University of Connecticut

Department of Economics Working Paper Series

Finite-sample Corrected Inference for Two-step GMM in Time Series

by

Jungbin Hwang
University of Connecticut

Gonzalo Valdés
Universidad de Tarapacá

Working Paper 2020-02
January 2020

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

Finite-sample Corrected Inference for Two-step GMM in Time Series *

Jungbin Hwang
Department of Economics,
University of Connecticut

Gonzalo Valdés
Departamento de Ingeniería Industrial y de Sistemas,
Universidad de Tarapacá

January, 2020

Abstract

This paper develops a finite-sample corrected and robust inference for efficient two-step generalized method of moments (GMM). One of the main challenges in efficient GMM is that we do not observe the moment process and have to use the estimated moment process to construct a GMM weighting matrix. We use a non-parametric long run variance estimator as the optimal GMM weighting matrix. To capture the estimation uncertainty embodied in the weight matrix, we extend the finite-sample corrected formula of Windmeijer (2005) to a heteroskedasticity autocorrelated robust (HAR) inference in time series setting. Using fixed-smoothing asymptotics, we show that our new test statistics lead to standard asymptotic F or t critical values and improve the finite sample performance of existing HAR robust GMM tests.

JEL Classification: C12, C13, C32

Keywords: Generalized Method of Moments, Heteroskedasticity Autocorrelated Robust, Finite-sample Correction, Fixed-smoothing Asymptotics, t and F tests.

*Email: jungbin.hwang@uconn.edu and gvaldes@uta.cl. Correspondence to: Jungbin Hwang, 365 Fairfield Way, U-1063, Storrs, CT 06269-1063.

1 Introduction

The generalized method of moments (GMM, Hansen, 1982) is one of the most widely applied methods in econometrics. In efficient GMM method, a feasible two-step estimator requires a consistent estimate of the variance-covariance matrix to weight the moment conditions. In the estimation of the weight matrix, the moment process is unobservable and has to be approximated by plugging an initial GMM estimator into the moment function. Windmeijer (2005) points out that the estimation uncertainty from the plugged-in estimator contributes to the finite-sample variability of the feasible two-step GMM estimator. He shows that the extra variation generated by the estimated weight matrix explains much of the difference between the estimated asymptotic variance and the actual finite sample variance of the GMM estimator. Windmeijer (2005) also proposes a finite-sample bias-corrected variance formula, which corrects the bias arising from the estimated efficient weight matrix instead of the true value. Windmeijer (2005)'s corrected variance formula has been popularly used in a wide variety of econometric models with high impact, e.g., Roodman (2009), Brown et al. (2009), Oberholzer-Gee and Strumpf (2007), and many others.

A fundamental assumption in Windmeijer (2005) is that the moment process is independent and identically distributed (i.i.d.). For time-series data, which is the focus of our paper, the i.i.d assumption in Windmeijer (2005) makes his corrected variance formula not applicable. This is because the optimal weight matrix is no longer a simple average of the estimated moment process. To extend Windmeijer's approach to time-series GMM, we need to consider the optimal weighting matrix as the long-run variance (LRV) of the true moment process, which is usually estimated by a non-parametric kernel or a series method. Due to the non-parametric nature of the LRV estimator, in time-series, the two-step GMM estimator is exposed to even higher variation from the estimated weight matrix. Consequently, the standard asymptotic variance formula without a finite-sample correction is severely biased, and the associated GMM tests suffer from excessive size distortions.

In this paper, we develop a finite-sample corrected and heteroskedasticity autocorrelated robust (HAR) inference for the efficient GMM method in the time-series setting. By explicitly considering the non-parametric LRV estimator, our finite-sample corrected variance formula extends Windmeijer (2005)'s one to none i.i.d data. Since our corrected variance formula is designed to take into account the extra variation due to the estimation of the LRV, it leads to a more accurate estimate of the actual finite-sample variance estimator. The key step of our approximation is that, instead of eliminating the estimation uncertainty of the initial estimator, which is of small stochastic order of magnitude, we explicitly derive the associated small order terms and use them to construct the finite-sample corrected variance formula. This paper formally shows that the finite-sample corrected variance can be consistently estimated. We show that this consistency does not depend on whether the smoothing parameter in the LRV estimator is fixed or is increasing with respect to the sample size.

With our finite-sample corrected variance estimator, we construct t and Wald statistics for the testing problem. To derive the asymptotic distributions of the finite-sample corrected statistics, we employ an alternative type of asymptotics from the HAR literature, which is called "fixed smoothing asymptotics." The fixed smoothing asymptotics holds the smoothing parameter in the LRV estimator fixed as the sample size increases. In the context of the efficient two-step GMM, Sun (2014b) and Hwang and Sun (2017) show that the alternative asymptotics yields to more accurate approximations, while the conventional normal and chi-square approximations poorly perform in finite samples. Together with the finite-sample variance corrected formula, the limiting distributions derived under the fixed-smoothing asymptotics provides a valid solution to the efficient GMM inference problem. The resulting fixed smoothing limiting distributions are highly non-standard; however, we modify the corrected test statistics and show that they are asymptotically standard t and F distributed. The standard t and F limits are very appealing in practical applications because, once practitioners apply the finite-sample corrected variance formula to construct the test statistics, the standard t and F critical values are readily available from any standard statistical table. No further simulations or re-sampling methods are needed.

Different approaches to the efficient GMM inference problem have been proposed in literature. For example, Hansen et al. (1996) proposes continuously updated GMM methods. A bootstrap approach for GMM is proposed in Hall and Horowitz (1996), Brown and Newey (2002) and Lee (2014, 2016). Numerical evidence on the finite sample performance of the asymptotic and bootstrap tests based on GMM estimators are provided in Bond and Windmeijer (2005). Newey and Smith (2004) and Anatolyev (2005) analyze higher-order properties for various class of GMM estimators. Hwang and Sun (2017) and Martínez-Iriarte et al. (2019) propose improved inferences for GMM methods using fixed smoothing asymptotics. A recent paper by Hwang et al. (2019) points out a connection between the finite-sample corrected and the misspecification robust asymptotic variance formula in i.i.d. data. This paper contributes to the literature by investigating the finite-sample properties of efficient GMM and its inferential problem in none i.i.d. time-series data.

This paper is also related to the HAR literature which is pioneered by Kiefer and Vogelsang (2002a, 2002b, 2005), Phillips (2005), Müller (2007) and Sun et al. (2008). Recent research along this line can be found in Sun (2014 a&b), Müller and Watson (2018) and Lazarus et al. (2019).

The rest of the paper is organized as follows. Section 2 describes the two-step GMM problem in a time-series setting. Section 3 explores the finite-sample corrected formula for the two-step GMM estimator. Section 4 establishes asymptotic distributions for the test statistics using the corrected variance formula. Section 5 presents discussions on the finite-sample adjustment and the finite-sample corrected variance formula for the iterated GMM estimator. Section 6 presents Monte Carlo simulation results and Section 7 concludes. Proofs are presented in the Appendix.

2 Two-step GMM in Time-series

We want to estimate a $d \times 1$ vector of parameter $\theta \in \Theta$ using a vector of observation $v_t \in \mathbb{R}^{d_x}$ at time t. The true parameter θ_0 is assumed to be an interior point of Θ. The moment condition is given as

$$Ef(v_t, \theta_0) = 0 \text{ iff } \theta = \theta_0,$$

where $f(v_t, \cdot)$ is an $m \times 1$ vector of twice continuously differentiable function and the process $f(v_t, \theta_0)$ is stationary with zero mean. We allow $f(v_t, \theta_0)$ to have general autocorrelation of unknown forms and satisfy $\sum_{j=-\infty}^{\infty} \|Ef(v_t, \theta_0) f(v_{t-j}, \theta_0)'\| < \infty$ and some mixing conditions for the time-series Functional Central Limit Theorem (FCLT) as follows

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\cdot]} f(v_t, \theta_0) \xrightarrow{d} \Lambda W(\cdot), \tag{1}$$

where $W(\cdot)$ is a m-dimensional standard Brownian motion and $\Omega = \Lambda\Lambda'$ is a positive definte long-run variance (LRV) of the moment process $f(v_t, \theta_0)$ which is defined

$$\Omega = \sum_{j=-\infty}^{\infty} Ef(v_t, \theta_0) f(v_{t-j}, \theta_0).$$

Also, we assume that $q = m - d > 0$, and the rank of $G = G(\theta_0) = E[\partial f(v_t, \theta_0)/\partial\theta']$ is d. So, the model is overidentified with a degree of overidentification q. Let

$$f_t(\theta) = \frac{1}{T}\sum_{s=1}^{t} f(v_s, \theta),$$

and define a one-step GMM estimator as

$$\begin{aligned}\hat{\theta}_1 &= \underset{\theta\in\Theta}{\arg\min} M_{\theta, W_T} \\ &= \underset{\theta\in\Theta}{\arg\min} f_T(\theta)' W_T^{-1} f_T(\theta),\end{aligned}$$

where W_T is an initial weight matrix whose components do not depend on the unknown parameter value θ_0 and $p\lim_{T\to\infty} W_T = W$. Using this one-step estimator, the feasible efficient two-step GMM estimator in Hansen(1982) is defined as

$$\hat{\theta}_2 = \underset{\theta\in\Theta}{\arg\min} M_{\theta, S_T(\hat{\theta}_1)} = \underset{\theta\in\Theta}{\arg\min} f_T(\theta)' S_T^{-1}(\hat{\theta}_1) f_T(\theta),$$

where

$$S_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h(\frac{t}{T}, \frac{s}{T})(f(v_t, \theta) - f_T(\theta))(f(v_s, \theta) - f_T(\theta))'.$$

Here, $Q_h(\frac{t}{T}, \frac{s}{T})$ is a symmetric weighting function with smoothing parameter h. For conventional kernel LRV estimators, $Q_h(r, s) = k((r-s)/b)$ and we take $h = 1/b$. For the orthonormal series (OS) LRV estimators, $Q_h(r, s) = K^{-1}\sum_{j=1}^{K}\Phi_j(r)\Phi_j(s)$ and we take $h = K$, where $\{\phi_j(r)\}$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \Phi_j(r)\, dr = 0$. We parametrize h in such a way that h indicates the level of smoothing for both types of LRV estimators.

By construction, $S_T(\hat{\theta}_1)$ is a quadratic heteroskedasticity autocorrelation robust (HAR) estimator for the LRV Ω. It is important to note that $S_T(\hat{\theta}_1)$ is a "centered" version of the LRV estimator, as it is based on the estimation of the demeaned moment process $f(v_t, \hat{\theta}_1) - f_T(\hat{\theta}_1)$. Under the conventional asymptotic theory, Hall (2000) shows that the demeaning procedure can potentially improve the power performance of the J-test using the HAR estimator. Also, this demeaned procedure plays an important role in fixed smoothing asymptotics as the random matrix limit of $S_T(\theta_0)$ is independent of the limiting distribution of $\sqrt{T} f_T(\theta_0)$ which is a normal distribution.

Now, to investigate the asymptotic behavior of $\hat{\theta}_2$, we look at the first order condition (FOC) for $\hat{\theta}_2$ given by

$$\frac{1}{2}\frac{\partial M_{\theta, S_T(\hat{\theta}_1)}}{\partial\theta}\bigg|_{\theta=\hat{\theta}_2} = G_T(\hat{\theta}_2)' S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2) = 0, \tag{2}$$

where

$$G_T\left(\hat{\theta}_2\right) = \frac{1}{T}\sum_{t=1}^{T} \left.\frac{\partial f(v_t,\theta)}{\partial \theta'}\right|_{\theta=\hat{\theta}_2}.$$

A standard approximation to characterize the distribution of $\hat{\theta}_2$ goes as follows. First, conditioning on $S_T(\hat{\theta}_1)$, we do a Taylor expansion of the FOC in (2)

$$\begin{aligned} 0 &= \left.\frac{1}{2}\frac{\partial M_{\theta,S_T(\hat{\theta}_1)}}{\partial \theta}\right|_{\theta=\hat{\theta}_2} \\ &= \left.\frac{1}{2}\frac{\partial M_{\theta,S_T(\hat{\theta}_1)}}{\partial \theta}\right|_{\theta=\theta_0} + \left.\frac{1}{2}\frac{\partial M_{\theta,S_T(\hat{\theta}_1)}}{\partial \theta \partial \theta'}\right|_{\theta=\theta_0} (\hat{\theta}_2-\theta_0) + O_p\left(\frac{1}{T}\right) \\ &= G_T(\theta_0)' S_T^{-1}(\hat{\theta}_1) f_T(\theta_0) + A(\theta_0, S_T(\hat{\theta}_1))(\hat{\theta}_2-\theta_0) + O_p\left(\frac{1}{T}\right). \end{aligned} \tag{3}$$

Here, $A(\theta_0, S_T(\hat{\theta}_1))$ is the second order derivative matrix of $M_{\theta,S_T(\hat{\theta}_1)}$ at $\theta=\theta_0$:

$$\begin{aligned} A(\theta_0, S_T(\hat{\theta}_1)) &= \left.\frac{1}{2}\frac{\partial Q_{\theta,S_T(\hat{\theta}_1)}}{\partial \theta \partial \theta'}\right|_{\theta=\theta_0} = \left.\frac{1}{2}\frac{\partial G_T(\theta)' S_T^{-1}(\hat{\theta}_1) f_T(\theta)}{\partial \theta'}\right|_{\theta=\theta_0} \\ &= G_T(\theta_0)' S_T^{-1}(\hat{\theta}_1) G_T(\theta_0) + H_T(\theta_0)' \left(I_d \otimes S_T^{-1}(\hat{\theta}_1) f_T(\theta_0)\right), \end{aligned}$$

where $H_T(\theta) \in \mathbb{R}^{dm\times d}$ is the second order derivative matrix of the moment process:

$$H_T(\theta_0) = \begin{bmatrix} \left.\frac{\partial G_T(\theta)}{\partial \theta_1}\right|_{\theta=\theta_0} \\ \dots \\ \left.\frac{\partial G_T(\theta)}{\partial \theta_d}\right|_{\theta=\theta_0} \end{bmatrix}.$$

Using the Taylor expansion of the FOC in (3), $\sqrt{T}(\hat{\theta}_2-\theta_0)$ is expanded as:

$$\sqrt{T}(\hat{\theta}_2-\theta_0) = -\left[A(\theta_0, S_T(\hat{\theta}_1))\right]^{-1} G_T(\theta_0)' S_T^{-1}(\hat{\theta}_1)\sqrt{T} f_T(\theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right), \tag{4}$$

and an estimator for the asymptotic variance of $\hat{\theta}_2$ is given by:

$$\widehat{var}(\hat{\theta}_2) = \frac{1}{T}\left[A(\hat{\theta}_2, S_T(\hat{\theta}_1))\right]^{-1}\left(G_T'(\hat{\theta}_2) S_T^{-1}(\hat{\theta}_1) G_T(\hat{\theta}_2)\right)\left[A(\hat{\theta}_2, S_T(\hat{\theta}_1))\right]^{-1\prime}. \tag{5}$$

Note that the sandwich form of $\widehat{var}(\hat{\theta}_2)$ in (4) is different from the standard asymptotic variance estimates $(G_T'(\hat{\theta}_2) S_T^{-1}(\hat{\theta}_1) G_T(\hat{\theta}_2))^{-1}$. This is because the second term, $\Gamma_T(\hat{\theta}_2)'(I_d \otimes S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2))$, in $A(\hat{\theta}_2, S_T(\hat{\theta}_1))$, which order is $O_p(T^{-1/2}) = o_p(1)$, is not zero in a finite sample and keeping this term could potentially improve $\widehat{var}(\hat{\theta}_2)$. When the moment conditions are linear in the parameter θ, the non-linear correction term in $A(\hat{\theta}_2, S_T(\hat{\theta}_1))$ is always zero as $\Gamma_T(\theta) = 0$ for all θ, and $A(\hat{\theta}_2, S_T(\hat{\theta}_1))$ can be simplified as $G_T'(\hat{\theta}_2) S_T^{-1}(\hat{\theta}_1) G_T(\hat{\theta}_2)$. As a result, the expression for $\widehat{var}(\hat{\theta}_2)$ coincides with the standard GMM variance estimates, i.e.

$$\widehat{var}(\hat{\theta}_2) = \frac{1}{T}(G_T' S_T^{-1}(\hat{\theta}_1) G_T)^{-1}. \tag{6}$$

From now on, we assume linearity in the moment conditions. This can be relaxed without any difficulties for any variance correction formula that we provide in this paper. This is because one can always replace (6) by (5) in order to obtain the non-linear correction term $\Gamma_T(\hat{\theta}_2)'(I_d \otimes S_T^{-1}(\hat{\theta}_1)f_T(\hat{\theta}_2))$ in $\widehat{var}(\hat{\theta}_2)$.

