CrossMark

*Econometric Theory*, **34**, 2018, 477–507.
doi:10.1017/S0266466617000123

# A GENERAL CLASS OF NON-NESTED TEST STATISTICS FOR MODELS DEFINED THROUGH MOMENT RESTRICTIONS

PAULO M.D.C. PARENTE
*Instituto Universitário de Lisboa* (*ISCTE-IUL*)
*Business Research Unit* (*BRU-IUL*)

In this article, we introduce a new class of Cox's non-nested test statistics for models defined through overidentifying moment restrictions that depend on a finite dimensional parameter vector. In addition to showing that the GEL test statistics proposed in Smith (1997, *Economic Journal* 107, 503–519) and Ramalho and Smith (2002, *Journal of Econometrics* 107, 99–125) are members of this class, we reveal that further members can be constructed using the artificial compound model approach, which was originally applied by Atkinson (1970, *Journal of the Royal Statistical Society, Series B* 32, 323–335) in the parametric setting. We investigate the asymptotic properties of the statistics and propose tests based on modified versions of these statistics that have correct asymptotic size in a uniform sense, a requirement not satisfied by existing Cox's non-nested tests. A Monte Carlo study examines the performance of the proposed tests.

## 1. INTRODUCTION

Since they were first introduced by Cox (1961, 1962), non-nested tests have become popular in the literature.[1] Given two competing models, the objective of Cox-type non-nested tests is twofold. In addition to allowing for inferences regarding whether the true distribution of the data lies in the model specified under the null hypothesis, non-nested tests are designed such that they are able to detect departures in the direction of the rival model evaluated at its pseudotrue value. This article addresses Cox-type non-nested test statistics that allow for the comparison of two rival models defined through overidentified moment restrictions that depend on a finite number of parameters. Our first contribution is to show that, given a particular generalized empirical likelihood estimator [GEL, (Smith, 1997)]

I am grateful for the constructive comments and criticisms by the Editor, a co-Editor and three anonymous referees on previous drafts. Earlier versions of this paper were presented at the 2008 European Meeting of the Econometric Society, Bocconi University, Milan and the Third Meeting of the Portuguese Economic Journal, Funchal. This paper has also benefited from the comments of R. Dupleich, J.J.S. Ramalho, J.M.C. Santos Silva and R.J. Smith. The usual disclaimer applies. This work was supported by Fundação para a Ciência e a Tecnologia. Address correspondence to Paulo M.D.C. Parente, BRU-IUL (Business Research Unit), ISCTE-IUL, Avenida das Forças Armadas, 1649-026 Lisboa, Portugal; e-mail: paulo.miguel.parente@iscte.pt.

© Cambridge University Press 2017 

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

for the parameters of the models, a number of Cox-type non-nested test statistics can be chosen. These statistics belong to a particular class that we introduce here and require the specification of certain functions by the econometrician. Additionally, since in particular instances the standard tests based on Cox-type non-nested test statistics can suffer from severe size distortions in finite samples, we further contribute by offering a new testing procedure, computed using modified versions of these statistics, that has correct asymptotic size in a uniform sense.

Cox's non-nested tests for models defined through moment restrictions were previously introduced by Singleton (1985) based on the two-step generalized method of moments (GMM) estimator (Hansen, 1982) for competing models based on the same number of moment restrictions. Ghysels and Hall (1990) proposed an alternative test that requires the specification of the distribution functions of some random variables.[2] R.J. Smith made a number of important contributions to this topic. In particular, he introduced non-nested test statistics based on GMM (Smith, 1992) and GEL (Smith, 1997 and Ramalho and Smith, 2002) that do not require parametric assumptions and allow the competing models to depend on a different number of moment restrictions. Note, however, that the tests proposed by R.J. Smith may also suffer from size distortions. This article complements and generalizes the research of R.J. Smith on GEL non-nested test statistics. In addition to revealing that the GEL non-nested test statistics of Smith (1997) and Ramalho and Smith (2002) are members of the novel class of test statistics proposed here, we show that our new testing procedure can also be computed using modified versions of his statistics.

Some members of the new class of non-nested test statistics are derived using the principle of artificial nesting, originally adopted by Atkinson (1970) to compare parametric models estimated using the maximum likelihood method.[3] Here, Atkinson's approach is adapted to the framework of models defined through moment restrictions using the empirical likelihood (EL) estimator of Qin and Lawless (1994) and Imbens (1997) and alternative estimators based on the minimization of divergence measures, such as those introduced by Corcoran (1998).

We study the asymptotic behavior of the proposed class of non-nested test statistics in a number of different situations. We first analyze the case in which one model is correctly specified while the alternative model is misspecified under the null hypothesis and show that the pointwise asymptotic distribution of the proposed statistics is standard normal. Additionally, we examine the consistency properties of the associated tests and the asymptotic local behavior of the statistics.

The standard Cox's non-nested test relies on pointwise asymptotic theory, and consequently, the required critical value is obtained from the tables of the standard normal distribution. Note, however, that the pointwise asymptotic distribution of the non-nested test statistic depends on the assumption that the model specified under the alternative is misspecified. When this model is misspecified, the pointwise asymptotic distribution is standard normal, while when it is correctly specified, it is a function of two weighted sums of chi-square distributions. The fact that the non-nested test statistic converges to different distributions in these

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

situations can lead to severe size distortions. For instance, if the model specified under the alternative is almost correctly specified, the finite sample distribution of the non-nested test statistic is far from the standard normal distribution, and consequently, the standard test produces incorrect inferences. This problem parallels that addressed by Shi (2015a) in the context of the Vuong test (Vuong, 1989), and we show here that a test procedure similar to the one she proposed in that context yields correct inferences using Cox's non-nested test statistics. This analysis also reveals that the power of this test does not depend on the functions specified to compute the test statistics when the degree of misspecification of the model specified under the alternative is small and only depends on the arguments of those functions.

This article is organized as follows. The minimum divergence and GEL classes of estimators are briefly reviewed in Section 2. The new class of non-nested test statistics based on EL and on other estimator-based discrepancy measures is described in Section 3, and some examples are discussed. Section 4 presents the asymptotic properties of the class of test statistics in different settings. Section 5 discusses the problem of possible size distortions in the standard Cox's non-nested tests, proposes the new modified test and shows that it has correct null rejection rates in a uniform sense. This section also investigates the power properties of the modified test. Section 6 reports a Monte Carlo study on the performance of the proposed statistics and offers some recommendations for practitioners. Finally, Section 7 concludes. The tables associated with the simulation experiments are presented in Appendix A. The proofs of the results given in the text and additional tables with the results of the Monte Carlo study are provided online at Cambridge Journals Online (journals.cambridge.org/ect) in supplementary material to this article.

## 2. FRAMEWORK

Let $z$ be a random vector with support $\mathbb{R}^k$, and consider a $m_g$-vector of indicator functions $g(z,\beta)$ that is known up to a $p_g$-vector $\beta$, where $m_g > p_g$ and the parameter space of $\beta$ is $\mathcal{B} \subset \mathbb{R}^{p_g}$. Let $\mathcal{M}$ denote the set of all probability measures on $\mathbb{R}^k$, and consider a statistical model $\mathcal{P}_g = \cup_{\beta\in\mathcal{B}}\mathcal{P}_g(\beta)$, where $\mathcal{P}_g(\beta) = \{F \in \mathcal{M} : \mathrm{E}_F[g(z,\beta)] = 0\}$ and $\mathrm{E}_F[\cdot]$ denotes the expectation taken with respect to $F$. Let $\mathrm{P}_0$ be the true probability measure of $z$, which is unknown. If $\mathrm{P}_0$ is in $\mathcal{P}_g$, the model $\mathcal{P}_g$ is denoted by Kitamura (2007, p. 180, para. 2) as a correctly specified model. A model $\mathcal{P}_g$ is misspecified if it does not contain $\mathrm{P}_0$ (Kitamura , 2007, p. 275, para. 3–4).

Let $\{z_i\}_{i=1}^n$ denote a sample of $n$ independent and identically distributed observations on the random vector $z$. Given $\beta$, the sample analogue of the population expectation $\mathrm{E}_{\mathrm{P}_0}[g(z,\beta)]$ is $\hat{g}(\beta) \equiv \sum_{i=1}^n g_i(\beta)/n$, where $g_i(\beta) \equiv g(z_i,\beta)$, $(i = 1,\ldots,n)$.

### 2.1. The Estimators

*2.1.1. Minimum Divergence Estimators.* Corcoran (1998) introduced a class of estimators based on the minimization of a discrepancy measure. Several

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

estimators belong to this class, namely EL, the continuous updated estimator (CUE) of Hansen, Heaton, and Yaron (1996), the exponential tilting estimator of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998) and Cressie-Read (CR) class of estimators (Cressie and Read, 1984). Corcoran's class of estimators solves the following problem:

$$\min_{p^n,\beta\in\mathcal{B}} \mathcal{I}\left(p^n,\iota^n\right), \tag{2.1}$$

$$\text{s.t. } p_i > 0, (i=1,\ldots,n), \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g_i(\beta) = 0,$$

where $p^n = (p_1,\ldots,p_n)'$, $\iota^n$ is an $n$-vector with all elements equal to $1/n$ and $\mathcal{I}(p^n,\iota^n)$ is a discrepancy measure. Kitamura (2007) applied the concept of $f$-divergence as defined by Csiszar (1963) and Ali and Silvey (1966) to introduce the generalized minimum contrast (GMC) class of estimators, which defines the discrepancy measure in (2.1) as

$$\mathcal{I}\left(p^n,\iota^n\right) = \frac{1}{n}\sum_{i=1}^{n}\phi\left(np_i\right), \tag{2.2}$$

where $\phi(.)$ is a convex function continuous at zero such that $\phi(1)=0$.[4,5]

*2.1.2. Generalized Empirical Likelihood.* The GEL class of estimators introduced by Smith (1997, 2001) also encompasses the EL, exponential tilting, CUE, and CR estimators. The GEL class of estimators is defined as follows. Let $\hat{P}_n^g(\beta,\lambda) \equiv \sum_{i=1}^{n}(\rho(\lambda' g_i(\beta)) - \rho_0)/n$, where $\rho(\cdot)$ is a concave function on its domain, an open interval $\mathcal{V}$ containing zero, $\rho_j(a) = \partial^j \rho(a)/\partial a^j$ with $\rho_j(0) = \rho_j$, $(j=0,1,\ldots)$, and it can be normalized without loss of generality such that $\rho_1 = \rho_2 = -1$. The GEL estimators are given by

$$\hat{\beta} = \arg\min_{\beta\in\mathcal{B}} \sup_{\lambda\in\hat{\Lambda}_n(\beta)} \hat{P}_n^g(\beta,\lambda), \tag{2.3}$$

where $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i=1,\ldots,n\}$.[6] The optimal auxiliary parameter estimator is defined as $\hat{\lambda} = \arg\max_{\lambda\in\hat{\Lambda}_n(\hat{\beta})} \hat{P}_n^g(\hat{\beta},\lambda)$, and the implied probabilities are defined as $\hat{p}_i = \rho_1(\hat{\lambda}' g_i(\hat{\beta})) \Big/ \sum_{i=1}^{n}\rho_1(\hat{\lambda}' g_i(\hat{\beta}))$.[7]

In general, the GEL class of estimators differs from GMC estimators. However, it is well known that for EL, exponential tilting, CUE and CR, the GEL program is the dual of the $f$-divergence program and thus yields the same estimators.[8] For simplicity, we consider estimators that belong both to GMC and GEL.

***Asymptotic Theory of GEL for Correctly Specified Models*:** The asymptotic properties of the non-nested test statistics depend on the large sample behavior of the GEL class of estimators, which is summarized here. Assume that $P_0$ is in $\mathcal{P}_g$ and in particular that there is a unique value $\beta_0 \in \mathcal{B}$ such that $E_{P_0}[g(z,\beta_0)] = 0$.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

Let $\Omega_{0,g} \equiv \mathrm{E}_{\mathrm{P}_0}[g(z,\beta_0)\,g(z,\beta_0)']$, $D_{0,g} \equiv \mathrm{E}_{\mathrm{P}_0}[\partial g(z,\beta_0)/\partial\beta]$ and denote the Euclidean norm of a matrix $W$ as $\|W\| = \mathrm{trace}\left(W'W\right)^{1/2}$. To prove consistency, Newey and Smith (2004) required the following assumptions.

**Assumption 2.1.** (a) $\beta_0 \in \mathcal{B}$ is the unique solution to $\mathrm{E}_{\mathrm{P}_0}[g(z,\beta)] = 0$; (b) $\mathcal{B}$ is compact; (c) $g(z,\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one; (d) $\mathrm{E}_{\mathrm{P}_0}[\sup_{\beta\in\mathcal{B}}\|g(z,\beta)\|^{\alpha}] < \infty$ for $\alpha > 2$; (e) $\Omega_{0,g}$ is nonsingular; and (f) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero.

Assumption 2.2 below is needed to prove the asymptotic normality of GEL.

**Assumption 2.2.** (a) $\beta_0 \in int(\mathcal{B})$; (b) $g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $\beta_0$, and $\mathrm{E}_{\mathrm{P}_0}\left[\sup_{\beta\in\mathcal{N}}\left\|\partial g(z,\beta)/\partial\beta'\right\|\right] < \infty$; and (c) $rank(D_{0,g}) = p_g$.

Let $\Sigma_{0,g} \equiv \left(D_{0,g}'\Omega_{0,g}^{-1}D_{0,g}\right)^{-1}$, $H_{0,g} \equiv \Sigma_{0,g}D_{0,g}'\Omega_{0,g}^{-1}$ and $P_{0,g} \equiv \Omega_{0,g}^{-1} - \Omega_{0,g}^{-1}D_{0,g}\Sigma_{0,g}D_{0,g}'\Omega_{0,g}^{-1}$. Newey and Smith (2004) proved that if Assumption 2.1 is satisfied, then $\hat{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\hat{\beta}) = O_p(n^{-1/2})$ and $\hat{\lambda} = O_p(n^{-1/2})$. Furthermore, they showed that $\sqrt{n}((\hat{\beta}-\beta_0)', \hat{\lambda}')' \xrightarrow{d} \mathcal{N}\left(0, \mathrm{diag}\left(\Sigma_{0,g}, P_{0,g}\right)\right)$, if Assumption 2.2 also holds.

***GEL for Misspecified Models*:** The asymptotic distribution of the non-nested tests also depends on the convergence of GEL estimators to the pseudo-true values in misspecified models. Consider a $m_h$-vector of indicator functions $h(z,\gamma)$ that is known up to a $p_h$-vector $\gamma$, where $m_h > p_h$ and the parameter space of $\gamma$ is $\mathcal{G} \subset \mathbb{R}^{p_h}$. Define a statistical model $\mathcal{P}_h = \cup_{\gamma\in\mathcal{G}}\mathcal{P}_h(\gamma)$, where $\mathcal{P}_h(\gamma) = \{F \in \mathcal{M} : \mathrm{E}_F[h(z,\gamma)] = 0\}$. We assume that there are no values of $\gamma \in \mathcal{G}$ that satisfy $\mathrm{E}_{\mathrm{P}_0}[h(z,\gamma)] = 0$. Since $\mathrm{P}_0$ is not in $\mathcal{P}_h$, this model is misspecified.

Let $h_i(\gamma) \equiv h(z_i,\gamma)$ and $\hat{P}_n^h(\gamma,\eta) = \sum_{i=1}^n(\rho(\eta'h_i(\gamma)) - \rho_0)/n$. The GEL estimators in this case are given by

$$\hat{\gamma} = \arg\min_{\gamma\in\mathcal{G}}\sup_{\eta\in\mathcal{H}}\hat{P}_n^h(\gamma,\eta), \tag{2.4}$$

where $\mathcal{H}$ is a compact parameter space (see Hong, Preston, and Shum, 2003). Let $\hat{\eta} = \arg\max_{\eta\in\mathcal{H}}\hat{P}_n^h\left(\hat{\gamma},\eta\right)$. Additionally, denote the GEL implied probabilities as $\hat{q}_i = \rho_1(\hat{\eta}'h_i(\hat{\gamma}))\big/\sum_{i=1}^n\rho_1(\hat{\eta}'h_i(\hat{\gamma}))$.

We assume consistency of $(\hat{\gamma}', \hat{\eta}')'$ for $(\gamma_0^{*\prime}, \eta_0^{*\prime})'$, the pseudo-true values of $(\gamma', \eta')'$ under the true probability measure $\mathrm{P}_0$, which are defined as $\gamma_0^* = \arg\min_{\gamma\in\mathcal{G}}\mathrm{E}_{\mathrm{P}_0}\left[\rho(\eta_0(\gamma)'h(z,\gamma))\right]$, where $\eta_0(\gamma) = \arg\max_{\eta\in\mathcal{H}}\mathrm{E}_{\mathrm{P}_0}\left[\rho\left(\eta'h(z,\gamma)\right)\right]$ and $\eta_0^* \equiv \eta_0(\gamma_0^*)$.

**Assumption 2.3.** $\hat{\gamma} \xrightarrow{p} \gamma_0^*$, $\hat{\eta} \xrightarrow{p} \eta_0^*$, where $\eta_0^* \neq 0$.

Hong et al. (2003) provide sufficient conditions for Assumption 2.3 to hold.[9,10]

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

## 3. A NEW CLASS OF TESTS FOR THE NON-NESTED HYPOTHESIS

Consider two competing hypotheses

$\mathrm{H}_g$ : There exists a unique value $\beta_0 \in \mathcal{B} : \mathrm{E}_{\mathrm{P}_0}[g(z, \beta_0)] = 0$,

$\mathrm{H}_h$ : There exists a unique value $\gamma_0 \in \mathcal{G} : \mathrm{E}_{\mathrm{P}_0}[h(z, \gamma_0)] = 0$.

The objective of this section is to propose Cox-type tests for the comparison of $\mathrm{H}_g$ and $\mathrm{H}_h$.

