Ann Inst Stat Math (2018) 70:605–630
https://doi.org/10.1007/s10463-017-0604-2

Bootstrap inference for misspecified moment condition models

Mihai Giurcanu[1] · **Brett Presnell**[2]

Received: 18 June 2016 / Revised: 11 January 2017 / Published online: 7 March 2017

Abstract We study the standard-bootstrap, the centered-bootstrap, and the empirical-likelihood bootstrap tests of hypotheses used in conjunction with generalized method of moments inference in correctly specified and misspecified moment condition models. We show that, under correct specification, the standard-bootstrap estimator of the null distribution of the J-test converges in distribution to a random distribution, verifying its inconsistency, while the centered and the empirical-likelihood bootstrap estimators are consistent. We provide higher-order expansions of the size distortions of the analytic and the bootstrap tests. We show that the standard-bootstrap parameter-tests are consistent under misspecification, while the centered-bootstrap parameter-tests are inconsistent. We propose a general bootstrap methodology which is highly accurate under correct specification and consistent under misspecification. In a simulation study, we explore the finite sample behavior of the analytic and the bootstrap tests for a panel data model and we apply our methodology on a real-world data set.

Electronic supplementary material The online version of this article (doi:10.1007/s10463-017-0604-2) contains supplementary material, which is available to authorized users.

✉ Mihai Giurcanu
giurcanu@uchicago.edu

Brett Presnell
presnell@ufl.edu

1 Department of Public Health Sciences, University of Chicago, 5841 S Maryland Ave, Room R325, Chicago, IL 60637, USA

2 Department of Statistics, University of Florida, 225 Griffin-Floyd Hall, Gainesville, FL 32611, USA

Keywords GMM inference · Standard-bootstrap · Centered-bootstrap · Empirical-likelihood bootstrap · Edgeworth expansions · Misspecified models

1 Introduction

Hansen (1982) introduced the generalized method of moments (GMM) to fit models in which the number of parameters is exceeded by the number of moment conditions identifying them. If the moment conditions are correctly specified, then under regularity conditions, the GMM estimators are consistent and asymptotically normally distributed, and the limiting null distributions of the (generalized) Wald test, likelihood-ratio test (LRT), and score test (ST) are chi-squared. We will refer to these tests collectively as parameter-tests. The limiting null distribution of the J-test of over-identifying restrictions, a goodness-of-fit test, is also chi-squared.

Monte Carlo experiments have shown that the analytic approximations suggested by large-sample theory often perform poorly in the GMM context (see the July 1996 special issue, Vol. 16, No. 3 of the *Journal of Business and Economic Statistics*). Bootstrap methods (Efron 1979) provide a computationally intensive alternative, and Hahn (1996) has shown that, for correctly specified models, the standard (nonparametric)-bootstrap (SB) distributions of GMM estimators are consistent. However, the moment conditions defining the GMM estimators do not hold in the "bootstrap world", a fact which portends difficulties for a naive bootstrap approach to GMM inference. As a remedy, Hall and Horowitz (1996) suggested the centered-bootstrap (CB), which centers the moment functions to have mean zero in the bootstrap world. In a similar vein, Brown and Newey (2002) proposed the empirical-likelihood bootstrap (ELB), in which resamples are drawn from a weighted empirical distribution, with the weights chosen in such a way that the model holds in the bootstrap world (see also Qin and Lawless 1994; Hall and Presnell 1999; Zhang 1999). Andrews (2002) proposed a computationally efficient bootstrap method for GMM estimators and showed higher-order accuracy of the CB symmetric t-test.

In empirical research, inference about parameters may be carried out in situations when the moment conditions are misspecified: possibly because the J-test or other tests of misspecification fails to reject a false null hypothesis; or because the data analyst fails to test for misspecification at all; or because the data analyst proceeds with the moment condition model in spite of finding evidence of misspecification (see, e.g., Imbens and Lancaster 1994). Thus, it is important to study the properties of GMM inference even when the moment conditions defining those parameters fail to hold. In a breakthrough paper on this topic, Hall and Inoue (2003) derived the limiting distributions of GMM estimators and of the null distributions of parameter-test statistics under misspecification. More recently, Lee (2014) used Hall and Inoue's misspecification-robust variance estimator of the GMM estimators and showed higher-order accuracy of the resulting SB symmetric percentile-t confidence intervals and t-tests under misspecification.

In this paper, we prove new results on the asymptotic properties of the analytic and bootstrap tests in correctly specified and misspecified moment condition models. First, we show that the SB estimator of the null distribution of the J-test converges *in* *distribution* to a random chi-squared distribution, showing its inconsistency even under correct specification, while the CB and the ELB distributions are consistent. We extend the results developed by Hall and Horowitz (1996) and Andrews (2002) and prove exact higher-order properties of the size distortions of the parameter-tests and the J-test under correct specification. In order to maintain both the *size* and *power*, we show that some care is required in constructing the bootstrap test statistics so that they consistently estimate the null distributions regardless of whether the null hypotheses are true. We propose an automatic method of selection of the weights of the ELB method which improves its properties under correct specification and under misspecification. Second, we extend the results of Hall and Inoue (2003) and show that under misspecification, the choice of the weight matrices may change not only the asymptotic covariance matrix of the GMM estimators, but also the rates of convergence and the limiting null distributions of the LRT and ST statistics. These results re-emphasize the critical role of the weight matrices in defining the estimators and the test statistics under possible model misspecification. Third, we study the asymptotic properties of the bootstrap tests under misspecification of the moment condition model. We demonstrate that the SB and the ELB parameter-tests are consistent while the CB parameter-tests are inconsistent. Here, we also prove that the asymptotic power of the SB J-test is zero and that the CB J-test is consistent under misspecification. These theoretical results enable us to propose a general bootstrap methodology which is highly accurate under correct specification and consistent under misspecification.

We close this section with an outline. In Sect. 2, we review the theory of GMM under correct specification and misspecification of the moment condition model. In Sect. 3, we study the consistency and higher-order properties of the bootstrap tests under correct specification. In Sect. 4, we extend our bootstrap consistency results for misspecified models. In Sect. 5, we propose a general bootstrap methodology which is highly accurate under correct specification and consistent under misspecification. In Sect. 6, the theoretical results are tested in a simulation study for a dynamic autoregressive panel data model (Arellano and Bond 1991). We have implemented the proposed methods of inference for autoregressive dynamic panel data models in the R language (R Core Team 2016) and created an R package which is available from the authors upon request. In Sect. 7, we apply the proposed bootstrap methodology to a data set tracking household electricity usage in a Florida residential area. Regularity conditions are included in an "Appendix", and detailed proofs of theoretical results are deferred to a supplementary appendix.

2 Background and notation

2.1 GMM inference

Let $X_{1:n} = \{X_1, \ldots, X_n\}$ be an i.i.d. sample from a distribution P_0 on a measurable space $\mathscr{X}$ and let X denote a generic random element with distribution P_0. An over-identified moment condition model assumes that the unknown parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^p$ is the unique solution of a system of equations:

$$\mathrm{E}\big(g(X, \theta_0)\big) = 0, \tag{1}$$

where $g : \mathscr{X} \times \Theta \to \mathbb{R}^q$ is a known *moment function* and $p < q$. The model (1) is said to be *misspecified* if there does not exist a value of θ for which the moment conditions $\mathrm{E}(g(X, \theta)) = 0$ hold. Let $\hat{Q}(\theta, W) = \bar{g}(\theta)^T W \bar{g}(\theta)$, where $\bar{g}(\theta) = n^{-1} \sum_{i=1}^n g(X_i, \theta)$ is the sample moment function and $W \in \mathbb{R}^{q \times q}$. For a positive definite weight matrix $\hat{W}$ (possibly random), the corresponding GMM estimator is defined as the minimizer over $\theta \in \Theta$ of $\hat{Q}(\theta, \hat{W})$.

Let $V(\theta) = \mathrm{Var}(g(X, \theta))$ and $V_0 = V(\theta_0)$. Hansen (1982) showed that choosing $\hat{W}$ to be any consistent estimator of V_0^{-1} minimizes the asymptotic covariance matrix of the GMM estimator in the Loewner ordering, and in this case, the GMM estimator is said to be *efficient*. Generally, one must first estimate θ_0 in order to consistently estimate V_0, and Hansen thus proposed the two-step GMM estimator obtained by first computing a *one-step* GMM estimator, $\hat{\theta}_{\mathrm{O}} = \mathrm{argmin}_{\theta \in \Theta}\ \hat{Q}(\theta, \hat{W})$, where $\hat{W}$ is an arbitrary positive definite weight matrix, often taken to be the identity matrix, although other choices are also possible (see, e.g., case (a) on p. 370 of Hall and Inoue 2003). The *two-step* GMM estimator is then defined as $\hat{\theta}_{\mathrm{T}} = \mathrm{argmin}_{\theta \in \Theta}\ \hat{Q}(\theta, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$, where $\hat{V}(\hat{\theta}_{\mathrm{O}})$ is a consistent estimator of V_0. Here, $\hat{V}(\theta)$ is either the sample covariance matrix $\hat{V}_{\mathrm{C}}(\theta) = n^{-1} \sum_{i=1}^n (g(X_i, \theta) - \bar{g}(\theta))(g(X_i, \theta) - \bar{g}(\theta))^T$, or its uncentered variant $\hat{V}_{\mathrm{U}}(\theta) = n^{-1} \sum_{i=1}^n g(X_i, \theta) g(X_i, \theta)^T$.

Obviously, this process may be iterated to obtain further multi-step GMM estimators, or one may use the *continuously-updated* GMM estimator (Hansen et al. 1996) given by $\hat{\theta}_{\mathrm{CU}} = \mathrm{argmin}_{\theta \in \Theta}\ \hat{Q}(\theta, \hat{V}(\theta)^{-1})$. When the moment conditions are correctly specified, the results of this paper apply with obvious modifications to any of these efficient GMM estimators. Under misspecification, however, our results change according to the specific efficient GMM estimator employed: for example, the results for $\hat{\theta}_{\mathrm{T}}$ depend on $\hat{W}$ used to define $\hat{\theta}_{\mathrm{O}}$ and on the choice of $\hat{V}(\theta)$, e.g., $\hat{V}_{\mathrm{C}}(\theta)$ or $\hat{V}_{\mathrm{U}}(\theta)$. For specificity, we focus on the commonly-used two-step estimator, which has the practical advantage of being relatively easy to calculate. However, similar results can be proved for other efficient GMM estimators.

Given a smooth function $h : \mathbb{R}^p \to \mathbb{R}^r$, we consider three classical parameter-tests for testing the null hypothesis $\mathrm{H}_0 : \theta_0 \in \Theta_0$ against the alternative hypothesis $\mathrm{H}_a : \theta_0 \notin \Theta_0$, where $\Theta_0 = \{\theta \in \Theta : h(\theta) = 0\}$. Let $\nabla g(x, \theta) \in \mathbb{R}^{q \times p}$ be the Jacobian matrix (with respect to θ) of $g(x, \theta)$, $G(\theta) = \mathrm{E}(\nabla g(X, \theta))$, $G_0 = G(\theta_0)$, and $H(\theta) = \nabla h(\theta) \in \mathbb{R}^{r \times p}$. Hansen (1982) showed that $n^{1/2}(\hat{\theta}_{\mathrm{T}} - \theta_0) \xrightarrow{d} N(0, D_0^{-1})$, where $D(\theta, W) = G(\theta)^T W G(\theta) \in \mathbb{R}^{p \times p}$, $D_0 = D(\theta_0, V_0^{-1})$, and $\xrightarrow{d}$ denotes convergence in distribution. This motivates the Wald test statistic given by

$$\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = n h(\hat{\theta}_{\mathrm{T}})^T \big(H(\hat{\theta}_{\mathrm{T}}) \hat{D}(\hat{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{T}}))^{-1} H(\hat{\theta}_{\mathrm{T}})^T\big)^{-1} h(\hat{\theta}_{\mathrm{T}}),$$

where $\hat{G}(\theta) = n^{-1} \sum_{i=1}^n \nabla g(X_i, \theta)$ and $\hat{D}(\theta, W) = \hat{G}(\theta)^T W \hat{G}(\theta)$. The (generalized) likelihood-ratio test (LRT) statistic is

$$\hat{L}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) = n \hat{Q}(\tilde{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) - n \hat{Q}(\hat{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}),$$

where $\tilde{\theta}_{\mathrm{T}} = \mathrm{argmin}_{\theta \in \Theta_0} \hat{Q}(\theta, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$ is the constrained two-step GMM estimator. Finally, the score test (ST) statistic is

$$\hat{S}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) = n\hat{C}(\tilde{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})^T \hat{D}(\hat{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{T}}))^{-1}\hat{C}(\tilde{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}),$$

where $\hat{C}(\theta, W) = \hat{G}(\theta)^T W \bar{g}(\theta) \in \mathbb{R}^p$.