3 Finite-sample Corrected Variance Formula

When the moment conditions are linear in the parameter θ, the higher order approximation error term $O_p(T^{-1/2})$ in equation (4) disappears and we have the following equation for $\sqrt{T}(\hat{\theta}_2 - \theta_0)$:

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) = -(G_T' S_T^{-1}(\hat{\theta}_1) G_T)^{-1} G_T' S_T^{-1}(\hat{\theta}_1) \sqrt{T} f_T(\theta_0). \tag{7}$$

Under Assumptions 1–3 in Section 4, we can apply Lemma 1 in Sun (2014), for any $\sqrt{T}$-consistent estimator $\hat{\theta}$, to obtain

$$S_T(\hat{\theta}) = S_T(\theta_0) + o_p(1). \tag{8}$$

Using this result, we can approximate (7) as

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) = \underbrace{-(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0)}_{:=\sqrt{T}\left(\tilde{\theta}_2 - \theta_0\right)} + o_p(1). \tag{9}$$

Any standard approximation of the two-step GMM estimator $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ is based on the first term in (9) which is the infeasible transformed moment condition. This term actually coincides with the first order expansion of the infeasible two-step GMM estimator $\sqrt{T}(\tilde{\theta}_2 - \theta_0)$ that uses the true parameter θ_0 to evaluate the weight matrix $S_T^{-1}(\theta_0)$. So, $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ is asymptotically equivalent to $\sqrt{T}(\tilde{\theta}_2 - \theta_0)$ and this implies that the estimation uncertainty of the initial one-step estimator $\hat{\theta}_1$ in $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ is ignored for any type of asymptotic analysis.

However, Windmeijer (2005) points out that the extra variation in $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ due to $\hat{\theta}_1$ can explain much of the finite sample behavior difference between $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ and $\sqrt{T}(\tilde{\theta}_2 - \theta_0)$. By estimating the part of the $o_p(1)$ term in (9), a finite-sample corrected variance estimate is obtained. Windmeijer (2005) shows that the corrected variance estimate approximates the finite sample variance well and leads to a more accurate inference. Windmeijer (2005) assumes that the moment process $f(v_t, \theta_0)$ is i.i.d, but this is problematic in a time-series set up. However, his idea of corrected variance estimate can be easily accommodated to our time-series set up. In doing so, the key step is, instead of eliminating the estimation uncertainty of $\hat{\theta}_1$ in (7), we further approximate the equation (7) by Taylor expansion, as a function of $\hat{\theta}_1$, in the estimated weight matrix $S_T(\hat{\theta}_1)$ as follows

$$\begin{aligned} \sqrt{T}(\hat{\theta}_2 - \theta_0) &= -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) \\ &\quad + D_{\theta_0, S_T(\theta_0)} \sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \tag{10}$$

where

$$D_{\theta_0, S_T(\theta_0)} = \left. \frac{\partial - (G_T' S_T^{-1}(\theta) G_T)^{-1} G_T' S_T^{-1}(\theta) \sqrt{T} f_T(\theta_0)}{\partial \theta'} \right|_{\theta=\theta_0}$$

is a $d \times d$ matrix. Using element by element differentiation of the $d \times 1$ vector $-(G_T' S_T^{-1}(\theta) G_T)^{-1}$ $G_T' S_T^{-1}(\theta)\sqrt{T} f_T(\theta_0)$ with respect to θ_j, for $1 \le j \le d$, we can express the j-th column of $D_{\theta_0, S_T(\theta_0)}$ as follows:

$$\begin{aligned} D_{\theta_0,S_T(\theta_0)}[.,j] &= \left.\frac{\partial - (G_T' S_T^{-1}(\theta) G_T)^{-1} G_T' S_T^{-1}(\theta) f_T(\theta_0)}{\partial \theta_j}\right|_{\theta=\theta_0} \\ &= (G_T' S_T^{-1}(\theta_0) G_T)^{-1} \left.\frac{\partial (G_T' S_T^{-1}(\theta) G_T)}{\partial \theta_j}\right|_{\theta=\theta_0} (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0) \\ &\quad - (G_T' S_T^{-1}(\theta) G_T)^{-1} \left.\frac{\partial (G_T' S_T^{-1}(\theta) f_T(\theta_0)}{\partial \theta_j}\right|_{\theta=\theta_0}, \end{aligned} \tag{11}$$

where the expression of the vector derivatives in (11) is given by

$$\begin{aligned} \left.\frac{\partial (G_T' S_T^{-1}(\theta) G_T)}{\partial \theta_j}\right|_{\theta=\theta_0} &= {G_T}' \left.\frac{\partial S_T^{-1}(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} G_T \\ &= -G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} S_T^{-1}(\theta_0) G_T; \\ \left.\frac{\partial G_T' S_T^{-1}(\theta) f_T(\theta_0)}{\partial \theta_j}\right|_{\theta=\theta_0} &= G_T' \left.\frac{\partial S_T^{-1}(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} f_T(\theta_0) \\ &= -G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} S_T^{-1}(\theta_0) f_T(\theta_0) \end{aligned}$$

The expression $D_{\theta_0,S_T(\theta_0)}[.,j]$ in (11) can be written as:

$$\begin{aligned} D_{\theta_0,S_T(\theta_0)}[.,j] &= -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} \\ &\quad \times S_T^{-1}(\theta_0) G_T (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0) \\ &\quad + (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\theta_0} S_T^{-1}(\theta_0) f_T(\theta_0), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial S_T(\theta)}{\partial \theta_j} &= \Upsilon_j(\theta) + \Upsilon_j'(\theta); \\ \Upsilon_j(\theta) &= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Q_h(\frac{t}{T}, \frac{s}{T}) \left(g_j(v_s, \theta) - \frac{1}{T} \sum_{s=1}^{T} g_j(v_s, \theta) \right) \\ &\quad \times \left(f(v_t, \theta_0) - \frac{1}{T} \sum_{s=1}^{T} f(v_s, \theta) \right)'; \\ g_j(v_s, \theta) &= \frac{\partial f(v_s, \theta)}{\partial \theta_j}. \end{aligned} \tag{12}$$

The expansion in (10) shows that the correction term $D_{\theta_0,S_T(\theta_0)}\sqrt{T}(\hat{\theta}_1 - \theta_0) = O_p(T^{-1/2})$ vanishes when the sample size T increases, but it is always non zero in finite samples. Therefore, taking into account the correction term $D_{\theta_0,S_T(\theta_0)}\sqrt{T}(\hat{\theta}_1 - \theta_0) = O_p(T^{-1/2})$ can improve the approximation of the variance of $\hat{\theta}_2$ in finite samples.

4 Asymptotics for Finite-sample Corrected Statistics

4.1 Formulation of finite-sample corrected variance

One important assumption in Windmeijer(2005)'s corrected variance estimate in (14) is approximating the distribution of $S_T(\theta_0)$ by the true population counterpart Ω. The approximation is based on the conventional increasing smoothing asymptotics which considers $h \to \infty$ as $T \to \infty$ such that $h/T \to 0$, as in Andrews (1991) and Sun (2014 a&b). Since $S_T(\theta_0)$ is treated as a consistent estimator of Ω, together with the CLT assumption in (1), the term $G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0)$ converges in distribution to $N(0, (G'\Omega^{-1}G))$. Keeping the Windmeijer's correction term $D_{\theta_0, S_T(\theta_0)}$ in (10), the asymptotically equivalent representation for (10) is given as:

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_2 - \theta_0) &= -(G'\Omega^{-1}G)^{-1}G'\Omega^{-1}\Lambda Z_1 - D_{\theta_0, S_T(\theta_0)}(G'W^{-1}G)^{-1}G'W^{-1}\Lambda Z_1 + o_p(1) \quad (13)\\
&= -\begin{pmatrix} (G'\Omega^{-1}G)^{-1} & D_{\theta_0, S_T(\theta_0)}(G'W^{-1}G)^{-1} \end{pmatrix}\begin{pmatrix} G'\Lambda'^{-1}Z_1 \\ G'W^{-1}\Lambda Z_1 \end{pmatrix} + o_p(1),
\end{aligned}$$

where $Z_1 \sim N(0, I_d)$ and $\Lambda\Lambda' = \Omega$. The sum of two normal distribution can be represented as a normal distribution as well, and this implies that we can obtain a normal representation, $N(0, \Xi)$, of the approximated distribution of $\sqrt{T}(\hat{\theta}_2 - \theta_0)$, which variance covariance matrix, Ξ, is given by:

$$\begin{aligned}
\Xi &= \begin{pmatrix} (G'\Omega^{-1}G)^{-1} & D_{\theta_0, S_T(\theta_0)}(G'W^{-1}G)^{-1} \end{pmatrix}\begin{pmatrix} G'\Omega^{-1}G & G'W^{-1}G \\ G'W^{-1}G & G'W^{-1}\Omega W^{-1}G \end{pmatrix}\\
&\quad \times \begin{pmatrix} (G'\Omega^{-1}G)^{-1} \\ (G'W^{-1}G)^{-1}D'_{\theta_0, S_T(\theta_0)} \end{pmatrix}\\
&= \left(G'\Omega^{-1}G\right)^{-1} + D_{\theta_0, S_T(\theta_0)}\left(G'\Omega^{-1}G\right)^{-1} + \left(G'\Omega^{-1}G\right)^{-1}D'_{\theta_0, S_T(\theta_0)}\\
&\quad + D_{\theta_0, S_T(\theta_0)}(G'W^{-1}G)^{-1}\left(G'W^{-1}\Omega W^{-1}G\right)(G'W^{-1}G)^{-1}D'_{\theta_0, S_T(\theta_0)}.
\end{aligned}$$

Motivated by this, the corrected variance estimate $\widehat{var}_c(\hat{\theta}_2)$ under the increasing smoothing asymptotics is given as:

$$\begin{aligned}
\widehat{var}_c(\hat{\theta}_2) &= \widehat{var}(\hat{\theta}_2) + \frac{1}{T}D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}\widehat{var}(\hat{\theta}_2) \qquad (14)\\
&\quad + \frac{1}{T}\widehat{var}(\hat{\theta}_2)D'_{\hat{\theta}_2, S_T(\hat{\theta}_1)} + D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}\widehat{var}(\hat{\theta}_1)D'_{\hat{\theta}_2, S_T(\hat{\theta}_1)},
\end{aligned}$$

where

$$\begin{aligned}
\widehat{var}(\hat{\theta}_1) &= \frac{1}{T}\left(G_T'W_T^{-1}G_T\right)^{-1}\left(G_T'W_T^{-1}S_T(\hat{\theta}_1)W_T^{-1}G_T\right)\left(G_T'W_T^{-1}G_T\right)^{-1};\\
\widehat{var}(\hat{\theta}_2) &= \frac{1}{T}\left(G_T'S_T^{-1}(\hat{\theta}_1)G_T\right)^{-1},
\end{aligned}$$

and

$$D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}[., j] = (G_T'S_T^{-1}(\hat{\theta}_1)G_T)^{-1}G_T'S_T^{-1}(\hat{\theta}_1)\left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat{\theta}_1} S_T^{-1}(\hat{\theta}_1)f_T(\hat{\theta}_2).$$

Using the FOC, $G_T'S_T^{-1}(\hat{\theta}_1)f_T(\hat{\theta}_2) = 0$, the first term of $D_{\theta_0, S_T(\theta_0)}[., j]$ in (11) is always equal to zero in the "estimated" correction term, $D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}[., j]$. The way we construct the corrected variance formula $\widehat{var}_c(\hat{\theta}_2)$ in (14) is the same as in Windmeijer (2005) in which it is assumed an i.i.d moment vector $\{f(v_t, \theta_0)\}$. The difference, in our setting, is that $S_T(\hat{\theta}_1)$ for us is a HAR estimator of the LRV which is robust to a non-i.i.d moment vector $\{f(v_t, \theta_0)\}$.

Now, suppose we want to test the linear null hypothesis $H_0 : R\theta_0 = r$ vs $H_0 : R\theta_0 \neq r$, where R is a $p \times d$ matrix with rank $p \leq d$. Then, the Wald test statistics using the finite-sample corrected variance $\widehat{var}_c(\hat{\theta}_2)$ can be constructed as follows:

$$F_c(\hat{\theta}_2) = \frac{1}{p}(R\hat{\theta}_2 - r)' \left[R\widehat{var}_c(\hat{\theta}_2)R'\right]^{-1} (R\hat{\theta}_2 - r), \tag{15}$$

The Wald statistic based on the conventional "sandwich" formula in (6) is given by

$$F(\hat{\theta}_2) = \frac{1}{p}(R\hat{\theta}_2 - r)' \left[R\widehat{var}(\hat{\theta}_2)R'\right]^{-1} (R\hat{\theta}_2 - r). \tag{16}$$

In constructing Wald statistics, we divide it by the number of hypothesis p. Although, this is not necessary, we use it only because we anticipate a more convenient F approximation in the following subsection.When $p = 1$ and for one-sided alternative hypotheses, we can construct the corresponding finite-sample corrected and uncorrected t statistics as

$$t_c(\hat{\theta}_2) = \frac{\sqrt{T}(R\hat{\theta}_T - r)}{\sqrt{R\widehat{var}_c(\hat{\theta}_2)R'}} \text{ and } t(\hat{\theta}_2) = \frac{\sqrt{T}(R\hat{\theta}_2 - r)}{\sqrt{R\widehat{var}(\hat{\theta}_2)R'}},$$

respectively. We assume the following.

Assumption 1 *(i) For kernel LRV estimators, the kernel function $k(\cdot)$ satisfies the following conditions: for any $b \in (0, 1]$, $k_b(x) = k(x/b)$ is symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable and $\int_{-\infty}^{\infty} k^2(x) < \infty$. (ii) For the OS-LRV variance estimator, the basis functions $\phi_j(\cdot)$ are piecewise monotonic, continuously differentiable and orthonormal in $L^2[0, 1]$ and $\int_0^1 \Phi_j(x)\, dx = 0$.*

Assumption 2 *As $T \to \infty$, $\hat{\theta}_2 = \theta_0 + o_p(1)$, $\hat{\theta}_1 = \theta_0 + o_p(1)$ for an interior point $\theta_0 \in \Theta$, where $\Theta \subseteq \mathbb{R}^d$ is a parameter space of interest.*

Assumption 3 *For any $\hat{\theta} = \theta_0 + o_p(1)$, $G_{[rT]}\left(\hat{\theta}\right) = T^{-1} \sum_{t=1}^{[rT]} \left.\frac{\partial f(v_t, \theta)}{\partial \theta'}\right|_{\theta=\hat{\theta}} = rG + o_p(1)$ uniformly in r where $G = G(\theta_0)$ has rank d and $G(\theta) = E[\partial f(v_t, \theta)/\partial \theta']$.*

Assumption 4 *For each $j = 1, \ldots, d$, and any $\theta_T = \theta_0 + o_p(1)$, $H_{j,[rT]}(\theta_T) = rH_j + o_p(1)$ uniformly in r where $H_{[rT],j}\left(\hat{\theta}\right) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_j(v_t, \hat{\theta})}{\partial \theta'}$ and $H_j = H_j(\theta_0)$ with $H_j(\theta) = E\left[\partial g_j(v_t, \theta)/\partial \theta'\right]$.*

Assumption 5 *For each $j = 1, \ldots, d$, $\{g_j(v_t, \theta_0)\}$ is a strict stationary process and $\sum_{i=-\infty}^{\infty} ||\Psi_{j,i}|| < \infty$ where $\Psi_{j,i} = E[g_j(v_t, \theta_0) g_j(v_{t-i}, \theta_0)']$, and $T^{-1/2} \sum_{t=1}^{[rT]} (g_j(v_t, \theta_0) - E[g_j(v_t, \theta_0)])$ satisfies the FCLT.*

Assumptions 1–3 are standard assumptions in the literature on HAR inference which are the same as in Sun (2014) and Hwang and Sun (2017 and 2018). Assumptions 4–5 are needed to prove the asymptotic validity of the finite-sample corrected variance formula. Assumptions 4–5 trivially holds if the moment conditions are linear in parameters.

