

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

May, 2021

Abstract

This paper develops a finite-sample corrected inference for the efficient generalized method of moments (GMM) in time series. To capture a higher-order uncertainty embodied in estimating the time series GMM weight matrix, we extend the finite-sample corrected variance formula of Windmeijer (2005) to heteroskedasticity autocorrelated robust (HAR) inference. Using fixed-smoothing asymptotics, we show that our finite-sample corrected test statistics lead to standard asymptotic t or F critical values and outperform existing GMM procedures on finite-samples, including continuously updating GMM. Not only does our finite-sample corrected variance formula correct for the bias arising from the plugged-in long-run variance estimation, but it is also not exposed to a potential side effect of Windmeijer's formula, which can introduce an additional type-I error after the correction.

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. We thank co-editor Serena Ng, an anonymous associate editor, and anonymous referees for valuable comments. We also thank Yixiao Sun for helpful comments and encouragement in the early stage of composing this paper. Gonzalo Valdés acknowledges financial support from Proyecto Mayor de Investigación Científica y Tecnológica UTA Mayor 2020 (8750-20).

1 Introduction

The generalized method of moments (GMM, Hansen, 1982) is one of the most widely applied methods in econometrics. In the 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 for the bias arising from the estimated efficient weight matrix. Windmeijer’s (2005) corrected variance formula has been popularly used in a wide variety of econometric models with high impact (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) renders his corrected variance formula inapplicable. This is because the GMM weighting matrix in time series is no longer a simple average of the estimated moment process. Instead, we need to consider the weight 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. Because of the non-parametric nature of the LRV estimator, in a 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’s (2005) formula to the case of non-i.i.d. Since our corrected variance formula is designed to take into account the extra variation due to the plugged-in estimation of the LRV, it leads to a less biased estimate of the actual finite-sample variance of the GMM 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. In addition, 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”. In the context of the efficient two-step GMM, Sun (2014b) and Hwang and Sun (2017) show that the alternative asymptotics yields more accurate approximations. Although their fixed-smoothing asymptotics captures the variance of the non-parametric LRV estimator, it ignores the finite-sample bias of LRV estimation. By correcting for the small-order bias of the plugged-in LRV estimator, this paper provides an improved fixed-smoothing asymptotic theory. Together with the finite-sample corrected variance formula, we derive standard t and F limiting distributions for our proposed test statistics, which provides a convenient

solution to the efficient GMM inference problem for finite samples.

This paper also provides an analytical expansion of the continuously updating (CU) GMM estimator proposed by Hansen et al. (1996) and compares it to our finite-sample corrected two-step GMM method. Our result indicates that the Wald inference in CU-GMM, using the standard sandwich variance formula, cannot reflect potentially large finite-sample variations embodied in the non-linear CU estimation. In contrast, our proposed GMM method does not involve the non-linearity in the CU-GMM estimation and instead uses the corrected variance formula that explicitly considers the finite-sample uncertainties in the estimated optimal weighting matrix. Also, via Monte Carlo simulations, we numerically compare the performance of the corrected two-step GMM test to that of the uncorrected CU-GMM test and show that the inferences drawn from our proposed GMM outperform those drawn from the CU-GMM on finite samples.

Our asymptotic framework is pointwise and can suffer from size distortion when the temporal dependence is strong (e.g., Müller (2014)). Preinerstorfer and Pötscher (2016) and Pötscher and Preinerstorfer (2018, 2019) point out that the size-distortion problem applies to any pointwise HAR test and provides the conditions under which the size of HAR tests is controlled. Similar problems can arise when there is weak identification (e.g., Stock and Wright (2000) and Guggenberger and Smith (2005)). The results of our Monte Carlo simulations show that our proposed finite-sample corrected GMM inference still performs better than the uncorrected version in these scenarios, by significantly reducing the size distortion.

