf{(40)}$$

$$T_S^U \xrightarrow{p} -\mathrm{E}^h \left[\frac{x_i^3 r(x_i)}{r(x_i)^2 + 4} \right] \neq 0. \qquad \textbf{(41)}$$

This is precisely what happened to the powers of these tests in Design II. On the other hand, our tests M_g, C_g, and S_g have nontrivial powers in all of the cases we considered. Among our tests, M_g and C_g appear to have better (size-corrected) power performances than S_g in Design II.

5. CONCLUSION

We propose three types of nonnested tests for competing conditional moment restriction models: the moment encompassing, Cox-type, and efficient score-encompassing tests. The test statistics are based on the conditional probabilities implied by conditional empirical likelihood. We investigate the asymptotic properties of the test statistics under the null, local alternative, and global alternative hypotheses. Our tests have power properties that are very distinct from some of the existing unconditional moment-based tests and are powerful against global alternatives that cannot be detected by the latter tests. In particular, if the support of the moment function is bounded and a mild regularity condition holds, we show that the Cox-type test is consistent against all departures from the null hypothesis toward the strictly nonnested alternative hypothesis. Simulation results illustrate that our tests have reasonable finite sample properties and, in some cases, dominate some of the existing tests based on unconditional moment restrictions. Although this paper focuses on the moment-encompassing, Cox-type, and efficient score encompassing tests, it is interesting to consider a general class of test statistics defined in the form of $T_a = n^{-1}\sum_{i=1}^{n} Q_a(x_i,\hat{\beta},\hat{\gamma})\hat{g}_i(\hat{\beta})$ as seen in Smith (1997) for nonnested unconditional hypotheses testing, and investigate more general properties and comparisons of the test statistics. We would like to leave this extension for future research.

NOTES

1. Kitamura et al.'s (2004) *smoothed* empirical likelihood and Zhang and Gijbels' (2003) *sieve* empirical likelihood are quite similar concepts. To avoid confusion, we follow Kitamura (2003) and adopt 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 nonnested and encompassing tests.

3. GEL is originally proposed by Smith (1997), and its higher order properties are investigated by Newey and Smith (2004).

4. See Owen (2001) for a comprehensive review of the empirical likelihood approach. See also Kitamura (2006) for a current update of the literature.

5. The hypotheses $\mathbf{H}_g$ and $\mathbf{H}_h$ should be restrictions on the same conditional distribution $z|x$. If the conditioning variables are different, i.e., $\mathbf{H}_g : \mathrm{E}[g(z,\beta_0)|x_g] = 0$ (a.s. x_g) and $\mathbf{H}_h : \mathrm{E}[h(z,\gamma_0)|x_h] = 0$ (a.s. x_h), our approach does not work. However, if the hypotheses are written as $\mathbf{H}_g : \mathrm{E}[g(z,\beta_0)|x_g,x_h] = 0$ (a.s. x_g,x_h) and $\mathbf{H}_h : \mathrm{E}[h(z,\gamma_0)|x_g,x_h] = 0$ (a.s. x_g,x_h), our approach is applicable. See the simulation Design I in Section 4 for an example.

6. Note that μ_i^g satisfies $\mu_i^g = 1$ for $i = 1,\ldots,n$.

7. If the trimming term is replaced with $I\{x_i \in \mathcal{X}_n\}$, where $\mathcal{X}_n$ converges to $\mathcal{X}$ in an adequate manner, then the CEL estimator is asymptotically efficient. Since this paper is concerned with specification testing, we consider the fixed trimming term I_i.

8. Under misspecification, the solution $p_{ji}^g(\hat{\beta})$ may not exist. In order for the solution to exist w.p.a.1, we assume that the origin is contained in the convex hull of $\{g(z_1,\hat{\beta}),\ldots,g(z_n,\hat{\beta})\}$ w.p.a.1. From Tripathi and Kitamura (2003, pp. 2067–2068), this assumption is satisfied if $\mathrm{E}[g(z,\beta_*)g(z,\beta_*)']$ has full rank and $\Pr\{z : \xi' g(z,\beta) = 0\} = 0$ for all fixed unit vectors ξ and all β in some compact neighborhood around β_*.

9. Although we may focus on the contrast of CEL based on $\hat{p}_{ji}^h(\hat{\gamma})$;

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

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 simplicity we analyze the contrast of the Euclidean likelihoods.

10. 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

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

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

11. We conjecture that it would be possible to allow discrete regressors by applying the trimming argument of Andrews (1995) and Kitamura et al. (2004). In this case we need to redefine the CEL weight as $w_{ji} = K\left(\frac{x_j^c - x_i^c}{b_n}\right) I\{x_i^d = x_j^d\} / \left(\sum_{j=1}^n K\left(\frac{x_j^c - x_i^c}{b_n}\right) I\{x_i^d = x_j^d\}\right)$, where x_j^c are continuous regressors and x_i^d are discrete ones.

12. A technical intuition for this point can be explained as follows: Asymptotic expansions of our test statistics are written as U-statistics with zero mean under the null hypothesis. Therefore, when we apply the U-statistic argument of Kitamura et al. (2004, Lem. B.2) or Powell, Stock, and Stoker (1989, Sect. 3.2) to derive the asymptotic normality of the test statistics, we can neglect bias terms in nonparametric estimation, which typically require us to impose some assumption on the lower bound of α.