2.2 GMM under correct specification

The J-test statistic is given by $\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) = n\hat{Q}(\hat{\theta}_{\mathrm{T}}, \hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$. Under correct specification of the model (1), Hansen (1982) showed that

$$n^{1/2}(\hat{\eta} - \eta_0) \xrightarrow{d} N(0, \Gamma_0), \tag{2a}$$

$$\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2, \quad \hat{L}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2, \quad \hat{S}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2, \tag{2b}$$

$$\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \xrightarrow{d} \chi_{q-p}^2, \tag{2c}$$

where $\hat{\eta} = (\hat{\theta}_{\mathrm{O}}^T, \hat{\theta}_{\mathrm{T}}^T)^T \in \mathbb{R}^{2p}$, $\eta_0 = (\theta_0^T, \theta_0^T)^T \in \mathbb{R}^{2p}$,

$$\Gamma_0 = \begin{pmatrix} D(\theta_0, W_0)^{-1} G_0^T W_0 V_0 W_0 G_0 D(\theta_0, W_0)^{-1} & D_0^{-1} \\ D_0^{-1} & D_0^{-1} \end{pmatrix}, \tag{3}$$

W_0 is the in-probability limit of $\hat{W}$, $\underset{\mathrm{H}_0}{\xrightarrow{d}}$ denotes convergence in distribution under the null hypothesis of the test, and χ_k^2 represents the chi-squared distribution with k degrees of freedom. Results (2b) and (2c) suggest the use of chi-squares as reference distributions, and we will refer to the resulting test procedures collectively as *analytic* chi-squared tests.

Lemma 1 improves on (2b) and (2c), and shows that under additional regularity conditions, the analytic tests with nominal level $\alpha \in (0, 1)$ have size distortions of exact order $O(n^{-1})$. Here $\chi_{r;\alpha}^2$ denotes the upper α-quantile of χ_r^2 and $\underset{\mathrm{H}_0}{=}$ denotes an equality under the corresponding null hypothesis of the test. Lemma 1 extends the result for the two-sided t-test of Andrews (2002, Eq. 2.1, p. 123) and shows that the same rate holds for all parameter-tests and the J-test.

Lemma 1 *Suppose the model* (1) *and conditions A–F in "Appendix" hold. Then*

$$\Pr\left(\hat{T}(\hat{V}(\hat{\theta}_T)^{-1}) > \chi_{r;\alpha}^2\right) \underset{\mathrm{H}_0}{=} \alpha + n^{-1}a(\alpha) + O(n^{-2}), \tag{4a}$$

$$\Pr\left(\hat{L}(\hat{V}(\hat{\theta}_O)^{-1}) > \chi_{r;\alpha}^2\right) \underset{\mathrm{H}_0}{=} \alpha + n^{-1}b(\alpha) + O(n^{-2}), \tag{4b}$$

$$\Pr\left(\hat{S}(\hat{V}(\hat{\theta}_O)^{-1}) > \chi_{r;\alpha}^2\right) \underset{\mathrm{H}_0}{=} \alpha + n^{-1}c(\alpha) + O(n^{-2}), \tag{4c}$$

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_O)^{-1}) > \chi^2_{q-p;\alpha}\big) = \alpha + n^{-1}d(\alpha) + O(n^{-2}), \tag{4d}$$

where $a(\cdot), b(\cdot), c(\cdot), d(\cdot) \neq 0$.

2.3 GMM under misspecification

The theory of GMM inference is more complicated under moment misspecification (Hall and Inoue 2003). Let $\mu(\theta) = \mathrm{E}(g(X, \theta))$ and $Q(\theta, W) = \mu(\theta)^T W \mu(\theta)$; the model (1) is misspecified if there is no $\theta \in \Theta$ for which $\mu(\theta) = 0$. Redefine $V(\theta)$ to be the in-probability limit of $\hat{V}(\theta)$, so that $V(\theta) = \mathrm{Var}(g(X, \theta))$ if $\hat{V}(\theta) = \hat{V}_{\mathrm{C}}(\theta)$, whereas $V(\theta) = \mathrm{Var}(g(X, \theta)) + \mu(\theta)\mu(\theta)^T$ if $\hat{V}(\theta) = \hat{V}_{\mathrm{U}}(\theta)$. The in-probability limits of $\hat{\theta}_{\mathrm{O}}$ and $\hat{\theta}_{\mathrm{T}}$ are θ_1 and θ_2, where

$$\theta_1 = \mathrm{argmin}_{\theta \in \Theta}\ Q(\theta, W_0) \quad \text{and} \quad \theta_2 = \mathrm{argmin}_{\theta \in \Theta}\ Q(\theta, V(\theta_1)^{-1}).$$

Under regularity conditions, $Q(\theta, W_0)$ and $Q(\theta, V(\theta_1)^{-1})$ are continuously differentiable functions of θ, and thus, θ_1 and θ_2 satisfy the equations

$$G(\theta_1)^T W_0 \mu(\theta_1) = 0 \quad \text{and} \quad G(\theta_2)^T V(\theta_1)^{-1} \mu(\theta_2) = 0. \tag{5}$$

Because we are focused here on the two-step estimator, we will treat the *implied* value θ_2 as the target parameter and take $\mathrm{H}_0 : \theta_2 \in \Theta_0$ as the null hypothesis of parameter-tests. Let $\mu_1 = \mu(\theta_1)$, $\mu_2 = \mu(\theta_2)$, $V_1 = V(\theta_1)$, $V_2 = V(\theta_2)$, $G_1 = G(\theta_1)$, $G_2 = G(\theta_2)$, and $H_2 = H(\theta_2)$. Let further $A_1 = A(\theta_1, W_0)$, $B = B(\theta_1, \theta_2)$, $A_2 = A(\theta_2, V(\theta_1)^{-1})$, and $C(\theta, W) = G(\theta)^T W \mu(\theta)$, where $A(\zeta_1, W) = (\partial_j C(\zeta_1, W) : 1 \le j \le p) \in \mathbb{R}^{p \times p}$,

$$B(\zeta_1, \zeta_2) = \big(-G(\zeta_2)^T V(\zeta_1)^{-1} \big(\partial_j V(\zeta_1)\big) V(\zeta_1)^{-1} \mu(\zeta_2) : 1 \le j \le p\big) \in \mathbb{R}^{p \times p},$$

and $\partial_j = \partial / \partial \zeta_{1,j}$ is the partial derivative with respect to the jth element of $\zeta_1 = (\zeta_{1,1}, \ldots, \zeta_{1,p})^T \in \mathbb{R}^p$. Let

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} & \Pi_{13} & O & O & O \\ O & O & O & \Pi_{24} & \Pi_{25} & \Pi_{26} \end{pmatrix}, \tag{6}$$

where O is a null matrix, $\Pi_{11} = G_1^T W_0$, $\Pi_{12} = (W_0 \mu_1)^T \otimes I_p$, $\Pi_{13} = \mu_1^T \otimes G_1^T$, $\Pi_{24} = -(V_1^{-1}\mu_2)^T \otimes (G_2^T V_1^{-1})$, $\Pi_{25} = G_2^T V_1^{-1}$, $\Pi_{26} = (V_1^{-1}\mu_2)^T \otimes I_p$, and $\otimes$ is the Kronecker product. Let $\Xi = \Pi \Sigma \Pi^T \in \mathbb{R}^{2p \times 2p}$, where Σ is defined at condition E in "Appendix". Let further $\Xi = (\Xi_{ij} : i, j = 1, 2)$, $\Xi_{ij} \in \mathbb{R}^{p \times p}$, $\Gamma = (\Gamma_{ij} : i, j = 1, 2) \in \mathbb{R}^{2p \times 2p}$ and $\Gamma_{ij} \in \mathbb{R}^{p \times p}$, with $\Gamma_{11} = A_1^{-1} \Xi_{11} A_1^{-1}$,

$$\begin{aligned} \Gamma_{12} = \Gamma_{21}^T &= -A_1^{-1} \Xi_{11} A_1^{-1} B^T A_2^{-1} + A_1^{-1} \Xi_{12} A_2^{-1}, \\ \Gamma_{22} &= A_2^{-1} B A_1^{-1} \Xi_{11} A_1^{-1} B^T A_2^{-1} - A_2^{-1} \Xi_{21} A_1^{-1} B^T A_2^{-1} \\ &\quad - A_2^{-1} B A_1^{-1} \Xi_{12} A_2^{-1} + A_2^{-1} \Xi_{22} A_2^{-1}. \end{aligned}$$

Hall and Inoue (2003, Theorems 2 and 5) showed that, under misspecification, $n^{1/2}(\hat{\theta}_{\mathrm{T}} - \theta_2) \xrightarrow{d} N(0, \Gamma_{22})$, and the limiting null distributions of parameter-tests are distributions of linear combinations of independent chi-squared variables. Specifically,

$$\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) \xrightarrow[\mathrm{H}_0]{d} \sum_{j=1}^{r} \lambda_j^W U_j^2, \tag{7a}$$

$$\hat{L}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \xrightarrow[\mathrm{H}_0]{d} \sum_{j=1}^{p} \lambda_j^L U_j^2, \tag{7b}$$

$$\hat{S}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) \xrightarrow[\mathrm{H}_0]{d} \sum_{j=1}^{p} \lambda_j^S U_j^2, \tag{7c}$$

where $U_j^2 \sim$ i.i.d. χ_1^2, $\lambda_1^W, \ldots, \lambda_r^W$ are the eigenvalues of

$$(H_2 \Gamma_{22} H_2^T)^{1/2} (H_2 (G_2^T V_2^{-1} G_2)^{-1} H_2^T)^{-1} (H_2 \Gamma_{22} H_2^T)^{1/2},$$

$\lambda_1^L, \ldots, \lambda_p^L$ are the eigenvalues of $\Omega_2^{1/2} A_2 \Omega_2^{1/2}$, $\Omega_2 = \Delta_2 A_2 \Gamma_{22} A_2 \Delta_2$,

$$\Delta_2 = A_2^{-1} H_2^T (H_2 A_2^{-1} H^T)^{-1} H_2 A_2^{-1},$$

and $\lambda_1^S, \ldots, \lambda_p^S$ are the eigenvalues of $\Omega_2^{1/2} A_2 (G_2^T V_2^{-1} G_2)^{-1} A_2 \Omega_2^{1/2}$. Let $\eta_0 = (\theta_1^T, \theta_2^T)^T$ and $\hat{\eta} = (\hat{\theta}_{\mathrm{O}}^T, \hat{\theta}_{\mathrm{T}}^T)^T$. Theorem 1 is an extension of Theorems 2 and 5 of Hall and Inoue (2003) and shows the limiting distribution of $\hat{\eta}$ and the limiting null distributions of $\hat{L}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1})$, $\hat{S}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1})$, and $\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$ under misspecification.

Theorem 1 *Suppose conditions A–E in "Appendix" hold. Then*

$$n^{1/2}(\hat{\eta} - \eta_0) \xrightarrow{d} N(0, \Gamma), \tag{8a}$$

$$n^{-1/2} \hat{L}(\hat{V}(\hat{\theta}_T)^{-1}) \xrightarrow[\mathrm{H}_0]{d} N(0, \sigma_L^2), \tag{8b}$$

$$n^{1/2}\left(n^{-1} \hat{S}(\hat{V}(\hat{\theta}_T)^{-1}) - \psi_2\right) \xrightarrow[\mathrm{H}_0]{d} N(0, \sigma_S^2), \tag{8c}$$

$$n\left(\bar{g}(\hat{\theta}_T) - \mu(\theta_2)\right)^T \hat{V}(\hat{\theta}_T)^{-1} \left(\bar{g}(\hat{\theta}_T) - \mu(\theta_2)\right) \xrightarrow{d} \sum_{j=1}^{q} \lambda_j^J U_j^2, \tag{8d}$$

$$n^{1/2}\left(n^{-1} \hat{J}(\hat{V}(\hat{\theta}_O)^{-1}) - Q(\theta_2, V(\theta_1)^{-1})\right) \xrightarrow{d} N(0, \sigma_J^2), \tag{8e}$$

where σ_L^2, ψ_2, σ_S^2, λ_j^J, σ_J^2 are defined in the proof.