Different approaches to the efficient GMM inference problem have been proposed in the literature. A bootstrap approach for GMM is developed in Hall and Horowitz (1996), Brown and Newey (2002), and Lee (2014), and numerical evidence of the finite-sample performance of the asymptotic and bootstrapped GMM tests is provided in Bond and Windmeijer (2005). Hwang et al. (2020) point out a connection between the finite-sample corrected variance formula and the misspecification-robust asymptotic variance formula. However, all these papers impose the i.i.d. assumption or restriction on the form of the time-series dependence. Newey and Smith (2004) and Anatolyev (2005) analyze higher-order properties for various classes of GMM estimators, including CU-GMM, but they mainly focus on point-estimation which differs from the testing problem considered in this paper. Hwang (2021) provides a finite-sample corrected inference for linear GMM in the presence of clustered dependence, but that formulation of the GMM weight matrix is a cluster-robust variance estimator which is applicable only to a specific form of clustered (or grouped) dependence structure.

This paper contributes to the literature by providing a comprehensive treatment of finite-sample corrected GMM methods in the presence of unknown forms of time-series dependence, which include both increasing and fixed-smoothing asymptotics, and finite-sample corrections for linear and non-linear moment conditions. Also, we note that using a deflated variance estimate after Windmeijer’s correction in testing can end up with a greater chance of falsely rejecting the null hypothesis than using the standard (uncorrected) sandwich variance estimate. This is because the Windmeijer correction requires estimating the smaller-order correction term whose estimation uncertainty is of the same order as the true correction term. To overcome this problem, we provide additional adjustments on top of Windmeijer’s formula, which makes our new formula is not exposed to Windmeijer’s formula’s side effect. The adjustments first compute the spectral decomposition of the difference between the corrected and uncorrected variance–covariance matrices and adjust the eigenvalues of the difference matrix to be non-negative.

The literature in HAR inference was 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 assumes that the smoothing parameter is fixed when the sample size grows. This tool is called fixed- b asymptotics in Kiefer and Vogelsang (2002, 2005) and, with more inclusive implications in both the kernel and OS LRV estimators, it is called fixed-smoothing asymptotics in Sun (2014 a&b). Recent research along this line can be found in Sun (2014 a&b), Müller and Watson (2018), Lazarus et al. (2019), and Martínez-Iriarte et al. (2020).

In a two-step GMM model, one can consider alternative testing-oriented selections of bandwidth (smoothing) parameters for HAR inference, as in Sun and Phillips (2009). However, their bandwidth formula is designed for kernel LRV, so it cannot be applied to the orthonormal series LRV estimator implemented in our proposed finite-sample corrected HAR inference. Also, the testing-oriented smoothing parameter does not always perform much better than the usual mean squared error (MSE)-optimal smoothing parameter choice (e.g., Sun and Phillips (2009) and Sun (2013)). The reason is that there is considerable estimation uncertainty in estimating the theoretically optimal smoothing parameter. For this reason, this paper opts for the more familiar and widely used MSE-optimal smoothing parameter in this paper, together with the finite-sample corrected variance formula.

The rest of the paper is organized as follows. Section 2 describes the two-step GMM problem in a time-series setting and explores the idea of the finite-sample correction in time-series two-step GMM. Section 3 establishes asymptotic distributions for the test statistics using the corrected variance formula. Section 4 presents analysis of the finite-sample distribution of the CU-GMM and compare it to the finite-sample corrected GMM. Section 5 presents Monte Carlo simulation results, and Section 6 concludes. Proofs of the main results are given in Appendix A, and Online Appendix B contains closed-form formulas for the non-linear moment case and for iterated GMM, as well as an IV regression example and proofs that are not given in Appendix A.

2 Finite-Sample Corrected GMM in Time Series

We want to estimate a d -component vector of parameters $\theta \in \Theta$ using a vector of observations $v_t \in \mathbb{R}^d$ at time t . The true parameter θ_0 is assumed to be an interior point of Θ . The moment condition is given by

$$E[f(v_t, \theta)] = 0 \text{ if and only if } \theta = \theta_0,$$

where $f(v_t, \cdot)$ is an m -component vector of twice continuously differentiable functions, 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 form, and we require it to satisfy $\sum_{j=-\infty}^{\infty} \|E[f(v_t, \theta_0)f(v_{t-j}, \theta_0)']\| < \infty$ and some mixing conditions so that the time-series functional central limit theorem (FCLT) holds:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T \rfloor} f(v_t, \theta_0) \xrightarrow{d} \Lambda B(\cdot), \quad (1)$$

where $B(\cdot)$ is an m -dimensional standard Brownian motion. $\Omega = \Lambda\Lambda'$ is an $m \times m$ strictly positive definite long-run variance (LRV) of the moment process $f(v_t, \theta_0)$ which is defined as $\Omega = \sum_{j=-\infty}^{\infty} E[f(v_t, \theta_0)f(v_{t-j}, \theta_0)']$. Also, we assume that $q = m - d > 0$, so the model is overidentified with degree of over-identification q . $G(\theta_0) = E[\partial f(v_t, \theta_0)/\partial \theta']$ is assumed to have full column rank.