### 3.1. The General Form of the Statistics

The non-nested test statistics introduced in this article are of the following form:

$$\mathcal{S}_{v(\cdot)} = \sqrt{n} \sum_{i=1}^{n} v_a\left(\hat{p}_i n\right) \hat{p}_i \left[ v_b\left(\hat{q}_i^{v_c} / \hat{p}_i^{v_c}\right) - \sum_{\ell=1}^{n} v_b\left(\hat{q}_\ell^{v_c} / \hat{p}_\ell^{v_c}\right) \hat{p}_\ell \right], \tag{3.1}$$

where $\hat{p}_i^{v_c} = v_c(\hat{\lambda}' g_i(\hat{\beta})) \Big/ \sum_{i=1}^{n} v_c(\hat{\lambda}' g_i(\hat{\beta}))$, $\hat{q}_i^{v_c} = v_c(\hat{\eta}' h_i(\hat{\gamma})) \Big/ \sum_{i=1}^{n} v_c(\hat{\eta}' h_i(\hat{\gamma}))$, $v(\cdot) = [v_a(\cdot), v_b(\cdot), v_c(\cdot)]$ and $v_a(\cdot), v_b(\cdot)$ and $v_c(\cdot)$ are real univariate functions chosen by the practitioner, depend on different arguments and satisfy the regularity conditions stated in Assumption 4.1(a) below. Examples of a number of test statistics that can be written in the form (3.1) for different functions $v_a(\cdot), v_b(\cdot)$ and $v_c(\cdot)$ are given in Section 3.3. We can also replace $(\hat{\gamma}', \hat{\eta}')'$ by $(\check{\gamma}', \check{\eta}')'$ where $\check{\gamma} = \arg\min_{\gamma \in \mathcal{G}} \sup_{\eta \in \mathcal{H}} \sum_{i=1}^{n} \hat{p}_i \rho(\eta' h_i(\gamma))$, and $\check{\eta} = \arg\max_{\eta \in \mathcal{H}} \sum_{i=1}^{n} \hat{p}_i \rho(\eta' h_i(\check{\gamma}))$. Henceforth, to simplify the notation, we replace the subscript $v(\cdot)$ by $v$, unless we explicitly specify this function; that is, we write $\mathcal{S}_{v(\cdot)}$ as $\mathcal{S}_v$. We also adopt this notation for the alternative statistics proposed below and for the matrices that depend on this function.

An alternative to $\mathcal{S}_v$ in which the ratio $\hat{q}_i^{v_c} / \hat{p}_i^{v_c}$ is replaced by $n\hat{q}_i^{v_c}$ is given by the statistic

$$\tilde{\mathcal{S}}_v = \sqrt{n} \sum_{i=1}^{n} v_a\left(\hat{p}_i n\right) \hat{p}_i \left[ v_b\left(n\hat{q}_i^{v_c}\right) - \sum_{\ell=1}^{n} v_b\left(n\hat{q}_\ell^{v_c}\right) \hat{p}_\ell \right]. \tag{3.2}$$

### 3.2. Interpretation of the New Class of Test Statistics

The test statistics (3.1) introduced here have a simple interpretation. Let $\hat{\mu}_i \equiv \hat{p}_i\left[v_b\left(\hat{q}_i^{v_c} / \hat{p}_i^{v_c}\right) - \sum_{\ell=1}^{n} v_b\left(\hat{q}_\ell^{v_c} / \hat{p}_\ell^{v_c}\right) \hat{p}_\ell\right]$, which satisfies the condition $\sum_{i=1}^{n} \hat{\mu}_i = 0$. Note that $\mathcal{S}_v = \sqrt{n} \sum_{i=1}^{n} v_a\left(\hat{p}_i n\right) \hat{\mu}_i$ and that $\hat{\mu}_i$ depends on both models, while $v_a\left(\hat{p}_i n\right)$ depends solely on the model specified under the null. We are testing asymptotic uncorrelateness between $v_a\left(\hat{p}_i n\right)$ and $\hat{\mu}_i$. If $\mathcal{P}_g$ is correctly specified and $\mathcal{P}_h$ is misspecified, the probability limit under $\mathrm{P}_0$ of $v_a\left(n\hat{p}_i\right)$ uniformly in $i$ is a nonstochastic constant; thus, $v_a\left(n\hat{p}_i\right)$ is asymptotically uncorrelated with $\hat{\mu}_i$ and $\mathcal{S}_v / \sqrt{n}$ converges to zero.[11] However, if $\mathcal{P}_g$ is misspecified, then $v_a\left(n\hat{p}_i\right)$ does not converge in probability (under $\mathrm{P}_0$) to a constant and

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

consequently is correlated with $\hat{\mu}_i$, as the latter depends on $(\hat{p}_i, \hat{p}_i^{\upsilon_c}, \hat{q}_i^{\upsilon_c})$. It follows that $\mathcal{S}_\upsilon/\sqrt{n}$ does not converge to zero.

### 3.3. Examples

*3.3.1. Smith's GEL Non-Nested Tests.* Let $\upsilon_a(\cdot) = -1/(\cdot)$, $\upsilon_b(\cdot) = \ln(\cdot)$ and $\upsilon_c(\cdot) = \exp(-\rho(\cdot))$. It follows that the statistics (3.1) and (3.2) simplify to $\mathcal{S}_{[-1/(\cdot),\ln(\cdot),\exp(-\rho(\cdot))]} = n^{-1/2}\{\sum_{i=1}^n [\rho(\hat{\eta}' h_i(\hat{\gamma})) - \rho(\hat{\lambda}' g_i(\hat{\beta}))] - \sum_{\ell=1}^n [\rho(\hat{\eta}' h_\ell(\hat{\gamma})) - \rho(\hat{\lambda}' g_\ell(\hat{\beta}))]n\hat{p}_\ell\}$ and $\tilde{\mathcal{S}}_{[-1/(\cdot),\ln(\cdot),\exp(-\rho(\cdot))]} = n^{-1/2}[\sum_{i=1}^n \rho(\hat{\eta}' h_i(\hat{\gamma})) - \sum_{\ell=1}^n \rho(\hat{\eta}' h_\ell(\hat{\gamma}))n\hat{p}_\ell]$, respectively, which correspond to the GEL statistics of Ramalho and Smith (2002) for the non-nested hypothesis based on the GEL objective function.[12] The non-nested test statistics of Smith (1997) are obtained by replacing $(\hat{\gamma}', \hat{\eta}')'$ with $(\check{\gamma}', \check{\eta}')'$.

*3.3.2. Atkinson's Approach .* We can apply Atkinson's approach to derive a non-nested test statistic for models defined through moment restrictions based on EL. Given a random sample $\{z_i, i = 1, \ldots, n\}$, consider the program

$$\max_{\theta, p_i q_i, i=1\ldots,n} \sum_{i=1}^n \ln(w_i(\theta)),$$

$$w_i(\theta) = \frac{p_i^{1-\theta} q_i^\theta}{\sum_{i=1}^n p_i^{1-\theta} q_i^\theta}, \tag{3.3}$$

$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n q_i = 1, \sum_{i=1}^n p_i g_i(\beta) = 0, \sum_{i=1}^n q_i h_i(\gamma) = 0, p_i > 0, q_i > 0,$ $i = 1, \ldots, n, \theta \in \mathbb{R}$. Note that setting $\theta = 0$ leads to the EL problem defined under the null hypothesis and choosing $\theta = 1$ yields the EL problem for the program defined under the alternative. Hence, the null hypothesis of the Cox non-nested tests to compare $\mathrm{H}_g$ with $\mathrm{H}_h$ is recast as a test for the parametric restriction $H_0$ : $\theta = 0$ vs $H_1 : \theta = 1$ (cf. Atkinson, 1970 and Gourieroux and Monfort, 1994, Sect. 4). Similar to Atkinson (1970) and Gourieroux and Monfort (1994, Sect. 4.3), a score-type statistic for the hypothesis $H_0 : \theta = 0$ is proposed, in which $\beta$ and $\gamma$ are replaced by their empirical likelihood estimators $\hat{\beta}$ and $\hat{\gamma}$ respectively: $\hat{S} = n^{-1/2}\sum_{i=1}^n [\ln(\hat{q}_i/\hat{p}_i) - \sum_{\ell=1}^n \ln(\hat{q}_\ell/\hat{p}_\ell)\hat{p}_\ell]$, where $\hat{q}_i$ and $\hat{p}_i$ correspond to the implied probabilities of both models. Note that

$$\hat{S} = \frac{1}{\sqrt{n}}\left\{\sum_{i=1}^n [\ln(1 - \hat{\lambda}' g_i(\hat{\beta})) - \ln(1 - \hat{\eta}' h_i(\hat{\gamma}))] - \sum_{\ell=1}^n n\hat{p}_\ell [\ln(1 - \hat{\lambda}' g_\ell(\hat{\beta})) - \ln(1 - \hat{\eta}' h_\ell(\hat{\gamma}))]\right\}$$

and, consequently, $-\hat{S}$ coincides with the Ramalho and Smith (2002) statistic $\mathcal{S}_{[-1/(\cdot),\ln(\cdot),\exp(-\ln(1-(\cdot)))]}$ based on EL.[13]

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

*3.3.3. Generalizations of Atkinson's Approach.* The program (3.3) is generalized here in two directions. First, a comprehensive class of estimators based on $f$-divergence functions to estimate the parameters is adopted (cf. Section 2.1.1). Second, a general compound model based on alternatives to (3.3) suggested by Santos Silva (2001) in a parametric framework is chosen. This alternative compound model in our framework is obtained by replacing (3.3) with

$$w_i(\theta) = \frac{\left[(1-\theta)p_i^{\alpha} + \theta q_i^{\alpha}\right]^{1/\alpha}}{\sum_{i=1}^{n}\left[(1-\theta)p_i^{\alpha} + \theta q_i^{\alpha}\right]^{1/\alpha}}. \tag{3.4}$$

Note that (3.3) corresponds to the case in which $\alpha \to 0$.[14] These ideas lead to the following general framework. Consider the program

$$\min_{w_i, i=1\ldots,n} \frac{1}{n}\sum_{i=1}^{n}\phi(nw_i(\theta)),$$

$\sum_{i=1}^{n} p_i = 1$, $\sum_{i=1}^{n} q_i = 1$, $\sum_{i=1}^{n} p_i g_i(\beta) = 0$, $\sum_{i=1}^{n} q_i h_i(\gamma) = 0$, $p_i > 0$, $q_i > 0$, $i = 1, \ldots, n$, $\theta \in \mathbb{R}$, where $w_i(\theta)$ can be given by (3.3) or (3.4) and $\phi(.)$ is defined in Section 2.1.1. Note that $\theta = 0$ renders the GMC problem defined under the null hypothesis and $\theta = 1$ yields the GMC problem defined under the alternative. As in the previous section, we replace $\beta$ and $\gamma$ by their GMC estimators $\hat{\beta}$ and $\hat{\gamma}$ and define the score-type statistic for $H_0 : \theta = 0$ computed as if $\hat{\beta}$ and $\hat{\alpha}$ were deterministic:

$$\hat{S} = \sqrt{n}\sum_{i=1}^{n}\phi_1\left(n\hat{p}_i\right)\bar{w}_{1,i}, \tag{3.5}$$

where $\phi_1(v) \equiv \partial\phi(v)/\partial v$, $\bar{w}_{1,i} = w_{1,i}(0, \hat{\beta}, \hat{\gamma})$, and $w_{1,i}(\theta, \beta, \gamma) \equiv \partial w_i(\theta)/\partial\theta$ in the notation of which the dependence of $\partial w_i(\theta)/\partial\theta$ on $\beta$ and $\gamma$ through $p_i$ and $q_i$ is made explicit. Note that $\phi_1(v) = -1/v$ for EL, $\phi_1(v) = \ln(v) + 1$ for exponential tilting, $\phi_1(v) = v - 1$ for CUE, and $\phi_1(v) = [v^{\tau} - 1]/\tau$ for CR. Additionally, the case (3.3) yields

$$\bar{w}_{1,i} = \ln(\hat{q}_i/\hat{p}_i)\hat{p}_i - \hat{p}_i\sum_{\ell=1}^{n}\ln(\hat{q}_\ell/\hat{p}_\ell)\hat{p}_\ell, \tag{3.6}$$

and for (3.4), one obtains

$$\bar{w}_{1,i} = \frac{1}{\alpha}\left[-1 + \left(\hat{q}_i/\hat{p}_i\right)^{\alpha}\right]\hat{p}_i - \hat{p}_i\frac{1}{\alpha}\sum_{\ell=1}^{n}\left[-1 + \left(\hat{q}_\ell/\hat{p}_\ell\right)^{\alpha}\right]\hat{p}_\ell. \tag{3.7}$$

These statistics are special cases of (3.1) with $v_a(\cdot) = \phi_1(\cdot)$. Additionally, for (3.6), $v_b(\cdot) = \ln(\cdot)$, and for (3.7), $v_b(\cdot) = \left[-1 + (\cdot)^{\alpha}\right]/\alpha$ and $v_c(\cdot) = \rho_1(\cdot)$. Thus, we obtain the statistics $\mathcal{S}_{[\phi_1(\cdot),\ln(\cdot),\rho_1(\cdot)]}$ and $\mathcal{S}_{[\phi_1(\cdot),[-1+(\cdot)^{\alpha}]/\alpha,\rho_1(\cdot)]}$.[15]

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

## 4. ASYMPTOTIC PROPERTIES OF THE NON-NESTED TEST STATISTICS

In this section, we derive the asymptotic distribution of the test statistics in different settings. First, we consider the case in which only $H_g$ holds and there are no vectors $\gamma \in \mathcal{G}$ such that $E_{P_0}[h(z,\gamma)] = 0$. Second, we prove the consistency of the class of test statistics. Finally, we investigate the properties of the test statistics under local alternatives.

### 4.1. Correctly Specified Model Versus Misspecified Model

We show that the test statistic (3.1) is asymptotically normal when $H_g$ holds and there are no vectors $\gamma \in \mathcal{G}$ such that $E_{P_0}[h(z,\gamma)] = 0$; that is, $\mathcal{P}_g$ is correctly specified, while $\mathcal{P}_h$ is misspecified.

We also introduce a Lagrange multiplier (LM) and GMM-type statistic, which are asymptotically equivalent to $\mathcal{S}_v$ and $\tilde{\mathcal{S}}_v$. These statistics require the estimation of the vector $A_{0,v} = E_{P_0}\left[v_{P_0}\left(\eta_0^{*\prime} h\left(z,\gamma_0^*\right)\right) g(z,\beta_0)'\right]$, where $v_{P_0}(\cdot) = v_b\left(v_c(\cdot)/E_{P_0}[v_c(\cdot)]\right)$. Given a consistent estimator for $A_{0,v}$, say $\hat{A}_v$, a LM statistic is given by $\mathcal{LM}_v = \hat{A}_v\sqrt{n}\hat{\lambda}$. Additionally, we propose the GMM-type statistic $\mathcal{J}_v = -\hat{A}_v\hat{\Omega}_g^{-1}\sqrt{n}\hat{g}(\hat{\beta})$, where $\hat{\Omega}_g$ is a consistent estimator of $\Omega_{0,g}$.

Assumption 4.1 is sufficient to prove the asymptotic normality of the test statistics proposed in this section. Let $v_{a,j}(a) = \partial^j v_a(a)/\partial a^j$, $j = 1,2$, $v_{c,1}(a) \equiv \partial v_c(a)/\partial a$.

**Assumption 4.1.** (a) $v_a(\cdot)$ is continuously differentiable at 1 with $v_{a,1}(1) = 1$, $v_b(\cdot)$ is a continuous function, and $v_c(\cdot)$ satisfies $v_c(0) \neq 0$ and is continuously differentiable at 0; (b) The first moments under $P_0$ of the following functions are finite for all $(\gamma', \eta', a)$ in a neighborhood of $\left(\gamma_0^{*\prime}, \eta_0^{*\prime}, E_{P_0}\left[v_c\left(\eta_0^{*\prime} h\left(z,\gamma_0^*\right)\right)\right]\right)'$: (i) $\left|v_c\left(\eta' h(z,\gamma)\right)\right|$, (ii) $v_b\left(v_c\left(\eta' h(z,\gamma)\right)/a\right))^2$; (c) $\sigma_0^2 = A_{0,v} P_{0,g} A_{0,v}' > 0$.

Assumption 4.1(a) is a smoothness condition on $v_a(\cdot)$, $v_b(\cdot)$, and $v_c(\cdot)$, while Assumptions 4.1(b) are standard domination conditions required to apply a uniform weak law of large numbers. Assumption 4.1(c) guarantees that the proposed statistics have a nondegenerate asymptotic distribution.

We prove the following theorem in the online supplementary material.

THEOREM 4.1. *If the data set $\{z_i\}_{i=1}^n$ is an independent and identically distributed sample drawn from $P_0$, $\hat{\Omega}_g \xrightarrow{p} \Omega_{0,g}$ and Assumptions 2.1, 2.2, 2.3, and 4.1 hold, then $\mathcal{S}_v$ converges in distribution to $\mathcal{N}\left(0,\sigma_0^2\right)$. Furthermore, $\tilde{\mathcal{S}}_v$, $\mathcal{LM}_v$ and $\mathcal{J}_v$ are asymptotically equivalent to $\mathcal{S}_v$.*

Let $\hat{D}_g$ be a consistent estimator of $D_{0,g}$, and define $\hat{\Sigma}_g \equiv \left(\hat{D}_g'\hat{\Omega}_g^{-1}\hat{D}_g\right)^{-1}$ and $\hat{P}_g \equiv \hat{\Omega}_g^{-1} - \hat{\Omega}_g^{-1}\hat{D}_g\hat{\Sigma}_g\hat{D}_g'\hat{\Omega}_g^{-1}$. Consider also the following alternative estimators for $A_{0,v}$: $\hat{A}_{v,1} = \sum_{i=1}^n v_b\left(n\hat{q}_i^{v_c}\right)\hat{g}_i'/n$, $\hat{A}_{v,2} = \sum_{i=1}^n v_b\left(\hat{q}_i^{v_c}/\hat{p}_i^{v_c}\right)\hat{g}_i'/n$, $\hat{A}_{v,3} = \sum_{i=1}^n v_b\left(\hat{q}_i^{v_c}/\hat{p}_i^{v_c}\right)\hat{g}_i'\hat{p}_i$ and $\hat{A}_{v,4} = \sum_{i=1}^n v_b\left(n\hat{q}_i^{v_c}\right)\hat{g}_i'\hat{p}_i$. The following

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

theorem proves that $\hat{\sigma}_j^2 = \hat{A}_{v,j}\hat{P}\hat{A}'_{v,j}$, $j = 1,\ldots,4$, are consistent estimators of $\sigma_0^2$.