Lemma 1 *Under Assumptions 1–3, we have*

$$D_{\hat{\theta}_2, S_T(\hat{\theta}_1)} = D_{\theta_0, S_T(\theta_0)}(1 + o_p(1)).$$

which holds when $h \to \infty$ such that h is fixed as $T \to \infty$, or $h \to \infty$ such that $h/T \to 0$.

Lemma 1 shows that the small order term $D_{\theta_0, S_T(\theta_0)}$ which motivates the formulation of the finite-sample corrected variance estimate is consistently estimated by $D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}$ in a relative sense. In the proof of Lemma 1, we show that the consistency of the estimated term $D_{\hat{\theta}_2, S_T(\hat{\theta}_1)}$ does not depend on whether the smoothing parameter h is fixed as $T \to \infty$, or $h \to \infty$ such that $h/T \to 0$. Since the variance correction terms in $\widehat{var}_c(\hat{\theta}_2)$ are of smaller order, the result of Lemma 1 indicates that the variance-corrected statistics are expected to have the same limiting distribution as the conventional Wald and t statistics. The following theorem formally proves this result.

Theorem 2 *Under Assumptions 1–5,*
(a) $t_c(\hat{\theta}_2) = t(\hat{\theta}_2) + o_p(1)$;
(b) $F_c(\hat{\theta}_2) = F(\hat{\theta}_2) + o_p(1)$,
where (a) and (b) hold when $h \to \infty$ such that $h/T \to 0$, or h is fixed as $T \to \infty$.

Under the increasing smoothing asymptotics, i.e. h goes to infinity but $h/T \to 0$, we have that $\widehat{var}(\hat{\theta}_2) \xrightarrow{p} (G'\Omega^{-1}G)^{-1}$, and thus $t(\hat{\theta}_2) \xrightarrow{d} N(0,1)$ and $F(\hat{\theta}_2) \xrightarrow{d} \chi_p^2/p$. The result in Theorem 2 justifies that the distribution of our finite-sample corrected t and Wald statistics can be approximated by

$$t_c(\hat{\theta}_2) \xrightarrow{d} N(0,1) \text{ and } F_c(\hat{\theta}_2) \xrightarrow{d} \frac{1}{p}\chi_p^2.$$

4.2 Fixed-smoothing asymptotics for finite-sample corrected variance

Although the conventional increasing smoothing asymptotics is a key device to prove the consistency of $S_T(\theta_0)$ and the conventional normal and chi-square approximations of the finite-sample corrected test statistics, the increasing smoothing asymptotics often fails to reflect finite sample variations of the nonparametric LRV estimation of $S_T(\theta_0)$. In fact, there is extensive numerical evidence that reports the poor finite sample performances of HAR inference using the increasing smoothing asymptotics, e.g., Kiefer and Vogelsang (2002 a&b, 2005), Sun et al. (2008), and Hwang and Sun (2017, 2019). This evidence opens a new stream of time-series research on HAR inference. The research along this line is pioneered by Kiefer and Vogelsang (2002, 2005), Phillips (2005), Müller (2007) and Sun et al. (2008). The HAR literature develops a new type of asymptotics that holds h fixed when $T \to \infty$. This tool is called "fixed smoothing asymptotics" in Sun (2014a), and it is called "fixed-b asymptotics" in Kiefer and Vogelsang (2002 a&b, 2005). In the context of the efficient GMM, Sun (2014b) and Hwang and Sun (2017) show that the new asymptotics yields to more accurate approximations, while the conventional increasing asymptotics poorly performs in finite samples.

In this section, we follow Hwang and Sun (2017) and derive the fixed-smoothing asymptotics of $t_c(\hat{\theta}_2)$ and $F_c(\hat{\theta}_2)$ using orthonormal series (OS) weighting function

$$Q_K\left(\frac{r}{T}, \frac{s}{T}\right) = \frac{1}{K}\sum_{j=1}^{K} \Phi_j\left(\frac{r}{T}\right)\Phi_j\left(\frac{s}{T}\right), \tag{17}$$

where $\{\Phi_j(r)\}_{j=1}^K$ are orthonormal basis functions on $L^2[0,1]$ satisfying $\int_0^1 \Phi_j(r)\,dr = 0$. The smoothing parameter h in $Q_h(r/T, s/T)$ is now equal to K, which is the number of terms in the OS-LRV estimator of $S_T(\theta_0)$. If K is even and $\{\Phi_j(r)\}_{j=1}^K = \{\Phi_{2j-1}(r) = \sqrt{2}\sin(2\pi jr),$ $\Phi_{2j}(r) = \sqrt{2}\cos(2\pi jr),\ j = 1, 2, \ldots, K/2\}$, then the OS-LRV estimator is

$$\begin{aligned} S_T(\theta_0) &= \frac{1}{K}\sum_{j=1}^{K} U_j(\theta_0)U_j(\theta_0)'; \\ U_j(\theta_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\left[f(v_t,\theta_0) - \frac{1}{T}\sum_{\tau=1}^{T} f(v_\tau,\theta_0)\right] \text{ for } j = 1, 2, \ldots, K. \end{aligned}$$

The OS-LRV estimator has gained considerable attention in recent HAR literature, e.g., Phillips (2005), Müller (2007), Sun (2013, 2014 a&b), Lazarus et al. (2018) and Lazarus et al. (2019). By construction, the OS-LRV estimator, which takes a simple average of the first $K/2$ periodograms., is proportional to an estimator of the spectral density at the origin.

For two stochastically bounded sequences of random vectors $\xi_n \in \mathbb{R}^\ell$ and $\eta_n \in \mathbb{R}^\ell$, let $\overset{a}{\sim}$ be a notion of asymptotic equivalence in distribution, $\xi_n \overset{a}{\sim} \eta_n$, that is, ξ_n and η_n converge in distribution to the same limits. The fixed smoothing approximation of $S_T(\theta_0)$ captures the finite sample variability of each periodgram by

$$U_j(\theta_0) \overset{a}{\sim} \Lambda \mathbb{U}_j \Lambda'$$

for each $j = 1, \ldots, K$, where

$$\mathbb{U}_j = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\left(e_t - \frac{1}{T}\sum_{s=1}^{T} e_s\right) \text{ and } e_t \overset{\text{i.i.d}}{\sim} N(0, I_m).$$

From the properties of the Fourier basis functions, $\sum_{t=1}^{T}\Phi_j(\frac{t}{T}) = 0$ and $T^{-1}\sum_{t=1}^{T}\Phi_i(\frac{t}{T})\Phi_j(\frac{t}{T}) = 1(i \neq j)$, it is easy to show that $\mathbb{U}_i = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_i(\frac{t}{T})e_t \overset{\text{i.i.d}}{\sim} N(0, I_m)$ over $j = 1, \ldots, K$. Then, holding K fixed, the OS-LRV is approximated by

$$S_T(\theta_0) \overset{a}{\sim} \mathbb{S} = \begin{pmatrix} \underset{d\times d}{\mathbb{S}_{11}} & \underset{d\times q}{\mathbb{S}_{22}} \\ \underset{q\times d}{\mathbb{S}_{22}} & \underset{q\times q}{\mathbb{S}_{22}} \end{pmatrix} := \Lambda\left(\frac{1}{K}\sum_{j=1}^{K}\mathbb{U}_j\mathbb{U}_j'\right)\Lambda',$$

Note that the approximated random variable $\mathbb{S} \sim K^{-1}\mathbb{W}_p(K, I_m)$ is a scaled Wishart random matrix with degree of freedom equal to K. The remaining question is whether $t_c(\hat{\theta}_2)$ and $F_c(\hat{\theta}_2)$ are asymptotically free of nuisance parameters including the correction term $D_{\theta_0, S_T(\theta_0)}$. From the result of Theorem 1 in Sun (2014), the standard t and Wald statistics, $t(\hat{\theta}_2)$ and $F(\hat{\theta}_2)$, do not depend on any nuisance parameters, and their limits are represented by :

$$t(\hat{\theta}_2) \xrightarrow{d} \mathbb{T} \overset{d}{=} \frac{Z_1 - \mathbb{S}_{12}\mathbb{S}_{22}^{-1}Z_2}{(\mathbb{S}_{11\cdot 2})^{1/2}}; \tag{18}$$

$$F(\hat{\theta}_2) \xrightarrow{d} \mathbb{F} \overset{d}{=} \frac{1}{p}\left(Z_1 - \mathbb{S}_{12}\mathbb{S}_{22}^{-1}Z_2\right)'\mathbb{S}_{11\cdot 2}^{-1}\left(Z_1 - \mathbb{S}_{12}\mathbb{S}_{22}^{-1}Z_2\right), \tag{19}$$

respectively, where $Z_1 \sim N(0, I_p)$, $Z_2 \sim N(0, I_p)$, $Z_1 \perp Z_2$, and $\mathbb{S}_{11\cdot 2} = \mathbb{S}_{11} - \mathbb{S}_{12}\mathbb{S}_{22}^{-1}\mathbb{S}_{21}$. From the results in Theorem 2, the limiting distributions of $t_c(\hat{\theta}_2)$ and $F_c(\hat{\theta}_2)$, under the fixed-smoothing asymptotics, are given by

$$t_c(\hat{\theta}_2) = t(\hat{\theta}_2) + o_p(1) \xrightarrow{d} \mathbb{T} \text{ and } F_c(\hat{\theta}_2) = F(\hat{\theta}_2) + o_p(1) \xrightarrow{d} \mathbb{F}.$$

The fixed-K limiting distributions in (18)–(19) are nonstandard. To investigate further, we use the well-known properties of the Wishart distribution from Proposition 7.9 in Bilodeau and Brenner (2008), and obtain that $\mathbb{S}_{11\cdot 2} \sim \mathbb{W}_p(K - p - q + 1, I_p)/G$, and $\mathbb{S}_{11\cdot 2}$ is independent of $\mathbb{S}_{12}$ and $\mathbb{S}_{22}$. This implies that conditioning on $\Delta := \mathbb{S}_{12}\mathbb{S}_{22}^{-1}Z_2$, the limiting distribution $\mathbb{F}$ satisfies

$$\frac{K - p - q + 1}{K}\mathbb{F} \overset{d}{=} \frac{K - p - q + 1}{K} \frac{(Z_1 + \Delta)'\mathbb{S}_{11\cdot 2}^{-1}(Z_1 + \Delta)}{p} \overset{d}{=} \mathcal{F}_{p,K-p-q+1}\left(\|\Delta\|^2\right), \tag{20}$$

where $\mathcal{F}_{p,K-p-q+1}(\|\Delta\|^2)$ is a noncentral F distribution with random noncentrality parameter $\|\Delta\|^2$. The random noncentrality parameter Δ is the source of non-standard limiting distribution $\mathbb{F}$, and in practice the critical values need be simulated. Considering our finite-sample corrected test statistics, it would be more convenient, in empirical applications, that we can additionally provide a correction for Δ. The modified t and Wald statistics are

$$\begin{aligned} \tilde{t}_c(\hat{\theta}_2) &= \frac{K - q}{K} \cdot \frac{t_c(\hat{\theta}_2)}{\sqrt{1 + \frac{1}{K}J(\hat{\theta}_2)}}; \\ \tilde{F}_c(\hat{\theta}_2) &= \frac{K - p - q + 1}{K} \cdot \frac{F_c(\hat{\theta}_2)}{1 + \frac{1}{K}J(\hat{\theta}_2)}, \end{aligned} \tag{21}$$

where $J(\hat{\theta}_2) = T f_T(\hat{\theta}_2)' S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2)$ is the standard J statistic for testing the over-identifying restrictions.

Assumption 6 *(a)* $T^{-1/2}\sum_{t=1}^T \Phi_j(t/T) f(v_t, \theta_0)$ *converges weakly to a continuous distribution, jointly over* $j = 0, 1, \ldots, K$. *(b) The following holds:*

$$\begin{aligned} &P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) f(v_t, \theta_0) \le x \text{ for } j = 0, 1, \ldots, K\right) \\ &= P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) \Lambda e_t \le x \text{ for } j = 0, 1, \ldots, K\right) + o(1) \text{ as } T \to \infty, \end{aligned}$$

where $e_t \overset{i.i.d}{-} N(0, I_m)$ *and* $x \in \mathbb{R}^m$.

Theorem 3 *Under Assumptions 1–6, for a fixed* K *as* $T \to \infty$*, we have*
(a) $\tilde{t}_c(\hat{\theta}_2) \xrightarrow{d} t_{K-q}$;
(b) $\tilde{F}_c(\hat{\theta}_2) \xrightarrow{d} \mathcal{F}_{p,K-p-q+1}$.

Theorem 3 shows that the finite sample variance corrections in $\tilde{t}_c(\hat{\theta}_2)$ and $\tilde{F}_c(\hat{\theta}_2)$ do not change the standard t and F limiting distributions found in Hwang and Sun (2017). Still, they can help improve the finite sample performance of our tests. Compared to the conventional normal and chi-square approximations, the fixed-smoothing asymptotics in Theorem 3 is expected to lead to a more accurate inference because the t and F limits are able to capture the estimation uncertainty of the non-parametric estimator $S_T(\theta_0)$ from the studentized HAR statistics. Also, the J-statistic modifications in our statistics can capture the estimation uncertainty of the two-step GMM estimator $\hat{\theta}_2$ arising from the random GMM weight $S_T(\theta_0)$, and thus remove the random noncentrality parameter Δ in the limit. The standard t and F limiting distributions are very appealing in empirical applications.

5 Discussions

5.1 Finite-sample adjustment of the corrected variance formula

Although our corrected variance formula $\widehat{var}_c(\hat{\theta}_2)$ is designed to have more variation than the uncorrected variance formula $\widehat{var}(\hat{\theta}_2)$, the randomness of $D_{\hat{\theta}_2,S_T(\hat{\theta}_1)}$ can possibly make $\widehat{var}_c(\hat{\theta}_2)$ smaller than $\widehat{var}(\hat{\theta}_2)$. For example, when we have that $\widehat{var}_c(\hat{\theta}_2)-\widehat{var}(\hat{\theta}_2)$ is negative semi-definite in finite samples, the corrected Wald test statistics $F_c(\hat{\theta}_2)$ could be larger than the uncorrected statistics $F(\hat{\theta}_2)$. Since the motivation of our corrected variance formula and corresponding t and Wald statistics is to reduce the size distortion, we set the gap between $\widehat{var}_c(\hat{\theta}_2)$ and $\widehat{var}(\hat{\theta}_2)$ to be positive. Thus, we propose to replace $\widehat{var}_c(\hat{\theta}_2)$ by $\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)$ where the adjusted variance correction $\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)$ satisfies

$$\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)-\widehat{var}(\hat{\theta}_2)\geq 0.$$

The adjustment step comes from looking at the matrix M_T, which is the difference between $\widehat{var}_c(\hat{\theta}_2)$ and $\widehat{var}(\hat{\theta}_2)$, calculate the spectral decomposition of $M_T=V_TL_TV_T'$, and replace the negative components of the diagonal eigenvalue matrix L_T with zeros. If we define this new eigenvalue matrix as $\tilde{L}_T$, the adjusted version of $\widehat{var}_c(\hat{\theta}_2)$ is constructed as.

$$\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)=\widehat{var}(\hat{\theta}_2)+\tilde{M}_T \text{ where } \tilde{M}_T=V_T\tilde{L}_TV_T'.$$

Using these adjusted variances, the corrected t and Wald statistics are defined as :

$$\begin{aligned}
\tilde{t}_c^{\mathrm{adj}}(\hat{\theta}_2) &= \frac{K-q}{K}\cdot\frac{t_c^{\mathrm{adj}}(\hat{\theta}_2)}{\sqrt{1+\frac{1}{K}J(\hat{\theta}_2)}};\\
\tilde{F}_c^{\mathrm{adj}}(\hat{\theta}_2) &= \frac{K-p-q+1}{K}\cdot\frac{F_c^{\mathrm{adj}}(\hat{\theta}_2)}{1+\frac{1}{K}J(\hat{\theta}_2)},
\end{aligned}$$

where

$$\begin{aligned}
t_c^{\mathrm{adj}}(\hat{\theta}_2) &= \frac{\sqrt{T}(R\hat{\theta}_T-r)}{\{R\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)R'\}^{1/2}};\\
F_c^{\mathrm{adj}}(\hat{\theta}_2) &= \frac{1}{p}\left(R\hat{\theta}_2-r\right)'\left(R\widehat{var}_c^{\mathrm{adj}}(\hat{\theta}_2)R'\right)^{-1}\left(R\hat{\theta}_2-r\right),
\end{aligned} \tag{22}$$