13. Since $\hat{g}_i(\hat{\beta}) \xrightarrow{p} \mathrm{E}[g(z_i,\beta_0)|x_i] = 0$ uniformly on $x_i \in \mathcal{X}_*$ under $\mathbf{H}_g$ (Lemma A.4), the second term of $\hat{M}(x_i,\hat{\beta},\hat{\gamma})' = 2\hat{h}_i(\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1} - \hat{g}_i(\hat{\beta})'\hat{V}_i(\hat{\beta})^{-1} J_i^h(\hat{\beta},\hat{\gamma})'\hat{V}_i^h(\hat{\gamma})^{-1}$ converges to 0 uniformly on $x_i \in \mathcal{X}_*$ under $\mathbf{H}_g$ and our assumptions.

14. If the moment functions $g(z,\beta)$ and $h(z,\gamma)$ have the same dimension, another way to formulate the local alternatives in the spirit of Singleton (1985, p. 402) would be

$$\mathbf{H}_{gn}^* : \left(1 - \frac{\eta}{\sqrt{n}}\right) \mathrm{E}\left[g(z,\beta_0)|x\right] + \frac{\eta}{\sqrt{n}} \mathrm{E}\left[h(z,\gamma_0)|x\right] = 0,$$

where $\eta \in R$ is a constant. This case can be treated similarly because $\mathbf{H}_{gn}^*$ now corresponds to $\mathbf{H}_{gn}$ with $\delta_h(x) = \eta\left\{\mathrm{E}\left[g(z,\beta_0)|x\right] - \mathrm{E}\left[h(z,\gamma_0)|x\right]\right\}$ and $\beta_{0n} = \beta_0$.

15. This result is a natural extension of Csiszar's (1975) analysis on the existence of the "I-projection" for unconditional probability measures.

16. Our result can be generalized to partly nonnested models (i.e., $\mathcal{G}_{z|x} \cap \mathcal{H}_{z|x}$ is nonempty). In this case we need to modify the definition of nonnested alternatives to guarantee that $(P_{z|x}^*)_{x\in\mathcal{X}} \notin \mathcal{H}_{z|x}$ holds.

17. By extending the results of Borwein and Lewis (1993) and Csiszár (1995) to the conditional moment set-up, we conjecture that this boundedness assumption can be reasonably weakened.

18. We expect that Corollary 3.1 can be extended to the GEL set-up. To this end, we need to apply a different entropy measure for each member of the GEL criterion function to obtain the best approximation like $(P_{z|x}^*)_{x\in\mathcal{X}}$.

19. For example, the GMM estimator satisfies this condition. It is hard to calculate the pseudo-true value of the CEL estimator because the Lagrange multiplier $\lambda_*^g(x,\beta)$ in (22) does not have an explicit solution. However, in our simulation the values of the GMM and CEL estimates are quite similar under the alternative hypothesis $\mathbf{H}_g$.

REFERENCES

Andrews, D.W.K. (1987) Asymptotic results for generalized Wald tests. *Econometric Theory* 3, 348–358.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Borwein, J.M. & A.S. Lewis (1993) Partially-finite programming in L_1 and the existence of maximum entropy estimates. *Siam Journal of Optimization* 3, 248–267.

Chen, Y. & C. Kuan (2002) The pseudo-true score encompassing test for non-nested hypotheses. *Journal of Econometrics* 106, 271–295.

Cox, D.R. (1961) Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 105–123. University of California Press.

Cox, D.R. (1962) Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B* 24, 406–424.

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

Csiszár, I. (1995) Generalized projections for non-negative functions. *Acta Mathematica Hungarica* 68, 161–185.

Davidson, R. & J. MacKinnon (1981) Several tests for model specification in the presence of alternative hypothesis. *Econometrica* 49, 781–793.

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

Donald, S.G., G.W. Imbens, & W.K. Newey (2003) Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.

Fisher, G. & M. McAleer (1981) Alternative procedures and associated tests of significance for non-nested hypotheses. *Journal of Econometrics* 16, 103–119.

Ghysels, E. & A. Hall (1990) Testing nonnested Euler conditions with quadrature-based methods of approximation. *Journal of Econometrics* 46, 273–308.

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.

Gourieroux, C. & A. Monfort (1994) Testing non-nested hypotheses. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2583–2637. Elsevier.

Gourieroux, C., A. Monfort, & A. Trognon (1983) Testing nested or non-nested hypotheses. *Journal of Econometrics* 21, 83–115.

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

Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69, 1661–1672.

Kitamura, Y. (2003) A Likelihood-Based Approach to the Analysis of a Class of Nested and Non-Nested Models. Manuscript, University of Pennsylvania.

Kitamura, Y. (2006) Empirical Likelihood Methods in Econometrics: Theory and Practice. Cowles Foundation Discussion Paper No. 1569, Yale University.

Kitamura, Y., G. Tripathi, & H. Ahn (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667–1714.

Loh, W. (1985) A new method for testing separate families of hypotheses. *Journal of the American Statistical Association* 80, 362–368.