THEOREM 4.2. *If the data set* $\{z_i\}_{i=1}^n$ *is an independent and identically distributed sample drawn from* $\mathrm{P}_0$, $\hat{\Omega}_g \xrightarrow{p} \Omega_{0,g}$, $\hat{D}_g \xrightarrow{p} D_{0,g}$ *and Assumptions 2.1, 2.3, and 4.1 hold, then* $\hat{\sigma}_j^2 \xrightarrow{p} \sigma_0^2$, $j = 1,\ldots 4$.

A test for the non-nested hypothesis can be implemented by using any of the statistics stated in Theorem 4.1 divided by any consistent estimator of $\sigma_0$. Theorem 4.2 shows that possible estimators for $\sigma_0$ are given by $\hat{\sigma}_j$, $j = 1,\ldots,4$. The asymptotic distribution of such a statistic is standard normal. Consequently, we reject the null hypothesis $\mathrm{H}_g$ if the absolute value of the observed value of the statistic is larger than a critical value, where the critical value is obtained from the standard normal distribution. In the Monte Carlo study in Section 6, the estimators for $\Omega_{0,g}$ and $D_{0,g}$ are $\hat{\Omega}_g \equiv \sum_{i=1}^n g_i(\hat{\beta})g_i(\hat{\beta})'/n$ and $\hat{D}_g \equiv n^{-1}\sum_{i=1}^n \partial g_i(\hat{\beta})/\partial\beta'$, respectively. These estimators are consistent under Assumptions 2.1 and 2.2.[16]

## 4.2. Consistency of the Test Statistics

In this subsection, we assume that there are no vectors $\beta \in \mathcal{B}$ such that $\mathrm{E}_{\mathrm{P}_0}[g(z,\beta)] = 0$; that is, the model $\mathcal{P}_g$ is misspecified. We assume also that the GEL estimator converges in probability for the pseudo-true value. The pseudo-true value is defined as $\beta_0^* = \arg\min_{\beta\in\mathcal{B}} \mathrm{E}_{\mathrm{P}_0}[\rho(\lambda_0(\beta)'g(z,\beta))]$, where $\lambda_0(\beta) = \arg\max_{\lambda\in\Lambda} \mathrm{E}_{\mathrm{P}_0}\left[\rho\left(\lambda' g(z,\beta)\right)\right]$, $\lambda_0^* \equiv \lambda_0(\beta_0^*)$ and $\Lambda$ is a compact parameter space (see Section 2.1.2).

**Assumption 4.2.** $\hat{\beta} \xrightarrow{p} \beta_0^*$ and $\hat{\lambda} \xrightarrow{p} \lambda_0^*$, where $\lambda_0^* \neq 0$.

The model $\mathcal{P}_h$ is allowed to be either correctly specified or misspecified.

**Assumption 4.3.** $\hat{\gamma} \xrightarrow{p} \gamma_0^*$, $\hat{\eta} \xrightarrow{p} \eta_0^*$.

In contradistinction to Assumption 2.3, Assumption 4.3 allows $\eta_0^* = 0$.

To simplify the notation, we define $\rho_1^{g,z} \equiv \rho_1(\lambda_0^{*\prime}g(z,\beta_0^*))/\mathrm{E}_{\mathrm{P}_0}[\rho_1(\lambda_0^{*\prime}g(z,\beta_0^*))]$, $v_c^{g,z} \equiv v_c(\lambda_0^{*\prime}g(z,\beta_0^*))/\mathrm{E}_{\mathrm{P}_0}[v_c(\lambda_0^{*\prime}g(z,\beta^*))]$ and $v_c^{h,z} \equiv v_c(\eta_0^{*\prime}h(z,\gamma_0^*))/\mathrm{E}_{\mathrm{P}_0}[v_c(\eta_0^{*\prime}h(z,\gamma_0^*))]$. The following conditions are required to prove the consistency of the tests based on $\mathcal{S}_v$ and $\tilde{\mathcal{S}}_v$.

**Assumption 4.4.** (a) The functions $v_a(.)$, $v_b(.)$, and $v_c(.)$ are continuous; (b) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero; (c) The first moments under $\mathrm{P}_0$ of the following functions are finite for all $(\beta',\lambda',\gamma',\eta',a,b,c)'$ in a neighborhood of $(\beta_0^{*\prime},\lambda_0^{*\prime},\gamma_0^{*\prime},\eta_0^{*\prime}$, $\mathrm{E}_{\mathrm{P}_0}[\rho_1(\lambda_0^{*\prime}g(z,\beta_0^*))]$, $\mathrm{E}_{\mathrm{P}_0}[v_c(\lambda_0^{*\prime}g(z,\beta_0^*))]$, $\mathrm{E}_{\mathrm{P}_0}[v_c(\eta_0^{*\prime}h(z,\gamma_0^*))])'$: (i) $|\rho_1(\lambda' g(z,\beta))|$, (ii) $|v_c(\lambda' g(z,\beta))|$, (iii) $|v_c(\eta' h(z,\gamma))|$, (iv) $\left|v_a\left(\frac{\rho_1(\lambda' g(z,\beta))}{a}\right)\frac{\rho_1(\lambda' g(z,\beta))}{a}v_b\left(\frac{v_c(\eta' h(z,\gamma))b}{cv_c(\lambda' g(z,\beta))}\right)\right|$, (v) $\left|v_a\left(\frac{\rho_1(\lambda' g(z,\beta))}{a}\right)\frac{\rho_1(\lambda' g(z,\beta))}{a}\right|$, (vi) $\left|\frac{\rho_1(\lambda' g(z,\beta))}{a}v_b\left(\frac{v_c(\eta' h(z,\gamma))b}{cv_c(\lambda' g(z,\beta))}\right)\right|$, (vii) $\left|v_a\left(\frac{\rho_1(\lambda' g(z,\beta))}{a}\right)\frac{\rho_1(\lambda' g(z,\beta))}{a}v_b\left(\frac{v_c(\eta' h(z,\gamma))}{c}\right)\right|$,

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

(viii) $\left|\frac{\rho_1(\lambda' g(z,\beta))}{a} v_b\left(\frac{v_c(\eta' h(z,\gamma))}{c}\right)\right|$; (d) $\mathrm{E}_{\mathrm{P}_0}\left[v_a\left(\rho_1^{g,z}\right)\rho_1^{g,z} v_b\left(v_c^{h,z}/v_c^{g,z}\right)\right] \neq \mathrm{E}_{\mathrm{P}_0}\left[v_a\left(\rho_1^{g,z}\right)\rho_1^{g,z}\right]\mathrm{E}_{\mathrm{P}_0}\left[\rho_1^{g,z} v_b\left(v_c^{h,z}/v_c^{g,z}\right)\right]$; and (e) $\mathrm{E}_{\mathrm{P}_0}\left[v_a\left(\rho_1^{g,z}\right)\rho_1^{g,z} v_b\left(v_c^{h,z}\right)\right] \neq \mathrm{E}_{\mathrm{P}_0}\left[v_a\left(\rho_1^{g,z}\right)\rho_1^{g,z}\right]\mathrm{E}_{\mathrm{P}_0}\left[\rho_1^{g,z} v_b\left(v_c^{h,z}\right)\right]$.

Conditions 4.4(a), (b), and (c) are conditions on smoothness and the existence of moments and ensure that some uniform weak law of large numbers holds. The crucial conditions that ensure the consistency of tests based on $\mathcal{S}_v$ and $\tilde{\mathcal{S}}_v$ are given by Assumptions 4.4(d) and (e), respectively.

Since $\mathcal{LM}_v$ and $\mathcal{J}_v$ depend on the estimators $\hat{A}_v$ and $\hat{\Omega}_g$, which are chosen by the practitioner, we provide high-level conditions, given by Assumption 4.5, that ensure the consistency of tests based on these statistics.

**Assumption 4.5.** (a) $\hat{A}_v \xrightarrow{p} A^*_{0,v}$, where $A^*_{0,v}\lambda^*_0 \neq 0$ (b) $\hat{\Omega}_g \xrightarrow{p} \Omega^*_{0,g}$, where $\Omega^*_{0,g}$ is positive definite and $\hat{g}(\hat{\beta}) \xrightarrow{p} \mathrm{E}_{\mathrm{P}_0}[g(z,\beta^*_0)]$, where $A^*_{0,v}\Omega^{*-1}_{0,g}\mathrm{E}_{\mathrm{P}_0}[g(z,\beta^*_0)] \neq 0$.

The following theorem guarantees the consistency of the tests based on the proposed non-nested test statistics provided that the estimator used for $\sigma_0^2$ converges in probability to a finite positive constant.

THEOREM 4.3. *If the data set $\{z_i\}_{i=1}^n$ is an independent and identically distributed sample drawn from* $\mathrm{P}_0$ *and Assumptions 4.2 and 4.3 hold, then* (*1*) $\mathcal{S}_v \xrightarrow{p} \pm\infty$ *if Assumptions 4.4*(*a*), (*b*) *and* (*c*) *are also satisfied;* (*2*) $\tilde{\mathcal{S}}_v \xrightarrow{p} \pm\infty$ *if Assumptions 4.4*(*a*), (*b*) *and* (*d*) *are also satisfied;* (*3*) $\mathcal{LM}_v \xrightarrow{p} \pm\infty$ *if Assumption 4.5*(*a*) *is also satisfied; and* (*4*) $\mathcal{J}_v \xrightarrow{p} \pm\infty$ *if Assumptions 4.5*(*a*) *and* (*b*) *are also satisfied.*

## 4.3. Distribution Under Local Alternatives

*4.3.1. Asymptotics for the Proposed Class of Test Statistics.* In this section, the asymptotic distribution of the proposed class of tests is obtained under a sequence of local alternatives. Henceforth, we use the notation $u_n \to u$ if a non-stochastic sequence $u_n$ converges to a limit $u$ as $n$ goes to infinity.

We assume that the data correspond to a triangular array $\{z_{in}\}_{i=1}^n$ that is independent and identically distributed across $i = 1, \ldots, n$ and satisfies

$$\mathrm{H}_{g,n} : \mathrm{E}_{\mathrm{P}_{0n}}\left[g\left(z_{i,n}, \beta_{0,n}\right)\right] = \delta_g/\sqrt{n},\ i = 1, \ldots, n, \tag{4.1}$$

where $\mathrm{P}_{0n}$ is the true probability measure of $z_{in}$ and $\beta_{0,n} \to \beta_0$. Local alternatives of the form (4.1) were also considered by Newey (1985) to derive the asymptotic distributions of GMM and the statistic based on GMM for overidentifying conditions under local alternatives.

The matrices $\Omega_{0,g}$, $D_{0,g}$, and $A_{0,v}$ are redefined below to accommodate the fact that we are now working with triangular arrays.

To prove consistency and asymptotic normality of the GEL estimator under (4.1) we consider the following assumption.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

**Assumption 4.6.** (a) $\mathrm{E}_{\mathrm{P}_{0n}}\left[g\left(z_{i,n},\beta_{0,n}\right)\right]=\delta_g/\sqrt{n}$, with $\beta_{0,n}\to\beta_0$; (b) $\mathrm{E}_{\mathrm{P}_{0n}}[g(z_{in},\beta)]\to g(\beta)$ for each $\beta\in\mathcal{B}$ and $g(\beta)$ has a unique zero at $\beta_0$ (c) $\mathcal{B}$ is compact; (d) $g(z_{i,n},\beta)$ is twice continuously differentiable in $\beta$ for all $z_{i,n}\in\mathbb{R}^k$; (e) For positive constants $\delta$ and $C$, $\mathrm{E}_{\mathrm{P}_{0n}}[\sup_{\beta\in\mathcal{B}}\|g(z_{in},\beta)\|^{2+\delta}]\leq C$, $\mathrm{E}_{\mathrm{P}_{0n}}\left[\sup_{\beta\in\mathcal{B}}\left\|\partial g(z_{i,n},\beta)/\partial\beta'\right\|^2\right]\leq C$, and $\mathrm{E}_{\mathrm{P}_{0n}}\left[\sup_{\beta\in\mathcal{B}}\left\|\partial^2 g(z_{i,n},\beta)/\partial\beta'\partial\beta'\right\|\right]\leq C$ for all $n\geq 1$; (f) $\mathrm{E}_{\mathrm{P}_{0n}}[g(z_{i,n},\beta_0)g(z_{i,n},\beta_0)']\to\Omega_{0,g}$ and $\Omega_{0,g}$ is nonsingular; (g) $\beta_0\in int(\mathcal{B})$; (h) $\mathrm{E}_{\mathrm{P}_{0n}}[\partial g(z_{i,n},\beta_0)/\partial\beta']\to D_{0,g}$ where $rank(D_{0,g})=p_g$; and (i) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero.

Recall from Section 2.1.2 that $\Sigma_{0,g}\equiv\left(D_{0,g}'\Omega_{0,g}^{-1}D_{0,g}\right)^{-1}$, $H_{0,g}\equiv\Sigma_{0,g}D_{0,g}'\Omega_{0,g}^{-1}$ and $P_{0,g}\equiv\Omega_{0,g}^{-1}-\Omega_{0,g}^{-1}D_{0,g}\Sigma_{0,g}D_{0,g}'\Omega_{0g}^{-1}$. Lemma SM1.9 in the online supplementary material reveals that under (4.1) and Assumption 4.6 $\sqrt{n}\hat{\lambda}\xrightarrow{d}\mathcal{N}(-P_{0,g}\delta_0,P_{0,g})$, which allows us to obtain the distribution of the non-nested test statistic.

The pseudo-true value now is redefined as $\gamma_{0,n}^*=\arg\min_{\gamma\in\mathcal{G}}\mathrm{E}_{\mathrm{P}_{0,n}}\left[\rho\left(\eta_{0,n}(\gamma)'h\left(z_{i,n},\gamma\right)\right)\right]$, where $\eta_{0,n}(\gamma)=\arg\max_{\eta\in\mathcal{H}}\mathrm{E}_{\mathrm{P}_{0,n}}\left[\rho\left(\eta'h\left(z_{i,n},\gamma\right)\right)\right]$ and $\eta_{0,n}^*\equiv\eta_{0,n}(\gamma_{0,n}^*)$. We assume that the sequence $(\gamma_{0,n}^{*\prime},\eta_{0,n}^{*\prime})'$ is convergent and that the estimators $(\hat{\gamma}',\hat{\eta}')$ converge to the same limit of that sequence.

**Assumption 4.7.** (a) $\mathcal{G}$ and $\mathcal{H}$ are compact sets; and (b) $\gamma_{0,n}^*\to\gamma^*$, $\eta_{0,n}^*\to\eta^*$, $\hat{\gamma}\xrightarrow{p}\gamma^*$ and $\hat{\eta}\xrightarrow{p}\eta^*$, where $\eta^*\neq 0$.

Let $A_{\mathrm{P}_{0,n},v}(\mu)=\mathrm{E}_{\mathrm{P}_{0,n}}[v_{\mathrm{P}_{0,n}}(\eta'h(z_{i,n},\gamma))g(z_{i,n},\beta)']$, where $v_{\mathrm{P}_{0,n}}(\cdot)=v_b(v_c(\cdot)/\mathrm{E}_{\mathrm{P}_{0,n}}[v_c(\cdot)])$ and $\mu=(\beta',\gamma',\eta')'$. The following assumptions are required to derive the asymptotic distribution of the proposed statistics under local alternatives.

**Assumption 4.8.** (a) $v_a(\cdot)$ is continuously differentiable at 1 with $v_{a,1}(1)=1$, $v_b(\cdot)$ is continuously differentiable on its domain and $v_c(\cdot)$ satisfies $v_c(0)\neq 0$ and is continuously differentiable on its domain; (b) $h(z_{i,n},\gamma)$ is continuously differentiable in $\gamma$ for all $z_{i,n}\in\mathbb{R}^k$; (c) For positive constants $\delta$ and $C$ and a compact set $\mathcal{A}$, $\mathrm{E}_{\mathrm{P}_{0,n}}\left[v_c\left(\eta'h\left(z_{i,n},\gamma^*\right)\right)\right]\in\mathcal{A}$, $n\geq 1$ and there exists a function $d(z_{i,n})$ such that $\mathrm{E}_{\mathrm{P}_{0,n}}\left[d(z_{i,n})\right]\leq C$ for all $n\geq 1$, where $f(z_{i,n},\gamma,\eta,a)\leq d(z_{i,n})$ for all $\gamma\in\mathcal{G}$, $\eta\in\mathcal{H}$, $a\in\mathcal{A}$, and $z_{i,n}\in\mathbb{R}^k$, and $f(z_{i,n},\gamma,\eta,a)$ is any of the functions $\left|v_c\left(\eta'h\left(z_{i,n},\gamma\right)\right)\right|^{1+\delta}$, $\left\|\partial v_c\left(\eta'h\left(z_{i,n},\gamma\right)\right)/\partial\theta'\right\|$, where $\theta=(\gamma',\eta')'$, $v_b\left(v_c\left(\eta'h\left(z_{i,n},\gamma\right)\right)/a\right)^2$ and $\left\|\partial v_b\left(v_c\left(\eta'h\left(z_{i,n},\gamma\right)\right)/a\right)/\partial\kappa'\right\|^2$, where $\kappa=(\gamma',\eta',a)'$; (d) $A_{\mathrm{P}_{0,n},v}(\mu^*)\to A_{0,v}$, where $\mu^*=(\beta_0',\gamma^{*\prime},\eta^{*\prime})'$; and (e) $\sigma_0^2=A_{0,v}P_{0,g}A_{0,v}'>0$.