Result (8a) determines the asymptotic covariance matrix of $\hat{\eta}$ under misspecification. Results (8b) and (8c) show that in the case of the LRT and ST tests, the rates of convergence depend on whether the test statistics are calculated with $\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}$ or $\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}$ [compare (7b) with (8b) and (7c) with (8c)]. Note that the case $\hat{W} = W_0$ (constant) is included here by taking the asymptotic covariance matrix of the vectorized $\hat{W}$ to be the null matrix. Using (8d), we can construct an analytic confidence region for μ_2, which may be helpful in identifying the specific moment conditions that have failed. However, since $\hat{\theta}_{\mathrm{T}}$ is not a consistent estimator of the true parameter, this method does not guarantee that we are able to identify the offending set of misspecified moment conditions. Lastly, (8e) describes the asymptotic power of the J-test under a fixed alternative.

3 Bootstrapping under correct specification

3.1 The standard-bootstrap

Conditional on $X_{1:n}$, a standard-bootstrap (SB) sample $X^*_{1:n} = \{X^*_1, \ldots, X^*_n\}$ satisfies $X^*_1, \ldots, X^*_n \sim$ i.i.d. $\mathbb{P}$, where $\mathbb{P} = n^{-1}\sum_{i=1}^n \delta_{X_i}$ denotes the empirical distribution of $X_{1:n}$ and δ_x represents the unit point mass at $x \in \mathscr{X}$. Let $g^*(x,\theta) = g(x,\theta)$ denote the SB version of $g(x,\theta)$; the reason for introducing this notation will become clear in the next subsection. The bootstrap versions of $\bar{g}(\theta)$, $\hat{Q}(\theta,\hat{W})$, $\hat{V}_{\mathrm{C}}(\theta)$, $\hat{V}_{\mathrm{U}}(\theta)$, and $\hat{G}(\theta)$ are then defined using the "plug-in principle", i.e., by replacing $g(x,\theta)$ with $g^*(x,\theta)$ and X_i with X^*_i. Specifically, let $\bar{g}^*(\theta) = n^{-1}\sum_{i=1}^n g^*(X^*_i,\theta)$, $\hat{Q}^*(\theta, W) = \bar{g}^*(\theta)^T W \bar{g}^*(\theta)$, $\hat{V}^*_{\mathrm{C}}(\theta) = n^{-1}\sum_{i=1}^n (g^*(X^*_i,\theta) - \bar{g}^*(\theta))(g^*(X^*_i,\theta) - \bar{g}^*(\theta))^T$, $\hat{V}^*_{\mathrm{U}}(\theta) = n^{-1}\sum_{i=1}^n g^*(X^*_i,\theta) g^*(X^*_i,\theta)^T$, and $\hat{G}^*(\theta) = \nabla \bar{g}^*(\theta)$. The bootstrap version of $\hat{\theta}_{\mathrm{O}}$ is $\hat{\theta}^*_{\mathrm{O}} = \operatorname{argmin}_{\theta\in\Theta} \hat{Q}^*(\theta, \hat{W}^*)$, where $\hat{W}^*$ is the bootstrap version of $\hat{W}$: if $\hat{W}$ is nonrandom, then $\hat{W}^* = \hat{W}$, and otherwise, $\hat{W}^*$ depends on $X^*_{1:n}$, $\mathbb{P}$, and g^* in the same way that $\hat{W}$ depends on $X_{1:n}$, P_0, and g. The bootstrap version of $\hat{\theta}_{\mathrm{T}}$ is $\hat{\theta}^*_{\mathrm{T}} = \operatorname{argmin}_{\theta\in\Theta} \hat{Q}^*(\theta, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1})$, where $\hat{V}^*(\theta)$ is the bootstrap version of $\hat{V}(\theta)$, i.e., either $\hat{V}^*_{\mathrm{C}}(\theta)$ or $\hat{V}^*_{\mathrm{U}}(\theta)$, according to whether $\hat{V}(\theta) = \hat{V}_{\mathrm{C}}(\theta)$ or $\hat{V}(\theta) = \hat{V}_{\mathrm{U}}(\theta)$.

Let $\hat{T}$ be a statistical quantity in $\mathbb{R}^k$ and let $\mathscr{L}(\hat{T})$ denote its distribution. The bootstrap estimator of $\mathscr{L}(\hat{T})$ is the conditional distribution given $X_{1:n}$ of its bootstrap version $\hat{T}^*$, denoted by $\mathscr{L}(\hat{T}^*|X_{1:n})$. Because it is determined by $X_{1:n}$, $\mathscr{L}(\hat{T}^*|X_{1:n})$ is a random distribution on $\mathbb{R}^k$, i.e., a random element in the space of distributions on $\mathbb{R}^k$, denoted by $\mathscr{P}^k$. We equip $\mathscr{P}^k$ with any distance that metrizes weak convergence, such as the Prohorov metric or the bounded-Lipschitz metric (see, e.g., Dudley 2002, pp. 393–399). We say that the bootstrap is consistent if the distance between $\mathscr{L}(\hat{T}^*|X_{1:n})$ and $\mathscr{L}(\hat{T})$ converges *in probability* to 0. If $\hat{T} \xrightarrow{d} T$, then the bootstrap is consistent if $\mathscr{L}(\hat{T}^*|X_{1:n})$ converges in probability to $\mathscr{L}(T)$, and write $\mathscr{L}(\hat{T}^*|X_{1:n}) \xrightarrow{P} \mathscr{L}(T)$. Note that our definition of bootstrap consistency $\mathscr{L}(\hat{T}^*|X_{1:n}) \xrightarrow{P} \mathscr{L}(T)$ is equivalent to the statement that $\mathscr{L}(T^*|X_{1:n})$ convergences weakly to $\mathscr{L}(T)$ in probability; the latter terminology was used by Hahn to prove the consistency of the standard-bootstrap distribution of GMM estimators under correct model specification (see Hahn 1996, p. 189).

In order to consistently estimate the null distributions of parameter-tests, their null hypotheses should hold in the bootstrap world, in which $\hat{\theta}_{\mathrm{T}}$ plays the role of the true parameter θ_0 (for a detailed discussion on this topic, see Hall and Wilson 1991). The simplest way to assure this, and the method we use here, is to define the bootstrap version of h and Θ_0 to be $h^*(\theta) = h(\theta) - h(\hat{\theta}_{\mathrm{T}})$ and $\Theta_0^* = \{\theta \in \Theta : h^*(\theta) = 0\}$. The bootstrap version of $\tilde{\theta}_{\mathrm{T}}$ is then $\tilde{\theta}_{\mathrm{T}}^* = \operatorname{argmin}_{\theta \in \Theta_0^*} \hat{Q}^*(\theta, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1})$, and the bootstrap parameter-test statistics are given by

$$\hat{T}^*(\hat{V}^*(\hat{\theta}_{\mathrm{T}}^*)^{-1}) = nh^*(\hat{\theta}_{\mathrm{T}}^*)^T \left(H(\hat{\theta}_{\mathrm{T}}^*)\hat{D}^*(\hat{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{T}}^*)^{-1})^{-1} H(\hat{\theta}_{\mathrm{T}}^*)^T\right)^{-1} h^*(\hat{\theta}_{\mathrm{T}}^*),$$
$$\hat{L}^*(\hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}) = n\hat{Q}^*(\tilde{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}) - n\hat{Q}^*(\hat{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}),$$
$$\hat{S}^*(\hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}) = n\hat{C}^*(\tilde{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1})^T \hat{D}^*(\hat{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{T}}^*)^{-1})^{-1} \hat{C}^*(\tilde{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}),$$

where $\hat{D}^*(\theta, W) = \hat{G}^*(\theta)^T W \hat{G}^*(\theta)$ and $\hat{C}^*(\theta, W) = \hat{G}^*(\theta)^T W \bar{g}^*(\theta)$. The bootstrap J-test statistic is $\hat{J}^*(\hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}) = n\hat{Q}^*(\hat{\theta}_{\mathrm{T}}^*, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1})$. Note that the null hypothesis of the J-test does not hold in the SB world since $\bar{g}(\hat{\theta}_{\mathrm{T}}) \neq 0$.

Let $\mathscr{L}_{\mathrm{S}}(\hat{T}^*|X_{1:n})$ denote the conditional distribution given $X_{1:n}$ of the SB version $\hat{T}^*$ of $\hat{T}$; then, $\mathscr{L}_{\mathrm{S}}(\hat{T}^*|X_{1:n})$ represents the SB estimator of $\mathscr{L}(\hat{T})$. Let $\hat{\eta}^* = (\hat{\theta}_{\mathrm{O}}^{*T}, \hat{\theta}_{\mathrm{T}}^{*T})^T$ be the bootstrap version of $\hat{\eta}$. Theorem 2 shows the asymptotic properties of SB tests under correct specification. Examination of the proof of Theorem 2 (as well as of Theorems 4 and 6) shows that these results continue to hold (under correct specification) if the test statistics are calculated using either $\hat{V}^*(\hat{\theta}_{\mathrm{T}}^*)^{-1}$ or $\hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1}$. Result (9a) extends the result of Theorem 1 of Hahn (1996), and we include it here for reference.

Theorem 2 *Suppose the model* (1) *and conditions A–E in "Appendix" hold. Then*

$$\mathscr{L}_{\mathrm{S}}\left(n^{1/2}(\hat{\eta}^* - \hat{\eta})\big|X_{1:n}\right) \xrightarrow{P} N(0, \Gamma_0), \tag{9a}$$
$$\mathscr{L}_{\mathrm{S}}\left(\hat{T}^*(\hat{V}^*(\hat{\theta}_T^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{9b}$$
$$\mathscr{L}_{\mathrm{S}}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{9c}$$
$$\mathscr{L}_{\mathrm{S}}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{9d}$$
$$\mathscr{L}_{\mathrm{S}}\left(\hat{J}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{d} \chi_{q-p}^2(U^2), \tag{9e}$$

where $U^2 \sim \chi_{q-p}^2$.

Comparing the conclusions of Theorem 2 with (2), we see that the SB consistently estimates the joint distribution of $\hat{\theta}_{\mathrm{O}}$ and $\hat{\theta}_{\mathrm{T}}$, and the null distributions of parameter-tests but not that of the J-test. Instead, the SB distribution of the J-test statistic converges *in distribution*, as a random element in the space of distributions on $\mathbb{R}$, to a random non-central chi-squared distribution with $q - p$ degrees of freedom and with a random noncentrality parameter which itself follows a (central) chi-squared distribution with $q - p$ degrees of freedom.

Let $\hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha}$, $\hat{\xi}^{\mathrm{L}}_{\mathrm{S},\alpha}$, and $\hat{\xi}^{\mathrm{S}}_{\mathrm{S},\alpha}$ denote the upper α-quantiles of the SB distributions of parameter-test statistics. Then, the SB Wald, LRT, and ST reject $\mathrm{H}_0 : \theta_0 \in \Theta_0$ at nominal level α whenever $\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) > \hat{\xi}^{\mathrm{T}}_{\mathrm{S},\alpha}$, $\hat{L}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) > \hat{\xi}^{\mathrm{L}}_{\mathrm{S},\alpha}$, and $\hat{S}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) > \hat{\xi}^{\mathrm{S}}_{\mathrm{S},\alpha}$, respectively. Let $\hat{\xi}^{\mathrm{J}}_{\mathrm{S},\alpha}$ denote the upper α-quantile of the SB distribution of the J-test statistic. Then, the SB J-test rejects the hypothesis of correct moment specification if $\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) > \hat{\xi}^{\mathrm{J}}_{\mathrm{S},\alpha}$. Theorem 2 shows that the SB parameter-tests are consistent irrespective of whether their null hypotheses are true or not. This provides assurance that the SB parameter-tests will maintain both the size and power for sufficiently large sample sizes. On the other hand, (9e) shows that the SB J-test is inconsistent even under correct specification.

Corollary 1 extends (9e) and shows that the SB J-test has asymptotic size equal to zero for nominal levels $\alpha \le 1/2$. This provides a more accurate description for the asymptotic size of the SB J-test (see paragraph 2, Lindsay and Qu 2003, p. 406) and provides a theoretical explanation for the empirical findings that the SB J-test may never reject a correctly specified model (see Brown and Newey 2002, p. 510).

Corollary 1 *Suppose the model* (1) *and conditions A–E in "Appendix" hold and that* $\alpha \le 1/2$. *Then*

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^J_{S,\alpha}\big) \to 0. \tag{10}$$

When the null hypotheses are true, these results can be refined. Specifically, Theorem 3 shows that the SB parameter-tests have size distortions of precise order $O(n^{-2})$, indicating an asymptotic improvement over their analytic versions [compare (11a)–(11c) with (4a)–(4c)], and that the size of the SB J-test converges to 0 with the precise rate n^{-1}. Closer examination of the proof shows that these sharp rates of parameter-tests continue to hold for any consistent estimator of the asymptotic covariance matrix of $\hat{\theta}_{\mathrm{T}}$ even under misspecification; hence, our results extend the higher-order improvements of the SB two-sided t-test developed by Lee (2014).