By construction, $t_c^{\mathrm{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\mathrm{adj}}(\hat{\theta}_2)$ always satisfy:

$$|\tilde{t}_c^{\mathrm{adj}}(\hat{\theta}_2)|\leq|\tilde{t}_c(\hat{\theta}_2)| \text{ and } \tilde{F}_c^{\mathrm{adj}}(\hat{\theta}_2)\leq\tilde{F}_c(\hat{\theta}_2),$$

and

$$t_c^{\mathrm{adj}}(\hat{\theta}_2)\xrightarrow{d}t_{K-q} \text{ and } \tilde{F}_c(\hat{\theta}_2)\xrightarrow{d}\mathcal{F}_{p,K-p-q+1}.$$

5.2 Finite-sample correction for iterated GMM

Another popular GMM estimator is the iterated GMM estimator studied in Hansen et al. (1996), which is designed to improve the finite sample performance of the two-step GMM estimator. For more discussion on the iterated GMM estimator, see Hansen and Lee (2019). Let us define the j-th iterated GMM estimator $\hat{\theta}^j_{\mathrm{IE}}$ as the solution to the following minimization problem

$$\hat{\theta}^j_{\mathrm{IE}} = \arg\min_{\theta\in\Theta} f_T(\theta)' S_T^{-1}(\hat{\theta}^{j-1}_{\mathrm{IE}}) f_T(\theta).$$

Under some regular conditions, Hansen and Lee (2019) shows that the loop of the iteration sequence $\hat{\theta}^j_{\mathrm{IE}}$ for $j = 1, 2, \ldots$ is a contraction mapping, which leads the iteration estimator $\hat{\theta}^\infty_{\mathrm{IE}}$ to a fixed point. Let $\hat{\theta}^0_{\mathrm{IE}}$ be the two-step estimator $\hat{\theta}_2$.Then, the asymptotic distribution of $\sqrt{T}(\hat{\theta}^1_{\mathrm{IE}} - \theta_0)$ can be represented as follows:

$$\begin{aligned}
\sqrt{T}(\hat{\theta}^1_{\mathrm{IE}} - \theta_0) &= -(G_T' S_T^{-1}(\hat{\theta}_2) G_T)^{-1} G_T' S_T^{-1}(\hat{\theta}_2)\sqrt{T} f_T(\theta_0) \qquad (23)\\
&= -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0) + D_{\theta_0, S_T(\theta_0)}\sqrt{T}(\hat{\theta}_2 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Recall that we have the expansion of the two-step GMM $\hat{\theta}_2$ as

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) = -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0) + D_{\theta_0, S_T(\theta_0)}\sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

Substituting the above expansion into (23), we can represent the first iteration estimator $\sqrt{T}(\hat{\theta}^1_{\mathrm{IE}} - \theta_0)$ as

$$\begin{aligned}
\sqrt{T}(\hat{\theta}^1_{\mathrm{IE}} - \theta_0) &= -\left(I_d + D_{\theta_0, S_T(\theta_0)}\right)(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0)\\
&\quad + D^2_{\theta_0, S_T(\theta_0)}\sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

The leading term in $\sqrt{T}(\hat{\theta}^1_{\mathrm{IE}} - \theta_0)$ consists of a (mixed) normal distribution part which is scaled by $I_d + D_{\theta_0, S_T(\theta_0)}$. Also, we notice the effect of the one-step estimator $\sqrt{T}(\hat{\theta}_1 - \theta_0)$ decays through the iteration procedure when we keep repeating this substitution until the j-th iteration:

$$\begin{aligned}
\sqrt{T}(\hat{\theta}^j_{\mathrm{IE}} - \theta_0) &= -\left[I_d + \sum_{i=1}^{j} D^i_{\theta_0, S_T(\theta_0)}\right](G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0)\\
&\quad + D^{j+1}_{\theta_0, S_T(\theta_0)}\sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

When the number of iterations j goes to infinity, $\hat{\theta}^j_{\mathrm{IE}}$ converges to $\hat{\theta}^\infty_{\mathrm{IE}}$. The impact of $\sqrt{T}(\hat{\theta}_1 - \theta_0)$ on $\sqrt{T}(\hat{\theta}^j_{\mathrm{IE}} - \theta_0)$ through $D^{j+1}_{\theta_0, S_T(\theta_0)} = O_p(T^{-(j+1)/2})$ can be perfectly removed and we have that

$$\begin{aligned}
\sqrt{T}(\hat{\theta}^\infty_{\mathrm{IE}} - \theta_0) &= -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} {G_T}' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0) + D_{\theta_0, S_T(\theta_0)}\sqrt{T}(\hat{\theta}^\infty_{\mathrm{IE}} - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right)\\
&= -\left(I_d - D_{\theta_0, S_T(\theta_0)}\right)^{-1}(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right)
\end{aligned}$$

Thus, the corrected variance estimate for $\hat{\theta}_{IE}^{\infty}$ can be constructed as follows :

$$\begin{aligned} \widehat{var}_c\left(\hat{\theta}_{\mathrm{IE}}^{\infty}\right) &= \left(I_d - D_{\hat{\theta}_{\mathrm{IE}}^{\infty},S_T(\hat{\theta}_{\mathrm{IE}}^{\infty})}\right)^{-1} \widehat{var}\left(\hat{\theta}_{\mathrm{IE}}^{\infty}\right) \left(I_d - D'_{\hat{\theta}_{\mathrm{IE}}^{\infty},S_T(\hat{\theta}_{\mathrm{IE}}^{\infty})}\right)^{-1} \qquad (24) \\ \widehat{var}\left(\hat{\theta}_{\mathrm{IE}}^{\infty}\right) &= \frac{1}{T}\left(G'_T S_T^{-1}(\hat{\theta}_{\mathrm{IE}}^{\infty}) G_T\right)^{-1}, \end{aligned}$$

The corresponding t and Wald statistics are

$$\begin{aligned} t(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{R\hat{\theta}_{\mathrm{IE}}^{\infty} - r}{\sqrt{R\widehat{var}(\hat{\theta}_{\mathrm{IE}}^{\infty})R'}}; \qquad (25) \\ F(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{1}{p}\left(R\hat{\theta}_{\mathrm{IE}}^{\infty} - r\right)'\left(R\widehat{var}(\hat{\theta}_{\mathrm{IE}}^{\infty})R'\right)^{-1}\left(R\hat{\theta}_{\mathrm{IE}}^{\infty} - r\right). \end{aligned}$$

Using the finite-sample corrected asymptotic variance estimates $\widehat{var}_c(\hat{\theta}_{\mathrm{IE}}^{\infty})$, one can also construct t and Wald statistics for $\hat{\theta}_{\mathrm{IE}}^{j}$ as

$$\begin{aligned} t_c(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{R\hat{\theta}_{\mathrm{IE}}^{\infty} - r}{\sqrt{R\widehat{var}_c(\hat{\theta}_{\mathrm{IE}}^{\infty})R'}}; \qquad (26) \\ F_c(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{1}{p}\left(R\hat{\theta}_{\mathrm{IE}}^{\infty} - r\right)'\left(R\widehat{var}_c(\hat{\theta}_{\mathrm{IE}}^{\infty})R'\right)^{-1}\left(R\hat{\theta}_{\mathrm{IE}}^{\infty} - r\right). \end{aligned}$$

The asymptotic distribution of $F_c(\hat{\theta}_{\mathrm{IE}}^{\infty})$ can be characterized as

$$\begin{aligned} F_c(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{1}{p} \times \left[R\left(I_d - D_{\theta_0,S_T(\theta_0)}\right)^{-1}(G'_T S_T^{-1}(\theta_0) G_T)^{-1} G'_T S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0)\right]' \\ &\times \begin{pmatrix} R\left(I_d - D_{\hat{\theta}_{\mathrm{IE}}^{\infty},S_T(\hat{\theta}_{\mathrm{IE}}^{\infty})}\right)^{-1}\left(G'_T S_T^{-1}(\hat{\theta}_{\mathrm{IE}}^{\infty}) G_T\right)^{-1} \\ \cdot\left(I_d - D'_{\hat{\theta}_{\mathrm{IE}}^{\infty},S_T(\hat{\theta}_{\mathrm{IE}}^{\infty})}\right)^{-1} R' \end{pmatrix}^{-1} \\ &\times \left[R\left(I_d - D_{\theta_0,S_T(\theta_0)}\right)^{-1}(G'_T S_T^{-1}(\theta_0) G_T)^{-1} G'_T S_T^{-1}(\theta_0)\sqrt{T} f_T(\theta_0)\right] + o_p(1). \end{aligned}$$

Under the fixed smoothing asymptotics, we have that $S_T(\theta_0) \overset{a}{\sim} \mathbb{S}_T$. The asymptotically equivalent distribution of $F_c(\hat{\theta}_{\mathrm{IE}}^{\infty})$ is then given by

$$\mathbb{F}_{\mathrm{IE}} = \frac{1}{P}\left[\tilde{R}(G'_T\mathbb{S}_T^{-1}G_T)^{-1}G'_T\mathbb{S}_T^{-1}\Lambda Z\right]'\left[\tilde{R}\left(G'_T\mathbb{S}_T^{-1}G_T\right)^{-1}\tilde{R}'\right]^{-1}\left[\tilde{R}(G'_T\mathbb{S}_T^{-1}G_T)^{-1}G'_T\mathbb{S}_T^{-1}\Lambda Z\right], \qquad (27)$$

where $\tilde{R} = R(I_d - D_{\theta_0,S_T(\theta_0)})^{-1}$ is a $p \times d$ matrix. Considering $\tilde{R} = R + o_p(1)$ and Theorem 1 in Sun(2014), we obtain $\mathbb{F}_{\mathrm{IE}} = \mathbb{F} + o_p(1)$. Thus, instead of approximating $F_c(\hat{\theta}_{\mathrm{IE}}^{\infty})$ by a conventional χ_d^2/p distribution, the standard t and F distributions can be used to obtain asymptotic critical values for $t_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty})$ and $F_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty})$ together with the corrected variance estimate $\widehat{var}_c(\hat{\theta}_{\mathrm{IE}}^{\infty})$, the J-statistic modification, and the finite-sample adjustments in Subsection 5.1:

$$\begin{aligned} \tilde{t}_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{K-q}{K}\cdot\frac{t_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty})}{\sqrt{1+\frac{1}{K}J(\hat{\theta}_{\mathrm{IE}}^{\infty})}} \xrightarrow{d} t_{K-q}; \qquad (28) \\ \tilde{F}_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty}) &= \frac{K-p-q+1}{K}\cdot\frac{F_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty})}{1+\frac{1}{K}J(\hat{\theta}_{\mathrm{IE}}^{\infty})} \xrightarrow{d} \mathcal{F}_{K-p-q+1}. \end{aligned}$$

6 Simulation Results

We follow the simulation design in Sun (2014b) and consider the following linear structural model:

$$y_t = \alpha + x_{1,t}\beta_1 + x_{2,t}\beta_2 + x_{3,t}\beta_3 + \epsilon_{y,t},$$

where $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are scalar regressors that are correlated with $\epsilon_{y,t}$. The unknown parameter vector is $\theta = (\alpha, \beta_1, \beta_2, \beta_3)' \in \mathbb{R}^d$ with $d = 4$ and there are m instruments $z_{0,t}, z_{1,t}, ..., z_{m-1,t}$ with $z_{0,t} \equiv 1$, and the reduced form equations for $x_{1,t}, x_{2,t}$ and $x_{3,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d-1}^{m-1} z_{i,t} + \epsilon_{x_j,t} \text{ for } j = 1, 2, 3.$$

We assume that $z_{i,t}$ for $i \geq 1$ follows an AR(1) process

$$z_{i,t} = \rho z_{i,t-1} + \sqrt{1-\rho^2} e_{z_i,t},$$

where

$$e_{z_i,t} = \frac{e^i_{zt} + e^0_{zt}}{\sqrt{2}}.$$

and $[e^0_{zt}, e^1_{zt}, ..., e^{m-1}_{zt}]' \overset{\text{iid}}{-} N(0, I_m)$. The DGP for $\epsilon_t = (\epsilon_{yt}, \epsilon_{x_1t}, \epsilon_{x_2t}, \epsilon_{x_3t})'$ is the same as the DGP for $(z_{1,t}, ..., z_{m-1,t})'$ except for the dimensionality difference. By construction, the two vectors, ϵ_t and $(z_{1,t}, ..., z_{m-1,t})'$, are independent. We consider the true parameters to be $\theta_0 = (0, 0, 0, 0)'$ and $\rho = 0.5$.

Define $x_t = (x_{1,t}, x_{2,t}, x_{3,t})'$ and $z_t = (z_{0,t}, z_{1,t}, ..., z_{m-1,t})'$, then we have the m-number of the moment conditions given by

$$E[f(v_t, \theta_0)] = E[z_t(y_t - x_t'\theta_0)] \in \mathbb{R}^m,$$

and provided a initial weight matrix W_T^{-1}, the one-step GMM estimator and its asymptotic variance estimator is as follows:

$$\begin{aligned}
\hat{\theta}_1 &= \left(X'ZW_T^{-1}Z'X\right)^{-1}\left(X'ZW_T^{-1}Z'y\right); \\
\widehat{var}(\hat{\theta}_1) &= T\left(X'ZW_T^{-1}Z'X\right)^{-1}\left(X'ZW_T^{-1}S_T\left(\hat{\theta}_1\right)W_T^{-1}Z'X\right)\left(X'ZW_T^{-1}Z'X\right)^{-1}
\end{aligned}$$

with $X = (x_1, ..., x_T)' \in \mathbb{R}^{T\times d}$, $Z = (z_1, ..., z_T)' \in \mathbb{R}^{T\times m}$, $y = (y_1, ..., y_T)'$ and

$$\begin{aligned}
S_T\left(\hat{\theta}_1\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T}, \frac{s}{T}\right)\left(z_t\epsilon_{y,t}(\hat{\theta}_1) - \frac{1}{T}\sum_{j=1}^{T} z_j\epsilon_{y,j}(\hat{\theta}_1)\right) \\
&\quad \times\left(z_s\epsilon_{y,s}(\hat{\theta}_1) - \frac{1}{T}\sum_{j=1}^{T} z_j\epsilon_{y,j}(\hat{\theta}_1)\right)'.
\end{aligned}$$

We choose the initial weight matrix W_T as $Z'Z/T$. This makes the initial one-step estimator $\hat{\theta}_1$ equivalent to the two-stage least square estimator (2SLS). Based on these formulations, we will check a variety of test statistics including both the two-step and iteration procedures.

1) Feasible two-step estimator $\hat{\theta}_2$: The efficient two-step estimator and its (uncorrected) variance estimator are as follows

$$\begin{aligned}
\hat{\theta}_2 &= \left(X'ZS_T^{-1}(\hat{\theta}_1)Z'X\right)^{-1} X'ZS_T^{-1}(\hat{\theta}_1)Z'y; \\
\widehat{var}(\hat{\theta}_2) &= T\left(X'ZS_T^{-1}(\hat{\theta}_1)Z'X\right)^{-1}; \\
\widehat{var}_c(\hat{\theta}_2) &= \widehat{var}(\hat{\theta}_2) + D_{\hat{\theta}_2,S_T(\hat{\theta}_1)}\widehat{var}(\hat{\theta}_2) + \widehat{var}(\hat{\theta}_2)D'_{\hat{\theta}_2,S_T(\hat{\theta}_1)} \\
&\quad + D_{\hat{\theta}_2,S_T(\hat{\theta}_1)}\widehat{var}(\hat{\theta}_1)D'_{\hat{\theta}_2,S_T(\hat{\theta}_1)},
\end{aligned}$$

where the j-th column of $D_{\hat{\theta}_{IE,1},S_N(\hat{\theta}_2)}$ is given by

$$D_{\hat{\theta}_{IE,1},S_N(\hat{\theta}_2)}[.,j] = -\left(X'ZS_T^{-1}(\hat{\theta}_1)Z'X\right)^{-1} X'ZS_T^{-1}(\hat{\theta}_1)\left.\frac{\partial S_T(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1} S_T^{-1}(\hat{\theta}_1)Z'\epsilon_y(\hat{\theta}_2)$$

with $\epsilon_y(\hat{\theta}_1) = \left(\epsilon_{y,1}(\hat{\theta}_1), ..., \epsilon_{y,T}(\hat{\theta}_1)\right)'$ and $\left.\frac{\partial S_T(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1}$ is defined by:

$$\left.\frac{\partial S_T(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_1} = \Upsilon_j(\hat{\theta}_1) + \Upsilon'_j(\hat{\theta}_1),$$

where

$$\Upsilon_j\left(\hat{\theta}_1\right) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T},\frac{s}{T}\right)\left(z_t x_{j,t} - \frac{1}{T}\sum_{i=1}^{T} z_i x_{j,i}\right)\left(\epsilon_{y,s}(\hat{\theta}_1)z'_s - \frac{1}{T}\sum_{i=1}^{T}\epsilon_{y,i}(\hat{\theta}_1)z'_i\right),$$

and the corresponding Wald test statistics, $F(\hat{\theta}_2)$, $\tilde{F}_c(\hat{\theta}_2)$, and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ are of the form shown in (16), (15), and (22), respectively.