The asymptotic distribution of the non-nested test statistics under local alternatives can now be obtained.

THEOREM 4.4. *If the data set $\{z_{in}\}_{i=1}^n$ is an independent and identically distributed sample across $i=1,\ldots,n$ and is drawn from* $\mathrm{P}_{0n}$ *and Assumptions 4.6,*

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

*4.7, and 4.8 hold, then* $\mathcal{S}_v$ *converges in distribution to* $\mathcal{N}\left(-A_{0,v}P_{0,g}\delta_g, \sigma_0^2\right)$. *Furthermore,* $\tilde{\mathcal{S}}_v$, $\mathcal{LM}_v$ *and* $\mathcal{J}_v$ *are asymptotically equivalent to* $\mathcal{S}_v$.

Proposition 3 of Newey (1985) reveals that $\mathcal{S}_v$ is asymptotically locally most powerful against (4.1) if and only if $\delta_g = -A'_{0,v}a$ for some scalar $a$ (cf. Sect. 3.3 of Ramalho and Smith, 2002). This result clarifies that different choices of $v_b(.)$ and $v_c(.)$ yield test statistics that are optimal against (4.1) for different vectors $\delta_g$ or, in other words, in different local directions. Hence, when choosing $v_b(.)$ and $v_c(.)$, the practitioner is implicitly deciding the direction against which the test is optimal.

We can also use the results obtained by Newey (1985) for GMM tests under local alternatives to compare the local power of the proposed Cox's non-nested tests with that of Hansen (1982)'s test for overidentifying restrictions. First note that the non-nested test statistics are asymptotically equivalent to a GMM test statistic (see Theorem 4.1). Because of the properties of the non-central chi-square distribution, the local power of a GMM test increases with the parameter of non-centrality and decreases with the degrees of freedom (see Newey, 1985, Sect. 3, p. 238). The number of degrees of freedom of the Hansen (1982) test statistic is $m_g - p_g$ and consequently it is larger or equal to that of the proposed test statistics which is 1. Proposition 6 of Newey (1985) implies that the proposed non-nested test statistics and Hansen (1982) test statistic for overidentifying restrictions have the same non-centrality parameter when $\delta_g = -A'_{0,v}a$. It follows that when $m_g - p_g > 1$ the proposed non-nested tests have larger local power than the Hansen (1982)'s test.[17] Suppose now that $\delta_g \neq -A'_{0,v}a$. In this case Proposition 6 of Newey (1985) shows that Hansen (1982)' test statistic has a parameter of non-centrality greater than or equal to that of any of the Cox's non-nested test statistics. Therefore, when $m_g - p_g > 1$, we cannot draw any conclusions about which is the test with largest local power. If $m_g - p_g = 1$, Hansen (1982) test has larger or equal local power than the proposed non-nested tests.

*4.3.2. Comparison with GMM Non-Nested Test Statistics.* In this section, we discuss and compare the properties of the proposed tests with the existing Cox non-nested tests based on GMM.

The first non-nested test statistic for overidentifying moment restrictions that was proposed in the literature and does not resort to any parametric assumption was introduced by Singleton (1985), who considered competing models based on the same number of moment restrictions, i.e., $m_g = m_h$. The Singleton (1985) statistic is given by $\mathcal{J}_{\text{KJS}} = (\hat{h}(\tilde{\gamma}) - \hat{g}(\tilde{\beta}))'\tilde{\Omega}_g^{-1}\sqrt{n}\hat{g}(\tilde{\beta})$, where $\tilde{\beta}$ is the two-step GMM estimator (Hansen, 1982) for $\beta_0$, $\tilde{\Omega}_g$ is a consistent estimator of $\Omega_{0,g}$, $\hat{h}(\gamma) \equiv \sum_{i=1}^n h(z_i, \gamma)/n$, $\tilde{\gamma}$ is the two-step GMM estimator for $\gamma^\dagger$, the unique minimizer of $\mathrm{E}_{\mathrm{P}_0}[h(z,\gamma)]'\Omega_{0,h}^{-1}(\gamma^\star)\mathrm{E}_{\mathrm{P}_0}[h(z,\gamma)]$, where $\Omega_{0,h}(\gamma) = \mathrm{E}_{\mathrm{P}_0}[h(z,\gamma)h(z,\gamma)']$, and $\gamma^\star$ is the pseudo-true value associated with the estimator used in the first step.[18]

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

Smith (1992) showed how one can allow the alternative models to have different numbers of moment restrictions by proposing the statistic $\mathcal{J}_{\text{RJS}} = \hat{h}(\tilde{\gamma})' \tilde{\Omega}_{h,\dagger}^{-1} \hat{C}_g \sqrt{n} \hat{g}(\tilde{\beta})$, where $\hat{C}_g$ is a $(m_h \times m_g)$ matrix chosen by the practitioner and converges to a matrix $C_g$, which is finite and has rank equal to $\min\{m_g, m_h\}$, and $\tilde{\Omega}_{h,\dagger}$ is a consistent estimator of $\Omega_{h,\dagger} \equiv \Omega_{0,h}(\gamma^\dagger)$. Choosing $\hat{C}_g = \tilde{\Omega}_{h,\dagger} \tilde{\Omega}_g^{-1}$ corresponds to a statistic that is asymptotically equivalent to Singleton (1985)'s statistic. In the simulations, we consider the matrix $\hat{C}_g = n^{-1} \sum_{i=1}^n h(z_i, \tilde{\gamma}) g(z_i, \tilde{\beta}) \tilde{\Omega}_g^{-1}$ that was proposed by Ramalho and Smith (2002, ftn. 10).[19]

Smith (1992) demonstrated that in the framework of Section 4.3.1, $\mathcal{J}_{\text{RJS}}$ converges in distribution to $\mathcal{N}(\mu_{\text{RJS}}, \sigma_{\text{RJS}}^2)$, where $\mu_{\text{RJS}} \equiv \mu_h^{\dagger\prime} \Omega_{h,\dagger}^{-1} C_g \Omega_{0,g} P_{0,g} \delta_0$, $\mu_h^\dagger \equiv \text{E}_{\text{P}_0}[h(z, \gamma^\dagger)]$ and $\sigma_{\text{RJS}}^2 \equiv \mu_h^{\dagger\prime} \Omega_{h,\dagger}^{-1} C_g \Omega_{0,g} P_{0,g} \Omega_{0,g} C_g' \Omega_{h,\dagger}^{-1} \mu_h^\dagger$ under standard regularity conditions. Consequently, this statistic is optimal against (4.1) if and only if $\delta_g = \Omega_{0,g} C_g' \Omega_{h,\dagger}^{-1} \mu_h^\dagger a$, for some scalar $a$. Therefore, different choices of $C_g$ lead to test statistics that are locally optimal against (4.1) for different vectors $\delta_g$. Comparing this result to that obtained in Section 4.3.1, we can conclude that this test and the new GEL non-nested tests are optimal against departures from the null in different directions. Additionally, we can see that the issue of choosing the matrix $\hat{C}_g$ is analogous to that of choosing $v_b(.)$ and $v_c(.)$.[20]

## 5. LOCAL ASYMPTOTIC THEORY

### 5.1. Framework and Asymptotic Results

Let $T_n$ denote a generic non-nested test statistic, such as those introduced in Sections 3.1 and 4.1. If model $\text{H}_g$ is true and if $\text{E}_{\text{P}_0}[h(z, \gamma_0^*)] \neq 0$, where $\gamma_0^*$ is the GEL pseudo-true value as defined in Section 2.1.2, then under some regularity conditions, the asymptotic distribution of $T_n$ is normal with mean zero and a particular asymptotic variance $\sigma_0^2$ (cf. Theorem 4.1), provided that $\sigma_0^2 > 0$. Let $V_n$ be a consistent estimator of $\sigma_0^2$ (e.g. $\hat{\sigma}_j^2$ $j = 1, \ldots 4$). The standardized non-nested test statistics have the form $W_n = T_n / \sqrt{V_n}$, and their asymptotic distribution is standard normal if $\sigma_0^2 > 0$. Note, however, that in finite samples $\text{E}_{\text{P}_0}[h(z, \gamma_0^*)]$ can be approximately zero leading the asymptotic variance $\sigma_0^2$ to be equal to zero. In such cases, the pointwise asymptotic theory developed in Section 4.1 under $\text{P}_0$ might lead to tests that are considerably oversized. In this section we show that an approach similar to that Shi (2015a) proposed in the context of Vuong tests leads to correct inferences.

Define the pseudo-true values under the probability measure P as $(\gamma_{\text{P}}^{*\prime}, \eta_{\text{P}}^{*\prime})'$, which are defined as $\gamma_{\text{P}}^* = \arg\min_{\gamma \in \mathcal{G}} \text{E}_{\text{P}}[\rho(\eta_{\text{P}}(\gamma)' h(z, \gamma))]$, where $\eta_{\text{P}}(\gamma) = \arg\max_{\eta \in \mathcal{H}} \text{E}_{\text{P}}[\rho(\eta' h(z, \gamma))]$ and $\eta_{\text{P}}^* \equiv \eta_{\text{P}}(\gamma_{\text{P}}^*)$. Denote also the maximum norm of a vector $a = (a_1, a_2, \ldots, a_m)'$ as $\|a\|_\infty = \max\{|a_1|, |a_2|, \ldots, |a_m|\}$. Let $\mathcal{E}_{\text{P},h} = \text{E}_{\text{P}}[h(z, \gamma_{\text{P}}^*)]$ and consider a sequence of probability measures, $\{\text{P}_n\}_{n=1}^\infty$, where $\sqrt{n} \mathcal{E}_{\text{P}_n,h}$ is approximately constant, in particular $\sqrt{n} \mathcal{E}_{\text{P}_n,h} \to \delta_h$, where

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

$\|\delta_h\|_\infty \in [0, +\infty]$.[21] Note that if $\|\delta_h\|_\infty < +\infty$, $\mathcal{E}_{P_n,h}$ converges to zero at rate $n^{1/2}$, while if $\|\delta_h\|_\infty = +\infty$, $\mathcal{E}_{P_n,h}$ is different from the zero vector.

To simplify the notation, denote $v_{b,1}(a) = \partial v_b(a)/\partial v$, $\kappa_{v_c} \equiv v_{c,1}(0)/v_c(0)$. We require the following regularity conditions.

**Assumption 5.1.** (a) $v_a(\cdot)$ is twice continuously differentiable at 1 with $v_{a,1}(1) = 1$, $v_b(\cdot)$ satisfies $v_b(1) = 0$ and is continuously differentiable on its domain with $v_{b,1}(1) = 1$ and $v_c(\cdot)$ satisfies $v_c(0) \neq 0$, is continuously differentiable on its domain and satisfies $\kappa_{v_c} = 1$; (b) $\mathcal{B}$ and $\mathcal{G}$ are compact sets, and $\mathcal{H}$ is a compact and convex set; (c) $g(z,\beta)$ and $h(z,\gamma)$ are twice continuously differentiable in $\beta$ and $\gamma$ respectively for all $z \in \mathbb{R}^k$; and (d) $\rho(v)$ is twice continuously differentiable in a neighborhood of zero, and $\rho_1 = \rho_2 = -1$.

Note that Assumption 5.1(a) is satisfied by all the examples considered in Section 3.3.

Let $s(z,\varphi) = (g(z,\beta)', h(z,\gamma)')'$, where $\varphi = (\beta', \gamma')'$. The maintained hypothesis is the following.

DEFINITION 5.1. *For positive constants $\delta$ and $C$ and a compact set $\mathcal{A}$, let $\mathcal{P}$ be the set of probability measures,* P, *on $\mathbb{R}^k$ such that the following statements hold*: (*a*) $H_n \subset \mathcal{H}$ *for* $n \geq 1$, *where* $H_n = \{\eta \in \mathbb{R}^{m_k} : \|\eta\| \leq n^{-\zeta}\}$ *with* $1/(2+\delta) < \zeta < 1/2$; (*b*) *There exists a function $d^*(z)$ such that* $E_P[d^*(z)] \leq C$, *where* $f^*(z,\beta,\gamma,\eta,a) \leq d^*(z)$ *for all* $\beta \in \mathcal{B}$, $\gamma \in \mathcal{G}$, $\eta \in \mathcal{H}$, $a \in \mathcal{A}$ *and* $z \in \mathbb{R}^k$, *and* $f^*(z,\beta,\gamma,\eta,a)$ *is any of the functions* $\|s(z,\varphi)\|^{2+\delta}$, $\|\partial s(z,\varphi)/\partial\varphi'\|^2$, $\|\partial^2 s(z,\varphi)/\partial\varphi\partial\varphi'\|$, $|v_c(\eta' h(z,\gamma))|^{1+\delta}$, $\|\partial v_c(\eta' h(z,\gamma))/\partial\theta'\|$, *where* $\theta = (\gamma', \eta')'$, $v_b(v_c(\eta' h(z,\gamma))/a))^2$ *and* $\|\partial v_b(v_c(\eta' h(z,\gamma))/a)/\partial\kappa'\|^2$, *where* $\kappa = (\gamma', \eta', a)'$; (*c*) $E_P[v_c(\eta' h(z,\gamma_P^*))] \in \mathcal{A}$; *and* (*d*) *the data set* $\{z_i\}_{i=1}^n$ *is an independent and identically distributed sample drawn from* P.

Condition (a) of Definition 5.1 is required to investigate the asymptotic properties of $\hat{\eta}$ and $\hat{\gamma}$ in the case in which $\|\delta_h\|_\infty < +\infty$. Condition (b) in Definition 5.1 presents the moment conditions needed to derive the properties of the statistics $W_n$ under drifting sequences of probability measures.

Definition 5.2 below defines the null hypothesis.

DEFINITION 5.2. *Let $\mathcal{P}_0$ be the set of probability measures,* P, *on $\mathbb{R}^k$ such that the following statements hold*: (*a*) $P \in \mathcal{P}$; (*b*) $E_P[g(z,\beta_{0,P})] = 0$, *for a unique value* $\beta_{0,P} \in \mathcal{B}$.

We derive the asymptotic distribution of $W_n$, when $\|\delta_h\|_\infty = +\infty$ and $\|\delta_h\|_\infty < +\infty$. Note that the standardized statistics can be written as $W_n = \sqrt{n}T_n / \sqrt{nV_n}$. We show that when $\|\delta_h\|_\infty < +\infty$, $\sqrt{n}T_n \xrightarrow{d} T$ and $nV_n \xrightarrow{d} V$ for some random variables $T$ and $V$, where $V$ is non-negative with probability approaching one, and consequently, $W_n \xrightarrow{d} T/\sqrt{V}$. The asymptotic behaviors of $\sqrt{n}T_n$ and $nV_n$ are investigated in this section for different choices of $T_n$ and $V_n$.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

The following specific forms of the statistic $\sqrt{n}T_n$ are considered: $\sqrt{n}\mathcal{S}_\upsilon$, $\sqrt{n}\tilde{\mathcal{S}}_\upsilon$,

$$\sqrt{n}\mathcal{LM}_{\upsilon,j} = \hat{A}_{\upsilon,j}\sqrt{n}\hat{\lambda}, \tag{5.1}$$

$$\sqrt{n}\mathcal{J}_{\upsilon,j} = -\hat{A}_{\upsilon,j}\hat{\Omega}^{-1}\sqrt{n}\hat{g}(\hat{\beta}), \tag{5.2}$$

$$j = 1,\ldots,4.$$

We show that all of these statistics are asymptotically equivalent if $\|\delta_h\|_\infty = +\infty$, but this equivalence is not maintained when $\|\delta_h\|_\infty < +\infty$. We also investigate the asymptotic distribution of the statistics $nV_n$: $n\hat{\sigma}_j^2$, $j = 1,\ldots,4$.

Let $\Omega_{\mathrm{P}}(\varphi) = \mathrm{E}_{\mathrm{P}}[s(z,\varphi)s(z,\varphi)']$, $D_{\mathrm{P},g}(\beta) = \mathrm{E}_{\mathrm{P}}[\partial g(z,\beta)/\partial\beta]$, $D_{\mathrm{P},h}(\gamma) = \mathrm{E}_{\mathrm{P}}[\partial h(z,\gamma)/\partial\gamma']$ and $A_{\mathrm{P},\upsilon}(\mu) = \mathrm{E}_{\mathrm{P}}[\upsilon_{\mathrm{P}}(\eta' h(z,\gamma))g(z,\beta)']$, where $\upsilon_{\mathrm{P}}(\cdot) = \upsilon_b(\upsilon_c(\cdot)/\mathrm{E}_{\mathrm{P}}[\upsilon_c(\cdot)])$ and $\mu = (\beta',\gamma',\eta')'$. We now define the sequences of probability measures along which we obtain the asymptotic distributions of the statistics.