Theorem 3 *Suppose the model* (1) *and conditions A–F in "Appendix" hold. Then*

$$\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_T)^{-1}) > \hat{\xi}^T_{S,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{11a}$$

$$\Pr\big(\hat{L}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^L_{S,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{11b}$$

$$\Pr\big(\hat{S}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^S_{S,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{11c}$$

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^J_{S,\alpha}\big) = n^{-1}d_S(\alpha) + O(n^{-2}), \quad \alpha \le 1/2, \tag{11d}$$

where $d_S(\cdot) \neq 0$.

In the following, we consider the behavior of the SB parameter-tests if the function h is not centered in the bootstrap world. To this end, consider the following *naive* SB parameter-test statistics:

$$\begin{aligned}
\hat{T}^\circ(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) &= nh(\hat{\theta}^*_{\mathrm{T}})^T\big(H(\hat{\theta}^*_{\mathrm{T}})\hat{D}^*(\hat{\theta}^*_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1})^{-1}H(\hat{\theta}^*_{\mathrm{T}})^T\big)^{-1}h(\hat{\theta}^*_{\mathrm{T}}),\\
\hat{L}^\circ(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) &= n\hat{Q}^*(\tilde{\theta}^\circ_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) - n\hat{Q}^*(\hat{\theta}^*_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}),\\
\hat{S}^\circ(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}) &= n\hat{C}^*(\tilde{\theta}^\circ_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1})^T\hat{D}^*(\hat{\theta}^*_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1})^{-1}\hat{C}^*(\tilde{\theta}^\circ_{\mathrm{T}}, \hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1}),
\end{aligned}$$

where $\tilde{\theta}_{\mathrm{T}}^{\circ} = \mathrm{argmin}_{\theta \in \Theta_0} \hat{Q}^*(\theta, \hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1})$. Corollary 2, whose proof is similar to the proof of (9e), shows that, had we failed to center h, then these *naive* SB estimators of the null distributions of parameter-tests would be inconsistent even under H_0.

Corollary 2 *Suppose the model* (1) *and conditions A–E in "Appendix" hold. Then*

$$\mathscr{L}_{\mathrm{S}}\left(\hat{T}^{\circ}(\hat{V}^*(\hat{\theta}_T^*)^{-1})\middle|X_{1:n}\right) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2(U^2), \tag{12a}$$

$$\mathscr{L}_{\mathrm{S}}\left(\hat{L}^{\circ}(\hat{V}^*(\hat{\theta}_O^*)^{-1})\middle|X_{1:n}\right) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2(U^2), \tag{12b}$$

$$\mathscr{L}_{\mathrm{S}}\left(\hat{S}^{\circ}(\hat{V}^*(\hat{\theta}_O^*)^{-1})\middle|X_{1:n}\right) \underset{\mathrm{H}_0}{\xrightarrow{d}} \chi_r^2(U^2), \tag{12c}$$

where $U^2 \sim \chi_r^2$.

A direct consequence of Corollary 2 is that these naive SB parameter-tests have asymptotic size equal to zero for nominal levels $\alpha \leq 1/2$. Using a similar method of proof as of (9e), it readily follows that these results continue to hold for the naive versions of the CB and ELB parameter-tests.

3.2 The centered-bootstrap

Let X^* denote a generic bootstrap observation, i.e., $X^* \sim \mathbb{P}$ conditional on $X_{1:n}$. In the case of the standard-bootstrap,

$$\mathrm{E}_{\mathrm{S}}\left(g^*(X^*, \theta)\middle|X_{1:n}\right) = n^{-1}\sum_{i=1}^{n} g(X_i, \theta) = \bar{g}(\theta) \neq 0 \quad \text{for all } \theta \in \Theta, \tag{13}$$

and in particular, $\mathrm{E}_{\mathrm{S}}(g^*(X^*, \hat{\theta}_{\mathrm{T}})|X_{1:n}) \neq 0$, where $\mathrm{E}_{\mathrm{S}}(\cdot|X_{1:n})$ denotes the (conditional) expectation under the SB method. Thus, the model (1) does not hold in the SB world, a fact which is reflected in the failure of the SB to consistently estimate the null distribution of the J-test according to (9e).

The centered-bootstrap (CB) of Hall and Horowitz (1996) replaces $g^* = g$ with $g^*(x, \theta) = g(x, \theta) - \bar{g}(\hat{\theta}_{\mathrm{T}})$, so that $\mathrm{E}_{\mathrm{C}}(g^*(X^*, \hat{\theta}_{\mathrm{T}})|X_{1:n}) = 0$. Hence, the model (1) holds in the CB world, with $\hat{\theta}_{\mathrm{T}}$ playing the role of θ_0. Let $\mathscr{L}_{\mathrm{C}}(\hat{T}^*|X_{1:n})$ denote the conditional distribution given $X_{1:n}$ of a CB quantity $\hat{T}^*$. Theorem 4 shows that this modification repairs the inconsistency of the SB J-test [compare (14e) with (9e)].

Theorem 4 *Suppose the model* (1) *and conditions A–E in "Appendix" hold. Then*

$$\mathscr{L}_{\mathrm{C}}\left(n^{1/2}(\hat{\eta}^* - \hat{\eta})\middle|X_{1:n}\right) \xrightarrow{P} N(0, \Gamma_0), \tag{14a}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{T}^*(\hat{V}^*(\hat{\theta}_T^*)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{14b}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{14c}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi_r^2, \tag{14d}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{J}^{*}(\hat{V}^{*}(\hat{\theta}_{O}^{*})^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi^{2}_{q-p}. \tag{14e}$$

When the null hypotheses are true these results can be refined. Theorem 5 shows that the size distortions of the CB tests are of precise order $O(n^{-2})$, indicating an asymptotic improvement over the analytic tests [compare (15a)–(15d) with (4a)–(4d)]. Here the subscript "C" indicates that the critical values of the tests are estimated using the CB method. Theorem 5 extends the higher-order result for the CB symmetric t-test of Andrews (2002, Theorem 2(c)) and shows that the same exact higher-order improvements continue to hold for the CB parameter-tests and the J-test.

Theorem 5 *Suppose the model* (1) *and conditions A–F in "Appendix" hold. Then*

$$\Pr\left(\hat{T}(\hat{V}(\hat{\theta}_{T})^{-1}) > \hat{\xi}^{T}_{C,\alpha}\right) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{15a}$$

$$\Pr\left(\hat{L}(\hat{V}(\hat{\theta}_{O})^{-1}) > \hat{\xi}^{L}_{C,\alpha}\right) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{15b}$$

$$\Pr\left(\hat{S}(\hat{V}(\hat{\theta}_{O})^{-1}) > \hat{\xi}^{S}_{C,\alpha}\right) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{15c}$$

$$\Pr\left(\hat{J}(\hat{V}(\hat{\theta}_{O})^{-1}) > \hat{\xi}^{J}_{C,\alpha}\right) = \alpha + O(n^{-2}). \tag{15d}$$

3.3 The empirical-likelihood bootstrap

Brown and Newey (2002) address (13) by leaving $g^{*} = g$ unchanged and modifying instead the conditional distribution of X^{*}. Specifically, the conditional distribution of X^{*} is taken to be a weighted empirical distribution $\mathbb{P}_{\hat{\theta}_{\mathrm{T}}}$, with the weights chosen to satisfy $\mathrm{E}_{\mathrm{E}}(g^{*}(X^{*}, \hat{\theta}_{\mathrm{T}})|X_{1:n}) = 0$ while minimizing the (forward) Kullback-Leibler divergence between $\mathbb{P}$ and $\mathbb{P}_{\hat{\theta}_{\mathrm{T}}}$, where $\mathrm{E}_{\mathrm{E}}(\cdot|X_{1:n})$ denotes the (conditional) expectation with respect to $\mathbb{P}_{\hat{\theta}_{\mathrm{T}}}$. In other words, $\mathbb{P}_{\hat{\theta}_{\mathrm{T}}} = \sum_{i=1}^{n} \hat{w}_i \delta_{X_i}$, where the weights $\hat{w}_1, \ldots, \hat{w}_n > 0$ minimize $-\sum_{i=1}^{n} \log(\hat{w}_i)$ subject to

$$\sum_{i=1}^{n} \hat{w}_i g(X_i, \hat{\theta}_{\mathrm{T}}) = 0 \quad \text{and} \quad \sum_{i=1}^{n} \hat{w}_i = 1. \tag{16}$$

Brown and Newey refer to this method as the empirical-likelihood bootstrap (ELB).

The empirical-likelihood weights may fail to exist for a given sample. Specifically, (16) requires that the interior of the convex hull of $\{g(X_1, \hat{\theta}_{\mathrm{T}}), \ldots, g(X_n, \hat{\theta}_{\mathrm{T}})\}$ contains $0 \in \mathbb{R}^{q}$. Divergences other than the forward Kullback–Leibler divergence can be used in specifying the weights (Hall and Presnell 1999), but all require that $0 \in \mathbb{R}^{q}$ be contained in the interior of the convex hull of $\{g(X_1, \hat{\theta}_{\mathrm{T}}), \ldots, g(X_n, \hat{\theta}_{\mathrm{T}})\}$. There is no guidance in the ELB literature on how to choose the weights if they fail to exist. More importantly, even when the weights do exist, their large sample behavior under misspecification is difficult or even impossible to describe. This motivates us to propose an automatic method of selection of the weights which improves the properties of the ELB under both correct specification and misspecification.

To this end, let $\hat{\ell}(\hat{\theta}_{\mathrm{T}})$ denote the empirical-likelihood-ratio test statistic for testing $\mathrm{H}_0 : \mu(\theta_0) = 0$ against $\mathrm{H}_a : \mu(\theta_0) \neq 0$ (Owen 1990; Qin and Lawless 1994). If the weights exist, then $\hat{\ell}(\hat{\theta}_{\mathrm{T}}) = -2\sum_{i=1}^{n}\log(n\hat{w}_i)$, and if the weights fail to exist, set $\hat{\ell}(\hat{\theta}_{\mathrm{T}}) = \infty$. Under correct specification, then $\hat{\ell}(\hat{\theta}_{\mathrm{T}}) \xrightarrow{d} \chi^2_{q-p}$. Let (α_n) be a sequence of nominal levels such that $\alpha_n \to 0$ and $\log(\alpha_n)/n \to 0$ as $n \to \infty$, and set

$$\hat{w}_i = n^{-1} \quad \text{if} \quad \hat{\ell}(\hat{\theta}_{\mathrm{T}}) > \chi^2_{q-p;\alpha_n}, \quad 1 \le i \le n. \tag{17}$$

Thus, if there is (strong) evidence against the moment condition model such that $\hat{\ell}(\hat{\theta}_{\mathrm{T}}) > \chi^2_{q-p;\alpha_n}$, then the ELB weights are set equal to the uniform weights, and in this situation, the ELB and the SB methods are equivalent. Let $\hat{\xi}^{\mathrm{J}}_{\mathrm{E},\alpha}$ denote the upper α-quantile of $\mathscr{L}_{\mathrm{E}}(\hat{J}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{O}})^{-1})|X_{1:n})$, where $\mathscr{L}_{\mathrm{E}}(\hat{T}^*|X_{1:n})$ denotes the conditional distribution of an ELB quantity $\hat{T}^*$. Since $\hat{\ell}(\hat{\theta}_{\mathrm{T}}) > \chi^2_{q-p;\alpha_n}$ constitutes strong evidence against the model, we define the ELB J-test to reject H_0 if

$$\hat{J}(\hat{\theta}_{\mathrm{T}}) > \hat{\xi}^{\mathrm{J}}_{\mathrm{E},\alpha} \quad \text{or} \quad \hat{\ell}(\hat{\theta}_{\mathrm{T}}) > \chi^2_{q-p;\alpha_n}. \tag{18}$$

Using the following result about the quantiles of chi-squared distributions (see Inglot and Ledwina 2006, p. 586)

$$0 < \liminf_{n\to\infty} \frac{\chi^2_{q-p;\alpha_n}}{-\log(\alpha_n)} < \limsup_{n\to\infty} \frac{\chi^2_{q-p;\alpha_n}}{-\log(\alpha_n)} < \infty, \tag{19}$$

then $\Pr(\hat{\ell}(\hat{\theta}_{\mathrm{T}}) > \chi^2_{q-p;\alpha_n}) \to 0$ under correct specification. Since the weights exist with probability tending to one (according to Lemma C.2), the results of the present section are not affected by this choice.