Mizon, G. & J. Richard (1986) The encompassing principle and its application to testing non-nested hypotheses. *Econometrica* 54, 657–678.

Newey, W.K. (1990) Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809–837.

Newey, W.K. (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10, 233–253.

Newey,W.K. & R.J. Smith (2004) Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.

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

Owen, A.B. (2001) *Empirical Likelihood*. Chapman and Hall.

Pesaran, M. & M. Weeks (2001) Non-nested hypothesis testing: An overview. In B. Baltagi (ed.), *A Companion to Econometric Theory*, Ch. 13, pp. 279–309. Blackwell.

Powell, J.L., J.L. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57, 1403–1430.

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

Ramalho, J.J.S. & R.J. Smith (2002) Generalized empirical likelihood non-nested tests. *Journal of Econometrics* 107, 99–125.

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

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

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

Tripathi, G. & Y. Kitamura (2003) Testing conditional moment restrictions. *Annals of Statistics* 31, 2059–2095.

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

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–26.

Wooldridge, J. (1990) An encompassing approach to conditional mean tests with application to testing nonnested hypotheses. *Journal of Econometrics* 45, 331–350.

Zhang, J. & I. Gijbels (2003) Sieve empirical likelihood and extensions of the generalized least squares. *Scandinavian Journal of Statistics* 30, 1–24.

APPENDIX

Notation. Denote

$$I_* = \{i : x_i \in \mathcal{X}_*, 1 \le i \le n\}, \qquad c_n = \sqrt{\frac{\log n}{nb_n^s}},$$

$$g_j(\beta) = g(z_j, \beta), \qquad h_j(\gamma) = h(z_j, \gamma), \qquad m_j(\beta, \gamma) = m\left(z_j, \beta, \gamma\right),$$

$$\hat{M}_i(\beta, \gamma) = \hat{M}(x_i, \beta, \gamma), \qquad M_i(\beta, \gamma) = M(x_i, \beta, \gamma),$$

$$K_{ji} = K\left(\frac{x_i - x_j}{b_n}\right), \qquad \hat{f}_i = \frac{1}{nb_n^s}\sum_{j=1}^{n} K_{ji},$$

$$\bar{V}_i(\beta) = \mathrm{E}\left[\frac{1}{nb_n^s}\sum_{j=1}^{n} K_{ji}\, g_j(\beta) g_j(\beta)' | x_i\right],$$

$$\bar{J}_i(\beta)' = \mathrm{E}\left[\frac{1}{nb_n^s}\sum_{j=1}^{n} K_{ji} m_j(\beta,\gamma) g_j(\beta)'|x_i\right],$$

$$\bar{G}_i(\beta) = \mathrm{E}\left[\frac{1}{nb_n^s}\sum_{j=1}^{n} K_{ji} \frac{\partial g_j(\beta)}{\partial \beta'}|x_i\right].$$

Proof of Theorem 3.1(i). An expansion of $\hat{p}_{ji}^g(\hat{\beta})$ around $\lambda_i^g(\hat{\beta}) = 0$ yields

$$\hat{p}_{ji}^g(\hat{\beta}) = \frac{w_{ji}}{1+\lambda_i^g(\hat{\beta})' g_j(\hat{\beta})} = w_{ji}(1 - \lambda_i^g(\hat{\beta})' g_j(\hat{\beta}) + r_{ji}), \tag{A.1}$$

where $r_{ji} = \frac{\lambda_i^g(\hat{\beta})' g_j(\hat{\beta}) g_j(\hat{\beta})' \lambda_i^g(\hat{\beta})}{(1+\tilde{\lambda}_i^{g\prime} g_j(\hat{\beta}))^3}$, and $\tilde{\lambda}_i^g$ is a point on the line joining $\lambda_i^g(\hat{\beta})$ and 0. Since $\hat{p}_{ji}^g(\hat{\beta}) - \hat{p}_{ji}^N = w_{ji}(-\lambda_i^g(\hat{\beta})' g_j(\hat{\beta}) + r_{ji})$, the definition of T_M in (11) implies

$$\begin{aligned} T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}_i(\hat{\beta},\hat{\gamma})' \hat{J}_i(\hat{\beta},\hat{\gamma})' \lambda_i^g(\hat{\beta}) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}_i(\hat{\beta},\hat{\gamma})' \left(\sum_{j=1}^{n} w_{ji} r_{ji} m_j(\hat{\beta},\hat{\gamma})\right) \qquad \text{(A.2)} \\ &= T^{(1)} + R^{(1)}. \end{aligned}$$

Here $R^{(1)}$ satisfies

$$\begin{aligned} \|R^{(1)}\| &\le \max_{i\in I_*} \|\hat{M}_i(\hat{\beta},\hat{\gamma})\| \max_{1\le j\le n} \|m_j(\hat{\beta},\hat{\gamma})\| \left(\max_{i\in I_*}\|\lambda_i^g(\hat{\beta})\|\right)^2 \\ &\quad \times \left\|\frac{1}{n}\sum_{i=1}^{n} I_i \sum_{j=1}^{n} w_{ji} \frac{g_j(\hat{\beta}) g_j(\hat{\beta})'}{(1+\tilde{\lambda}_i^{g\prime} g_j(\hat{\beta}))^3}\right\|. \qquad \text{(A.3)} \end{aligned}$$