2) Iterated estimator $\hat{\theta}_\infty$: The iteration estimator and its (uncorrected) variance estimator is given by

$$\begin{aligned}
\hat{\theta}_{\text{IE}}^{\infty} &= \left(X'ZS_T^{-1}(\hat{\theta}_{IE}^{\infty})Z'X\right)^{-1} X'ZS_T^{-1}(\hat{\theta}_{IE}^{\infty})Z'y; \\
\widehat{var}_c(\hat{\theta}_{\text{IE}}^{\infty}) &= (I_d - D_{\hat{\theta}_{IE}^{\infty},S_T(\hat{\theta}_{IE}^{\infty})})^{-1}\widehat{var}(\hat{\theta}_{IE}^{\infty})(I_d - D'_{\hat{\theta}_{IE}^{\infty},S_T(\hat{\theta}_{IE}^{\infty})})^{-1}; \\
\widehat{var}(\hat{\theta}_{\text{IE}}^{\infty}) &= T\left(X'ZS_T^{-1}(\hat{\theta}_{IE}^{\infty})Z'X\right)^{-1},
\end{aligned}$$

where

$$D_{\hat{\theta}_\infty,S_T(\hat{\theta}_\infty)}[.,j] = -\left(X'ZS_T^{-1}(\hat{\theta}_{IE}^{\infty})Z'X\right)^{-1} X'ZS_T^{-1}(\hat{\theta}_{IE}^{\infty})\left.\frac{\partial S_T(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_{IE}^{\infty}} S_T^{-1}(\hat{\theta}_{IE}^{\infty})Z'\epsilon_y(\hat{\theta}_{IE}^{\infty})$$

and $\left.\frac{\partial S_T(\theta)}{\partial\theta_j}\right|_{\theta=\hat{\theta}_{IE}^{\infty}} = \Upsilon_j(\hat{\theta}_{\text{IE}}^{\infty}) + \Upsilon_j(\hat{\theta}_{\text{IE}}^{\infty})'$. The corresponding Wald test statistics, $F(\hat{\theta}_{\text{IE}}^{\infty})$, $\tilde{F}_c(\hat{\theta}_{\text{IE}}^{\infty})$, and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$ are shown in (25), (26), and (28), respectively.

6.1 Point estimation

We construct the uncorrected and corrected asymptotic variance estimates by employing the commonly used Bartlett kernel. For the basis functions in the OS-HAR estimation, we use the following orthonormal basis functions $\{\Phi_j\}_{j=1}$ in $L^2[0,1]$:

$$\Phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x) \text{ and } \Phi_{2j}(x) = \sqrt{2}\sin(2j\pi x) \text{ for } j = 1, ..., K/2.$$

For the choice of K in the OS-LRV estimation, we employ the following AMSE-optimal formula in Phillips (2005):

$$K_{MSE} = 2 \times \left\lceil 0.5 \left(\frac{tr\left[(I_{m^2} + \mathbb{K}_{mm})(\Omega^* \otimes \Omega^*)\right]}{4vec(B^*)'vec(B^*)} \right)^{1/5} T^{4/5} \right\rceil,$$

where $\lceil \cdot \rceil$ is the ceiling function, $\mathbb{K}_{mm}$ is $m^2 \times m^2$ commutation matrix and

$$B^* = -\frac{\pi^2}{6} \sum_{j=-\infty}^{\infty} j^2 E u_t^* u_{t-j}^{*\prime}.$$

Similarly, in the case of kernel LRV estimation, we select the smoothing parameter b according to the AMSE-optimal formula in Andrews (1991). The unknown parameters in the AMSE are either calibrated or data-driven using the VAR(1) plug-in approach. The qualitative messages remain the same regardless of how the unknown parameters are obtained.

We consider $m \in \{5, 7, 9\}$ and the corresponding degrees of overidentification are $q = \{1, 3, 5\}$. We look at a finite sample performance of GMM estimators, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_{\text{IE}}^{\infty}$, and corresponding asymptotic variance estimates proposed in the paper. The number of replication is $10,000$ in all of our Monte Carlo simulations.Tables 1–6 show the results which can be summarized as follows.

First, the asymptotic variance estimates of the one-step estimator are close to the actual finite sample variances. This is because the one-step estimator does not require a non-parametric LRV estimate as its GMM weight matrix. This is consistent with Hansen et al. (1996) and in Windmeijer (2005). Second, in contrast to the one-step GMM, the asymptotic variance estimates of the two-step GMM estimators are severely affected by a downward bias in finite samples. The bias is more serious when the sample size is relatively small. For example, when $T = 100$ and $q = 3$, Table 1 indicates that the asymptotic variance estimates, $\widehat{Var}(\hat{\theta}_2)$, is about 35% downward biased from the true finite sample variance $Var(\hat{\theta}_2)$. When $T = 200$, the downward bias decreases by 13%. The bias becomes larger as the degree of overidentification q increases. Table 4 shows the same quantitative messages using the OS-LRV and the iterated GMM estimator.

Table 1 also shows that the finite-sample corrected variance estimate, $\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$, proposed in this paper, successfully reduces the downward biases of $\widehat{Var}(\hat{\theta}_2)$. For instance, when $T = 100$ and $q = 3$, $\widehat{Var}^{\text{adj}}(\hat{\theta}_2)$ reduces the bias of $\widehat{Var}(\hat{\theta}_2)$ by 11%. The improving bias correction of our estimator increases as q increases. We find the same quantitative messages using the OS-LRV and the iterated GMM estimator in Table 4.

Although our corrected variance formula successfully improves the finite-sample behaviors of the asymptotic variance estimates for the GMM estimators, there is still a notable difference between the finite-sample corrected asymptotic variance estimate and the actual finite sample variances, especially when ρ increases. This is not surprising given that the time-series GMM method bears a large amount of finite-sample variability from the non-parametric LRV estimator.

Moreover, the finite-sample variability increases as the time series dependence increases. This can be seen by comparing the results for different values of ρ in Tables 1–6. Our findings are consistent with recent HAR literature, e.g. Sun (2014b) and Hwang and Sun (2017, 2018). Considering the non-parametric LRV estimator, we construct t and Wald statistics for the testing problem using $\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$ and investigate the finite sample performances of the standard t and F tests proposed in this paper.

6.2 Hypothesis testing

We consider the following null hypothesis of interest

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0,$$

where the number of restricted parameters in R is $p = 3$. The significance level α is 5% . For the Bartlett-LRV, we examine the empirical size of the standard Wald statistics, $F(\hat{\theta}_2)$, and the finite-sample corrected Wald statistics, $F_c^{\text{adj}}(\hat{\theta}_2)$, using the conventional chi-square critical values. For the OS-LRV, we examine the empirical size of the Wald statistics, $\tilde{F}(\hat{\theta}_2)$, the finite-sample corrected and the modified Wald statistics, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, using the asymptotic F critical values derived under the fixed smoothing asymptotics. The same test statistics with the iterated GMM estimators, $F(\hat{\theta}_{\text{IE}}^{\infty})$, $F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$, $\tilde{F}(\hat{\theta}_{\text{IE}}^{\infty})$, and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$, are considered.

Table 7 reports the empirical size of the two-step and the iteration procedures based on the conventional chi-square critical values with the Bartlett-LRV. It is clear that the tests based on the uncorrected variance estimates, $F(\hat{\theta}_2)$ and $F(\hat{\theta}_{\text{IE}}^{\infty})$, suffer from severe size distortions. For example, when $T = 100$ and $\rho = 0.50$, the empirical sizes of $F(\hat{\theta}_2)$ and $F(\hat{\theta}_{\text{IE}}^{\infty})$ are reported to be around 21%–38% and these size distortions increase up to 34%–57% when ρ becomes 0.70. As we point out, one possible reason for the failure of chi-square test is the different behavior between the asymptotic variance estimate and the actual finite sample variance of the two-step GMM estimators. The difference can be reflected on the corrected variance estimates that we provide. When the corrected versions of test statistics $F_c^{\text{adj}}(\hat{\theta}_2)$ and $F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$ are used, they can reduce the size distortions significantly. As an example, with the Bartlett kernel estimates, the empirical size distortions of $F(\hat{\theta}_2)$ and $F(\hat{\theta}_{\text{IE}}^{\infty})$, which are around 13%–22% when $T = 100$ and $\rho = 0.50$, are reduced to 12%–16%. Lastly, the results in Table 7 also indicate that the size distortions become larger as the degree of overidentification q increases.

Although the simulation results with the chi-square critical values suggest that the variance corrections can improve the finite sample inferences, the empirical sizes in Table 7 indicate the limitations of the chi-square test. This is because the chi-square critical value from the increasing smoothing asymptotics cannot capture the estimation uncertainty in the nonparametric weight matrix $S_T(\theta_0)$. To reflect the estimation uncertainty of $S_T(\theta_0)$ and make a further improvement on the finite sample inference, we employ the F critical values using the test statistics driven by the fixed smoothing asymptotics. The results are provided in Table 8. We first observe that the size distortions of all testing procedures are substantially reduced. For example, the empirical sizes of $\tilde{F}(\hat{\theta}_2)$ and $\tilde{F}(\hat{\theta}_{\text{IE}}^{\infty})$ are reported to be between 10%–18% when $T = 100$ and $\rho = 0.50$. Thus, the F tests clearly improve the empirical sizes from the previous chi-square tests reducing by 16%. This agrees with the previous literature such as Hwang and Sun (2017) and Kiefer and Vogelsang (2005) which highlight the accuracy of the fixed smoothing asymptotics.

Moreover, the tests with corrected variance estimates, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$, can further improve the empirical size distortions. This is shown in Table 8 which indicates that the empirical size distortions are further reduced to 8%–11% when $T = 100$ and $\rho = 0.50$.The rest of the results in Table 8 exhibit similar quantitative and qualitative interpretations. In sum, our empirical findings are consistent with the theoretical results developed in this paper which indicate that $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$ procedures are able to further refine the fixed smoothing asymptotics by capturing the initial estimation uncertainty from the non-parametric LRV estimator. Also, it is interesting to notice that the amount of size improvement using $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$ is increasing as the degree of overidentification q increases. But there is no clear evidence for the advantage of iteration procedures $\hat{\theta}_{\text{IE}}^{\infty}$, as the performance of $\hat{\theta}_{\text{IE}}^{\infty}$ and $\hat{\theta}_2$ are very close to each other in all cases.

7 Conclusion

This paper develops a more accurate heteroskedasticity autocorrelated robust (HAR) inference as well as a more accurate asymptotic variance estimate for efficient GMM estimators in time-series. We extend Windmeijer (2005)'s approach to the time series setting by explicitly considering the non-parametric LRV estimator in his corrected variance formula. The finite-sample corrected variance estimate in our paper successfully corrects the bias arising from the estimated LRV. We formally show the consistency of the finite-sample corrected variance estimate, and prove that this consistency does not depend on whether the smoothing parameter in the LRV estimator is fixed or is increasing with respect to the sample size.

With our finite-sample corrected variance estimator, this paper constructs t and Wald statistics using the fixed smoothing asymptotics developed in recent HAR literature. The standard t and F limiting distributions derived under the fixed-smoothing asymptotics provide a valid solution to the efficient GMM inference problem. Our results are very appealing to practitioners because they can apply the finite-sample corrected variance formula and the corresponding tests using the standard t and F critical values. Thus, no further simulations or re-sampling methods are needed.

Our Monte Carlo result sow that the asymptotic t and F tests developed in this paper further reduce the empirical size distortions compared to the existing tests in Sun (2014b) and Hwang and Sun (2017). Also, our numerical findings show that the amount of size improvement increases as the degree of overidentification increases or the time series dependence increases.

Table 1: Finite sample performance of GMM estimators and asymptotic variance estimates using Bartlett LRV where $\rho = 0.30$

$\rho = 0.30$ with Bartlett-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0011	0.0015	0.0012	0.0015	0.0009	-0.0004
$Var(\hat{\theta}_1)$	0.0174	0.0088	0.0168	0.0081	0.0162	0.0081
$\widehat{Var}(\hat{\theta}_1)$	0.0141	0.0075	0.0135	0.0072	0.0131	0.0070
$\hat{\theta}_2$	0.0013	0.0018	0.0018	0.0020	0.0008	-0.0006
$Var(\hat{\theta}_2)$	0.0181	0.0089	0.0185	0.0086	0.0193	0.0089
$\widehat{Var}(\hat{\theta}_2)$	0.0137	0.0073	0.0120	0.0067	0.0107	0.0063
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0146	0.0076	0.0139	0.0073	0.0133	0.0071
$\hat{\theta}_{\text{IE}}^{\infty}$	0.0015	0.0017	0.0014	0.0019	0.0014	-0.0007
$Var(\hat{\theta}_{\text{IE}})$	0.0181	0.0089	0.0188	0.0086	0.0200	0.0090
$\widehat{Var}(\hat{\theta}_{\text{IE}})$	0.0137	0.0073	0.0121	0.0067	0.0107	0.0063
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_{\text{IE}})$	0.0147	0.0076	0.0142	0.0073	0.0139	0.0072

Table 2: Finite sample performance of GMM estimators and asymptotic variance estimates using Bartlett LRV where $\rho = 0.50$

$\rho = 0.50$ with Bartlett-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0015	0.0018	0.0016	0.0016	0.0003	0.0001
$Var(\hat{\theta}_1)$	0.0244	0.0122	0.0231	0.0113	0.0221	0.0111
$\widehat{Var}(\hat{\theta}_1)$	0.0170	0.0094	0.0160	0.0089	0.0153	0.0087
$\hat{\theta}_2$	0.0014	0.0019	0.0025	0.0019	-0.0008	-0.0002
$Var(\hat{\theta}_2)$	0.0254	0.0125	0.0258	0.0122	0.0269	0.0125
$\widehat{Var}(\hat{\theta}_2)$	0.0162	0.0091	0.0136	0.0081	0.0117	0.0074
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0179	0.0096	0.0169	0.0091	0.0161	0.0088
$\hat{\theta}_{\text{IE}}^{\infty}$	0.0018	0.0017	0.0019	0.0016	-0.0005	0.0001
$Var(\hat{\theta}_{\text{IE}}^{\infty})$	0.0255	0.0125	0.0264	0.0123	0.0284	0.0128
$\widehat{Var}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0162	0.0091	0.0137	0.0081	0.0117	0.0074
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0181	0.0097	0.0173	0.0093	0.0167	0.0090

Table 3: Finite sample performance of GMM estimators and asymptotic variance estimates using Bartlett LRV where $\rho = 0.70$

$\rho = 0.70$ with Bartlett-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005
$Var(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194
$\widehat{Var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127
$\hat{\theta}_2$	0.0039	0.0030	-0.0012	0.0016	0.0004	0.0014
$Var(\hat{\theta}_2)$	0.0463	0.0222	0.0453	0.0223	0.0465	0.0226
$\widehat{Var}(\hat{\theta}_2)$	0.0214	0.0133	0.0166	0.0111	0.0132	0.0096
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0254	0.0147	0.0241	0.0138	0.0229	0.0130
$\hat{\theta}_{\text{IE}}^{\infty}$	0.0028	0.0031	-0.0010	0.0011	0.0012	0.0006
$Var(\hat{\theta}_{\text{IE}}^{\infty})$	0.0469	0.0223	0.0473	0.0227	0.0508	0.0236
$\widehat{Var}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0214	0.0133	0.0165	0.0111	0.0131	0.0096
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0258	0.0148	0.0243	0.0141	0.0218	0.0136