Theorem 6 shows that the ELB parameter-tests are consistent and that the ELB repairs the inconsistency of the SB J-test under correct specification.

Theorem 6 *Suppose the model* (1) *and conditions A–E in "Appendix" hold,* $\alpha_n \to 0$*, and* $\log(\alpha_n)/n \to 0$*. Then*

$$\mathscr{L}_{\mathrm{E}}\left(n^{1/2}(\hat{\eta}^* - \hat{\eta})\middle|X_{1:n}\right) \xrightarrow{P} N(0, \Gamma_0), \tag{20a}$$

$$\mathscr{L}_{\mathrm{E}}\left(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_T)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{20b}$$

$$\mathscr{L}_{\mathrm{E}}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{20c}$$

$$\mathscr{L}_{\mathrm{E}}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{20d}$$

$$\mathscr{L}_{\mathrm{E}}\left(\hat{J}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1})\middle|X_{1:n}\right) \xrightarrow{P} \chi^2_{q-p}. \tag{20e}$$

Theorem 7 shows that, under additional regularity conditions, the ELB, like the CB, reduces the size distortions of the tests to the order $O(n^{-2})$. The subscript "E" indicates that the critical values of the tests are estimated using the ELB method. Brown and Newey (2002, Eq. 9) showed that the size distortions of the ELB symmetric t-test and J-test are of order $o(n^{-1})$. Theorem 7 shows a sharper rate; specifically, the size distortions of the ELB parameter-tests and the J-test are of exact order $O(n^{-2})$.

Theorem 7 *Suppose the model* (1) *and conditions A–F hold,* $\alpha_n = \exp(-n^{\beta})$ *with* $\beta \in (0, 1)$*, and* $\log \mathrm{E}(\exp(t^T f(X, \theta_0))) < \infty$ *for* t *in a neighborhood of* $0 \in \mathbb{R}^k$*. Then*

$$\Pr\big(\hat{T}(\hat{V}(\hat{\theta}_T)^{-1}) > \hat{\xi}^T_{E,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{21a}$$

$$\Pr\big(\hat{L}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^L_{E,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{21b}$$

$$\Pr\big(\hat{S}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^S_{E,\alpha}\big) \underset{\mathrm{H}_0}{=} \alpha + O(n^{-2}), \tag{21c}$$

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_O)^{-1}) > \hat{\xi}^J_{E,\alpha}\big) = \alpha + O(n^{-2}). \tag{21d}$$

4 Bootstrapping under misspecification

4.1 The standard-bootstrap

Under moment misspecification, the SB versions of θ_1 and θ_2 are $\hat{\theta}_\mathrm{O}$ and $\hat{\theta}_\mathrm{T}$, respectively. Theorem 8 shows the asymptotic properties of the SB under misspecification.

Theorem 8 *Suppose conditions A–E in "Appendix" hold. Then*

$$\mathscr{L}_\mathrm{S}\left(n^{1/2}(\hat{\eta}^* - \hat{\eta})\big|X_{1:n}\right) \xrightarrow{P} N(0, \Gamma), \tag{22a}$$

$$\mathscr{L}_\mathrm{S}\left(\hat{T}^*(\hat{V}^*(\hat{\theta}^*_T)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \mathscr{L}\Big(\sum_{j=1}^{r} \lambda^W_j U^2_j\Big), \tag{22b}$$

$$\mathscr{L}_\mathrm{S}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \mathscr{L}\Big(\sum_{j=1}^{p} \lambda^L_j U^2_j\Big), \tag{22c}$$

$$\mathscr{L}_\mathrm{S}\left(n^{-1/2}\hat{L}^*(\hat{V}^*(\hat{\theta}^*_T)^{-1})\big|X_{1:n}\right) \xrightarrow{P} N(0, \sigma^2_L), \tag{22d}$$

$$\mathscr{L}_\mathrm{S}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \mathscr{L}\Big(\sum_{j=1}^{p} \lambda^S_j U^2_j\Big), \tag{22e}$$

$$\mathscr{L}_\mathrm{S}\left(n^{-1/2}\big(\hat{S}^*(\hat{V}^*(\hat{\theta}^*_T)^{-1}) - \hat{S}(\hat{V}(\hat{\theta}_T)^{-1})\big)\big|X_{1:n}\right) \xrightarrow{P} N(0, \sigma^2_S), \tag{22f}$$

$$\mathscr{L}_\mathrm{S}\left(n(\bar{g}^*(\hat{\theta}^*_T) - \bar{g}(\hat{\theta}_T))^T \hat{V}^*(\hat{\theta}^*_T)^{-1}(\bar{g}^*(\hat{\theta}^*_T) - \bar{g}(\hat{\theta}_T))\big|X_{1:n}\right) \xrightarrow{P} \mathscr{L}\Big(\sum_{j=1}^{q} \lambda^J_j U^2_j\Big), \tag{22g}$$

$$\mathscr{L}_\mathrm{S}\left(n^{-1/2}(\hat{J}^*(\hat{V}^*(\hat{\theta}^*_O)^{-1}) - \hat{J}(\hat{V}(\hat{\theta}_O)^{-1}))\big|X_{1:n}\right) \xrightarrow{P} N(0, \sigma^2_J). \tag{22h}$$

Result (22a) shows that the SB estimator of the joint distribution of $\hat{\theta}_\mathrm{O}$ and $\hat{\theta}_\mathrm{T}$ is consistent [compare (22a) with (8a)]. Equations (22b)–(22f) show that the SB estimators of the null distributions of parameter-tests are consistent irrespective of the weight matrix used in the definition of the test statistics [compare (22b)–(22f) with (7a)–(7c) and (8b)–(8c)]. An immediate consequence of these results is that, if the model is misspecified, then the asymptotic sizes of the SB parameter-tests are equal to their nominal levels. Using (22g), we can construct a consistent SB confidence region for μ_2, which, similarly to its analytic version, it may be helpful in identifying the specific moment conditions that have failed. The result (22h) shows that, under misspecification, $\mathscr{L}_{\mathrm{S}}(\hat{J}^*(\hat{V}^*(\hat{\theta}_{\mathrm{O}}^*)^{-1})|X_{1:n})$ is approximately a normal distribution centered at $\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1})$. Since the upper α-quantile of a normal distribution is greater than (or equal to) its mean for $\alpha \leq 1/2$, then

$$\Pr\big(\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{O}})^{-1}) > \hat{\xi}^{\mathrm{J}}_{\mathrm{S},\alpha}\big) \to 0 \quad \text{for} \quad \alpha \leq 1/2. \tag{23}$$

Hence, the power of the SB J-test converges to 0 for nominal levels $\alpha \leq 1/2$. Therefore, (10) and (23) together imply that the asymptotic size and power of the SB J-test are equal to zero for nominal levels $\alpha \leq 1/2$.

4.2 The centered-bootstrap

Since the moment condition model is constrained to hold in the CB world irrespective of whether or not it holds in the population, the CB versions of θ_1 and θ_2 are both equal to $\hat{\theta}_{\mathrm{T}}$. Let $\hat{\eta}_{\mathrm{c}} = (\hat{\theta}_{\mathrm{T}}^T, \hat{\theta}_{\mathrm{T}}^T)^T \in \mathbb{R}^{2p}$, $A_3 = G_2^T W_0 G_2$, $A_4 = G_2^T V_2^{-1} G_2$, and

$$\Gamma_2 = \begin{pmatrix} A_3^{-1} G_2^T W_0 V_2 W_0 G_2 A_3^{-1} & A_4^{-1} \\ A_4^{-1} & A_4^{-1} \end{pmatrix}.$$

Theorem 9 shows the asymptotic properties of the CB under misspecification.

Theorem 9 *Suppose conditions A–E in "Appendix" hold. Then*

$$\mathscr{L}_{\mathrm{C}}\left(n^{1/2}(\hat{\eta}^* - \hat{\eta}_c)\big|X_{1:n}\right) \xrightarrow{P} N(0, \Gamma_2), \tag{24a}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{T}^*(\hat{V}^*(\hat{\theta}_T^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{24b}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{24c}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{L}^*(\hat{V}^*(\hat{\theta}_T^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{24d}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{24e}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{S}^*(\hat{V}^*(\hat{\theta}_T^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_r, \tag{24f}$$

$$\mathscr{L}_{\mathrm{C}}\left(\hat{J}^*(\hat{V}^*(\hat{\theta}_O^*)^{-1})\big|X_{1:n}\right) \xrightarrow{P} \chi^2_{q-p}. \tag{24g}$$

Result (24a) shows that the CB estimator of the joint distribution of $\hat{\theta}_{\mathrm{O}}$ and $\hat{\theta}_{\mathrm{T}}$ is inconsistent [compare (24a) with (8a)]. Results (24b)–(24f) show that the CB estimators of the null distributions of parameter-tests are inconsistent [compare (24b)–(24f) with (7a)–(7c) and (8b)–(8c)]. Result (24g) shows that the CB estimator of the null distribution of the J-test is consistent under H_a; indeed, under misspecification, the null hypothesis of the J-test is false, and thus, the CB J-test is consistent.

4.3 The empirical-likelihood bootstrap

The ELB versions of θ_1 and θ_2 are $\hat{\theta}_\mathrm{O}$ and $\hat{\theta}_\mathrm{T}$, respectively. By (19), it readily follows that $\Pr(\hat{\ell}(\hat{\theta}_\mathrm{T}) > \chi^2_{q-p;\alpha_n}) \to 1$ under misspecification, and thus, the ELB and the SB methods are equivalent with probability tending to one. By Theorem 8, the ELB joint distribution of $\hat{\theta}_\mathrm{O}$ and $\hat{\theta}_\mathrm{T}$ and of the parameter-test statistics are consistent. Theorem 8 also implies that the ELB J-test statistic is inconsistent in this case. However, by (18) and (19), the power of the ELB J-test tends to 1; this is an improvement compared to the SB J-test whose power tends to 0 according to (23). By imposing additional constraints on the first- and second-order moments of the weighted empirical distribution, one may develop a modified version of the ELB, which, under mild additional conditions, provides a consistent estimator of the null distribution of the J-test. However, in practice, the weights may fail to exist, and thus, although theoretically interesting, this ELB version may not be feasible/useful.

5 Summary of theoretical results

Higher-order expansions of the sizes of the analytic and the bootstrap tests show that, if the model is correctly specified, then the size distortions of the bootstrap parameter-tests are of smaller order than those of the analytic chi-squared tests. On the other hand, the SB distribution of the J-test statistic is inconsistent, while the CB and the ELB J-tests are consistent. Under misspecification, the SB and the ELB parameter-tests are consistent while the corresponding CB parameter-tests are inconsistent. Lastly, the CB distribution of the J-test is consistent, while the corresponding SB and ELB distributions are inconsistent.

These results, summarized in Table 1, enable us to propose an approach which improves accuracy over the analytic chi-squared tests under correct specification while achieving consistency under misspecification. If misspecification is suspected, we first

Table 1 Summary of theoretical results

Test	SB	CB	ELB
Wald	✓✓✓	✓✓✗	✓✓✓
LRT	✓✓✓	✓✓✗	✓✓✓
ST	✓✓✓	✓✓✗	✓✓✓
J-test	✗✗✗	✓✓✓	✓✓✗

The first mark indicates whether the bootstrap estimators of the null distributions of the tests are consistent under correct specification. The second mark indicates whether the size distortions of the bootstrap tests are of smaller order than of the analytic tests under correct specification. The third mark indicates whether the bootstrap estimators of the null distributions of the tests are consistent under misspecification perform the CB or the ELB J-test. If the CB (or the ELB) J-test does not reject the model, then we proceed under the assumption that the model is correctly specified, and in this case, we can use either the SB, or the CB, or the ELB to test hypotheses and to construct bootstrap confidence regions for θ_0. If the CB (or the ELB) J-test rejects the moment condition model, then we use either the SB or the ELB to test statements and to construct confidence regions for the target parameter under misspecification θ_2. It is important to note, however, that in this situation, the evidence suggests that θ_2 is not a solution of the original moment conditions and, indeed, that no such solution exists. As a result, the parameter itself does not carry the same practical interpretation that it might have had the moment conditions to hold.