Assumption 3.2(iv) implies

$$\max_{i\in I_*} \|\hat{M}_i(\hat{\beta},\hat{\gamma})\| = O_p(1). \tag{A.4}$$

From Assumption 3.2(i) and (iv) and Tripathi and Kitamura (2003, Lem. C.4),

$$\max_{1\le j\le n} \|g_j(\hat{\beta})\| = o\left(n^{1/\zeta}\right), \qquad \max_{1\le j\le n} \|m_j(\hat{\beta},\hat{\gamma})\| = o\left(n^{1/\zeta_m}\right). \tag{A.5}$$

From Lemmas A.1 and A.4,

$$\max_{i\in I_*} \|\lambda_i^g(\hat{\beta})\| = O_p(c_n) + o_p\left(n^{-1/2+1/\eta}\right). \tag{A.6}$$

Since (A.5) and (A.6) imply that $\max_{i\in I_*,1\le j\le n} |\tilde{\lambda}_i^{g\prime} g_j(\hat{\beta})| = o_p(1)$, we have $\left\|\frac{1}{n}\sum_{i=1}^{n} I_i \sum_{j=1}^{n} w_{ji} \frac{g_j(\hat{\beta}) g_j(\hat{\beta})'}{\left(1+\tilde{\lambda}_i^{g\prime} g_j(\hat{\beta})\right)^3}\right\| \le O_p(1)$ by Lemma A.1. Thus, from (A.3)–(A.6),

$$\|R^{(1)}\| \le O_p(1) o(n^{1/\zeta_m}) \left\{O_p(c_n) + o_p(n^{-1/2+1/\eta})\right\}^2 O_p(1) = o_p(n^{-1/2}), \tag{A.7}$$

where the equality follows from $\alpha < \frac{1}{3s} \le \frac{1}{s}\left(1 - \frac{4}{\zeta_m}\right)$ and $\frac{1}{\zeta_m} + \frac{2}{\eta} \le \frac{1}{2}$. From (A.2) and Lemma A.4,

$$\begin{aligned} T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}_i(\hat{\beta},\hat{\gamma})'\hat{J}_i(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) - \frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}_i(\hat{\beta},\hat{\gamma})'\hat{J}_i(\hat{\beta},\hat{\gamma})' r_i^g \\ &\quad + o_p(n^{-1/2}) \\ &= T^{(2)} + R^{(2)} + o_p(n^{-1/2}). \end{aligned} \tag{A.8}$$

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

$$\begin{aligned} \|R^{(2)}\| &\le \max_{i \in I_*} \|\hat{M}_i(\hat{\beta},\hat{\gamma})\| \max_{i \in I_*} \|r_i^g\| \left\| \frac{1}{n}\sum_{i=1}^{n} I_i \hat{J}_i(\hat{\beta},\hat{\gamma}) \right\| \\ &= O_p(1)\, o_p(n^{1/\zeta}) \left\{ O_p(c_n^2) + o_p(n^{-1+2/\eta}) \right\} O_p(1) = o_p(n^{-1/2}), \end{aligned} \tag{A.9}$$

where the last equality follows from $\alpha < \frac{1}{3s} \le \frac{1}{s}\left(1 - \frac{4}{\zeta}\right)$ and $\frac{1}{\zeta} + \frac{2}{\eta} \le \frac{1}{2}$. Thus, from (A.8),

$$\begin{aligned} T_M &= -\frac{1}{n}\sum_{i=1}^{n} I_i \hat{M}_i(\hat{\beta},\hat{\gamma})'\hat{J}_i(\hat{\beta},\hat{\gamma})'\hat{V}_i(\hat{\beta})^{-1}\hat{g}_i(\hat{\beta}) + o_p(n^{-1/2}) \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_i M_i(\beta_0,\gamma_*)'\,\hat{J}_i(\beta_0,\gamma_*)'\,\hat{V}_i(\beta_0)^{-1}\,\hat{g}_i(\hat{\beta}) + R^{(3)} + o_p(n^{-1/2}). \end{aligned} \tag{A.10}$$

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

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

From Assumption 3.2(iv) and a similar argument to derive (A.15) 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

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

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 $||R_c^{(3)}|| = o_p\left(n^{-1/2}\right)$. Thus, from (A.10),

$$
\begin{aligned}
T_M &= -\frac{1}{n}\sum_{i=1}^n I_i M_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1} \hat{g}_i(\hat{\beta}) + o_p(n^{-1/2}) \\
&= -\frac{1}{n}\sum_{i=1}^n I_i M_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1} \{\hat{g}_i(\beta_0) + \hat{G}_i(\tilde{\beta})(\hat{\beta}-\beta_0)\} + o_p(n^{-1/2}) \\
&= -\frac{1}{n}\sum_{i=1}^n I_i M_i(\beta_0,\gamma_*)' \hat{J}_i(\beta_0,\gamma_*)' \hat{V}_i(\beta_0)^{-1} \hat{g}_i(\beta_0) \\
&\quad + \hat{H}_M(\beta_0,\gamma_*)\,\Delta \frac{1}{n}\sum_{i=1}^n \psi(x_i,z_i,\beta_0) \\
&\quad + R^{(4)} + o_p(n^{-1/2}) \\
&= T_{Ma} + T_{Mb} + R^{(4)} + o_p(n^{-1/2}), \qquad \textbf{(A.11)}
\end{aligned}
$$