Table 4: Finite sample performance of GMM estimators and asymptotic variance estimates using OS-LRV where $\rho = 0.30$

$\rho = 0.30$ with OS-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005
$Var(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194
$\widehat{Var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127
$\hat{\theta}_2$	0.0011	0.0017	0.0016	0.0016	0.0016	-0.0009
$Var(\hat{\theta}_2)$	0.0182	0.0089	0.0191	0.0087	0.0205	0.0091
$\widehat{Var}(\hat{\theta}_2)$	0.0142	0.0075	0.0123	0.0069	0.0106	0.0064
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0153	0.0078	0.0142	0.0075	0.0133	0.0071
$\hat{\theta}_{IE}^{\infty}$	0.0013	0.0017	0.0018	0.0017	0.0024	-0.0007
$Var(\hat{\theta}_{IE}^{\infty})$	0.0183	0.0089	0.0195	0.0087	0.0215	0.0092
$\widehat{Var}(\hat{\theta}_{IE}^{\infty})$	0.0143	0.0075	0.0124	0.0069	0.0107	0.0064
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_{IE}^{\infty})$	0.0154	0.0078	0.0145	0.0075	0.0135	0.0072

Table 5: Finite sample performance of GMM estimators and asymptotic variance estimates using OS-LRV where $\rho = 0.50$

$\rho = 0.50$ with OS-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005
$Var(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194
$\widehat{Var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127
$\hat{\theta}_2$	0.0019	0.0020	0.0017	0.0022	-0.0013	-0.0010
$Var(\hat{\theta}_2)$	0.0257	0.0125	0.0270	0.0124	0.0296	0.0130
$\widehat{Var}(\hat{\theta}_2)$	0.0174	0.0097	0.0142	0.0085	0.0117	0.0076
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0192	0.0102	0.0177	0.0095	0.0162	0.0090
$\hat{\theta}_{\text{IE}}^{\infty}$	0.0022	0.0016	0.0015	0.0020	-0.0005	-0.0003
$Var(\hat{\theta}_{\text{IE}}^{\infty})$	0.0258	0.0125	0.0280	0.0125	0.0314	0.0133
$\widehat{Var}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0175	0.0097	0.0144	0.0085	0.0117	0.0077
$\widehat{Var}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0194	0.0102	0.0179	0.0096	0.0161	0.0091

Table 6: Finite sample performance of GMM estimators and asymptotic variance estimates using OS-LRV where $\rho = 0.70$

$\rho = 0.70$ with OS-LRV						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$\hat{\theta}_1$	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005
$Var(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194
$\widehat{Var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127
$\hat{\theta}_2$	0.0038	0.0032	-0.0015	0.0015	-0.0017	0.0010
$Var(\hat{\theta}_2)$	0.0468	0.0225	0.0487	0.0231	0.0551	0.0242
$\widehat{Var}(\hat{\theta}_2)$	0.0240	0.0146	0.0177	0.0120	0.0131	0.0100
$\widehat{Var}_c^{\mathrm{adj}}(\hat{\theta}_2)$	0.0284	0.0160	0.0256	0.0147	0.0238	0.0136
$\hat{\theta}_{\mathrm{IE}}^{\infty}$	0.0022	0.0032	-0.0015	0.0014	0.0012	0.0002
$Var(\hat{\theta}_{\mathrm{IE}}^{\infty})$	0.0476	0.0226	0.0511	0.0236	0.0578	0.0254
$\widehat{Var}(\hat{\theta}_{\mathrm{IE}}^{\infty})$	0.0242	0.0147	0.0176	0.0120	0.0128	0.0101
$\widehat{Var}_c^{\mathrm{adj}}(\hat{\theta}_{\mathrm{IE}}^{\infty})$	0.0286	0.0162	0.0244	0.0149	0.0194	0.0135

Table 7: Empirical size of two step and iteration procedures based on the Bartlett kernel and asymptotic Chi square test under the increasing smoothing asymptotics

Bartlett-LRV using $\chi_p^{1-\alpha}/p$						
$\rho = 0.30$						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$F(\hat{\theta}_2)$	0.1404	0.1014	0.1949	0.1253	0.2533	0.1508
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.1186	0.0917	0.1399	0.0975	0.1643	0.1094
$F(\hat{\theta}_{\text{IE}}^{\infty})$	0.1412	0.1011	0.1945	0.1256	0.2604	0.1521
$F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.1192	0.0904	0.1417	0.0978	0.1651	0.1092
$\rho = 0.50$						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$F(\hat{\theta}_2)$	0.2085	0.1416	0.2793	0.1816	0.3608	0.2241
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.1725	0.1211	0.1944	0.1357	0.2300	0.1539
$F(\hat{\theta}_{\text{IE}}^{\infty})$	0.2078	0.1406	0.2841	0.1830	0.3769	0.2258
$F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.1718	0.1211	0.1988	0.1359	0.2436	0.1556
$\rho = 0.70$						
	$q = 1$		$q = 3$		$q = 5$	
	$T = 100$	$T = 200$	$T = 100$	$T = 200$	$T = 100$	$T = 200$
$F(\hat{\theta}_2)$	0.3408	0.2223	0.4420	0.2977	0.5478	0.3615
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.2704	0.1882	0.3008	0.2118	0.3500	0.2410
$F(\hat{\theta}_{\text{IE}}^{\infty})$	0.3401	0.2231	0.4570	0.3034	0.5731	0.3774
$F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.2743	0.1872	0.3260	0.2194	0.4033	0.2516

Table 8: Empirical size of two step and iteration procedures based on the OS-LRV and asymptotic F test under the fixed smoothing asymptotics

OS-LRV using $\mathcal{F}^{1-\alpha}_{p,K-p-q+1}/p$						
$\rho = 0.30$						
	$q=1$		$q=3$		$q=5$	
	$T=100$	$T=200$	$T=100$	$T=200$	$T=100$	$T=200$
$\tilde{F}(\hat{\theta}_2)$	0.0821	0.0735	0.1076	0.0840	0.1287	0.0938
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.0682	0.0657	0.0708	0.0650	0.0750	0.0650
$\tilde{F}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0818	0.0729	0.1087	0.0840	0.1364	0.0957
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0671	0.0656	0.0712	0.0636	0.0828	0.0632
$\rho = 0.50$						
	$q=1$		$q=3$		$q=5$	
	$T=100$	$T=200$	$T=100$	$T=200$	$T=100$	$T=200$
$\tilde{F}(\hat{\theta}_2)$	0.1024	0.0828	0.1344	0.1022	0.1712	0.1223
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.0838	0.0710	0.0828	0.0730	0.0909	0.0785
$\tilde{F}(\hat{\theta}_{\text{IE}}^{\infty})$	0.1042	0.0838	0.1368	0.1043	0.1815	0.1253
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.0830	0.0704	0.0883	0.0743	0.1050	0.0812
$\rho = 0.70$						
	$q=1$		$q=3$		$q=5$	
	$T=100$	$T=200$	$T=100$	$T=200$	$T=100$	$T=200$
$F(\hat{\theta}_2)$	0.1346	0.1062	0.1779	0.1340	0.2200	0.1653
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.1007	0.0883	0.0984	0.0892	0.1010	0.0975
$F(\hat{\theta}_{\text{IE}}^{\infty})$	0.1354	0.1073	0.1864	0.1367	0.2414	0.1732
$F_c^{\text{adj}}(\hat{\theta}_{\text{IE}}^{\infty})$	0.1045	0.0883	0.1167	0.0925	0.1444	0.1089

8 Appendix

Before we prove Lemma 1, we need technical results provided in the following lemmas. Let us define

$$\Upsilon_j^*(\hat{\theta}_1) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h^*\left(\frac{t}{T},\frac{s}{T}\right) g_j(v_t,\hat{\theta}_1) f(v_s,\hat{\theta}_1)',$$

where

$$Q_h^*(r,s) = Q_h(r,s) - \int_0^1 Q_h(\tau_1,s)d\tau_1 - \int_0^1 Q_h(r,\tau_2)d\tau_2 + \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2. \qquad (29)$$

Lemma 4 *Under Assumptions 1 and $h\to\infty$ such that $h/T\to 0$.*
(a) $T^{-2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} Q_h(\frac{\tau_1}{T},\frac{\tau_2}{T}) - \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2 = o(1)$.
(b) $T^{-1}\sum_{\tau=1}^{T} Q_h\left(\frac{t}{T},\frac{\tau}{T}\right) - \int_0^1 Q_h\left(\frac{t}{T},\tau_2\right)d\tau_2 = o(1)$.
(c) $T^{-1}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) = o(1)$.

Proof of 4. We start by proving the results for the case when $Q_h(r,s) = k((r-s)/b)$. Denote $B_T = bT$. For part (a),

$$\begin{aligned}
\frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} Q_h(\frac{\tau_1}{T},\frac{\tau_2}{T}) &= \frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} k\left(\frac{\tau_1-\tau_2}{B_T}\right) = \frac{1}{T}\sum_{j=-T+1}^{T-1}\frac{(T-1-|j|)}{T}k\left(\frac{j}{B_T}\right) \\
&= \left(\frac{B_T}{T}\right)\frac{1}{B_T}\sum_{j=-T+1}^{T-1} k\left(\frac{j}{B_T}\right) - \frac{B_T}{T}\frac{1}{T}\sum_{j=-T+1}^{T-1}\frac{(1+|j|)}{B_T}k\left(\frac{j}{B_T}\right) \\
&= \left(\frac{B_T}{T}\right)\underbrace{\frac{1}{B_T}\sum_{j=-T+1}^{T-1} k\left(\frac{j}{B_T}\right)}_{\to\int_{-\infty}^{\infty}k(x)<\infty}\left[1-\frac{1}{T}\right] - \left(\frac{B_T}{T}\right)^2\underbrace{\frac{1}{B_T}\sum_{j=-T+1}^{T-1}\frac{|j|}{B_T}k\left(\frac{j}{B_T}\right)}_{\to\int_{-\infty}^{\infty}|x|k(x)|<\infty} \\
&= o(1),
\end{aligned}$$

since $B_T\to\infty$ such that $B_T/T\to 0$. By Assumption 1, $k\left((\tau_1-\tau_2)/b\right)\to 0$ for any τ_1 and τ_2, and this enables us to apply dominated convergence theorem and obtain

$$\begin{aligned}
\int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2 &= \int_0^1\int_0^1 k\left(\frac{\tau_1-\tau_2}{b}\right)d\tau_1 d\tau_2 \\
&= o(1).
\end{aligned}$$

Similarly, for part (b), we have

$$\begin{aligned}
\frac{1}{T}\sum_{\tau=1}^{T} Q_h\left(\frac{t}{T},\frac{\tau}{T}\right) - \int_0^1 Q_h\left(\frac{t}{T},\tau_2\right)d\tau_2 &= \frac{1}{T}\sum_{\tau=1}^{T} k\left(\frac{t-\tau}{S_T}\right) - \int_0^1 k\left(\frac{t/T-\tau_2}{b}\right)d\tau_2 \\
&= \frac{S_T}{T}\times\frac{1}{S_T}\sum_{\tau=1}^{T} k\left(\frac{t-\tau}{S_T}\right) - \int_0^1 k\left(\frac{t}{Tb}-\frac{\tau_2}{b}\right)d\tau_2 \\
&\le \frac{S_T}{T}\times\frac{1}{S_T}\sum_{j=-\infty}^{\infty}\left|k\left(\frac{j}{S_T}\right)\right| + o(1) \\
&= o(1).
\end{aligned}$$

For part (c),

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) &= \frac{1}{T}\sum_{t=1}^{T} k\left(\frac{t-\tau}{S_T}\right) - \int_0^1 k\left(\frac{\tau_1-\tau/T}{b}\right)d\tau_1 - \frac{1}{T}\sum_{t=1}^{T}\int_0^1 k\left(\frac{t/T-\tau_2}{b}\right)d\tau_2 \\
&\quad + \int_0^1\int_0^1 k\left(\frac{\tau_1-\tau_2}{b}\right)d\tau_1 d\tau_2 \\
&= \frac{1}{T}\sum_{t=1}^{T} k\left(\frac{t-\tau}{S_T}\right) - \int_0^1 k\left(\frac{\tau_1-\tau/T}{b}\right)d\tau_1 + o(1) \\
&= o(1),
\end{aligned}
$$

where the last equation follows by the proof of (b).

Next, we consider the case of the OS-LRV with $Q_h(r,s) = K^{-1}\sum_{j=1}^{K}\Phi_j(r)\Phi_j(s)$ and $K\to\infty$ such that $K/T\to 0$. Then, the result of part (a) follows by

$$
\begin{aligned}
& \frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} Q_h(\frac{\tau_1}{T},\frac{\tau_2}{T}) - \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2 \\
=\ & \frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T}\left[\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{\tau_1}{T}\right)\Phi_j\left(\frac{\tau_2}{T}\right)\right] - \frac{1}{K}\sum_{j=1}^{K}\left(\int_0^1\Phi_j(\tau_1)d\tau_1\right)\left(\int_0^1\Phi_j(\tau_2)d\tau_2\right) \\
=\ & \frac{1}{K}\sum_{j=1}^{K}\left(\frac{1}{T}\sum_{\tau_1=1}^{T}\Phi_j\left(\frac{\tau_1}{T}\right)\right)\left(\frac{1}{T}\sum_{\tau_2=1}^{T}\Phi_j\left(\frac{\tau_2}{T}\right)\right) \\
=\ & \frac{1}{K}\sum_{j=1}^{K}\left(\int_0^1\Phi_j(\tau_1)\,d\tau_1 + O\left(\frac{1}{T}\right)\right)\left(\int_0^1\Phi_j(\tau_2)\,d\tau_2 + O\left(\frac{1}{T}\right)\right) \\
=\ & O\left(\frac{1}{T^2}\right) = o(1),
\end{aligned}
$$

since $\int_0^1\Phi_j(\tau)\,d\tau = 0$ by Assumption 1. Part (b) follows in a similar manner since

$$
\begin{aligned}
\frac{1}{T}\sum_{\tau=1}^{T} Q_h\left(\frac{t}{T},\frac{\tau}{T}\right) - \int_0^1 Q_h\left(\frac{t}{T},\tau_2\right)d\tau_2 &= \frac{1}{T}\sum_{\tau=1}^{T}\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{\tau}{T}\right) - \int_0^1\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{t}{T}\right)\Phi_j(\tau_2)\,d\tau_2 \\
&= \frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{t}{T}\right)\left(\frac{1}{T}\sum_{\tau=1}^{T}\Phi_j\left(\frac{\tau}{T}\right)\right) \\
&\quad -\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{t}{T}\right)\left(\int_0^1\Phi_j(\tau_2)\,d\tau_2\right) \\
&= \frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{t}{T}\right)\left(\int_0^1\Phi_j(r)dr + O\left(\frac{1}{T}\right)\right) \\
&= O\left(\frac{1}{T}\right) = o(1).
\end{aligned}
$$

Lastly, it is straightforward to check that $Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)=Q_h\left(\frac{t}{T},\frac{\tau}{T}\right)$ due to $\int_0^1\Phi_j\left(\tau\right)d\tau=0$. Therefore,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{K}\sum_{j=1}^{K}\Phi_j\left(\frac{t}{T}\right)\Phi_j\left(\frac{\tau}{T}\right)\right)\\
&= \frac{1}{K}\sum_{j=1}^{K}\left(\frac{1}{T}\sum_{t=1}^{T}\Phi_j\left(\frac{t}{T}\right)\right)\Phi_j\left(\frac{\tau}{T}\right)\\
&= O\left(\frac{1}{T}\right)=o(1).
\end{aligned}$$

■

Lemma 5 *Under Assumptions 1–5, for any* $\hat{\theta}=\theta_0+O_p(1/\sqrt{T})$,

$$\Upsilon_j(\hat{\theta})=\Upsilon_j^*(\hat{\theta})+o_p(1)$$

holding h *fixed as* $T\to\infty$, *or* $h\to\infty$ *such that* $h/T\to 0$.