6 Empirical results

6.1 A dynamic panel data model

In this section, we present the results of an extensive simulation study of the finite sample performance of the analytic and the bootstrap GMM tests for a dynamic autoregressive panel data model with individual effects (see, e.g., Arellano and Bond 1991; Blundell and Bond 1998; Hsiao 2003)

$$X_{i,j+p} - \eta_i = \theta_1(X_{i,j+p-1} - \eta_i) + \cdots + \theta_p(X_{i,j} - \eta_i) + \epsilon_{i,j+p}, \tag{25}$$

where $X_{i,j}$ is the response of the ith unit at time j, with $1 \le i \le n$ and $1 \le j \le m-p$, $\theta_1, \ldots, \theta_p$ are the autoregressive parameters, $\eta_1, \ldots, \eta_n$ are the individual (random) effects, with $\eta_i \sim$ i.i.d.$(0, \sigma_\eta^2)$, and $\epsilon_{i,j}$ are the random errors, with $\epsilon_{i,j} \sim$ i.i.d.$(0, \sigma_\epsilon^2)$. We rewrite (25) in an equivalent form as follows

$$X_{i,j+p} = \theta_1 X_{i,j+p-1} + \cdots + \theta_p X_{i,j} + (1 - \theta_1 - \cdots - \theta_p)\eta_i + \epsilon_{i,j+p}. \tag{26}$$

Assume that for each i, $\{X_{i,j} : 1 \le j \le m\}$ is a causal sequence, i.e., the current values $X_{i,j}$ are independent of the future errors $\{\epsilon_{i,k} : k \ge j+1\}$. Applying the lag-one difference operator Δ to the both sides of (26), we obtain

$$\mathrm{E}\big(X_{i,k}(\Delta X_{i,l} - \theta_1 \Delta X_{i,l-1} - \cdots - \theta_p \Delta X_{i,l-p})\big) = 0, \tag{27}$$

for all $1 \le i \le n$, where $\Delta X_{i,j} = X_{i,j} - X_{i,j-1}$, $1 \le k \le m-2$ and $p+2 \le l \le m$, with $k \le l-2$. These moment conditions have been proposed by Arellano and Bond (1991). Blundell and Bond (1998) augmented (27) with moment conditions obtained by substituting $\Delta X_{i,l-2}$ for $X_{i,k}$ in (27).

By aggregating the moment functions corresponding to each given lag $(l-k)$ between the instrument $X_{i,k}$ and the vector of differences $(\Delta X_{i,l}, \ldots, \Delta X_{i,l-p})$, we obtain $m-2$ moment conditions. Specifically, let $g(x, \theta) : \mathbb{R}^m \times \Theta \mapsto \mathbb{R}^{m-2}$, where $g(x, \theta) = (g_j(x, \theta) : 1 \le j \le m-2)^T$ and $\theta = (\theta_1, \ldots, \theta_p)^T \in \mathbb{R}^p$, with

$$g_k(x,\theta) = \frac{1}{m-p-1}\sum_{l=p}^{m-2} x_{l-k+1}(\Delta x_{l+2} - \theta_1 \Delta x_{l+1} - \cdots - \theta_p \Delta x_{l-p+2}) \quad \text{(28a)}$$

for $1 \leq k \leq p$, and

$$g_k(x,\theta) = \frac{1}{m-k-1}\sum_{l=1}^{m-k-1} X_l(\Delta x_{l+k+1} - \theta_1 \Delta x_{l+k} - \cdots - \theta_p \Delta x_{l+k-p+1}) \quad \text{(28b)}$$

for $p+1 \leq k \leq m-2$. Preliminary simulation results show that the GMM estimator defined by the moment conditions (28) performs similarly to the GMM estimators proposed by Arellano and Bond (1991) and Blundell and Bond (1998) implemented in the R package plm (Croissant and Millo 2008, p. 21–24) for smaller panels (e.g., $n = 100$ and $m = 10$) and has improved finite sample properties for larger panels (e.g., $n = 100$ and $m = 50$).

6.2 Simulation results

Computations were done in the R language (R Core Team 2016), and the R package emplik (Zhou 2015) was used to calculate the ELB weights. We have implemented the analytic and the bootstrap GMM inference for dynamic autoregressive panel data models in an R package which is available from the authors upon request. Simulations were run on the HiPerGator, a cluster of servers hosted by the High-Performance Computing Center at University of Florida.

In the first simulation study, we generate the samples from the panel data model

$$X_{i,j+1} = \theta_1 X_{i,j} + (1-\theta_1)\eta_i + \epsilon_{i,j+1}, \quad (29)$$

where $\theta_1 = \nu$ and ν belongs to a grid of 101 equally spaced values in $[-0.50, 0.50]$. The random effects η_i and the errors $\epsilon_{i,j}$ are generated as i.i.d. $\sim N(0, 1)$. For each value of ν on the grid, we generate $S = 1000$ samples from the model (29) with $\theta_1 = \nu$, and for each simulated sample, $B = 999$ bootstrap samples are used to approximate the critical values of the bootstrap tests. In the simulation study and data analysis, we set the nominal levels $\alpha_n = n^{-3/2}$ for the ELB method (given by (17)). For sample sizes $n \geq 50$, we have $n^{-3/2} \leq 0.003$, and thus, this modification has essentially no practical effect on the size of the tests under correct specification. The null hypothesis of parameter-tests is $\mathrm{H}_0 : \theta_1 = 0$, with nominal level $\alpha = 0.05$. The empirical rejection rates of the tests are calculated as the proportion of samples for which the null hypothesis is rejected.

In the second simulation study, we generate the samples from the panel model

$$X_{i,j+2} = \theta_1 X_{i,j+1} + \theta_2 X_{i,j} + (1-\theta_1-\theta_2)\eta_i + \epsilon_{i,j+2}, \quad (30)$$

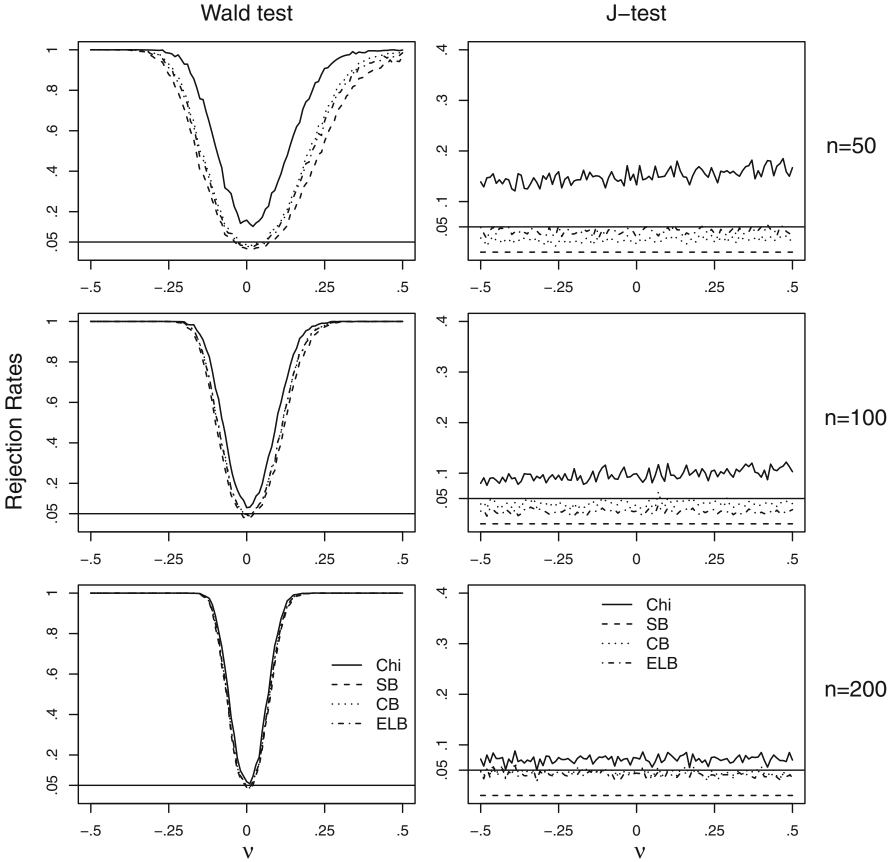

Fig. 1 Empirical rejection rates of the analytic (Chi), the SB, the CB, and the ELB Wald and J-tests under correct specification of the dynamic panel data model (29). The number of units is $n = 50, 100, 200$, the number of time points is $m = 10$, the number of simulated samples is $S = 1000$, and the number of bootstrap samples is $B = 999$. The null hypothesis of the Wald test is $\mathrm{H}_0 : \theta_1 = 0$, where $\theta_1 = \nu$

where $\theta_1 = -0.10$, $\theta_2 = \nu(\nu - 0.10)$, and ν belongs to a grid of 101 equally spaced values in $[-0.50, 0.50]$. In this case, we incorrectly fit the panel data model given by (29). Let $\theta_1^{(2)}$ denote the target parameter corresponding to θ_1, where recall that the target parameter corresponding to θ_1 is defined as the limiting value of its $\hat{\theta}_{\mathrm{T}}$. The null hypothesis of parameter-tests is set as $\mathrm{H}_0 : \theta_1^{(2)} = \theta_1^{(2)}$, so that their null hypotheses are true for all values of ν.

Figure 1 shows the empirical rejection rates of the analytic and the bootstrap tests under correct specification of the panel data model (29) plotted against the values of ν. The rejection rates of the LRT and ST are similar to the Wald test and are thus not included here. The rejection rates of the Wald test are simulation estimates of its size when $\nu = 0$ and of its power when $\nu \neq 0$. Since (29) is correctly specified for all values of ν, the rejection rates of the J-test are simulation estimates of its size. Figure 2 shows the empirical rejection rates of the analytic and the bootstrap tests for

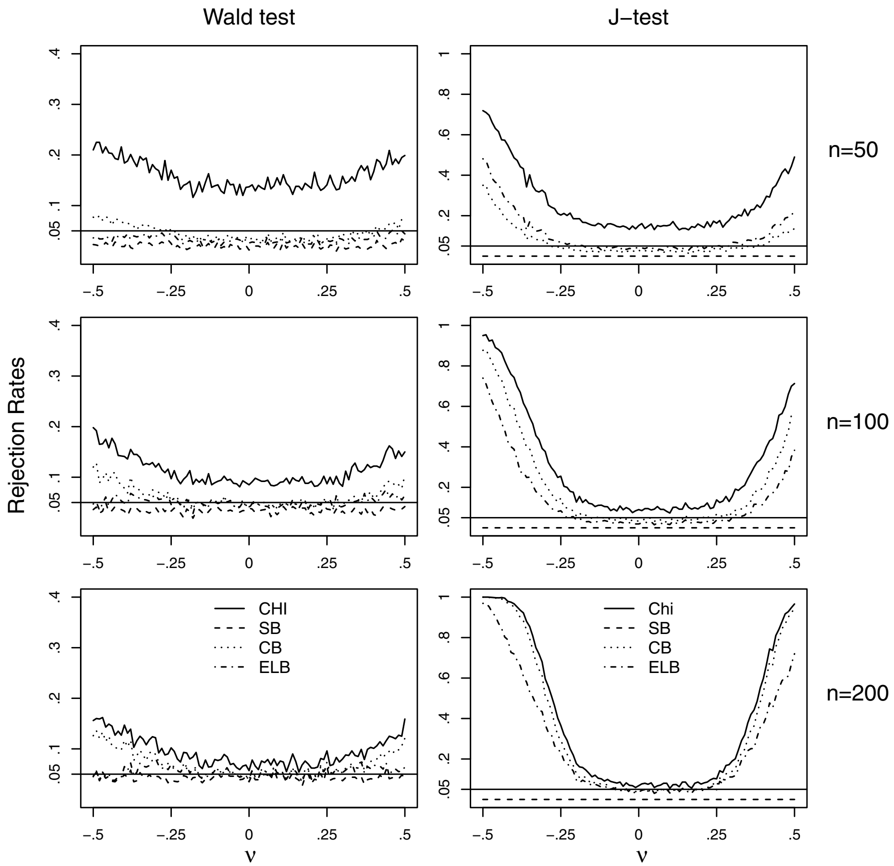

Fig. 2 Empirical rejection rates of the analytic (Chi), the SB, the CB, and the ELB Wald and J-tests under misspecification of the dynamic panel data model (30). The number of units is $n = 50, 100, 200$, the number of time points is $m = 10$, the number of simulated samples is $S = 1000$, and the number of bootstrap samples is $B = 999$. The null hypothesis of the Wald test is $\mathrm{H}_0 : \theta_1^{(2)} = \theta_1^{(2)}$, where $\theta_1^{(2)}$ is the implied value of θ_1

the misspecified model (30) plotted against the values of ν. Since $\mathrm{H}_0 : \theta_1^{(2)} = \theta_1^{(2)}$ is true for all values of ν, the rejection rates of the Wald test are simulation estimates of its size. The model (30) is correctly specified for $\nu = 0$ and $\nu = 0.10$ and misspecified for all other values of ν; thus, the rejection rates of the J-test are simulation estimates of its size when $\nu = 0$ and $\nu = 0.10$, and of its power otherwise. Table 2 shows the rejection rates of the analytic and the bootstrap tests under correct specification (model (29) with $\nu = 0$) and under misspecification (model (30) with $\nu = -0.50$).