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. Now $R^{(4)}$ is implicitly defined and satisfies

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

$$\times \left\| \frac{1}{n} \sum_{i=1}^{n} I_i \{\hat{G}_i(\tilde{\beta}) - \hat{G}_i(\beta_0)\} \right\| ||\hat{\beta} - \beta_0||$$

$$+ \max_{i \in I_*} ||M_i(\beta_0, \gamma_*)|| \max_{i \in I_*} ||\hat{J}_i(\beta_0, \gamma_*)|| \max_{i \in I_*} ||\hat{V}_i(\beta_0)^{-1}|| \max_{i \in I_*} ||\hat{G}_i(\beta_0)||$$

$$\times o_p(n^{-1/2})$$

$$= o_p(n^{-1+1/\eta_2}) + o_p(n^{-1/2}) = o_p(n^{-1/2}),$$

where the equality follows from Assumption 3.2(iv) and Lemmas A.1, A.2, and A.3. Thus, from (A.11), we have $T_M = T_{Ma} + T_{Mb} + o_p\left(n^{-1/2}\right)$. Now T_{Ma} is written as

$$T_{Ma} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i \mathrm{E}[\hat{f}_i | x_i]^{-1} M_i\left(\beta_0, \gamma^*\right)' \bar{J}_i(\beta_0, \gamma_*)' \bar{V}_i(\beta_0)^{-1} \frac{1}{nb_n^s} K_{ji} g_j(\beta_0) + R_a^{(5)}$$

$$= \bar{T}_{Ma} + R_a^{(5)}, \tag{A.12}$$

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

$$||R_a^{(5)}|| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} I_i M_i(\beta_0, \gamma_*)' \left\{ \hat{J}_i(\beta_0, \gamma_*) - \mathrm{E}[\hat{f}_i | x_i]^{-1} \bar{J}_i(\beta_0, \gamma_*) \right\}' \hat{V}_i(\beta_0)^{-1} \hat{g}_i(\beta_0) \right\|$$

$$+ \left\| \frac{1}{n} \sum_{i=1}^{n} I_i \mathrm{E}[\hat{f}_i | x_i]^{-1} M_i(\beta_0, \gamma_*)' \bar{J}_i(\beta_0, \gamma_*)' \right.$$

$$\left. \times \left\{ \hat{V}_i(\beta_0)^{-1} - \mathrm{E}[\hat{f}_i | x_i] \bar{V}_i(\beta_0)^{-1} \right\} \hat{g}_i(\beta_0) \right\|$$

$$+ \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i \left\{ \hat{f}_i^{-1} - \mathrm{E}[\hat{f}_i | x_i]^{-1} \right\} M_i(\beta_0, \gamma_*)' \bar{J}_i(\beta_0, \gamma_*)' \right.$$

$$\left. \times \bar{V}_i(\beta_0)^{-1} \frac{1}{nb_n^s} K_{ji} g_j(\beta_0) \right\|$$

$$= ||R_{aa}^{(5)}|| + ||R_{ab}^{(5)}|| + ||R_{ac}^{(5)}||.$$

From Assumption 3.2(iv), Lemmas A.1 and A.2, and Tripathi and Kitamura (2003, Lem. C.1), we have $||R_{aa}^{(5)}|| \leq O_p(c_n^2) = o_p(n^{-1/2})$ from $\alpha < \frac{1}{3s}$. Similarly, we have $||R_{ab}^{(5)}|| \leq O_p(c_n^2) = o_p(n^{-1/2})$. Moreover, Assumption 3.2(iv), Lemmas A.1 and A.2, and Tripathi and Kitamura (2003, eqn. (C.1)) imply $||R_{ac}^{(5)}|| \leq O_p(c_n^2) = o_p(n^{-1/2})$. Thus, from (A.12), we have $T_{Ma} = \bar{T}_{Ma} + o_p(n^{-1/2})$. By applying the U-statistic arguments of Kitamura et al. (2004, pp. 1696–1698) and Powell et al. (1989, Lem. 3.1), we have the asymptotic linear form for $\bar{T}_{Ma}$:

$$\sqrt{n}\bar{T}_{Ma} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_i M_i(\beta_0, \gamma_*)' J_i(\beta_0, \gamma_*)' V_i(\beta_0)^{-1} g_i(\beta_0) + o_p(1). \tag{A.13}$$

From Lemmas A.1, A.2, and A.3 and a weak law of large numbers, we can show that $\hat{H}_M(\beta_0, \gamma_*) \xrightarrow{p} \mathrm{E}[I_i M_i(\beta_0, \gamma_*)' J_i(\beta_0, \gamma_*)' V_i(\beta_0)^{-1} G_i(\beta_0)] = H_M(\beta_0, \gamma_*)$. Therefore T_{Mb} satisfies

$$\sqrt{n} T_{Mb} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} H_M(\beta_0, \gamma_*) \Delta \psi(x_i, z_i, \beta_0) + o_p(1). \tag{A.14}$$