DEFINITION 5.3. *Let* $\mathrm{Seq}\left(\beta_0,\gamma^*,\eta^*,\delta_h,\Omega,D_g,D_h,A_\upsilon\right)$ *be the set of sequences* $\{\mathrm{P}_n\}_{n=1}^\infty$ *such that* $\mathrm{P}_n \in \mathcal{P}$ *for every* $n$, $\beta_{0,\mathrm{P}_n} \to \beta_0$, *where* $\beta_0 \in int(\mathcal{B})$, $\gamma^*_{\mathrm{P}_n} \to \gamma^*$, $\eta^*_{\mathrm{P}_n} \to \eta^*$, $\sqrt{n}\mathcal{E}_{\mathrm{P}_n,h} \to \delta_h$, *where* $\|\delta_h\|_\infty \in [0,+\infty]$, $\Omega_{\mathrm{P}_n}(\varphi^*) \to \Omega$, *where* $\varphi^* = (\beta_0',\gamma^{*\prime})'$, $D_{\mathrm{P}_n,g}(\beta_0) \to D_g$, $D_{\mathrm{P}_n,h}(\gamma^*) \to D_h$, $A_{\mathrm{P}_n,\upsilon}(\mu^*) \to A_\upsilon$, *where* $\mu^* = (\beta_0',\gamma^{*\prime},\eta^{*\prime})'$, $\mathrm{E}_{\mathrm{P}_n}(g(z,\beta)) \to g(\beta)$ *for each* $\beta \in \mathcal{B}$, *where* $g(\beta)$ *has a unique zero at* $\beta_0$, *and* $\mathrm{E}_{\mathrm{P}_n}(h(z,\gamma)) \to h(\gamma)$ *for each* $\gamma \in \mathcal{G}$. *The matrix* $\Omega$ *is positive definite, its upper-left block* $m_g \times m_g$, *denoted* $\Omega_g$ *is positive definite, and* $rank(D_g) = p_g$. *Furthermore, if* $\|\delta_h\|_\infty \in [0,+\infty)$, *then* $\gamma^* \in int(\mathcal{G})$, *the lower-right block* $m_h \times m_h$ *of* $\Omega$, *denoted as* $\Omega_h$, *is positive definite, and* $rank(D_h) = p_h$, $h(\gamma)$ *has a unique zero at* $\gamma^*$ *and if* $\|\delta_h\|_\infty = +\infty$, *then* $h(\gamma) \neq 0$ *for* $\forall\gamma \in \mathcal{G}$, $\eta^*_{\mathrm{P}_n} \to \eta^* \neq 0$ *and* $\sigma^2 \equiv A_\upsilon' P_g A_\upsilon > 0$, *where* $P_g \equiv \Omega_g^{-1} - \Omega_g^{-1}D_g\Sigma_g D_g'\Omega_g^{-1}$ *and* $\Sigma_g \equiv \left(D_g'\Omega_g^{-1}D_g\right)^{-1}$.

The following assumption is a high-level assumption similar to Assumption 6.2(a) of Shi (2015a).

**Assumption 5.2.** Under any sequence $\{\mathrm{P}_n\}_{n=1}^\infty \in \mathrm{Seq}\left(\beta_0,\gamma^*,\eta^*,\delta_h,\Omega,D_g,D_h,A_\upsilon\right)$ for some $\left(\beta_0,\gamma^*,\eta^*,\delta_h,\Omega,D_g,D_h,A_\upsilon\right)$, we have $\hat{\eta} \xrightarrow{p} \eta^*$ and $\hat{\gamma} \xrightarrow{p} \gamma^*$ if $\|\delta_h\|_\infty = +\infty$.

Let $S_g = (I_{m_g}, 0_{m_g,m_h})$ and $S_h = (0_{m_h,m_g}, I_{m_h})$ be selection matrices such that $S_g s(z,\varphi) = g(z,\beta)$ and $S_h s(z,\varphi) = h(z,\gamma)$. Consider the Cholesky decomposition of the matrix $\Omega = C_\Omega C_\Omega'$; let $m = m_g + m_h$ and $z$ be a $m$-dimensional random vector that is multivariate normally distributed with mean 0 and covariance matrix $I_m$, i.e., $z \sim \mathcal{N}(0, I_m)$. Define $w(\delta,\Omega) = \delta + C_\Omega z$, where $\delta$ is a $m$-vector and write $\mathcal{T}(\delta,\Omega,Q) = w(\delta,\Omega)'Qw(\delta,\Omega)$. Consider the matrices $Q_1 \equiv S_h' P_h \Omega_{hg} P_g S_g - S_g' P_g S_g$, where $\Omega_{hg}$ is the lower-left block $m_h \times m_g$ of $\Omega$, $P_h \equiv \Omega_h^{-1} - \Omega_h^{-1}D_h\Sigma_h D_h'\Omega_h^{-1}$ and $\Sigma_h \equiv \left(D_h'\Omega_h^{-1}D_h\right)^{-1}$, $Q_2 \equiv S_h' P_h \Omega_{hg} P_g S_g$, $Q_3 \equiv K_h P_g K_h'$ and $Q_4 \equiv (K_h - K_g)P_g(K_h - K_g)'$ where $K_h \equiv -S_h' P_h \Omega_{hg}$ and $K_g \equiv -S_g' P_g \Omega_g$.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

The following theorem presents the asymptotic distributions of the statistics described above under each sequence in $\text{Seq}\left(\beta_0, \gamma^*, \eta^*, \delta_h, \boldsymbol{\Omega}, D_g, D_h, A_v\right)$.

THEOREM 5.1. *Suppose that Assumption 5.1 holds. Under a sequence* $\{\mathrm{P}_n\}_{n=1}^{\infty} \in \text{Seq}\left(\beta_0, \gamma^*, \eta^*, \delta_h, \boldsymbol{\Omega}, D_g, D_h, A_v\right)$ *for some* $\left(\beta_0, \gamma^*, \eta^*, \delta_h, \boldsymbol{\Omega}, D_g, D_h, A_v\right)$ *with* $\mathrm{P}_n \in \mathcal{P}_0$ *for every n and satisfying Assumption 5.2, the following statements hold*:

1. *If* $\|\delta_h\|_\infty = +\infty$, *then* $\mathcal{S}_v$ *converges in distribution to* $\mathcal{N}\left(0, \sigma^2\right)$ *and* $\tilde{\mathcal{S}}_v$, $\mathcal{LM}_v$ *and* $\mathcal{J}_v$ *are asymptotically equivalent to* $\mathcal{S}_v$. *Additionally,* $\hat{\sigma}_j^2$, $j = 1, \ldots, 4$ *are consistent estimators of* $\sigma^2$.
2. *If* $\|\delta_h\|_\infty < +\infty$, *then*
   (a) $\sqrt{n}\mathcal{S}_v \xrightarrow{d} \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_1\right)$ *and* $\sqrt{n}\mathcal{J}_{v,2}$, $\sqrt{n}\mathcal{J}_{v,3}$, $\sqrt{n}\mathcal{LM}_{v,2}$ *and* $\sqrt{n}\mathcal{LM}_{v,3}$ *are asymptotically equivalent to* $\sqrt{n}\mathcal{S}_v$.
   (b) $\sqrt{n}\tilde{\mathcal{S}}_v \xrightarrow{d} \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_2\right)$ *and* $\sqrt{n}\mathcal{J}_{v,1}$, $\sqrt{n}\mathcal{J}_{v,4}$, $\sqrt{n}\mathcal{LM}_{v,1}$ *and* $\sqrt{n}\mathcal{LM}_{v,4}$ *are asymptotically equivalent to* $\sqrt{n}\tilde{\mathcal{S}}_v$.
   (c) $n\hat{\sigma}_1^2 \xrightarrow{d} \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_3\right)$ *and* $n\hat{\sigma}_4^2$ *is asymptotically equivalent to* $n\hat{\sigma}_1^2$.
   (d) $n\hat{\sigma}_2^2 \xrightarrow{d} \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_4\right)$ *and* $n\hat{\sigma}_3^2$ *is asymptotically equivalent to* $n\hat{\sigma}_2^2$.
   (e) $n\hat{\sigma}_j^2$ *is non-negative with probability approaching one for* $j = 1, 2, 3, 4$.

Hence, under each sequence in $\text{Seq}\left(\beta_0, \gamma^*, \eta^*, \delta_h, \boldsymbol{\Omega}, D_g, D_h, A_v\right)$, it follows from the continuous mapping theorem that the standardized non-nested test statistics $W_n = T_n \big/ \sqrt{V_n}$ converge to a standard normal variate when $\|\delta_h\|_\infty = +\infty$, and if $\|\delta_h\|_\infty < +\infty$, it converges to one of the following random variables depending on the statistic $T_n$ and estimator $V_n$ chosen: $W_{i,j}^*(\delta_h) = \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_i\right) \big/ \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_j\right)^{1/2}$ $i = 1,2$, $j = 3,4$.[22] In Section SM2.2 of the online supplementary material we investigate the limit of the random variable $W_{i,j}^*(\delta_h)$ as some elements of $\delta_h$ approach infinity. We reveal in that section that this limit is path dependent as if we take the limit along some paths, $W_{i,j}^*(\delta_h)$ converges to $\mathcal{N}(0,1)$, but if we take the limit along different paths the random variable $W_{i,j}^*(\delta_h)$ is not affected.[23]

## 5.2. Shi-Type Statistics

Shi (2015a, Sects. 3.2 and 3.3) noted that the tails of distribution functions of random variables such as $\mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_i\right) \big/ \mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_j\right)^{1/2}$ $i = 1,2$, $j = 3,4$ are very different from those of the standard normal not only because $\text{E}[\mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_i\right)]$ is different from zero but also because $\mathcal{T}\left(S_h'\delta_h, \boldsymbol{\Omega}, Q_j\right)$ can take values close to zero with significant probability. Consequently, using critical values obtained from the tables of the standard normal random variable can lead to severe size distortions of statistical tests (see Shi, 2015a, Thm. 3.2). In this

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

section, we propose Shi-type statistics to solve this problem (see Shi, 2015a, Sect. 4.1).[24,25]

The general form of this statistic is $W_n^s = (T_n - \widehat{bias}/\sqrt{n})/(V_n + c \cdot \widehat{af}/n)^{0.5}$, where $\widehat{bias}$ is an estimator of $\mathrm{E}[\mathcal{T}(S_h'\delta_h, \Omega, Q_i)]$, $c$ is a non-negative constant, and $\widehat{af}$ is an adjustment factor that is added because, as mentioned above, the asymptotic distribution $nV_n$ can take values close to zero with some probability. A consequence of Theorem 5.1 is that the precise mathematical expressions of $\widehat{bias}$ and $\widehat{af}$ depend on the choice of the statistic $T_n$ and the estimator $V_n$ of $\sigma^2$. Let $T_n(1)$ be one of the statistics $\mathcal{S}_v$, $\mathcal{J}_{v,2}$, $\mathcal{J}_{v,3}$, $\mathcal{LM}_{v,2}$ or $\mathcal{LM}_{v,3}$, and $T_n(2)$ denote one of the statistics $\tilde{\mathcal{S}}_v$, $\mathcal{J}_{v,1}$, $\mathcal{J}_{v,4}$, $\mathcal{LM}_{v,1}$ and $\mathcal{LM}_{v,4}$. Additionally, define $V_n(3)$ the estimators $\hat{\sigma}_1^2$ or $\hat{\sigma}_4^2$ and $V_n(4)$ the estimators $\hat{\sigma}_2^2$ or $\hat{\sigma}_3^2$.

To propose the test statistics, we require consistent estimators for $Q_i$, $Q_j$, and $\Omega$. An estimator for $\Omega$ is given by $\hat{\Omega} \equiv \sum_{i=1}^n s_i(\hat{\varphi}) s_i(\hat{\varphi})'/n$, where $\hat{\varphi} = (\hat{\beta}', \hat{\eta}')'$. Let $\hat{\Omega}_g \equiv \sum_{i=1}^n g_i(\hat{\beta}) g_i(\hat{\beta})'/n$, $\hat{D}_g \equiv n^{-1} \sum_{i=1}^n \partial g_i(\hat{\beta})/\partial \beta'$, $\hat{\Omega}_h \equiv \sum_{i=1}^n h_i(\hat{\gamma}) h_i(\hat{\gamma})'/n$, $\hat{D}_h \equiv n^{-1} \sum_{i=1}^n \partial h_i(\hat{\gamma})/\partial \gamma'$, $\hat{\Omega}_{hg} \equiv \sum_{i=1}^n h_i(\hat{\gamma}) g_i(\hat{\beta})'/n$, $\hat{P}_g \equiv \hat{\Omega}_g^{-1} - \hat{\Omega}_g^{-1} \hat{D}_g \hat{\Sigma}_g \hat{D}_g' \hat{\Omega}_g^{-1}$, $\hat{\Sigma}_g \equiv (\hat{D}_g' \hat{\Omega}_g^{-1} \hat{D}_g)^{-1}$, $\hat{P}_h \equiv \hat{\Omega}_h^{-1} - \hat{\Omega}_h^{-1} \hat{D}_h \hat{\Sigma}_h \hat{D}_h' \hat{\Omega}_h^{-1}$ and $\hat{\Sigma}_h \equiv (\hat{D}_h' \hat{\Omega}_h^{-1} \hat{D}_h)^{-1}$. Consequently, consistent estimators for these matrices are given by $\hat{Q}_1 \equiv S_h' \hat{P}_h \hat{\Omega}_{hg} \hat{P}_g S_g - S_g' \hat{P}_g S_g$, $\hat{Q}_2 \equiv S_h' \hat{P}_h \hat{\Omega}_{hg} \hat{P}_g S_g$, $\hat{Q}_3 \equiv \hat{K}_h \hat{P}_g \hat{K}_h'$ and $\hat{Q}_4 \equiv (\hat{K}_h - \hat{K}_g) \hat{P}_g (\hat{K}_h - \hat{K}_g)'$, where $K_h \equiv -S_h' \hat{P}_h \hat{\Omega}_{hg}$ and $\hat{K}_g \equiv -S_g' \hat{P}_g \hat{\Omega}_g$.

Note that under $\mathrm{E}_{\mathrm{P}_n}[g(z, \beta_{\mathrm{P}_n})] = 0$, and if $\|\delta_h\|_\infty < +\infty$, the mean of the statistic asymptotic distribution of $\sqrt{n} T_n(i)$ is equal to $\mathrm{tr}(Q_i \Omega)$, for $i = 1, 2$, while the mean of the asymptotic distribution of $nV_n(j)$ is $\mathrm{tr}(Q_j \Omega) + \delta_h' S_h Q_j S_h' \delta_h$, for $j = 3, 4$. Therefore, we consider the following Shi-type statistics

$$W_n^s(i,j) = \frac{T_n(i) - \mathrm{tr}(\hat{Q}_i \hat{\Omega})/\sqrt{n}}{\sqrt{V_n(j) + c_{ij} \cdot \mathrm{tr}(\hat{Q}_j \hat{\Omega})/n}}, i = 1, 2, j = 3, 4,$$

where $c_{ij}$ are non-negative constants that depend on the choices of the statistic $T_n(i)$ and estimator $V_n(j)$. We describe a data-dependent method to choose $c_{ij}$ below. Note that $\mathrm{tr}(\hat{Q}_j \hat{\Omega})$ is an estimator of $\mathrm{tr}(Q_j \Omega)$, which is the only component of the mean of the asymptotic distribution of $nV_n(j)$ that is estimable.[26]

Let $cv(1 - \tau, S_h'\delta_h, \Omega, Q_i, Q_j, c_{ij})$ denote the $1 - \tau$ quantile of $|\mathcal{W}_{ij}(S_h'\delta_h, \Omega, Q_i, Q_j)|$, where $\mathcal{W}_{ij}(S_h'\delta_h, \Omega, Q_i, Q_j) = [\mathcal{T}(S_h'\delta_h, \Omega, Q_i) - \mathrm{tr}(Q_i \Omega)]/[\mathcal{T}(S_h'\delta_h, \Omega, Q_j) + c_{ij} \cdot \mathrm{tr}(Q_j \Omega)]^{1/2}$, $i = 1, 2$; $j = 3, 4$ and let $c_h$ be a large positive constant. To implement the statistical test, we define the critical value as

$$cv^*(1 - \tau, \hat{\Omega}, \hat{Q}_i, \hat{Q}_j) = \max\left\{ \max_{\|\delta_h\|_\infty \in [0, c_h]} cv(1 - \tau, S_h'\delta_h, \hat{\Omega}, \hat{Q}_i, \hat{Q}_j, c_{ij}), z_{1-\tau/2} \right\},$$

where $z_{1-\tau/2}$ corresponds to the quantile $1 - \tau/2$ of the standard normal random variable. The reason that we take the maximum of critical values over $\|\delta_h\|_\infty \in [0, c_h]$ rather than $\|\delta_h\|_\infty \in [0, +\infty)$ is technical. Note

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

that $\max_{\|\delta_h\|_\infty \in [0,c_h]} cv\left(1-\tau, S_h'\delta_h, \Omega, Q_i, Q_j, c_{ij}\right)$ is continuous in $\Omega, Q_i$ and $Q_j$ by the maximum theorem because the set $[0, c_h]$ is compact. This result is used in the proof of Theorem 5.2 below. The new test rejects $H_0$: $E_P\left[g(z,\beta_{0,P})\right] = 0$ if $\left|W_n^s(i,j)\right| > cv^*(1-\tau, \hat{\Omega}, \hat{Q}_i, \hat{Q}_j)$. Let $cv_{ij}(1-\tau,\Delta) = \max_{\|\delta_h\|_\infty \in \Delta} cv\left(1-\tau, S_h'\delta_h, \Omega, Q_i, Q_j, c_{ij}\right)$ and $\Delta \subseteq [0,+\infty)$, $i = 1,2$; $j = 3,4$. To be able to prove that the proposed test has well-controlled asymptotic size, we require the following assumption.

**Assumption 5.3.** $c_h$ is chosen such that $cv_{ij}(1-\tau,[0,c_h]) = cv_{ij}(1-\tau,[0,+\infty))$, $i = 1,2$; $j = 3,4$.

Assumption 5.3 is a high-level assumption that requires the constant $c_h$ to be chosen large enough such that the maximum of the critical values over $\|\delta_h\|_\infty \in [0,+\infty)$ does not change if we restrict $\|\delta_h\|_\infty$ to be in $[0,c_h]$. Theorem 5.2 provides the desired result.