The results of the simulation study follow the theoretical results fairly closely. Specifically, the analytic chi-squared tests have empirical sizes above the nominal levels, with sizes nearly double the nominal level even for fairly large sample sizes ($n = 100$). The CB and the ELB tests do a better job of maintaining the nominal levels when the model is correctly specified, while the SB tests tend to have rejection rates

Table 2 Empirical rejection rates of the analytic (Chi), the SB, the CB, and the ELB Wald, LRT, ST, and J-test at nominal level $\alpha = 0.05$, under correct specification (model (29) with $\nu = 0$) and under misspecification (model (30) with $\nu = -0.50$)

Test	n	Correct specification				Misspecification			
		Chi	SB	CB	ELB	Chi	SB	CB	ELB
Wald	50	0.16	0.02	0.04	0.03	0.21	0.02	0.08	0.04
	100	0.08	0.04	0.04	0.05	0.20	0.04	0.12	0.05
	200	0.07	0.04	0.04	0.04	0.16	0.05	0.13	0.04
LRT	50	0.16	0.02	0.04	0.03	0.23	0.03	0.08	0.04
	100	0.08	0.04	0.04	0.04	0.22	0.04	0.14	0.05
	200	0.07	0.04	0.04	0.04	0.18	0.05	0.15	0.05
ST	50	0.16	0.02	0.04	0.03	0.25	0.03	0.09	0.04
	100	0.08	0.04	0.04	0.04	0.25	0.04	0.17	0.05
	200	0.07	0.04	0.05	0.04	0.20	0.05	0.17	0.05
J	50	0.13	0.00	0.02	0.03	0.72	0.00	0.35	0.48
	100	0.10	0.00	0.04	0.03	0.95	0.00	0.88	0.74
	200	0.07	0.00	0.05	0.04	1.00	0.00	1.00	0.97

The number of units is $n = 50, 100, 200$, the number of time points is $m = 10$, the number of simulated samples is $S = 1000$, and the number of bootstrap samples is $B = 999$. The null hypothesis of parameter-tests under correct model specification is $\mathrm{H}_0 : \theta_1 = 0$ and under misspecification is $\mathrm{H}_0 : \theta_1^{(2)} = \theta_1^{(2)}$, where $\theta_1^{(2)}$ is the implied value of θ_1

below the nominal level, and extremely so in the case of the SB J-test, which never rejected the null hypothesis. Conversely, the CB parameter-tests have empirical sizes above the nominal level under misspecification, while the SB and the ELB parameter-tests maintain the nominal level quite well. The power of the SB J-test is 0, the CB and the ELB J-tests maintain the nominal level fairly closely, and while having similar power for $n = 50$ and $n = 100$, the CB J-test has better power for $n = 200$.

7 Electricity data

The data considered here comprise of the monthly average electricity usage (measured in KWh) for 254 households in a residential area in the state of Florida over 12 months from January 2011 to December 2011. The response variable $Y_{i,j}$ is the monthly average electricity usage of the ith household in month j, $1 \le i \le 254$ and $1 \le j \le 12$. To remove the overall trend, we center the data by subtracting the monthly average electricity usage from each observation; that is, let $X_{i,j} = Y_{i,j} - (1/254)\sum_{k=1}^{254} Y_{k,j}$. We fit the dynamic panel data model (29) to the data $\{X_{i,j}\}$, with $n = 254$ units and $m = 12$ time points.

We first test for the goodness-of-fit of the model. The J-test statistic is 66.866, and the p-values of the analytic, the SB, the CB, and the ELB J-tests are 0.000, 0.853, 0.001, and 0.000, respectively. Since both the CB and the ELB J-tests strongly reject the moment condition model, there is statistical evidence that the dynamic panel data

Table 3 Equal-tailed and symmetric analytic (Norm), SB, CB, and ELB lower and upper confidence limits of $\theta_1^{(2)}$ at 95% nominal confidence level

		Equal-tailed CI			Symmetric CI		
	Norm	SB	CB	ELB	SB	CB	ELB
Lower	0.795	0.768	0.754	0.769	0.712	0.754	0.710
Upper	0.851	0.956	0.892	0.956	0.935	0.893	0.936

model is misspecified. Next, let us suppose that we continue inference in spite of finding evidence of misspecification of the model. Let $\theta_1^{(2)}$ denote the target of inference (the implied value of θ_1). Using our bootstrap GMM methodology, we could use either the SB or the ELB methods to test statements and to construct confidence intervals about $\theta_1^{(2)}$. The Wald, the LRT, and the ST statistics for testing the null hypothesis $\mathrm{H}_0 : \theta_1^{(2)} = 0$ are 757.11, 671.50, and 595.56, respectively, and the p-values of the analytic and the bootstrap parameter-tests are all equal to 0.000. Thus, there is statistical evidence that $\theta_1^{(2)} \neq 0$. Table 3 shows the 95% analytic (Norm), SB, CB, and ELB equal-tailed and symmetric confidence limits for $\theta_1^{(2)}$. Note that the SB and the ELB confidence intervals are similar, as expected, and while the CB confidence intervals are approximately 30% shorter, they do not maintain the nominal level (according to Theorem 9).

Appendix: Regularity conditions

Let $\hat{\Psi}(\zeta) : \mathbb{R}^{2p} \mapsto \mathbb{R}^{2p}$ be given by $\hat{\Psi}(\zeta) = (\hat{C}(\zeta_1, \hat{W})^T, \hat{C}(\zeta_2, \hat{V}(\zeta_1)^{-1})^T)^T$, where $\zeta = (\zeta_1^T, \zeta_2^T)^T \in \mathbb{R}^{2p}$. We assume the following regularity conditions:

A. $\Theta \subset \mathbb{R}^p$ is compact, θ_1 and θ_2 are unique minimizers and interior points of Θ, i.e., $Q(\theta_1, W_0) < Q(\zeta_1, W_0)$ for all $\zeta_1 \neq \theta_1$ and $Q(\theta_2, V_1^{-1}) < Q(\zeta_2, V_1^{-1})$ for all $\zeta_2 \neq \theta_2$.
B. $g(x, \zeta_1)$ is measurable in x for each $\zeta_1 \in \Theta$ and twice continuously differentiable in ζ_1 for P_0-almost all $x \in \mathscr{X}$. There exists a measurable function $\kappa(x)$ with $\mathrm{E}(\kappa(X)^4) < \infty$ such that $\|g(x, \zeta_1)\| \leq \kappa(x)$ and $\|\partial^k_{i_1 \ldots i_k} g(x, \zeta_1)\| \leq \kappa(x)$ for P_0-almost all $x \in \mathscr{X}$ and $k = 1, 2$, where $\|\cdot\|$ is the Euclidean norm and $\partial^k_{i_1 \ldots i_k} g(x, \zeta_1) = \partial^k g(x, \zeta_1)/\partial\zeta_{1,i_1} \ldots \partial\zeta_{1,i_k}$.
C. V_1, V_2, A_1, and A_2 are non-singular, and $\operatorname{rank}(G_1) = \operatorname{rank}(G_2) = p$. The function $h : \mathbb{R}^p \to \mathbb{R}^r$ is continuously differentiable, with $\operatorname{rank}(H_2) = r \leq p$.
D. $\hat{W} \xrightarrow{P} W_0$ and $\hat{W}^* \xrightarrow{P} W_0$, where W_0 is nonrandom, symmetric, and positive definite.
E. The following asymptotic results hold:

$$n^{1/2}\big(\hat{\psi}(\eta_0) - \psi(\eta_0)\big) \xrightarrow{d} N(0, \Sigma), \tag{31a}$$

$$\mathscr{L}_{\mathrm{S}}\left(n^{1/2}(\hat{\psi}^*(\hat{\eta}) - \hat{\psi}(\hat{\eta}))\middle|X_{1:n}\right) \xrightarrow{P} N(0, \Sigma), \tag{31b}$$

where $\eta_0 = (\theta_1^T, \theta_2^T)^T$,

$$\psi(\eta_0) = \begin{pmatrix} \mu(\theta_1) \\ \mathrm{vec}(G(\theta_1)^T) \\ \mathrm{vec}(W_0) \\ \mathrm{vec}(V(\theta_1)) \\ \mu(\theta_2) \\ \mathrm{vec}(G(\theta_2)^T) \end{pmatrix}, \hat{\psi}(\eta_0) = \begin{pmatrix} \bar{g}(\theta_1) \\ \mathrm{vec}(\hat{G}(\theta_1)^T) \\ \mathrm{vec}(\hat{W}) \\ \mathrm{vec}(\hat{V}(\theta_1)) \\ \bar{g}(\theta_2) \\ \mathrm{vec}(\hat{G}(\theta_2)^T) \end{pmatrix},$$

$$\hat{\psi}^*(\hat{\eta}) = \begin{pmatrix} \bar{g}^*(\hat{\theta}_{\mathrm{O}}) \\ \mathrm{vec}(\hat{G}^*(\hat{\theta}_{\mathrm{O}})^T) \\ \mathrm{vec}(\hat{W}^*) \\ \mathrm{vec}(\hat{V}^*(\hat{\theta}_{\mathrm{O}})) \\ \bar{g}^*(\hat{\theta}_{\mathrm{T}}) \\ \mathrm{vec}(\hat{G}^*(\hat{\theta}_{\mathrm{T}})^T) \end{pmatrix}.$$

F. There exists a measurable map $f : \mathbb{R}^d \times \Theta \to \mathbb{R}^k$ such that the random variables $Z_1, \ldots, Z_n$ are i.i.d. in $\mathbb{R}^k$, where $Z_i = f(X_i, \theta_0)$, $q < k$, $\mathrm{E}(Z_i) = 0$, $\mathrm{Var}(Z_i) = I_k$, and

$$\begin{aligned} \hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = \hat{Z}^T \big(&\Upsilon_0 + n^{-1/2}\Upsilon_1(\hat{Z} \otimes I_k) + n^{-1}\Upsilon_2(\hat{Z} \otimes I_{k^2})(\hat{Z} \otimes I_k) \\ &+ n^{-3/2}\Upsilon_3(\hat{Z} \otimes I_{k^3})(\hat{Z} \otimes I_{k^2})(\hat{Z} \otimes I_k)\big)\hat{Z} + O_P(n^{-2}), \end{aligned} \tag{32}$$