From (A.11), (A.13), and (A.14), a central limit theorem yields

$$\sqrt{n} T_M = \sqrt{n} \bar{T}_{Ma} + \sqrt{n} T_{Mb} + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i^M(\beta_0, \gamma_*) + o_p(1)$$

$$\xrightarrow{d} N(0, \Phi_M). \tag{A.15}$$

Since $\hat{\Phi}_M$ is consistent for Φ_M, we have

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

Proof of Theorem 3.1(ii). From (A.1) and Lemma A.4, T_C in (14) is written as

$$T_C = \frac{1}{n} \sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} (\hat{p}_{ji}^g(\hat{\beta}) + \hat{p}_{ji}^N) h_j(\hat{\gamma}) \right\}' \hat{V}_i^h(\hat{\gamma})^{-1} \left\{ \sum_{j=1}^{n} (\hat{p}_{ji}^g(\hat{\beta}) - \hat{p}_{ji}^N) h_j(\hat{\gamma}) \right\}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} (2w_{ji} - w_{ji} \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})) h_j(\hat{\gamma}) \right\}'$$

$$\times \hat{V}_i^h(\hat{\gamma})^{-1} \left\{ \sum_{j=1}^{n} (w_{ji} \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})) h_j(\hat{\gamma}) \right\} + R^{(1c)},$$

where $R^{(1c)}$ is implicitly defined. From a similar argument to derive (A.7), $R^{(1c)}$ satisfies

$$||R^{(1c)}|| \le \left\| \frac{1}{n} \sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} (2w_{ji} - w_{ji} \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})) h_j(\hat{\gamma}) \right\}' \times \hat{V}_i^h(\hat{\gamma})^{-1} \left\{ \sum_{j=1}^{n} w_{ji} r_{ji} h_j(\hat{\gamma}) \right\} \right\|$$

$$+ \left\| \frac{1}{n} \sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} w_{ji} r_{ji} h_j(\hat{\gamma}) \right\}' \hat{V}_i^h(\hat{\gamma})^{-1} \left\{ \sum_{j=1}^{n} (w_{ji} \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})) h_j(\hat{\gamma}) \right\} \right\|$$

$$+ \left\| \frac{1}{n} \sum_{i=1}^{n} I_i \left\{ \sum_{j=1}^{n} w_{ji} r_{ji} h_j(\hat{\gamma}) \right\}' \hat{V}_i^h(\hat{\gamma})^{-1} \left\{ \sum_{j=1}^{n} w_{ji} r_{ji} h_j(\hat{\gamma}) \right\} \right\|$$

$$\leq o(n^{1/\zeta_m})\left\{O_p(c_n)+o_p(n^{-1/2+1/\eta})\right\}^2$$
$$+o(n^{1/\zeta_m})\left\{O_p(c_n)+o_p(n^{-1/2+1/\eta})\right\}^3$$
$$+o(n^{2/\zeta_m})\left\{O_p(c_n)+o_p(n^{-1/2+1/\eta})\right\}^4$$
$$=o_p(n^{-1/2}).$$

Thus, from Lemma A.4, we have

$$T_C=-\frac{1}{n}\sum_{i=1}^{n}I_i\left\{\sum_{j=1}^{n}(2w_{ji}-w_{ji}\lambda_i^g(\hat\beta)'g_j(\hat\beta))h_j(\hat\gamma)\right\}'$$
$$\times\hat V_i^h(\hat\gamma)^{-1}\left\{\sum_{j=1}^{n}(w_{ji}\lambda_i^g(\hat\beta)'g_j(\hat\beta))h_j(\hat\gamma)\right\}+o_p(n^{-1/2})$$
$$=-\frac{1}{n}\sum_{i=1}^{n}I_i\left\{2\hat h_i(\hat\gamma)-\hat J_i^h(\hat\beta,\hat\gamma)'\hat V_i(\hat\beta)^{-1}\hat g_i(\hat\beta)\right\}'$$
$$\times\hat V_i^h(\hat\gamma)^{-1}\left\{\hat J_i^h(\hat\beta,\hat\gamma)'\hat V_i(\hat\beta)^{-1}\hat g_i(\hat\beta)\right\}+R^{(2c)}+o_p(n^{-1/2}),\qquad\textbf{(A.16)}$$

where $R^{(2c)}$ is implicitly defined. A similar argument to show (A.9) yields that $||R^{(2c)}||=o_p(n^{-1/2})$. By setting

$$\hat M_i(\beta,\gamma)'=\{2\hat h_i(\gamma)-\hat J_i^h(\beta,\gamma)'\hat V_i(\beta)^{-1}\hat g_i(\beta)\}'\hat V_i^h(\gamma)^{-1},$$
$$M_i(\beta,\gamma)'=2\mathrm{E}\left[h(z_i,\gamma)|x_i\right]'V_i^h(\gamma)^{-1},\qquad m(z_j,\beta,\gamma)=h(z_j,\gamma),$$

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

$$\sqrt{n}T_C=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^C(\beta_0,\gamma_*)+o_p(1)\xrightarrow{d}N(0,\phi_C).\qquad\textbf{(A.17)}$$