Proof of Lemma 5. We first consider the case that h is fixed when $T\to\infty$. For each $j=1,\ldots,d$,

$$\begin{aligned}
&\left\|\Upsilon_j^*(\hat{\theta})-\Upsilon_j(\hat{\theta})\right\| \qquad (30)\\
=&\left\|-\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}e_T(s)g_j(v_t,\hat{\theta})f(v_s,\hat{\theta})'\right]-\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}e_T(t)g_j(v_t,\hat{\theta})f(v_s,\hat{\theta})'\right]\right.\\
&\left.+\left[\frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T}Q_h(\frac{\tau_1}{T},\frac{\tau_2}{T})-\int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2\right]\left[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}g_j(v_t,\hat{\theta})f(v_s,\hat{\theta})'\right]\right\|\\
\leq&\underbrace{\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_j(v_t,\hat{\theta})\right\|}_{:=A}\underbrace{\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}e_T(s)f(v_s,\hat{\theta})\right\|}_{:=B}+\underbrace{\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}f(v_s,\hat{\theta})\right\|}_{:=C}\underbrace{\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_T(t)g_j(v_t,\hat{\theta})\right\|}_{:=D}\\
&+O\left(\frac{1}{T}\right)\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_j(v_t,\hat{\theta})\right\|\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{T}f(v_s,\hat{\theta})'\right\|,
\end{aligned}$$

where

$$e_T(t)=\frac{1}{T}\sum_{\tau=1}^{T}Q_h\left(\frac{t}{T},\frac{\tau}{T}\right)-\int_0^1 Q_h\left(\frac{t}{T},\tau_2\right)d\tau_2.$$

Note that $e_T(t)=O(1/T)=o(1)$ uniformly over t for fixed h from Assumption 1. From the proof of Lemma 1-(a) in Sun (2014), we obtain

$$B=O_p\left(\frac{1}{T}\right)\text{ and }C=O_p\left(1\right). \qquad (31)$$

For each $j = 1, \ldots, d$, we apply mean value theorem to the term A and obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(v_t, \breve{\theta}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(v_t, \theta_0) + H_{T,j}(\bar{\theta}^*_{T,j})\sqrt{T}(\breve{\theta}_T - \theta_0) = O_p(1) \tag{32}$$

for some $\bar{\theta}^*_{T,j}$ which is between $\hat{\theta}$ and θ_0. For the term D,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t) g_j(v_t, \hat{\theta}) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t) g_j(v_t, \theta_0) + \frac{1}{T}\sum_{t=1}^{T} e_T(t)\frac{\partial g_j(v_t, \bar{\theta}^*_{T,j})}{\partial \theta'}\sqrt{T}(\hat{\theta} - \theta_0) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t) g_j(v_t, \theta_0) + \frac{1}{T}\sum_{t=1}^{T}[e_T(t) - e_T(t+1)] \qquad (33)\\
&\quad \times H_{t,j}(\bar{\theta}^*_{T,j})\sqrt{T}(\hat{\theta} - \theta_0) + e_T(T) H_{T,j}(\bar{\theta}^*_{T,j})\sqrt{T}(\hat{\theta} - \theta_0),
\end{aligned}$$

where the second equation follows from summation by parts. For the last term in (33),

$$e_T(T) H_{T,j}(\bar{\theta}^*_{T,j})\sqrt{T}(\hat{\theta} - \theta_0) = O_p\left(\frac{1}{T}\right)$$

by Assumptions 1 and 4. For any m-dimensional vector a,

$$\begin{aligned}
& E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t) a' g_j(v_t, \theta_0)\right)^2\right] \\
= \;& \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} e_T(t) a' E\left[g_j(v_t, \theta_0) g_j(v_s, \theta_0)'\right] a e_T(s) \\
\leq \;& \left(\sup_{1\leq t\leq T} e_T(t)\right)^2 \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} a' E\left[g_j(v_t, \theta_0) g_j(v_s, \theta_0)'\right] a \\
\leq \;& O\left(\frac{1}{T^2}\right)\sum_{i=-\infty}^{\infty} |a'\Psi_{j,i} a| \leq O\left(\frac{1}{T^2}\right)||a||^2 \sum_{i=-\infty}^{\infty} ||\Psi_{j,i}|| = O\left(\frac{1}{T^2}\right)
\end{aligned}$$

by Assumption 5. Together with Markov inequality, this leads us to get

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_T(t) g_j(v_t, \theta_0) = O_p\left(\frac{1}{T}\right).$$

Let us define $\epsilon_{j,t} = H_{t,j}(\bar{\theta}^*_{T,j}) - (t/T)H_j$ for each $j = 1, \ldots, d$. By Assumption 4, $\epsilon_{j,t}$ is $o_p(1)$ uniformly over t. The second term in (33) can be written as

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}[e_T(t) - e_T(t+1)] H_{T,j}(\bar{\theta}^*_{T,j})\sqrt{T}(\hat{\theta} - \theta_0) \\
= \;& \frac{1}{T}\sum_{t=1}^{T}[e_T(t) - e_T(t+1)]\epsilon_{j,t}\sqrt{T}(\hat{\theta} - \theta) + \frac{1}{T}\sum_{t=1}^{T-1}[e_T(t) - e_T(t+1)] t H_j \sqrt{T}(\hat{\theta} - \theta_0) \\
= \;& \frac{1}{T}\sum_{t=1}^{T}[e_T(t) - e_T(t+1)]\epsilon_{j,t}\sqrt{T}(\hat{\theta} - \theta) + \frac{1}{T}\sum_{t=1}^{T-1} e_T(t) H_j \sqrt{T}(\hat{\theta} - \theta_0) - e_T(T) H_j \sqrt{T}(\hat{\theta} - \theta_0) \\
= \;& O_p\left(\frac{1}{T}\right),
\end{aligned}$$

where the last equation follows by $\sqrt{T}(\hat{\theta}-\theta) = O_p(1)$, $\sup_{1\leq t\leq T}\epsilon_{j,t} = o_p(1)$, and $\sup_{1\leq t\leq T} e_T(t) = O(1/T)$, which leads us to $D = O_p(1/T)$. Combining this together with (31) and (32) into (30),

$$\Upsilon_j(\hat{\theta}) = \Upsilon_j^*(\hat{\theta}) + O_p\left(\frac{1}{T}\right) \tag{34}$$

$$= \Upsilon_j^*(\hat{\theta}) + o_p(1) \tag{35}$$

which is the desired result.

Now, we consider the case when $h \to \infty$ such that $h/T \to 0$. Using parts (a) and (b) in Lemma 4, we obtain that

$$\frac{1}{T^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T} Q_h(\frac{\tau_1}{T},\frac{\tau_2}{T}) - \int_0^1\int_0^1 Q_h(\tau_1,\tau_2)d\tau_1 d\tau_2 = o(1); \tag{36}$$

$$\sup_{1\leq t\leq T} e_T(t) = o(1) \tag{37}$$

also holds when $h \to \infty$ such that $h/T \to 0$. A careful inspection of the proof when h is fixed as $T \to \infty$ indicates that (36) and (37) are the necessary conditions for 35 to hold when $h \to \infty$ such that $h/T \to 0$, which completes the proof. ∎

Lemma 6 *Define* $S_t(C) = T^{-1}\sum_{s=1}^{t} C_s$ $t = 0, 1, \ldots, T$ *with* $S_0(C) = 0$ *for a generic sequence of matrices* $\{C_t\}$*. Then, for any two sequence of matrices* $\{A_t\}$ *and* $\{B_t\}$*,*

$$\begin{aligned}
&\frac{1}{T^2}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t B_\tau' \\
=\;& \sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1} \triangledown Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) S_t(A)S_\tau'(B) + \sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right) - Q_h^*\left(\frac{t+1}{T},1\right)\right] S_t(A)S_T(B)' \\
&+\sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau}{T}\right) - Q_h^*\left(1,\frac{\tau+1}{T}\right)\right] S_T(A)S_\tau'(B) + Q_h^*(1,1)\, S_T(A)S_T'(B),
\end{aligned}$$

where

$$\triangledown Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) := Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau+1}{T}\right) + Q_h^*\left(\frac{t+1}{T},\frac{\tau+1}{T}\right).$$

Proof Lemma of 6. We use the formula of summation by part:

$$\frac{1}{T}\sum_{t=1}^{T} a_t b_t' = \frac{1}{T}a_T C_T' - \frac{1}{T}\sum_{t=1}^{T-1}(a_{t+1}-a_t)C_t' \text{ where } C_t = \sum_{s=1}^{t} b_s \tag{38}$$

for any conformable vectors a_t and b_t. Consider

$$\frac{1}{T^2}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t B_\tau' = \frac{1}{T}\sum_{\tau=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t\right) B_\tau'. \tag{39}$$

We first apply the formula in (38) to the term inside the bracket by setting $a_t = Q_h^*(t/T, \tau/T)$, $b_t = A_t$, and $C_t = \sum_{s=1}^t A_s$. Then,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^T Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t \\
= \;& Q_h^*\left(1,\frac{\tau}{T}\right)\left(\frac{1}{T}\sum_{s=1}^T A_s\right) - \sum_{t=1}^{T-1}\left(\frac{1}{T}\sum_{s=1}^T A_s\right)\left[Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right]
\end{aligned}$$

for each $\tau = 1, \ldots, T$. Combining this into (39), we have

$$\begin{aligned}
&\frac{1}{T^2}\sum_{\tau=1}^T\sum_{t=1}^T Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t B_\tau' \\
= \;& \frac{1}{T}\sum_{\tau=1}^T Q_h^*\left(1,\frac{\tau}{T}\right) S_T(A) B_\tau' - \sum_{t=1}^{T-1} S_t(A)\left[\frac{1}{T}\sum_{\tau=1}^T\left\{Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right\} B_\tau'\right]
\end{aligned} \quad (40)$$

We repeat to apply the formula in (38) to the terms in the right-handside. For the first term, setting $a_\tau = Q_h^*\left(1,\frac{\tau}{T}\right) S_T(A)$, $b_\tau = B_\tau'$, and $C_\tau = \sum_{s=1}^\tau B_s'$, we obtain

$$\begin{aligned}
&\frac{1}{T}\sum_{\tau=1}^T Q_h^*\left(1,\frac{\tau}{T}\right) S_T(A) B_\tau' \\
= \;& Q_h^*(1,1) S_T(A) S_T(B)' - \sum_{\tau=1}^{T-1} Q_h^*\left(1,\frac{\tau+1}{T}\right) - Q_h^*\left(1,\frac{\tau}{T}\right) S_T(A) S_\tau(B)'
\end{aligned}$$

For the terms inside the bracket of the second term, set $a_\tau = Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)$, $b_\tau = B_\tau'$, and $C_\tau = \sum_{s=1}^\tau B_s'$, we get

$$\begin{aligned}
&\frac{1}{T}\sum_{\tau=1}^T\left\{Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right\} B_\tau' \\
= \;& \left[Q_h^*\left(\frac{t+1}{T},1\right) - Q_h^*\left(\frac{t}{T},1\right)\right] S_T(B)' - \sum_{\tau=1}^{T-1}\left[Q_h^*\left(\frac{t+1}{T},\frac{\tau+1}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau+1}{T}\right)\right. \\
&\left. - Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) + Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right] S_\tau(B)'.
\end{aligned}$$

Combining these results into (40), we obtain

$$
\begin{aligned}
& \frac{1}{T^2}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) A_t B_\tau' \\
= \;& Q_h^*\left(1,1\right) S_T(A)S_T(B)' - \sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau+1}{T}\right) - Q_h^*\left(1,\frac{\tau}{T}\right)\right] S_T(A)S_\tau(B)' \\
& -\sum_{t=1}^{T-1} S_t(A)\left[Q_h^*\left(\frac{t+1}{T},1\right) - Q_h^*\left(\frac{t}{T},1\right)\right] S_T(B)' + \sum_{t=1}^{T-1} S_t(A)\sum_{\tau=1}^{T-1}\left[Q_h^*\left(\frac{t+1}{T},\frac{\tau+1}{T}\right) - Q_h^*\left(\frac{t}{T},\frac{\tau+1}{T}\right)\right. \\
& \left. -Q_h^*\left(\frac{t+1}{T},\frac{\tau}{T}\right) + Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right)\right] S_\tau(B)' \\
= \;& \sum_{t=1}^{T-1}\sum_{\tau=1}^{T-1} \triangledown Q_h^*\left(\frac{t}{T},\frac{\tau}{T}\right) S_t(A)S_\tau(B)' + \sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T},1\right) - Q_h^*\left(\frac{t+1}{T},1\right)\right] S_t(A)S_T(B)' \\
& + \sum_{\tau=1}^{T-1}\left[Q_h^*\left(1,\frac{\tau}{T}\right) - Q_h^*\left(1,\frac{\tau+1}{T}\right)\right] S_T(A)S_\tau(B)' + Q_h^*\left(1,1\right) S_T(A)S_T(B)',
\end{aligned}
$$

as desired. ■

Proof of Lemma 1. For each $j = 1, ..., d$, we have

$$
\begin{aligned}
D_{\hat\theta_2, S_T(\hat\theta_1)}[., j] &= (G_T' S_T^{-1}(\hat\theta_1) G_T)^{-1} G_T' S_T^{-1}(\hat\theta_1) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1} S_T^{-1}(\hat\theta_1) f_T(\hat\theta_2) \\
&= (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1} S_T^{-1}(\theta_0) f_T(\hat\theta_2)(1+o_p(1)),
\end{aligned}
$$

where the second equality holds by (8). Using a Taylor expansion, we have

$$
f_T(\hat\theta_2) = f_T(\theta_0) - G_T\left\{G_T' S_T^{-1}(\theta_0) G_T\right\}^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0)(1+o_p(1))
$$

Thus,

$$
\begin{aligned}
D_{\hat\theta_2, S_T(\hat\theta_1)}[., j] = & \left\{G_T' S_T^{-1}(\theta_0) G_T\right\}^{-1} G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1} S_T^{-1}(\theta_0) f_T(\theta_0)(1+o_p(1)) \\
& - \left\{G_T' S_T^{-1}(\theta_0) G_T\right\}^{-1} G_T' S_T^{-1}(\theta_0) \left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1} S_T^{-1}(\theta_0) G_T \\
& \times \left\{G_T' S_T^{-1}(\theta_0) G_T\right\}^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0)\Big\} (1+o_p(1)),
\end{aligned}
$$

for each $j = 1, ..., d$. For the term, $\left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1}$, recall that

$$
\begin{aligned}
\left.\frac{\partial S_T(\theta)}{\partial \theta_j}\right|_{\theta=\hat\theta_1} &= \Upsilon_j(\hat\theta_1) + \Upsilon_j'(\hat\theta_1); \\
\Upsilon_j(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q_h\left(\frac{t}{T},\frac{s}{T}\right)\left(g_j(v_t,\theta) - \frac{1}{T}\sum_{\tau=1}^{T} g_j(v_\tau,\theta)\right)\left(f(v_s,\theta) - \frac{1}{T}\sum_{\tau=1}^{T} f(v_\tau,\theta)\right)'; \\
g_j(v_s,\theta) &= \frac{\partial f(v_s,\theta)}{\partial \theta_j}.
\end{aligned}
$$