$\Upsilon_j \in \mathbb{R}^{k \times k^j}$, $0 \le j \le 3$, $\Upsilon_0 \in \mathbb{R}^{k \times k}$ is idempotent of rank r, $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$, and $\hat{Z} = n^{1/2}\bar{Z}$. We assume that an Edgeworth expansion of $F_n(x) = \Pr(\hat{Z} \le x)$ holds, i.e.,

$$F_n(x) = \int_{(-\infty,x)} \big(1 + n^{-1/2}p_1(z) + n^{-1}p_2(z) + n^{-3/2}p_3(z)\big)\phi(z)\,\mathrm{d}z + O(n^{-2}),$$

where $\phi(z)$ is the density of $N(0, I_k)$ and $p_j(z)$ are odd/even polynomials of z for odd/even $1 \le j \le 3$, i.e., $p_j(z) = (-1)^j p_j(-z)$. We further assume that

$$\begin{aligned} \hat{T}^*(\hat{V}^*(\hat{\theta}_{\mathrm{T}}^*)^{-1}) = \hat{Z}^{*T}\big(&\hat{\Upsilon}_0 + n^{-1/2}\hat{\Upsilon}_1(\hat{Z}^* \otimes I_k) + n^{-1}\hat{\Upsilon}_2(\hat{Z}^* \otimes I_{k^2})(\hat{Z}^* \otimes I_k) \\ &+ n^{-3/2}\hat{\Upsilon}_3(\hat{Z}^* \otimes I_{k^3})(\hat{Z}^* \otimes I_{k^2})(\hat{Z}^* \otimes I_k)\big)\hat{Z}^* + O_P(n^{-2}), \end{aligned} \tag{33}$$

where $\hat{Z}^* = n^{1/2}(\bar{Z}^* - \bar{Z})$, $\hat{\Upsilon}_j$ are the sample versions of Υ_j, and a (conditional) Edgeworth expansion of $\hat{F}(x) = \Pr(\hat{Z}^* \le x | X_{1:n})$ holds, i.e.,

$$\hat{F}(x) = \int_{(-\infty,x)} \big(1 + n^{-1/2}\hat{p}_1(z) + n^{-1}\hat{p}_2(z) + n^{-3/2}\hat{p}_3(z)\big)\phi(z)\,\mathrm{d}z + O_P(n^{-2}),$$

where $\hat{p}_j(z)$ are the sample versions of $p_j(z)$. Similar expansions hold for the LRT and the ST statistics. In the case of the J-test statistic, assume that

$$\hat{J}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = \hat{Z}^T\big(\Phi_0 + n^{-1/2}\Phi_1(\hat{Z}\otimes I_k) + n^{-1}\Phi_2(\hat{Z}\otimes I_{k^2})(\hat{Z}\otimes I_k) \\ + n^{-3/2}\Phi_3(\hat{Z}\otimes I_{k^3})(\hat{Z}\otimes I_{k^2})(\hat{Z}\otimes I_k)\big)\hat{Z} + O_P(n^{-2}), \tag{34}$$

where $\Phi_j \in \mathbb{R}^{k\times k^j}$, $0 \le j \le 3$, and $\Phi_0 \in \mathbb{R}^{k\times k}$ is idempotent of rank $q - p$. We further assume that

$$\begin{aligned}\hat{J}^*(\hat{V}^*(\hat{\theta}^*_{\mathrm{T}})^{-1}) = (\hat{Z}^* + \hat{Z})^T\big(&\hat{\Phi}_0 + n^{-1/2}\hat{\Phi}_1(\hat{Z}^*\otimes I_k) \\ &+n^{-1}\hat{\Phi}_2(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k) \\ &+n^{-3/2}\hat{\Phi}_3(\hat{Z}^*\otimes I_{k^3})(\hat{Z}^*\otimes I_{k^2})(\hat{Z}^*\otimes I_k)\big)(\hat{Z}^* + \hat{Z}) \\ &+O_P(n^{-2}),\end{aligned} \tag{35}$$

where $\hat{\Phi}_j$ are the sample versions of Φ_j.

Remark 1 Conditions A–D are standard regularity conditions for identification of η_0 and consistency of $\hat{\eta}$. Note that condition E is similar to condition (12) of Hall and Inoue (2003). The main difference between the two regularity conditions consists of that Hall and Inoue studied the asymptotic properties of efficient GMM estimators which are defined in terms of one weight matrix. In our case, the two-step GMM estimator is defined in terms of two weight matrices; specifically, the pilot weight matrix $\hat{W}$, which is used to compute $\hat{\theta}_{\mathrm{O}}$, and $\hat{V}(\hat{\theta}_{\mathrm{O}})$, which is used to calculate $\hat{\theta}_{\mathrm{T}}$. Note further that if $\hat{W} = W_0$, then (31a) holds by the central limit theorem and (31b) holds by the central limit theorem for the bootstrap (see, e.g., Van der Vaart 1998, Theorem 23.4).

Remark 2 If $g(x, \theta)$ is six times continuously differentiable (in θ) for P_0-almost all $x \in \mathscr{X}$ and $h(\theta)$ is five times continuously differentiable, then (32) holds. To this end, Taylor expansion of $\hat{\Psi}(\hat{\eta})$ about η yields

$$\begin{aligned}\hat{\Psi}(\hat{\eta}) = \hat{\Psi}(\eta_0) &+ \sum_{i=1}^{2p}\partial_i\hat{\Psi}(\eta_0)(\hat{\eta}_i - \eta_{0,i}) + (1/2)\sum_{i,j=1}^{2p}\partial^2_{ij}\hat{\Psi}(\eta_0)(\hat{\eta}_i - \eta_{0,i})(\hat{\eta}_j - \eta_{0,j}) \\ &+(1/6)\sum_{i,j,k=1}^{2p}\partial^3_{ijk}\hat{\Psi}(\eta_0)(\hat{\eta}_i - \eta_{0,i})(\hat{\eta}_j - \eta_{0,j})(\hat{\eta}_k - \eta_{0,k}) \\ &+(1/24)\sum_{i,j,k,l=1}^{2p}\partial^4_{ijkl}\hat{\Psi}(\eta_0)(\hat{\eta}_i - \eta_{0,i})(\hat{\eta}_j - \eta_{0,j})(\hat{\eta}_k - \eta_{0,k})(\hat{\eta}_l - \eta_{0,l}) \\ &+O_P(n^{-5/2}),\end{aligned}$$

where $\eta_0 = (\eta_{0,1}, \ldots, \eta_{0,2p})^T \in \mathbb{R}^{2p}$. Hence

$$\begin{aligned}0 &= \hat{\Psi}(\eta_0) + \nabla\hat{\Psi}(\eta_0)(\hat{\eta} - \eta_0) + (1/2)\nabla^2\hat{\Psi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) \\ &\quad +(1/6)\nabla^3\hat{\Psi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^2}\big)\big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) \\ &\quad +(1/24)\nabla^4\hat{\Psi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^3}\big)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^2}\big) \\ &\quad \big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) + O_P(n^{-5/2}),\end{aligned} \tag{36}$$

where $\nabla\hat{\Psi}(\eta_0) = (\partial_i\hat{\Psi}(\eta_0) : 1 \leq i \leq 2p) \in \mathbb{R}^{2p\times 2p}$, $\nabla^2\hat{\Psi}(\eta_0) = (\partial_i\nabla\hat{\Psi}(\eta_0) : 1 \leq i \leq 2p) \in \mathbb{R}^{2p\times(2p)^2}$, $\nabla^3\hat{\Psi}(\eta_0) = (\partial_i\nabla^2\hat{\Psi}(\eta_0) : 1 \leq i \leq 2p) \in \mathbb{R}^{2p\times(2p)^3}$, and $\nabla^4\hat{\Psi}(\eta_0) = (\partial_i\nabla^3\hat{\Psi}(\eta_0) : 1 \leq i \leq 2p) \in \mathbb{R}^{2p\times(2p)^4}$.
Let

$$\hat{\eta} = \eta_0 + n^{-1/2}\hat{a} + n^{-1}\hat{b} + n^{-3/2}\hat{c} + n^{-2}\hat{d} + O_P(n^{-5/2}) \tag{37}$$

be an asymptotic expansion of $\hat{\eta}$ about η_0, where $\hat{a} = \hat{b} = \hat{c} = \hat{d} = O_P(1)$. Replacing $\hat{\eta} - \eta_0$ given by (37) in (36), we obtain the following expressions for $\hat{a}$, $\hat{b}$, $\hat{c}$, and $\hat{d}$:

$$\begin{aligned}\hat{a} &= -n^{1/2}\nabla\hat{\Psi}(\eta_0)^{-1}\hat{\Psi}(\eta_0), \\ \hat{b} &= -(1/2)\nabla\hat{\Psi}(\eta_0)^{-1}\nabla^2\hat{\Psi}(\eta_0)(\hat{a} \otimes I_{2p})\hat{a}, \\ \hat{c} &= -\nabla\hat{\Psi}(\eta_0)^{-1}\big\{(1/2)\nabla^2\hat{\Psi}(\eta_0)\big((\hat{a} \otimes I_{2p})\hat{b} + (\hat{b} \otimes I_{2p})\hat{a}\big) \\ &\qquad + (1/6)\nabla^3\hat{\Psi}(\eta_0)(\hat{a} \otimes I_{(2p)^2})(\hat{a} \otimes I_{2p})\hat{a}\big\}, \\ \hat{d} &= -\nabla\hat{\Psi}(\eta_0)^{-1}\big\{(1/2)\nabla^2\hat{\Psi}(\eta_0)\big((\hat{a} \otimes I_{2p})\hat{c} + (\hat{b} \otimes I_{2p})\hat{b} + (\hat{c} \otimes I_{2p})\hat{a}\big) \\ &\qquad + (1/6)\nabla^3\hat{\Psi}(\eta_0)\big((\hat{a} \otimes I_{(2p)^2})(\hat{b} \otimes I_{2p})\hat{a} + (\hat{a} \otimes I_{(2p)^2})(\hat{a} \otimes I_{2p})\hat{b} \\ &\qquad + (\hat{b} \otimes I_{(2p)^2})(\hat{a} \otimes I_{2p})\hat{a}\big) \\ &\qquad + (1/24)\nabla^4\hat{\Psi}(\eta_0)(\hat{a} \otimes I_{(2p)^3})(\hat{a} \otimes I_{(2p)^2})(\hat{a} \otimes I_{2p})\hat{a}\big\}.\end{aligned}$$

Next, write $\hat{T}(\hat{V}(\hat{\theta}_{\mathrm{T}})^{-1}) = n\hat{\varphi}(\hat{\eta})$, where $\hat{\varphi}(\zeta) : \mathbb{R}^{2p} \to \mathbb{R}$ is defined as

$$\hat{\varphi}(\zeta) = h(\zeta_2)^T\big(H(\zeta_2)\hat{D}(\zeta_2, \hat{V}(\zeta_2)^{-1})^{-1}H(\zeta_2)^T\big)^{-1}h(\zeta_2),$$

with $\zeta = (\zeta_1^T, \zeta_2^T)^T \in \mathbb{R}^{2p}$. Taylor expansion of $\hat{\varphi}(\hat{\eta})$ about η_0 yields

$$\begin{aligned}\hat{\varphi}(\hat{\eta}) &= \hat{\varphi}(\eta_0) + \nabla\hat{\varphi}(\eta_0)(\hat{\eta} - \eta_0) + (1/2)\nabla^2\hat{\varphi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) \\ &\quad + (1/6)\nabla^3\hat{\varphi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^2}\big)\big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) \\ &\quad + (1/24)\nabla^4\hat{\varphi}(\eta_0)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^3}\big)\big((\hat{\eta} - \eta_0) \otimes I_{(2p)^2}\big) \\ &\quad \times \big((\hat{\eta} - \eta_0) \otimes I_{2p}\big)(\hat{\eta} - \eta_0) + O_P(n^{-5/2}).\end{aligned}$$

Since $\hat{\varphi}(\eta_0) = 0$ and $\nabla\hat{\varphi}(\eta_0) = 0$, by substituting $\hat{\eta}$ given by (37), then the higher-order expansion (32) holds. The analytic and conditional Edgeworth expansions hold under additional finite moment and the Cramer condition $\limsup_{||t||\to\infty}$

$|\mathrm{E}(\exp(it^T Z))| < 1$ For more details, see Theorem 2 of Bhattacharya and Ghosh (1978) and Theorem 5.1 of Hall (1992).

References

Andrews, D. W. K. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, *70*, 119–162.

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*, 277–297.

Bhattacharya, R. N., Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *The Annals of Statistics*, *6*, 434–451.

Blundell, R., Bond, S. (1998). Initial conditions and moment conditions in dynamic panel data models. *Journal of Econometrics*, *87*, 115–143.

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

Croissant, Y., Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2), 21–24.

Dudley, R. M. (2002). *Real analysis and probability*. Cambridge: Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at jackknife. *The Annals of Statistics*, *7*, 1–26.

Hahn, J. (1996). A note on bootstrapping generalized method of moments estimators. *Econometric Theory*, *12*, 187–197.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer-Verlag.

Hall, P., Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, *64*, 891–916.

Hall, A. R., Inoue, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, *114*, 361–394.

Hall, P., Presnell, B. (1999). Intentionally biased bootstrap methods. Journal of the Royal Statistical Society. *Series B (Statistical Methodology)*, *61*, 143–158.

Hall, P., Wilson, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, *47*, 757–762.

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

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

Hsiao, C. (2003). *Analysis of panel data*. New York: Cambridge University Press.

Imbens, G. W., Lancaster, T. (1994). Combining micro and macro data in microeconomic models. *The Review of Economic Studies*, *61*, 655–680.

Inglot, T., Ledwina, T. (2006). Asymptotic optimality of new adaptive test in regression model. *Annales de l'Institute Henri Poincaré*, *42*, 579–590.

Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. *Journal of Econometrics*, *178*, 398–413.

Lindsay, B. G., Qu, A. (2003). Inference functions and quadratic score tests. *Statistical Science*, *18*, 394–410.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.

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

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. New York: Springer.

Zhang, B. (1999). Bootstrapping with auxiliary information. *The Canadian Journal of Statistics*, *27*, 237–249.

Zhou, M. (2015). *Emplik: Empirical likelihood ratio for censored/truncated data*. R package version 1.0-2.