Since $\hat\phi_C$ is consistent for ϕ_C, we have

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

■

Proof of Theorem 3.1(iii). From (A.1) and Lemma A.4, we have

$$T_S=\frac{1}{n}\sum_{i=1}^{n}I_i\hat G_i^h(\hat\gamma)'\hat V_i^h(\hat\gamma)^{-1}\left\{\sum_{j=1}^{n}\{\hat p_{ji}^g(\hat\beta)-\hat p_{ji}^N\}h_j(\hat\gamma)\right\}$$
$$=-\frac{1}{n}\sum_{i=1}^{n}I_i\hat G_i^h(\hat\gamma)'\hat V_i^h(\hat\gamma)^{-1}\left\{\sum_{j=1}^{n}(w_{ji}\lambda_i^g(\hat\beta)'g_j(\hat\beta))h_j(\hat\gamma)\right\}+R^{(1s)}$$
$$=-\frac{1}{n}\sum_{i=1}^{n}I_i\hat G_i^h(\hat\gamma)'\hat V_i^h(\hat\gamma)^{-1}\{\hat J_i^h(\hat\beta,\hat\gamma)'\hat V_i(\hat\beta)^{-1}\hat g_i(\hat\beta)\}+R^{(1s)}+R^{(2s)},\qquad\textbf{(A.18)}$$

where $R^{(1s)}$ and $R^{(2s)}$ are implicitly defined. Similar arguments to derive (A.7) and (A.9) yield $||R^{(1s)}|| = o_p(n^{-1/2})$ and $||R^{(2s)}|| = o_p(n^{-1/2})$, respectively. By setting

$$\hat{M}_i(\beta,\gamma)' = \hat{G}_i^h(\gamma)'\hat{V}_i^h(\gamma)^{-1}, \qquad M_i(\beta,\gamma)' = G_i^h(\gamma)'V_i^h(\gamma)^{-1},$$

$$m\left(z_j,\beta,\gamma\right) = h\left(z_j,\gamma\right),$$

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

$$\sqrt{n}T_S = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i^S(\beta_0,\gamma_*) + o_p(1) \xrightarrow{d} N(0,\Phi_S).$$

Since $\hat{\Phi}_S$ is consistent for Φ_S, we have

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

■

Proof of Theorem 3.2(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, Lem. 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 $\left\|\tilde{r}_i^g\right\| = o_p(n^{1/\zeta})\{\left(\max_{i\in I_*}\left\|\sum_{j=1}^n w_{ji}g_j(\beta_{0n})\right\|\right)^2 + ||\hat{\beta}-\beta_{0n}||^2\sum_{j=1}^n w_{ji}d_1\left(z_j\right)^2\}$, and the $o_p(n^{1/\zeta})$ term does not depend on $i \in I_*$. From the continuity of $\delta_h(x)$ and $f(x)$, and the compactness of $\mathcal{X}_*$, an adapted version of Tripathi and Kitamura (2003, Lem. 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 $\mathbf{H}_{gn}$,

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

Therefore the conclusion is obtained. ■

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

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

Therefore the conclusion is obtained. ■

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

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

Therefore the conclusion is obtained. ■

Proof of Theorem 3.3(i). Let $\tilde{J}_i(\beta,\gamma)'=\sum_{j=1}^{n}w_{ji}\frac{m_j(\beta,\gamma)g_j(\beta)'}{1+\lambda_i^g(\beta)'g_j(\beta)}$. By the definitions of $\hat{p}_{ji}^g(\beta)$ in (7) and T_M in (11),

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

$$= -\frac{1}{n}\sum_{i=1}^{n} I_i M_i(\beta_*, \gamma_0)' \tilde{J}_i(\hat{\beta}, \hat{\gamma})' \lambda_*^g(x_i, \beta_*) + o_p(1)$$

$$= -\frac{1}{n}\sum_{i=1}^{n} I_i M_i(\beta_*, \gamma_0)' J_{i*}(\beta_*, \gamma_0)' \lambda_*^g(x_i, \beta_*) + o_p(1)$$

$$= \mu_{hM} + o_p(1),$$

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 the fourth equality follows by applying similar arguments as Lemma A.2 and Newey (1994, Lem. 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 Theorem 3.3(ii). Observe that under $\mathbf{H}_h : \mathrm{E}\left[h_i(\gamma_0)|x_i\right] = 0$,

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

$$\leq \left(\sup_{x_i \in \mathcal{X}_*} ||\hat{h}_i(\hat{\gamma}) - \mathrm{E}\left[h_i(\gamma_0)|x_i\right]||\right)^2 \sup_{x_i \in \mathcal{X}_*} ||\hat{V}_i^h(\hat{\gamma})^{-1}|| \left(\frac{1}{n}\sum_{i=1}^{n} I_i\right) = o_p(1),$$

where the equality follows from the same argument as Lemmas A.1 and A.4 (replace $g_i(\beta)$ with $h_i(\gamma)$), and $\frac{1}{n}\sum_{i=1}^{n} I_i = O_p(1)$ (by a law of large numbers). Also, from the definition of $\hat{p}_{ji}^g(\beta)$ in (7),