We want to show that

$$\Upsilon_j(\hat{\theta}_1) = \Upsilon_j(\theta_0) + o_p(1). \tag{41}$$

We first consider the case in which h is fixed as $T \to \infty$. For some $\bar{\theta}^*_{T,j}$ and $\breve{\theta}^*_T$ which are between $\hat{\theta}_1$ and θ_0,we have

$$\begin{aligned}
\Upsilon^*_j(\hat{\theta}_1) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_h\left(\frac{t}{T},\frac{s}{T}\right) g_j(v_t,\hat{\theta}_1) f(v_s,\hat{\theta}_1)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} Q^*_h\left(\frac{t}{T},\frac{s}{T}\right)\left(g_j(v_t,\theta_0) + \frac{\partial g_j(v_t,\bar{\theta}^*_{T,j})}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right) \\
&\quad \times \left(f(v_s,\theta_0) + \frac{\partial f(v_s,\breve{\theta}^*_T)}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right)' \\
&= \Upsilon^*_j(\theta_0) + I_1 + I_2 + I_3,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\left(\frac{\partial g_j(v_t,\bar{\theta}^*_{T,j})}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right) f(v_\tau,\theta_0)'; \\
I_2' &= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\left(\frac{\partial f(v_t,\breve{\theta}^*_T)}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right) g_j(v_\tau,\theta_0)'; \\
I_3 &= \frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\left(\frac{\partial g_j(v_t,\bar{\theta}^*_{T,j})}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right)\left[\frac{\partial f(v_\tau,\breve{\theta}^*_T)}{\partial\theta'}(\hat{\theta}_1-\theta_0)\right]'
\end{aligned}$$

for each $j = 1,\ldots,d$. Let us define $\epsilon_{j,t} = H_{t,j}(\bar{\theta}^*_{T,j}) - (t/T)H_j$ for each $j = 1,\ldots,d$. By Assumption 4, $\epsilon_{j,t}$ is $o_p(1)$ uniformly over t. From Lemma 6, we can write I_1, setting $A_t = \frac{\partial g_j(v_t,\bar{\theta}^*_{T,j})}{\partial\theta'}(\hat{\theta}_1-\theta_0)$ and $B_\tau = u_\tau := f(v_\tau,\theta_0)$, as follows

$$\begin{aligned}
I_1 &= H_j(\hat{\theta}_1-\theta_0)\frac{1}{T}\sum_{\tau=1}^{T}\sum_{t=1}^{T} Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right) u_\tau' + TQ^*_h(1,1)\epsilon_{j,T}(\hat{\theta}_1-\theta_0)S_T'(u) \\
&\quad + T\sum_{\tau=1}^{T-1}\left[Q^*_h\left(1,\frac{\tau}{T}\right) - Q^*_h\left(1,\frac{\tau+1}{T}\right)\right]\epsilon_{j,T}(\hat{\theta}_1-\theta_0)S_\tau'(u) \\
&\quad + T\sum_{t=1}^{T-1}\left[Q^*_h\left(\frac{t}{T},1\right) - Q^*_h\left(\frac{t+1}{T},1\right)\right]\epsilon_{j,t}(\hat{\theta}_1-\theta_0)S_T'(u) \\
&\quad + T\sum_{\tau=1}^{T-1}\sum_{t=1}^{T-1} \triangledown Q^*_h\left(\frac{t}{T},\frac{\tau}{T}\right)\epsilon_{j,t}(\hat{\theta}_1-\theta_0)S_\tau'(u) \\
I_1 &: = I_{11} + I_{12} + I_{13} + I_{14} + I_{15}.
\end{aligned}$$

We want to show that each of I_{1i} for $i = 1, \ldots 5$ is $o_p(1)$ as $T \to \infty$ holding h fixed. For I_{11}, there exists a finite $M > 0$ which does not depend on t such that

$$\begin{aligned}
\|I_{11}\| &= \left\| H_j \sqrt{T}(\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} \left[\frac{1}{T} \sum_{t=1}^{T} Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] u_\tau' \right\| \\
&= \left\| H_j \sqrt{T}(\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} \left[\int_0^1 Q_h^* \left(r, \frac{\tau}{T} \right) dr + O\left(\frac{1}{T} \right) \right] u_\tau' \right\| \\
&\leq \frac{M}{T} \left\| \left(H_j \sqrt{T}(\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} u_\tau' \right) \right\| \\
&= O_p\left(\frac{1}{T} \right) = o_p(1),
\end{aligned} \tag{42}$$

where the inequality follows by $\int_0^1 Q_h^* \left(r, \frac{\tau}{T} \right) dr = 0$. It is easy to check $I_{12} = o_p(1)$ from $\epsilon_{j,T} = O_p(1)$. Next, we consider I_{13}. By summation by parts,

$$\begin{aligned}
I_{13} &= T \sum_{\tau=1}^{T-1} \left[Q_h^* \left(1, \frac{\tau}{T} \right) - Q_h^* \left(1, \frac{\tau+1}{T} \right) \right] \epsilon_{j,T} (\hat{\theta}_1 - \theta_0) S_\tau'(u) \\
&= \epsilon_{j,T} \sqrt{T}(\hat{\theta}_1 - \theta_0) \left[\frac{1}{\sqrt{T}} \sum_{\tau=1}^{T} Q_h^* \left(1, \frac{\tau}{T} \right) u_\tau \right] \\
&\quad - \epsilon_{j,T} \sqrt{T}(\hat{\theta}_1 - \theta_0) Q_h^* (1, 1)\, S_T'(u) \\
&= o_p(1),
\end{aligned}$$

which follows by the boundedness of the function $Q_h^*(\cdot, \cdot)$ and $\epsilon_{j,T} = o_p(1)$. For I_{14},

$$\begin{aligned}
\|I_{14}\| &= \left\| T \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \epsilon_{j,t} (\hat{\theta}_1 - \theta_0) S_T'(u) \right\| \\
&\leq \left\| \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \epsilon_{j,t} \right\| \left\| \sqrt{T}(\hat{\theta}_1 - \theta_0) \right\| \left\| \sqrt{T} S_T'(u) \right\| \\
&\leq \sup_{1 \leq t \leq T} ||\epsilon_{j,t}|| \left\| \sum_{t=1}^{T-1} Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right\| \times O_p(1) \\
&= o_p(1) \left| Q_h^* \left(\frac{1}{T}, 1 \right) - Q_h^* (1, 1) \right| = o_p(1).
\end{aligned}$$

Lastly, we re-express the term I_{15}

$$\begin{aligned}
\|I_{15}\| &= \left\| T \sum_{\tau=1}^{T-1} \sum_{t=1}^{T-1} \triangledown Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \epsilon_{j,t} (\hat{\theta}_1 - \theta_0) S_\tau'(u) \right\| \\
&\leq \sup_{1 \leq t \leq T} ||\epsilon_{j,t}|| \left\| \sum_{\tau=1}^{T-1} \left[\sum_{t=1}^{T-1} \triangledown Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] \sqrt{T}(\hat{\theta}_1 - \theta_0) \sqrt{T} S_\tau'(u) \right\|.
\end{aligned} \tag{43}$$

For each $\tau = 1, \ldots, T$,

$$\begin{aligned}
& \sum_{t=1}^{T-1} \triangledown Q_h^*\left(\frac{t}{T}, \frac{\tau}{T}\right) \\
= & \sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T}, \frac{\tau}{T}\right) - Q_h^*\left(\frac{t+1}{T}, \frac{\tau}{T}\right)\right] - \sum_{t=1}^{T-1}\left[Q_h^*\left(\frac{t}{T}, \frac{\tau+1}{T}\right) - Q_h^*\left(\frac{t+1}{T}, \frac{\tau+1}{T}\right)\right] \\
= & \left[Q_h^*\left(\frac{1}{T}, \frac{\tau}{T}\right) - Q_h^*\left(1, \frac{\tau}{T}\right)\right] - \left[Q_h^*\left(\frac{1}{T}, \frac{\tau+1}{T}\right) - Q_h^*\left(1, \frac{\tau+1}{T}\right)\right] \\
= & \left[Q_h^*\left(\frac{1}{T}, \frac{\tau}{T}\right) - Q_h^*\left(\frac{1}{T}, \frac{\tau+1}{T}\right)\right] - \left[Q_h^*\left(1, \frac{\tau}{T}\right) - Q_h^*\left(1, \frac{\tau+1}{T}\right)\right].
\end{aligned}$$

Using this, we re-express the upper bound of $\|I_{15}\|$ in (43) by

$$\begin{aligned}
& \sup_{1\le t\le T} ||\epsilon_{j,t}|| \left\| \sum_{\tau=1}^{T-1}\left[\sum_{t=1}^{T-1} \triangledown Q_h^*\left(\frac{t}{T}, \frac{\tau}{T}\right)\right] \sqrt{T}(\hat{\theta}_1 - \theta_0)\sqrt{T} S_\tau'(u) \right\| \\
\le & \sup_{1\le t\le T} ||\epsilon_{j,t}|| \times \left| \sum_{\tau=1}^{T-1} \left[Q_h^*\left(\frac{1}{T}, \frac{\tau}{T}\right) - Q_h^*\left(\frac{1}{T}, \frac{\tau+1}{T}\right)\right] - \left[Q_h^*\left(1, \frac{\tau}{T}\right) - Q_h^*\left(1, \frac{\tau+1}{T}\right)\right] \right| \\
& \times \left\| \sqrt{T}(\hat{\theta}_1 - \theta_0) \right\| \left\| \max_{1\le \tau\le T} \sqrt{T} S_\tau'(u) \right\| \\
= & o_p(1) \times \left| \left[Q_h^*\left(\frac{1}{T}, \frac{1}{T}\right) - Q_h^*\left(\frac{1}{T}, 1\right)\right] - \left[Q_h^*\left(1, \frac{1}{T}\right) - Q_h^*(1,1)\right] \right| \times O_p(1) \\
= & o_p(1),
\end{aligned}$$

where the second last equation follows by the FCLT and continuous mapping theorem. We have therefore showed that $I_1 = o_p(1)$. The proofs of $I_2 = o_p(1)$ and $I_3 = o_p(1)$ can be done in a very similar manner, and we omit the details. Finally, we obtain that

$$\begin{aligned}
\Upsilon_j^*(\hat{\theta}_1) &= \Upsilon_j^*(\theta_0) + o_p(1) \\
&= \Upsilon_j^*(\theta_0)(1 + o_p(1)).
\end{aligned} \tag{44}$$

Using the result in Lemma 5, we obtain

$$\Upsilon_j^*(\hat{\theta}_1) = \Upsilon_j(\hat{\theta}_1)(1 + o_p(1)) \text{ and } \Upsilon_j^*(\theta_0) = \Upsilon_j(\theta_0)(1 + o_p(1)),$$

which implies $\Upsilon_j(\hat{\theta}_1) = \Upsilon_j(\theta_0)(1 + o_p(1))$ for each $j = 1, \ldots, d$. From this result, it is straightforward to obtain

$$D_{\hat{\theta}_2, S_T(\hat{\theta}_1)} = D_{\theta_0, S_T(\theta_0)}(1 + o_p(1)), \tag{45}$$

which is the desired result.

Now, we consider the case when $h \to \infty$ such that $h/T \to 0$. A careful inspection of the proof when h is fixed as $T \to \infty$ indicates that the necessary condition for (44) to hold is provided in part (c) of Lemma 4,which can be used in (42) to obtain the desired result. ■

Proof of Theorem 2. We only prove part (b), as the proof of (a) can be done in the same way. Now, define the infeasible finite-sample corrected variance

$$\begin{aligned}
\widehat{var}_c^{\inf}(\hat{\theta}_2) &= \widehat{var}(\hat{\theta}_2) + \frac{1}{T} D_{\theta_0, S_T(\theta_0)} \widehat{var}(\hat{\theta}_2) \\
&\quad + \frac{1}{T}\widehat{var}(\hat{\theta}_2) D'_{\theta_0, S_T(\theta_0)} + D_{\theta_0, S_T(\theta_0)} \widehat{var}(\hat{\theta}_1) D'_{\theta_0, S_T(\theta_0)},
\end{aligned}$$

with corresponding statistics

$$F_c^{\inf}(\hat{\theta}_2) = \frac{1}{p}(R\hat{\theta}_2 - r)' \left[R\widehat{var}_c^{\inf}(\hat{\theta}_2)R'\right]^{-1}(R\hat{\theta}_2 - r).$$

Then, the result in (45) implies $\widehat{var}_c(\hat{\theta}_2) = \widehat{var}_c^{\inf}(\hat{\theta}_2)(1+o_p(1))$. Thus the corresponding infeasible Wald statistic satisfies

$$F_c(\hat{\theta}_2) = F_c^{\inf}(\hat{\theta}_2)(1 + o_p(1))$$

Also, $D_{\theta_0,S_T(\theta_0)} = o_p(1)$ implies

$$\widehat{var}_c^{\inf}(\hat{\theta}_2) = \widehat{var}(\hat{\theta}_2)(1 + o_p(1)),$$

and this leads us to get

$$\begin{aligned} F_c(\hat{\theta}_2) &= F_c^{\inf}(\hat{\theta}_2)(1 + o_p(1)) \\ &= F(\hat{\theta}_2) + o_p(1), \end{aligned}$$

as desired. ■

Proof of Theorem 3. Define the modified t and Wald statistics using $t(\hat{\theta}_2)$ and $F(\hat{\theta}_2)$ as

$$\begin{aligned} \tilde{t}(\hat{\theta}_2) &:= \frac{K-q}{K} \cdot \frac{t(\hat{\theta}_2)}{\sqrt{1 + \frac{1}{K}J(\hat{\theta}_2)}}; \\ \tilde{F}(\hat{\theta}_2) &:= \frac{K-p-q+1}{K} \cdot \frac{F(\hat{\theta}_2)}{1 + \frac{1}{K}J(\hat{\theta}_2)}, \end{aligned}$$

Under Assumptions 2–6, we can apply Theorem 1in Hwang and Sun (2017), and have that

$$\tilde{t}(\hat{\theta}_2) \xrightarrow{d} t_{p,K-p-q+1} \text{ and } \tilde{F}(\hat{\theta}_2) \xrightarrow{d} \mathcal{F}_{p,K-p-q+1}$$

for a fixed K as $T \to \infty$. By Theorem 2,

$$\tilde{t}_c(\hat{\theta}_2) = \tilde{t}(\hat{\theta}_2) + o_p(1) \text{ and } \tilde{F}_c(\hat{\theta}_2) = \tilde{F}(\hat{\theta}_2) + o_p(1),$$

and this gives us the desired results

$$\tilde{t}_c(\hat{\theta}_2) \xrightarrow{d} t_{K-p-q+1} \text{ and } \tilde{F}_c(\hat{\theta}_2) \xrightarrow{d} \mathcal{F}_{p,K-p-q+1}.$$

■

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–854.

[2] Anatolyev, S. (2005). GMM, GEL, serial correlation, and asymptotic bias. Econometrica, 73(3), 983-1002.

[3] Bilodeau, M., & Brenner, D. (2008). Theory of multivariate statistics. Springer Science & Business Media.

[4] Bond, S., & Windmeijer, F. (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. Econometric Reviews, 24(1), 1-37.

[5] Brown, J. R., Fazzari, S. M., & Petersen, B. C. (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom. The Journal of Finance, 64(1), 151-185.

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

[7] Hall, P., & Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica: Journal of the Econometric Society, 891-916.

[8] Hansen, B. E., & Lee, S. (2019). Inference for Iterated GMM Under Misspecification. Working paper.

[9] Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics, 14(3), 262-280.

[10] Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica: Journal of the Econometric Society, 1269-1286.

[11] Hwang, J., & Sun, Y. (2017). Asymptotic F and t Tests in an Efficient GMM Setting. Journal of econometrics, 198(2), 277-295.

[12] Hwang, J., & Sun, Y. (2018). Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework. Journal of econometrics, 207(2), 381-405.

[13] Hwang, J., Kang, B., & Lee, S. (2019). A Doubly Corrected Robust Variance Estimator for Linear GMM. Working paper, Department of Economics, University of Connecticut.

[14] Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. Econometrica, 70(5), 2093-2095.

[15] Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory, 18(6), 1350-1366.

[16] Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory, 21(6), 1130-1164.

[17] Lazarus, E., Lewis, D. J., & Stock, J. H. (2019). The size-power tradeoff in HAR inference. Available at SSRN 3436372.

[18] Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR Inference: Recommendations for Practice. Journal of Business & Economic Statistics, 36(4), 541-559.

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

[20] Lee, S. (2016). Asymptotic refinements of a misspecification-robust bootstrap for GEL estimators. Journal of econometrics, 192(1), 86-104.

[21] Martínez-Iriarte, J., Sun, Y., & Wang, X. (2019). Asymptotic F Tests under Possibly Weak Identification. Working paper, Department of Economics, UC San Diego.

[22] Müller, U. K. (2007). A theory of robust long-run variance estimation. Journal of Econometrics, 141(2), 1331-1352.

[23] Müller, U. K., & Watson, M. W. (2018). Long-Run Covariability. Econometrica, 86(3), 775-804.

[24] Newey, W. K., & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica, 72(1), 219-255.

[25] Oberholzer-Gee, F., & Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. Journal of political economy, 115(1), 1-42.

[26] Phillips, P. C. (2005). Automated discovery in econometrics. Econometric Theory, 21(1), 3-20.

[27] Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. The stata journal, 9(1), 86-136.

[28] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. The Econometrics Journal, 16(1), 1-26.

[29] Sun, Y. (2014a). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. Journal of Econometrics, 178, 659-677.

[30] Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. Econometrica, 82(6), 2327-2370.

[31] Sun, Y., Phillips, P. C., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. Econometrica, 76(1), 175-194.

[32] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of econometrics, 126(1), 25-51.