THEOREM 5.2. *Suppose that Assumptions 5.1, 5.2, and 5.3 hold. Then, for any* $c_{ij} \geq 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} \Pr\left(\left|W_n^s(i,j)\right| > cv^*(1-\tau,\hat{\Omega},\hat{Q}_i,\hat{Q}_j)\right) \leq \tau,\ i=1,2;\ j=3,4.$$

## 5.3. Implementation Issues

The following steps describe how to compute the critical value $cv^*(1-\tau,\hat{\Omega},\hat{Q}_i,\hat{Q}_j)$ in practice based on the statistic $W_n^s(i,j)$ for a particular choice of $i$ and $j$ using the software MATLAB.[27]

1. Obtain the GEL estimators $(\hat{\beta}', \hat{\lambda}')'$, $(\hat{\gamma}', \hat{\eta}')'$ and the estimators $\hat{Q}_i$, $\hat{Q}_j$ and $\hat{\Omega}$.
2. Choice of the constant $c_{ij}$: Fix $\delta_h = 0$, and choose a level $\tau$. We use the MATLAB implementation of the golden section method to find $c_{ij}^* = \arg\min_{c_{ij}\in[0,100]} \left|cv\left(1-\tau,0,\hat{\Omega},\hat{Q}_i,\hat{Q}_j,c_{ij}\right) - z_{1-\tau/2}\right|$, where $cv\left(1-\tau,0,\hat{\Omega},\hat{Q}_i,\hat{Q}_j,c_{ij}\right)$ is obtained via simulation for each $c_{ij}$.
3. Obtain $\delta_h^* = \arg\max_{\|\delta_h\|_\infty\in[0,c_h]} cv\left(1-\tau,S_h'\delta_h,\hat{\Omega},\hat{Q}_i,\hat{Q}_j,c_{ij}^*\right)$, where $c_h = 10$, using the MATLAB implementation of the genetic algorithm, where for each $\delta_h$, $cv\left(1-\tau,S_h'\delta_h,\hat{\Omega},\hat{Q}_i,\hat{Q}_j,c_{ij}^*\right)$ is obtained via simulation. We use $\delta_h^*$ to initialize the implementation of the MATLAB simplex search algorithm to ensure local optima are obtained. The optimum obtained is denoted as $\delta_h^{**}$.[28]
4. Define $cv^*(1-\tau,\hat{\Omega},\hat{Q}_i,\hat{Q}_j) = \max\left\{cv\left(1-\tau,S_h'\delta_h^{**},\hat{\Omega},\hat{Q}_i,\hat{Q}_j,c_{ij}^*\right), z_{1-\tau/2}\right\}$.

## 5.4. Power Properties

The proof of consistency of the modified tests is similar to that of standard non-nested tests (see Section 4.2) because the corrections made to the statistics con-

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

verge to zero if the alternatives are fixed under standard regularity conditions. Hence, we only analyze here the power properties under local alternatives of the following form:

**Assumption 5.4.** The sequence of true probability measures $\{\mathrm{P}_n\}_{n=1}^{\infty} \in \mathrm{Seq}\left(\beta_0, \gamma^*, \eta^*, \delta_h, \Omega, D_g, D_h, A_v\right)$ for some $\left(\beta_0, \gamma^*, \eta^*, \delta_h, \Omega, D_g, D_h, A_v\right)$, weakly converges to a $\mathrm{P}_0 \in \mathcal{P}_0$ and satisfies $\sqrt{n}\mathrm{E}_{\mathrm{P}_n}(g(z, \beta_{0,\mathrm{P}_n})) \to \delta_g \neq 0$ and $\|\delta_g\|_{\infty} < \infty$.

THEOREM 5.3. *Suppose that Assumptions 5.1, 5.2, and 5.4 hold. Then, for any* $c_{ij} \geq 0$,

$$\lim_{n\to\infty} \Pr_{\mathrm{P}_n}\left(\left|W_n^s(i,j)\right| > cv^*(1-\tau, \hat{\Omega}, \hat{Q}_i, \hat{Q}_j)\right)$$
$$= \begin{cases} 1 - F_{|\mathcal{W}_{ij}|}(cv_{ij}^*) & \text{if } \|\delta_h\|_{\infty} < +\infty \\ \Phi\left(-cv_{ij}^* - \frac{A_v P_g \delta_g}{\sigma}\right) + \Phi\left(-cv_{ij}^* + \frac{A_v P_g \delta_g}{\sigma}\right) & \text{if } \|\delta_h\|_{\infty} = +\infty \end{cases},$$

*where* $cv_{ij}^* = cv^*(1-\tau, \Omega, Q_i, Q_j)$ *and* $F_{|\mathcal{W}_{ij}|}(\cdot)$ *is the distribution function of* $\left|\mathcal{W}_{ij}(\delta, \Omega, Q_i, Q_j)\right|$ *with* $\delta = (\delta_g', \delta_h')$ *and* $\mathcal{W}_{ij}(\delta, \Omega, Q_i, Q_j) = [\mathcal{T}(\delta, \Omega, Q_i) - \mathrm{tr}(Q_i\Omega)] / \left[\mathcal{T}\left(\delta, \Omega, Q_j\right) + c_{ij} \cdot \mathrm{tr}(Q_j\Omega)\right]^{1/2}$, $i = 1, 2$, $j = 3, 4$ *and* $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution.*

Several conclusions can be drawn from the results of Theorem 5.3. First, since $\mathrm{E}[\mathcal{T}(\delta, \Omega, Q_i)] = \mathrm{tr}(Q_i\Omega) + \delta' Q_i \delta$, there is bias in the numerator that is not eliminated by the modified statistic, and consequently, the tests are expected to have some power when $\|\delta_h\|_{\infty} < +\infty$. Second, the random variables $\mathcal{W}_{ij}(\delta, \Omega, Q_i, Q_j)$, $i = 1, 2$, $j = 3, 4$ do not depend on $v_a(\cdot)$, $v_b(\cdot)$ and $v_c(\cdot)$. Therefore, when $\|\delta_h\|_{\infty} < +\infty$, the choice of the functions $v_a(\cdot)$, $v_b(\cdot)$ and $v_c(\cdot)$ does not affect the power of the tests. Finally, Theorems 5.2 and 5.3 show that the asymptotic distributions of the statistics $W_n^s(i, j)$ under the null and local alternatives when $\|\delta_h\|_{\infty} < +\infty$ depend crucially on the argument of the function $v_b(\cdot)$ and not on the form of this function. Specifically, these asymptotic distributions are different if the argument of the function $v_b(\cdot)$ is $\hat{q}_i^{v_c} / \hat{p}_i^{v_c}$ or $n\hat{q}_i^{v_c}$.

## 6. MONTE CARLO STUDY

This section reports the results of a Monte Carlo study on the performance of the non-nested test statistics. We consider the following different settings: (*i*) model $\mathrm{H}_g$ is true and for $\forall \gamma \in \mathcal{G}$, $\mathrm{E}_{\mathrm{P}_0}[h(z, \gamma)] \neq 0$, (*ii*) both models $\mathrm{H}_g$ and $\mathrm{H}_h$ are true, and (*iii*) there is a local departure from $\mathrm{H}_g$. In settings (*i*) and *(ii),* we evaluate the size of the tests based on the proposed statistics and their modified versions. In setting (*iii*), we assess the power of the tests and compare it to that of the optimal test.

In the simulation study on the size and power of the tests, 5,000 replications were performed and two instrumental variable models were compared. The model

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

specified under the null is

$$y = \beta_g x_g + u_g, \tag{6.1}$$

where the vector of instruments is $w_g = (w_{g1}, w_{g2}, w_{g3})'$. The alternative model is

$$y = \beta_h x_h + u_h \tag{6.2}$$

and the vector of instruments in this case is given by $w_h = (w_{h1}, w_{h2}, w_{h3})'$. We consider competing models that depend on a single parameter to ease the computational burden associated with GEL estimation.

## 6.1. Experimental Design

To evaluate the size and power of the tests, it is necessary to generate the vector $(x_g, u_g, w_g', x_h, u_h, w_h')'$. In the reported simulations, this vector is multivariate normal distributed with mean zero and a positive definite variance covariance matrix that has ones along the main diagonal. The off-diagonal elements, corresponding to the correlation coefficients, are defined below. Henceforth, the correlation between the random variables $a$ and $b$ is denoted as $\rho_{a,b}$.

*6.1.1. Size.* To evaluate the size of the statistical tests, the true model is considered to be (6.1) with $\beta_g = 1$. The following values for the correlations are chosen: $\rho_{u_g,x_g} = \rho_{u_g,x_h} = 0.2$, and $\rho_{x_g,w_{hj}} = 0.4$, $j = 1, 2, 3$. Because the instruments under the null are valid, the correlation between the errors and instruments is zero, i.e., $\rho_{u_g,w_{gj}} = 0$ for all $j$.

The correlation between $u_g$ and the elements of $w_h$ satisfies the condition $\rho_{u_g,w_{h2}} = \rho_{u_g,w_{h3}} = -\rho_{u_g,w_{h1}}$. Consequently, if $\rho_{u_g,w_{h2}} \neq 0$, there is no value of $\gamma$ that satisfies the moment restrictions $\mathrm{E}[(y - \gamma x_h)w_h] = 0$ under model (6.1). To compare the results on the size of the tests for different degrees of misspecification of the alternative model, three values for $\rho_{u_g,w_{h2}}$ are considered: $\rho_{u_g,w_{h2}} = 0$ (no misspecification), $\rho_{u_g,w_{h2}} = 0.2$ and $\rho_{u_g,w_{h2}} = 0.4$.

The correlations between $x_g$ and $w_{gj}$ are identical for all $j$, i.e., $\rho_{x_g,w_{gj}} = 0.45$ for all $j$. To ensure that there are no values of $\beta$ that satisfy the condition $\mathrm{E}((y - x_g\beta)w_g) = 0$ under (6.2), we set $\rho_{u_h,w_{g1}} = \rho_{u_h,w_{g2}} = 0.3$ and $\rho_{u_h,w_{g3}} = 0$. In the experiments, we define $\rho_{x_g,x_h} = 0.15$, $\rho_{w_{gj},w_{hj}} = 0.4$ for all $j$ and $\rho_{w_{gi},w_{hj}} = 0$ for $i \neq j$. We also set $\rho_{x_h,w_{hj}} = 0.35$ for all $j$, $\rho_{x_h,u_h} = 0.2$, $\rho_{x_g,u_h} = 0.25$ and $\rho_{u_g,u_h} = 0.4$. The remaining correlations not defined above are set to zero.

To evaluate the empirical size of the tests, we consider random samples of size $n$ from the distribution of $(y, x_g, w_g', x_h, w_h')$. The models specified under the null and the alternative are estimated with $\{y_i, x_{gi}, w_{gi}'\}_{i=1}^n$ and $\{y_i, x_{hi}, w_{hi}'\}_{i=1}^n$, respectively.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

*6.1.2. Power.* To study the empirical power of the tests, we set $\rho_{u_g,w_{h2}} = 0.4$ and define the remaining correlations as above. We generate a local departure from the model (6.1) by considering a sequence $\lambda_n = \omega n^{-1/2}$, where $\omega$ is a positive constant, and the model

$$y = \beta_* x^* + u^*, \tag{6.3}$$

where $x^* = \left(\lambda_n x_h + (1-\lambda_n)x_g\right)/\sigma_{x,n}$, $\sigma^2_{x,n} = \text{var}(\lambda_n x_h + (1-\lambda_n)x_g)$, $u^* = \left(\lambda_n u_h + (1-\lambda_n)u_g\right)/\sigma_{u,n}$, $\sigma^2_{u,n} = \text{var}(\lambda_n u_h + (1-\lambda_n)u_g)$, and the instruments are given by $w^* = \left(\lambda_n w_h + (1-\lambda_n)w_g\right)/\sigma_{w,n}$, $\sigma^2_{w,n} = \text{var}(\lambda_n w_{hj} + (1-\lambda_n)w_{gj})$. Note that this model satisfies the moment restrictions

$$\text{E}((y - \beta_* x^*)w^*) = a_n\delta/\sqrt{n} + O(1/n), \tag{6.4}$$

where $\delta = \text{E}(u_h w_g') + \text{E}(u_g w_h')$ and $a_n = \omega/(\sigma_{u,n}\sigma_{w,n})$, $\sigma_{u,n}\sigma_{w,n} \to 1$. To investigate the power properties of the tests, we consider different values for $\omega$ and generate $y$ via (6.3) with $\beta_* = 1$. In this case, to evaluate the empirical power of the tests, we obtain random samples of size $n$ from the distribution of $(y^*, x^*, w^{*\prime}, x_h, w_h')$. The model specified under the null is estimated with the sample $\left\{y_i, x_i^*, w_i^{*\prime}\right\}_{i=1}^n$, while the model specified under the alternative is estimated with the observations $\{y_i, x_{hi}, w_{hi}'\}_{i=1}^n$.

## 6.2. Test Statistics

Given the large number of the statistics and estimators for the asymptotic variance proposed here, only a selection of the results is presented. We report the results for the statistics $\mathcal{S}_\upsilon$ (eq 3.1), $\tilde{\mathcal{S}}_\upsilon$ (eq 3.2), $\mathcal{LM}_{\upsilon,1}$ (eq 5.1), and $\mathcal{J}_{\upsilon,1}$ (eq 5.2). To simplify the notation, we write the last two statistics as $\mathcal{LM}_\upsilon$ and $\mathcal{J}_\upsilon$.

We present the results for the proposed statistics based on EL and the function $\upsilon(\cdot) = [\phi_1(\cdot), \upsilon_b(\cdot), \rho_1(\cdot)]$ with $\upsilon_b(\cdot) = \left[-1 + (\cdot)^\alpha\right]/\alpha$ (see Section 3.3.3) for particular choices of $\alpha$.[29] To simplify the notation of the statistics for this choice of $\upsilon(\cdot)$, we refer to the statistics $\mathcal{S}_\upsilon$, $\tilde{\mathcal{S}}_\upsilon$, $\mathcal{LM}_\upsilon$ and $\mathcal{J}_\upsilon$ as $\mathcal{S}_\alpha$, $\tilde{\mathcal{S}}_\alpha$, $\mathcal{LM}_\alpha$ and $\mathcal{J}_\alpha$, respectively. We use the subscript $\alpha$ to refer to the value of $\alpha$ chosen in the transformation $\upsilon_b(\cdot) = \left[-1 + (\cdot)^\alpha\right]/\alpha$. The subscript 0 refers to the case in which $\alpha \to 0$ (i.e., $\upsilon_b(\cdot) = \ln(\cdot)$). The functions $\phi_1(\cdot)$ and $\rho_1(\cdot)$ are those implied by EL (see Sections 2.1.2 and 3.3.3). We consider the results for the cases of $\alpha \to 0$, $\alpha = 1, 1.5, 2, 3$ for illustrative purposes.

We present the results on the size and power of the statistical tests based on standardized statistics, i.e., the statistics divided by the square root of an estimator of its asymptotic variance, which we denote in the Tables A1 and A2 in the Appendix as "original", and on Shi-type versions of these statistics. The critical values for the tests based on the Shi-type statistics were obtained using 500 replications.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

For conciseness, we present here only the results for the statistics based on the estimator of its asymptotic variance that provided the best results in terms of the size of the tests in the Monte Carlo study. These were $\hat{\sigma}_3^2$ for $\mathcal{S}_v$ and $\hat{\sigma}_1^2$ for $\tilde{\mathcal{S}}_v$, $\mathcal{LM}_v$ and $\mathcal{J}_v$ (see Section 4.1).[30]

We also report the results for the Cox tests based on the GMM statistics of Singleton (1985) and Smith (1992) (see Section 4.3.2 and ftn. 26). For comparison purposes, we consider tests for overidentifying restrictions based on the GMM statistic (Hansen, 1982) and the likelihood ratio statistic computed with EL (Qin and Lawless, 1994), which are labelled $\mathcal{J}_{\text{GMM}}$ and $\mathcal{LR}_{\text{EL}}$ respectively.

Additionally, we present the results for the tests based on the unfeasible optimal GMM statistic against (6.4) (see Prop. 3 of Newey, 1985), which is given by $\mathcal{J}_{\text{OPT}} = \delta' \tilde{\Omega}_g^{-1} \sqrt{n} \hat{g}(\tilde{\beta})$, where $\delta = \text{E}(u_h w_g') + \text{E}(u_g w_h')$ and $\tilde{\Omega}_g = \sum_{i=1}^n g_i(\tilde{\beta}) g_i(\tilde{\beta})'/n$.[31] Note that the power of a test based on this statistic provides a theoretical upper bound for the power of any specification test.[32]

## 6.3. Summary and Recommendations

We summarize here the results obtained below. The original GMM tests computed using $\mathcal{J}_{\text{KJS}}$ performed well in terms of size and satisfactorily in terms of power. The tests based on $\mathcal{J}_{\text{RJS}}$ are considerably oversized, although the tests based on the Shi-type version of this statistic performed well not only in terms of size but also power.

The empirical size of the tests based on $\tilde{\mathcal{S}}_\alpha$, $\mathcal{LM}_\alpha$, and $\mathcal{J}_\alpha$ more closely approximates the nominal size than that of $\mathcal{S}_\alpha$. The tests based on the Shi-type versions of these statistics produced rejection frequencies close to or smaller than the nominal size under the null. However, their empirical power is considerably lower than that of the original test in the case of $\tilde{\mathcal{S}}_\alpha$ and $\mathcal{J}_\alpha$. Consequently, we advocate the use of the Shi-type statistic based on $\mathcal{LM}_\alpha^{\text{EL}}$ for low values of $\alpha$ ( $\alpha \to 0$ or $\alpha = 1$) to perform Cox-type non-nested tests, given their good empirical power properties.

In our study the empirical size and power properties of the tests for overidentifying restrictions are good.

## 6.4. Analysis of Size Results

Table A1 in Appendix A presents the empirical size for the non-nested test statistics detailed above for sample sizes 200 and 400. The nominal size for all tests reported is 0.05. The results obtained are discussed here.

*6.4.1. Correctly Specified Model Versus Misspecified Model.* We consider here the cases where $\rho_{u_g, w_{h2}} > 0$, which corresponds to a setting in which the model specified under the null is correctly specified and the model specified under the alternative is misspecified. We begin by analyzing the behavior of the original statistics, and we then focus on the results of the tests based on the Shi-type statistics.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

Concerning the GMM non-nested tests, we can see that those based on the Smith (1992) statistic are oversized and that those based on the Singleton (1985) statistic display rejection rates close to 0.05.