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

$$= \frac{1}{n}\sum_{i=1}^{n} I_i \left\{\sum_{j=1}^{n} \frac{w_{ji} h_j(\hat{\gamma})}{1 + \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})}\right\}' \hat{V}_i^h(\hat{\gamma})^{-1} \left\{\sum_{j=1}^{n} \frac{w_{ji} h_j(\hat{\gamma})}{1 + \lambda_i^g(\hat{\beta})' g_j(\hat{\beta})}\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} I_i \mathrm{E}\left[\left.\frac{h(z_i, \gamma_0)}{1 + \lambda_*^g(x_i, \beta_*)' g(z_i, \beta_*)}\right| x_i\right]' \hat{V}_i^h(\hat{\gamma})^{-1}$$

$$\times \mathrm{E}\left[\left.\frac{h(z_i, \gamma_0)}{1 + \lambda_*^g(x_i, \beta_*)' g(z_i, \beta_*)}\right| x_i\right] + o_p(1)$$

$$= \frac{1}{n}\sum_{i=1}^{n} I_i \mathrm{E}\left[\left.\frac{h(z_i, \gamma_0)}{1 + \lambda_*^g(x_i, \beta_*)' g(z_i, \beta_*)}\right| x_i\right]' V_i^h(\gamma_0)^{-1}$$

$$\times \mathrm{E}\left[\left.\frac{h(z_i, \gamma_0)}{1 + \lambda_*^g(x_i, \beta_*)' g(z_i, \beta_*)}\right| x_i\right] + o_p(1)$$

$$= \mu_{hC} + o_p(1),$$

where the second equality follows from Assumption 3.2(iv) and the third equality follows from the same argument as Lemma A.1. Combining these results, we have $T_C = \mu_{hC} + o_p(1)$ and thus $C_g/\sqrt{n} \xrightarrow{p} \mu_{hC}/\sqrt{\phi_{hC}}$ under $\mathbf{H}_h$. The conclusion is obtained. ■

Proof of Theorem 3.3(iii). By the definitions of $\hat{p}^g_{ji}(\beta)$ in (7) and T_S in (16),

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

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, Lem. 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. ■

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

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

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

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

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

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

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

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

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

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

$$\sup_{x_i\in\mathcal{X}_*} ||\hat{J}_i(\hat{\beta},\hat{\gamma})' - \hat{J}_i(\beta_0,\gamma_*)'||$$

$$= \sup_{x_i\in\mathcal{X}_*}\left\| \sum_{j=1}^n w_{ji}\left(m_j(\beta_0,\gamma_*) + \frac{\partial m_j(\tilde{\beta},\tilde{\gamma})}{\partial(\beta',\gamma')}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\gamma}-\gamma_*\end{pmatrix}\right)\right.$$

$$\left.\times\left(g_j(\beta_0) + \frac{\partial g_j(\tilde{\beta})}{\partial\beta'}(\hat{\beta}-\beta_0)\right)' - \sum_{j=1}^n w_{ji} m_j(\beta_0,\gamma_*) g_j(\beta_0)'\right\|$$

$$\le ||\hat{\beta}-\beta_0|| \max_{1\le j\le n}\left\|m_j(\beta_0,\gamma_*)\right\| \sup_{x_i\in\mathcal{X}_*}\left\|\sum_{j=1}^n w_{ji} d_1(z_j)\right\|$$

$$+\left\|\begin{matrix}\hat{\beta}-\beta_0\\ \hat{\gamma}-\gamma_*\end{matrix}\right\| \max_{1\le j\le n}\left\|g_j(\beta_0)\right\| \sup_{x_i\in\mathcal{X}_*}\left\|\sum_{j=1}^n w_{ji} d_m(z_j)\right\|$$

$$+\left\|\hat{\beta}-\beta_0\right\|\left\|\begin{matrix}\hat{\beta}-\beta_0\\ \hat{\gamma}-\gamma_*\end{matrix}\right\| \max_{1\le j\le n}\left\|g_j(\beta_0)\right\| \sup_{x_i\in\mathcal{X}_*}\left\|\sum_{j=1}^n w_{ji} d_m(z_j)\right\|$$

$$= R_a^J + R_b^J + R_c^J,$$

where $(\tilde{\beta},\tilde{\gamma})$ is a point on the line joining $(\hat{\beta},\hat{\gamma})$ and (β_0,γ_*). From (A.5), Assumption 3.1(ii) and (iii), and Tripathi and Kitamura (2003, Lem. C.6), we have

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

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

From $\eta \ge 6$ and $\eta_m \ge 6$, R_c^J is negligible. Therefore the first part is obtained.

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

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

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

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

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

$$\sup_{x_i\in\mathcal{X}_*}\left\|\sum_{j=1}^{n} w_{ji}\frac{\partial g_j^{(k)}(\hat{\beta})}{\partial\beta^{(\ell)}}-\sum_{j=1}^{n} w_{ji}\frac{\partial g_j^{(k)}(\beta_0)}{\partial\beta^{(\ell)}}\right\| \leq \sup_{x_i\in\mathcal{X}_*}\left\|\sum_{j=1}^{n} w_{ji}d_2\left(z_j\right)\right\| ||\hat{\beta}-\beta_0||$$

$$= o(n^{1/\eta_2})O_p(n^{-1/2}),$$

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

(Second part) The second part is obtained from the proof of Newey (1994, Lem. 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* $\mathbf{H}_g$

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

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

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

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