In general, the EL tests based on $\mathcal{S}_\alpha$ are oversized. The results improve slightly, but not substantially, when the sample size is equal to 400. In some simulations, not reported here, with different experimental designs, we found that the tests based on the statistic $\mathcal{S}_\alpha$ have size close to 0.05 for relatively high positive values of $\alpha$ and are oversized for other values of $\alpha$. Consequently, since the results of the tests based on $\mathcal{S}_\alpha$ are quite sensitive to the value of $\alpha$, we do not recommend them.[33]

Among the tests computed with $\tilde{\mathcal{S}}_\alpha$, $\mathcal{LM}_\alpha$, and $\mathcal{J}_\alpha$, those based on the latter statistic displayed the best size properties. However, the tests based on $\tilde{\mathcal{S}}_\alpha$ and $\mathcal{LM}_\alpha$ are slightly oversized.

The tests calculated with the Shi-type statistics perform better than those based on the original statistics in that their rejection frequencies are closer to or smaller than the nominal rate 0.05. This holds not only for the EL but also for the GMM statistics. Note, however, that tests computed using the Shi-type statistics based on $\tilde{\mathcal{S}}_\alpha$ are still slightly oversized for $\alpha \to 0$ and $\alpha = 1$.

In general, the tests for overidentifying restrictions perform well in terms of size.

*6.4.2. Competing Correctly Specified Models.* We focus here on the case in which both models are correctly specified, that is, $\rho_{u_g, w_{h2}} = 0$. In this case, apart from the tests computed using the statistics $\mathcal{S}_\alpha$, and those of Singleton (1985) and Smith (1992), all tests based on unmodified statistics, unexpectedly, have rejection frequencies close to 0.05.

In general, tests based on Shi (2015a) modifications of all statistics perform well in terms of size, although surprisingly, the rejection frequencies of the tests computed using the Shi-type statistics based $\mathcal{LM}_\alpha$ are marginally higher than those calculated using the original statistic.

The tests for overidentifying restrictions have good size properties, which is expected because they do not depend on the alternative model.

## 6.5. Analysis of the Power Results

Table A2 in Appendix A reports the empirical power of the tests at the nominal level of 0.05 for sample sizes of 200 and 400.

The results indicate that the tests based on the statistics $\mathcal{S}_v$ and $\mathcal{J}_{\text{RJS}}$ tend to have high power, although these tests are oversized. We also note that the test computed using Singleton's statistic $\mathcal{J}_{\text{KJS}}$ displays satisfactory empirical power, albeit lower than that of the optimal test.

Among the tests that performed well in terms of size, we can see that the rejection frequencies under the alternative of the tests based on the statistics $\tilde{\mathcal{S}}_\alpha$ increase with $\alpha$, while those based on $\mathcal{LM}_\alpha$ and $\mathcal{J}_\alpha$ decrease with $\alpha$.

We can note in Table A2 that the rejection frequencies of the tests based on the Shi-type statistics are much lower than those computed using the original

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

statistics in the case of $\tilde{\mathcal{S}}_\alpha$ and $\mathcal{J}_\alpha$. In contrast, in the case of $\mathcal{LM}_\alpha$, these rejection frequencies are very similar. Note that for low values of $\alpha$, the empirical power of these tests is very close to that of the optimal test.

Interestingly, the tests based on the Shi-type versions of the Singleton (1985) and Smith (1992) statistics yield rejection rates close to the original tests. In particular, the rejection rates of the tests calculated with the Shi-type version of the Smith (1992) statistic are very similar to those obtained with the optimal test.

The overidentifying moment tests possess good power characteristics in our study. Their empirical power is higher than that of most tests based on non-nested test statistics apart from the case of $\mathcal{LM}_0$ and $\mathcal{J}_0$.

## 7. CONCLUSION

In this article, new test statistics are proposed for non-nested hypothesis testing in models defined through overidentifying moment restrictions. It is shown that Ramalho and Smith (2002) test statistics are members of this class. Additional members can be derived using an artificial program that depends on a parameter $\theta$. This program nests the GMC programs under the null hypothesis and the alternative that are obtained by taking $\theta$ equal to zero and one, respectively. Consequently, a test for the non-nested hypothesis is equivalent to a test for the parametric restriction that $\theta$ is equal to zero. A score-type test is used to derive a test statistic in this context.

The asymptotic properties of the new non-nested test statistics are also investigated. We consider the case in which only one model is correctly specified under the null hypothesis and give conditions that ensure the consistency of the tests proposed. We obtain the distribution of the test statistics under local alternatives.

In addition, we introduce Cox's non-nested tests based on Shi-type statistics (Shi, 2015a) that have correct asymptotic size in a uniform sense and investigate their power properties under local alternatives.

The statistics that produce tests with good size and power properties are revealed in a Monte Carlo study. Based on the results obtained, we recommend using Shi-type versions of an LM statistic computed with EL and the Smith (1992) statistic.

### *NOTES*

1. See, for instance, the survey articles of MacKinnon (1983), Gouriéroux and Monfort (1994) and Pesaran and Weeks (2001).

2. The Ghysels and Hall (1990) statistic depends on a mean of a function computed under the null hypothesis. To evaluate this mean, the specification of the distribution functions of some random variables is required.

3. Atkinson's approach to develop non-nested test statistics was further explored by Quandt (1974) and Santos Silva (2001) who considered different artificial models.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

4. For EL $\phi(v) = -\ln(v)$, for exponential tilting $\phi(v) = v\ln(v)$, for CUE $\phi(v) = (v-1)^2/2$ and for CR, $\phi(v) = \left[v^{\tau+1} - v - \tau(v-1)\right]/[\tau(1+\tau)]$. Other forms for $\phi(v)$ can be found in Pardo (2006, p. 6).

5. For two discrete probability distributions, $p = \{p_1, p_2, \ldots\}$ and $q = \{q_1, q_2 \ldots\}$, the $f$-divergence between them is defined to be $D_\phi(p,q) = \sum_{i=1}^{\infty} p_i \phi(q_i/p_i)$, where $\phi(.)$ is a convex function continuous at zero such that $\phi(1) = 0$.

6. The EL and exponential tilting estimators arise with $\rho(v) = \ln(1-v)$, where the domain is $\mathcal{V} = (-\infty, 1)$ and $\rho(v) = -\exp(v)$, respectively. CUE arises with $\rho(v) = -(v+1)^2/2$, and CR arises with $\rho(v) = -(\tau+1)^{-1}(1+\tau v)^{(1+\tau)/\tau}$.

7. The implied probabilities were given by Owen (1988) for EL, by Kitamura and Stutzer (1997) for exponential tilting, by Back and Brown (1993) for quadratic $\rho(.)$, and by Brown and Newey (2002) for the general case.

8. See Qin and Lawless (1994), Imbens (1997), Kitamura and Stutzer (1997), Imbens et al. (1998) and Newey and Smith (2004). Smith (2007) and Kitamura (2007) demonstrated that this result holds if the inverse of $\phi(.)$ is homogenous.

9. See Lemma 1 of Hong et al. (2003) and the discussion on p. 937 of that article.

10. Note that the computation of GEL in (2.3) and (2.4) requires zero to be in the convex hull of both $\{g_i(\hat{\beta})\}_{i=1}^n$ and $\{h_i(\hat{\gamma})\}_{i=1}^n$ (see Kitamura, 2007, Sect. 8.1). Since this condition might not be satisfied in the case of EL, we suggest replacing this estimator by the adjusted EL estimator of Chen, Variyath, and Abraham (2008) in such cases.

11. See Lemma SA.1 in the supplemental appendix.

12. See equation (4.4) in Ramalho and Smith (2002, p. 108).

13. Note that since our test procedures are based on the absolute value of the statistics, any inference based on $\hat{S}$ or $-\hat{S}$ leads to the same decision.

14. The case of $\alpha = 1$ was also considered by Atkinson (1970) and Quandt (1974) and advocated by Santos Silva (2001).

15. There is a close relationship between the proposed non-nested test statistics in Section 3.3.3 and discrepancy measures. To observe this relationship, write $\hat{P} = (\hat{p}_1, \ldots, \hat{p}_n)$ and $\hat{Q} = (\hat{q}_1, \ldots, \hat{q}_n)$, and note that in the case of (3.6), we obtain $\bar{\omega}_{1,i} = -\hat{p}_i[-\ln(\hat{q}_i/\hat{p}_i) - \mathcal{MDI}(\hat{P}||\hat{Q})]$, where $\mathcal{MDI}$ is the minimum discrimination information measure (see Pardo, 2006, p. 6): $\mathcal{MDI}(\hat{P}||\hat{Q}) = -\sum_{\ell=1}^n \ln(\hat{q}_\ell/\hat{p}_\ell)\,\hat{p}_\ell$ and the terms $-\ln(\hat{q}_i/\hat{p}_i)$ are the individual components of this measure. For $\alpha \neq 1$, (3.7) can be re-written as $\bar{\omega}_{1,i} = (\alpha-1)\,\hat{p}_i[[-1 + (\hat{q}_i/\hat{p}_i)^\alpha]/[\alpha(\alpha-1)] - \mathcal{CR}_{\alpha-1}(\hat{P}||\hat{Q})]$, where $\mathcal{CR}_{\alpha-1}(\hat{P}||\hat{Q}) = \sum_{\ell=1}^n [-1 + (\hat{q}_\ell/\hat{p}_\ell)^\alpha]p_\ell / [\alpha(\alpha-1)]$ is the Cressie-Read discrepancy measure with parameter $\alpha - 1$. The term $[-1 + (\hat{q}_i/\hat{p}_i)^\alpha]/[\alpha(\alpha-1)]$ comprises the individual components of this measure. The link between discrepancy measures and the compound model approach does not hold if $\alpha = 1$, which corresponds to the case advocated by Quandt (1974) and Santos Silva (2001). Note that in this case, $\bar{\omega}_{1,i} = \hat{q}_i - \hat{p}_i$, which does not depend on any divergence measure.

16. In the cases in which the computation of GEL estimators $\hat{\beta}$ and $\hat{\gamma}$ is difficult, we can replace them by two-step GMM estimators $\tilde{\beta}$ and $\tilde{\gamma}$ and the corresponding Lagrange multipliers by solving $\tilde{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\tilde{\beta})} \hat{P}_n^g(\tilde{\beta}, \lambda)$, $\tilde{\eta} = \arg\max_{\eta \in M} \hat{P}_n^h(\tilde{\gamma}, \eta)$ as in Brown and Newey (2002). Since in the case of EL the convex-hull condition might not be satisfied, one should compute the Lagrange multipliers via the adjusted-EL of Chen et al. (2008). Although this approach leads to valid Cox non-nested test statistics (see Santos Silva, 2001, p. 579), these are not asymptotically equivalent to the proposed statistics based on GEL because GEL and GMM estimators converge to different pseudo-true values in the case of misspecified models.

17. See the discussion of Proposition 6 in Newey, 1985, p. 242.

18. If the estimator used in the first step is $\tilde{\gamma}^\star = \arg\min_{\gamma \in \mathcal{G}} \hat{h}(\gamma)' W_n \hat{h}(\gamma)$ and $W_n \xrightarrow{p} W$, where $W$ is a non-singular matrix, the pseudo-true value associated with this estimator is given by $\gamma^\star = \arg\min_{\gamma \in \mathcal{G}} \mathrm{E}[h(z,\gamma)]' W \mathrm{E}[h(z,\gamma)]$.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

19. Smith (1992) considered a number of possible alternative choices for the matrix $\hat{C}_g$ in the context of instrumental variable regression.

20. The Singleton (1985) statistic in the framework of Section 4.1 converges in distribution to $\mathcal{N}(0, \sigma^2_{\text{KJS}})$ with $\sigma^2_{\text{KJS}} = \mu_h^{\dagger\prime} P_{0,g} \mu_h^{\dagger}$. In the same framework, the asymptotic distribution of Smith (1992) is $\mathcal{N}(0, \sigma^2_{\text{RJS}})$ with $\sigma^2_{\text{RJS}} = \mu_h^{\dagger\prime} \Omega_{h,\dagger}^{-1} C_g \Omega_{0,g} P_{0,g} \Omega_{0,g} C_g' \Omega_{h,\dagger}^{-1} \mu_h^{\dagger}$.

In the Monte Carlo study, $\sigma^2_{\text{KJS}}$ is estimated by $\hat{\sigma}^2_{\text{RJS}} = \big(\hat{h}(\tilde{\gamma}) - \hat{g}(\tilde{\beta})\big)' \tilde{P}_g \big(\hat{h}(\tilde{\gamma}) - \hat{g}(\tilde{\beta})\big)$ and $\sigma^2_{\text{RJS}}$ is estimated by $\hat{\sigma}^2_{\text{RJS}} = \big(\hat{h}(\tilde{\gamma}) - \hat{C}_n \hat{g}(\tilde{\beta})\big)'_{\dagger} \tilde{\Omega}_{h,\dagger}^{-1} \hat{C}_g \tilde{\Omega}_g \tilde{P}_g \tilde{\Omega}_g \hat{C}_g' \tilde{\Omega}_{h,\dagger}^{-1} \big(\hat{h}(\tilde{\gamma}) - \hat{C}_n \hat{g}(\tilde{\beta})\big)$, where $\tilde{P}_g$ and $\tilde{\Omega}_g$ are defined similarly to $\hat{P}_g$ and $\hat{\Omega}_g$, with $\hat{\beta}$ replaced by two-step GMM estimator $\tilde{\beta}$ (see Section 4.1) and $\tilde{\Omega}_{h,\dagger} = \sum_{i=1}^n h_i(\tilde{\gamma}) h_i(\tilde{\gamma})'/n$.

21. The maximum norm $\|\cdot\|_\infty$ can be replaced by any equivalent norm in this analysis.

22. Note that we can also obtain results of this type for the Singleton (1985) and Smith (1992) statistics (see ftn. 20). We present here the results without proofs. In the case in which $\|\delta_h\|_\infty = \infty$, $J^{\text{GMM}}_{\text{KJS}}/\hat{\sigma}_{\text{KJS}} \xrightarrow{d} \mathcal{N}(0,1)$ and $J^{\text{GMM}}_{\text{RJS}}/\hat{\sigma}_{\text{RJS}} \xrightarrow{d} \mathcal{N}(0,1)$. On the other hand, when $\|\delta_h\|_\infty \leq c_h$, $\sqrt{n} J^{\text{GMM}}_{\text{KJS}} \xrightarrow{d} \mathcal{T}\big(S_h'\delta_h, \Omega, Q_{1,\text{KJS}}\big)$, where $Q_{1,\text{KJS}} = S_h' P_h \Omega_h P_g S_g - S_g' P_g S_g$, $n\hat{\sigma}^2_{\text{KJS}} \xrightarrow{d} \mathcal{T}\big(S_h'\delta_h, \Omega, Q_{2,\text{KJS}}\big)$, where $Q_{2,\text{KJS}} = K_{\text{KSJ}} P_g K_{\text{KSJ}}'$, with $K_{\text{KSJ}} = S_h' P_h \Omega_h - S_g' P_g \Omega_g$, $\sqrt{n} J^{\text{GMM}}_{\text{RJS}} \xrightarrow{d} \mathcal{T}\big(S_h'\delta_h, \Omega, Q_{1,\text{RJS}}\big)$, where $Q_{1,\text{RJS}} = S_h' P_h C_g \Omega_g P_g S_g$ and $n\hat{\sigma}^2_{\text{RJS}} \xrightarrow{d} \mathcal{T}\big(S_h'\delta_h, \Omega, Q_{2,\text{RJS}}\big)$, where $Q_{2,\text{RJS}} = K_{\text{KSJ}} \Omega_h^{-1} C \Omega_g P_g \Omega_g C_g' \Omega_h^{-1} K_{\text{KSJ}}'$, where $K_{\text{KSJ}} = S_h' P_h \Omega_h - S_g' P_g \Omega_g C_g'$.

23. See Lemma SM2.14 in the online supplementary material for details.

24. We are grateful to an anonymous referee for suggesting this approach.

25. See also Hall and Pelletier (2011), Hsu and Shi (2017), Shennach and Wilhelm (2017) and Shi (2015b) for alternative solutions to this problem in the context of Vuong tests.

26. Similarly, we can define Shi-type statistics based on the Singleton (1985) and Smith (1992) statistics. These are given by $W_n^{\text{KJS}} = \big[\mathcal{J}_{\text{KJS}} - \text{tr}(\tilde{Q}_{1,\text{KJS}}\tilde{\Omega})/\sqrt{n}\big]/\big[\hat{\sigma}_{\text{KJS}} + c_{\text{KJS}} \cdot \text{tr}(\tilde{Q}_{2,\text{KJS}}\tilde{\Omega})/n\big]^{1/2}$ and $W_n^{\text{RJS}} = \big[\mathcal{J}_{\text{RJS}} - \text{tr}(\tilde{Q}_{1,\text{RJS}}\tilde{\Omega})/\sqrt{n}\big]/\big[\hat{\sigma}_{\text{RJS}} + c_{\text{RJS}} \cdot \text{tr}(\tilde{Q}_{2,\text{RJS}}\tilde{\Omega})/n\big]^{1/2}$, where $c_{\text{KJS}}$ and $c_{\text{RJS}}$ are non-negative constants and the estimators $\tilde{\Omega}$, $\tilde{Q}_{1,\text{KJS}}$, $\tilde{Q}_{2,\text{KJS}}$, $\tilde{Q}_{1,\text{RJS}}$, and $\tilde{Q}_{2,\text{RJS}}$ are consistent for $\Omega$, $Q_{1,\text{KJS}}$, $Q_{2,\text{KJS}}$, $Q_{1,\text{RJS}}$ and $Q_{2,\text{RJS}}$, respectively. In the Monte Carlo study, these estimators are obtained using two-step GMM estimators and replacing expected values by averages.

27. This procedure is similar to that considered by Shi (2015a, Sect. 4.2).

28. Note that $\|\delta_h^{**}\|_\infty$ might not be in $[0, c_h]$.

29. Additional results for the statistical tests computed using exponential tilting are provided online at Cambridge Journals Online in the supplementary material to this article.

30. We computed the statistical tests using the following estimators for the asymptotic variance of GEL non-nested test statistics: $\hat{\sigma}_j^2$, $j = 1, \ldots, 4$, (see Section 4.1). The full set of results is available upon request.

31. Note that under the null, the asymptotic variance of $\mathcal{J}_{\text{OPT}}$ is given by $\sigma^2_{\text{OPT}} = \delta' P_g \delta$, which is estimated via $\hat{\sigma}^2_{\text{OPT}} = \delta' \tilde{P}_g \delta$ (see ftn. 20).

32. We compute the two-step GMM estimators (where the two-stage least squares is the first-step estimator) using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm of `MATLAB`. Because the EL estimates obtained via BFGS in some cases violate the convex hull condition (see ftn. 12), we compute EL using the `matElike` `MATLAB` package with the optional Zipsolver package (Zedlewski, 2008). `matElike`, rather than solving the dual EL problem, solves the primal EL problem. In the case of non-convergence, EL is computed employing BFGS applied to the EL dual problem. Since logarithms do not admit negative arguments the `MATLAB` code of Owen, available at http://www-stat.stanford.edu/∼owen/empirical/, was used in which logarithms are replaced by $f(x) = \log(\xi) - 1.5 + 2(x/\xi) - 0.5(x/\xi)^2$ if $x < \xi$ and $f(x) = \log(x)$ if $x \geq \xi$. However, we only consider EL estimates obtained via this procedure to be valid solutions if the convex hull condition is satisfied; otherwise, no solution is reported. We consider that the convex hull condition is satisfied

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

if $\left\| \sum_{i=1}^{n} \hat{p}_i g(z_i, \hat{\beta}) \right\| < 10^{-4}$. Note that in all simulation experiments conducted here, all the EL estimates were valid solutions.

33. The tests based on the statistics $\mathcal{J}_{v,2}$, $\mathcal{J}_{v,3}$, $\mathcal{LM}_{v,2}$, and $\mathcal{LM}_{v,3}$ exhibited similar performance to that based on $\mathcal{S}_\alpha$.

## REFERENCES


Ali, S.M. & D.D. Silvey (1966) A general class of coefficient of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B* 28(1), 131–142.

Atkinson, A.C. (1970) A method for discriminating between models. *Journal of the Royal Statistical Society, Series B* 32, 323–335.

Back, K. & D. Brown (1993) Implied probabilities in GMM estimators. *Econometrica* 61, 971–976.

Brown, B.W. & W.K. Newey (2002) Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Business and Economic Statistics* 20(4), 507–517.

Chen, J.A.M. Variyath, & B. Abraham (2008) Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics* 17(2), 426–443.

Corcoran, S. (1998) Bartlett adjustment of empirical discrepancy statistics. *Biometrika* 85, 965–972.

Cox, D.R. (1961) Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics* 1, 105–123.

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

Cressie, N. & T. Read (1984) Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B* 46, 440–464.

Csiszar, I. (1963) Eine informationstheoretische Ungleichungen und ihre Anwendung auf den Beweis der Ergodicitat von Markoffschen Ketten. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 8, 85–108.

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

Gouriéroux, C. & A. Monfort (1994) Testing non-nested hypotheses. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, 1st ed., vol. 4, chapter 44, pp. 2583–2637. Elsevier.

Hall, A.R. & D. Pelletier (2011) Nonnested testing in models estimated via generalized method of moments. *Econometric Theory* 27, 443–456.

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

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

Hong, H., B. Preston, & M. Shum (2003) Generalized empirical likelihood based model selection criteria for moment based models. *Econometric Theory* 19, 923–943.

Hsu, Y. & X. Shi (2017) Model-selection tests for conditional moment restriction models. *The Econometrics Journal*, forthcoming.

Imbens, G. (1997) One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies* 64, 359–383.

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

Kitamura, Y. (2007) Empirical likelihood methods in econometrics: Theory and practice. In R.W. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3, Econometric Society Monograph, ESM 43, pp. 174–237. Cambridge University Press.


https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

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

MacKinnon, J.G. (1983) Model specification tests against non-nested alternatives. *Econometric Reviews* 2, 85–110.

Newey, W.K. (1985) Generalized method of moments specification testing. *Journal of Econometrics* 29, 229–256.

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

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

Pardo, L. (2006) *Statistical Inference Based on Divergence Measures*. Chapman & Hall.

Pesaran, M.H. & M. Weeks (2001) Non-nested hypothesis testing: An overview. In B.H. Baltagi (ed.), *Companion to Theoretical Econometrics*, pp. 279–309. Basil Blackwell.

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

Quandt, R.E. (1974) A comparison of methods for testing non-nested hypotheses. *Review of Economics and Statistics* 56, 251–255.

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

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

Schennach, S. & D. Wilhelm (2017) A simple parametric model selection test. *Journal of the American Statistical Association*, forthcoming.

Shi, X. (2015a) A nondegenerate Vuong test. *Quantitative Economics* 6(1), 85–121, 03.

Shi, X. (2015b) Model selection tests for moment inequality models. *Journal of Econometrics* 187, 1–17.

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.

Smith, R.J. (2001) GEL Methods for Moment Condition Models. Working paper, University of Bristol. Revised version CeMMAP working papers *CWP 19/04*, cemmap, I.F.S. and U.C.L. Available at http://cemmap.ifs.org.uk/wps/cwp0419.pdf.

Smith, R.J. (2007) Weak instruments and empirical likelihood: A discussion of the papers by D.W.K. Andrews and J.H. Stock and Y. Kitamura. In R.W. Blundell, W.K. Newey, & T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, vol. 3, Econometric Society Monograph Series, ESM 43, pp. 238–260. Cambridge University Press.

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

Zedlewski, J. (2008) Practical empirical likelihood estimation using matElike, Manuscript, Harvard University.

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

# APPENDIX A: Empirical Size and Power

**TABLE A1.** Rejection frequencies of the tests under the null

| $n$ | | 200 | | | | | | 400 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Statistic | | Original | | | Shi-type | | | Original | | | Shi-type | | |
| Estimator | $\rho_{u_g,w_{h2}}$ | 0.0 | 0.2 | 0.4 | 0.0 | 0.2 | 0.4 | 0 | 0.2 | 0.4 | 0 | 0.2 | 0.4 |
| GMM | $\mathcal{J}_{\text{GMM}}$ | 4.8 | 4.6 | 5.1 | – | – | – | 5.1 | 5.0 | 5.3 | – | – | – |
| | $\mathcal{J}_{\text{RJS}}$ | 14.6 | 9.2 | 9.5 | 3.7 | 3.8 | 5.1 | 13.6 | 6.8 | 7.3 | 3.6 | 3.9 | 4.1 |
| | $\mathcal{J}_{\text{KJS}}$ | 6.8 | 4.7 | 5.4 | 2.9 | 2.7 | 3.1 | 6.3 | 5.0 | 4.8 | 2.5 | 3.2 | 3.0 |
| | $\mathcal{J}_{\text{OPT}}$ | 5.0 | 4.8 | 5.2 | – | – | – | 5.3 | 4.4 | 4.7 | – | – | – |
| EL | $\mathcal{LR}_{\text{EL}}$ | 5.8 | 5.6 | 6.2 | – | – | – | 5.6 | 5.3 | 5.8 | – | – | – |
| | $\mathcal{S}_0$ | 16.7 | 14.9 | 14.6 | 1.2 | 1.2 | 1.6 | 15.0 | 13.2 | 13.6 | 1.3 | 1.1 | 1.3 |
| | $\tilde{\mathcal{S}}_0$ | 6.7 | 6.8 | 9.3 | 4.1 | 5.8 | 8.5 | 5.6 | 5.3 | 8.1 | 3.2 | 4.2 | 6.9 |
| | $\mathcal{LM}_0$ | 6.8 | 6.2 | 7.9 | 7.5 | 4.8 | 6.2 | 6.1 | 5.3 | 6.2 | 6.3 | 3.8 | 5.0 |
| | $\mathcal{J}_0$ | 5.5 | 4.9 | 7.0 | 2.7 | 2.5 | 3.6 | 5.5 | 5.1 | 6.2 | 2.9 | 2.8 | 3.2 |
| | $\mathcal{S}_1$ | 16.7 | 14.6 | 14.1 | 1.2 | 1.0 | 1.4 | 15.0 | 13.1 | 12.8 | 1.2 | 0.9 | 1.2 |
| | $\tilde{\mathcal{S}}_1$ | 7.0 | 6.8 | 8.6 | 4.1 | 5.4 | 7.3 | 5.9 | 5.1 | 7.4 | 3.4 | 3.5 | 5.8 |
| | $\mathcal{LM}_1$ | 6.6 | 5.9 | 6.9 | 6.9 | 4.6 | 5.2 | 5.9 | 5.2 | 5.9 | 6.1 | 3.5 | 4.6 |
| | $\mathcal{J}_1$ | 5.3 | 4.9 | 5.9 | 2.7 | 2.6 | 3.3 | 5.4 | 4.9 | 5.6 | 2.7 | 2.8 | 3.1 |
| | $\mathcal{S}_{1.5}$ | 16.7 | 14.4 | 13.2 | 1.2 | 1.2 | 1.7 | 15.0 | 12.8 | 11.6 | 1.2 | 0.9 | 1.4 |
| | $\tilde{\mathcal{S}}_{1.5}$ | 7.2 | 6.8 | 7.8 | 4.1 | 5.1 | 6.1 | 6.0 | 5.6 | 7.0 | 3.4 | 3.6 | 5.2 |
| | $\mathcal{LM}_{1.5}$ | 6.4 | 5.8 | 6.5 | 6.8 | 4.7 | 4.7 | 5.9 | 4.9 | 6.0 | 6.0 | 3.5 | 4.4 |
| | $\mathcal{J}_{1.5}$ | 5.3 | 4.9 | 5.6 | 2.6 | 2.6 | 3.1 | 5.4 | 4.7 | 5.6 | 2.6 | 2.6 | 3.2 |
| | $\mathcal{S}_2$ | 16.7 | 13.9 | 11.8 | 1.3 | 1.4 | 2.2 | 15.0 | 12.4 | 10.1 | 1.2 | 1.1 | 1.9 |
| | $\tilde{\mathcal{S}}_2$ | 7.3 | 6.8 | 7.5 | 4.3 | 4.9 | 5.4 | 6.4 | 5.7 | 6.4 | 3.4 | 3.6 | 5.0 |
| | $\mathcal{LM}_2$ | 6.2 | 5.9 | 6.5 | 6.7 | 4.4 | 4.5 | 5.8 | 5.1 | 6.0 | 5.9 | 3.6 | 4.1 |
| | $\mathcal{J}_2$ | 5.3 | 4.8 | 5.5 | 2.7 | 2.7 | 3.2 | 5.3 | 4.8 | 5.9 | 2.7 | 2.6 | 3.8 |
| | $\mathcal{S}_3$ | 16.5 | 13.2 | 9.2 | 1.7 | 2.0 | 3.1 | 14.9 | 11.7 | 7.5 | 1.4 | 1.7 | 2.6 |
| | $\tilde{\mathcal{S}}_3$ | 8.0 | 6.8 | 7.1 | 4.5 | 4.9 | 4.9 | 6.7 | 6.3 | 6.5 | 3.4 | 3.8 | 4.8 |
| | $\mathcal{LM}_3$ | 6.0 | 5.6 | 6.1 | 6.3 | 4.0 | 4.0 | 5.5 | 5.3 | 5.9 | 5.6 | 3.4 | 3.7 |
| | $\mathcal{J}_3$ | 5.1 | 4.6 | 5.3 | 2.7 | 2.7 | 3.4 | 5.2 | 4.9 | 5.8 | 2.6 | 3.0 | 3.5 |

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press

**TABLE A2.** Rejection frequencies of the tests under the alternative

| $n$ | | 200 | | | | | | 400 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Statistic | | Original | | | Shi-type | | | Original | | | Shi-type | | |
| Estimator | $\omega$ | 3 | 4 | 5 | 3 | 4 | 5 | 3 | 4 | 5 | 3 | 4 | 5 |
| GMM | $\mathcal{J}_{\text{GMM}}$ | 32.1 | 49.6 | 63.2 | – | – | – | 33.4 | 54.1 | 72.3 | – | – | – |
| | $\mathcal{J}_{\text{RJS}}$ | 53.5 | 72.0 | 83.0 | 39.6 | 56.6 | 68.7 | 50.3 | 69.5 | 84.8 | 38.3 | 58.1 | 75.6 |
| | $\mathcal{J}_{\text{KJS}}$ | 29.7 | 42.0 | 50.9 | 26.1 | 37.6 | 45.4 | 32.9 | 50.6 | 65.5 | 28.1 | 45.5 | 60.9 |
| | $\mathcal{J}_{\text{OPT}}$ | 39.4 | 59.1 | 72.9 | – | – | – | 41.6 | 63.5 | 81.0 | – | – | – |
| EL | $\mathcal{LR}_{\text{EL}}$ | 34.9 | 52.9 | 66.2 | – | – | – | 34.3 | 55.2 | 73.4 | – | – | – |
| | $\mathcal{S}_0$ | 48.3 | 63.1 | 70.4 | 13.2 | 23.0 | 34.3 | 48.6 | 66.4 | 79.4 | 11.7 | 23.3 | 38.6 |
| | $\tilde{\mathcal{S}}_0$ | 14.1 | 26.7 | 44.3 | 6.7 | 11.9 | 20.7 | 13.1 | 26.5 | 45.6 | 6.3 | 13.0 | 25.6 |
| | $\mathcal{LM}_0$ | 40.6 | 58.2 | 69.7 | 38.6 | 56.7 | 69.0 | 41.4 | 60.5 | 77.1 | 37.2 | 56.8 | 74.5 |
| | $\mathcal{J}_0$ | 37.5 | 55.0 | 67.9 | 22.1 | 35.2 | 44.4 | 40.3 | 60.3 | 77.0 | 28.1 | 45.1 | 62.6 |
| | $\mathcal{S}_1$ | 47.5 | 62.2 | 67.2 | 12.3 | 20.7 | 30.8 | 47.8 | 65.9 | 78.1 | 11.2 | 22.0 | 35.2 |
| | $\tilde{\mathcal{S}}_1$ | 21.1 | 34.6 | 52.0 | 12.3 | 19.8 | 31.1 | 19.9 | 32.3 | 50.4 | 13.5 | 20.6 | 33.2 |
| | $\mathcal{LM}_1$ | 34.3 | 51.8 | 65.3 | 32.9 | 51.3 | 65.9 | 32.7 | 51.6 | 70.8 | 29.3 | 48.3 | 68.1 |
| | $\mathcal{J}_1$ | 30.7 | 48.7 | 63.4 | 16.6 | 27.0 | 37.8 | 31.6 | 50.7 | 69.7 | 20.9 | 35.3 | 52.5 |
| | $\mathcal{S}_{1.5}$ | 45.8 | 60.1 | 64.8 | 12.9 | 20.9 | 30.5 | 45.1 | 62.6 | 74.7 | 12.0 | 22.9 | 35.9 |
| | $\tilde{\mathcal{S}}_{1.5}$ | 23.2 | 37.7 | 54.7 | 15.8 | 24.1 | 36.3 | 21.9 | 36.0 | 51.8 | 16.0 | 25.7 | 38.5 |
| | $\mathcal{LM}_{1.5}$ | 30.3 | 47.4 | 61.4 | 28.5 | 46.2 | 61.3 | 28.3 | 44.7 | 63.5 | 24.0 | 40.4 | 59.5 |
| | $\mathcal{J}_{1.5}$ | 26.9 | 44.3 | 58.9 | 14.8 | 23.6 | 33.8 | 27.4 | 43.6 | 62.6 | 18.4 | 30.9 | 46.8 |
| | $\mathcal{S}_2$ | 42.7 | 57.8 | 61.7 | 14.1 | 22.7 | 31.4 | 41.0 | 58.1 | 70.1 | 13.4 | 24.5 | 38.3 |
| | $\tilde{\mathcal{S}}_2$ | 24.3 | 39.5 | 56.2 | 17.6 | 27.8 | 41.0 | 23.1 | 36.6 | 52.5 | 17.3 | 28.1 | 42.1 |
| | $\mathcal{LM}_2$ | 27.0 | 43.0 | 57.6 | 24.0 | 40.8 | 55.6 | 26.1 | 41.6 | 57.8 | 20.7 | 35.3 | 52.5 |
| | $\mathcal{J}_2$ | 24.7 | 39.9 | 54.0 | 15.5 | 23.7 | 32.3 | 25.2 | 40.0 | 56.8 | 18.4 | 30.4 | 43.9 |
| | $\mathcal{S}_3$ | 37.0 | 51.7 | 57.3 | 16.9 | 26.7 | 35.1 | 34.3 | 49.5 | 62.7 | 15.5 | 27.6 | 41.6 |
| | $\tilde{\mathcal{S}}_3$ | 24.4 | 39.9 | 56.5 | 18.6 | 31.0 | 46.2 | 23.0 | 35.7 | 52.3 | 17.5 | 28.8 | 44.1 |
| | $\mathcal{LM}_3$ | 26.6 | 40.5 | 52.9 | 21.2 | 34.9 | 48.7 | 27.0 | 40.4 | 54.9 | 20.5 | 33.2 | 47.7 |
| | $\mathcal{J}_3$ | 23.4 | 36.9 | 49.7 | 16.4 | 25.5 | 34.5 | 25.9 | 39.6 | 54.0 | 19.8 | 31.6 | 45.0 |

https://doi.org/10.1017/S0266466617000123 Published online by Cambridge University Press