Let $f_T(\theta) = T^{-1} \sum_{s=1}^T f(v_s, \theta)$ and $M(\theta, S_T(\hat{\theta}_1)) = f_T(\theta)' S_T^{-1}(\hat{\theta}_1) f_T(\theta)$. The feasible efficient two-step GMM estimator in Hansen(1982) is defined as

$$\hat{\theta}_2 = \arg \min_{\theta \in \Theta} M(\theta, S_T(\hat{\theta}_1)),$$

where $S_T(\theta)$ is an $m \times m$ two-step GMM weight matrix defined as

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

In the two-step GMM weight matrix, $\hat{\theta}_1$ is the one-step (initial) GMM estimator which minimizes the quadratic form in the sample moments, but its weight matrix W_T^{-1} does not depend on the unknown parameter value θ_0 . We assume that $p \lim_{T \rightarrow \infty} W_T = W$.

By construction, $S_T(\hat{\theta}_1)$ is a quadratic heteroskedasticity autocorrelation robust (HAR) estimator for Ω which uses a symmetric weighting function, $Q_h(t/T, s/T)$, with smoothing parameter h . For conventional kernel LRV estimators, with $Q_h(r, s) = k((r-s)/b)$, we take $h = 1/b$. For orthonormal series (OS) LRV estimators with $Q_h(r, s) = K^{-1} \sum_{j=1}^K \Phi_j(r)\Phi_j(s)$, we take $h = K$, where $\{\Phi_j(r)\}$ is a set of 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.

Note that we use a ‘‘centered’’ version of the LRV estimator, $S_T(\hat{\theta}_1)$, which is based on the estimation of the demeaned moment process $f(v_t, \hat{\theta}_1) - f_T(\hat{\theta}_1)$. The demeaned procedure plays an advantageous role in our finite-sample corrected GMM tests because it allows the denominator of the test statistics (the finite-sample corrected variance) to be asymptotically independent of the numerator (the sample moments).

To understand the asymptotic approximation of $\hat{\theta}_2$, we look at the first-order condition (FOC) for $\hat{\theta}_2$ given by

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

where $G_T(\hat{\theta}_2) = T^{-1} \sum_{t=1}^T \partial f(v_t, \theta) / \partial \theta' |_{\theta=\hat{\theta}_2}$. A first-order approximation to characterize the distribution of $\hat{\theta}_2$ can be constructed as follows. First, conditioning on $S_T(\hat{\theta}_1)$, we do a Taylor expansion of the FOC in (2):

$$0 = \frac{1}{2} \frac{\partial M(\theta, S_T(\hat{\theta}_1))}{\partial \theta} \Bigg|_{\theta=\hat{\theta}_2} = G_T(\theta_0)' S_T^{-1}(\hat{\theta}_1) f_T(\theta_0) + A(\theta_0, S_T(\hat{\theta}_1))(\hat{\theta}_2 - \theta_0) + O_p \left(\frac{1}{T} \right), \quad (3)$$

where $A(\theta_0, S_T(\hat{\theta}_1))$ is the matrix of second-order derivatives of $M(\theta, S_T(\hat{\theta}_1))$ at $\theta = \theta_0$, and $H_T(\theta) \in \mathbb{R}^{dm \times d}$ is the matrix of second-order derivatives of the moment process. The closed-form expressions for $A(\theta_0, S_T(\hat{\theta}_1))$ and $H_T(\theta)$ are provided in Online Appendix B.1. Using the Taylor expansion of the FOC in (3), we can expand $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ as

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

assuming $A(\theta_0, S_T(\hat{\theta}_1))$ is invertible.

For simplicity of exposition, we illustrate the main idea of our finite-sample correction by assuming that the moment conditions are linear in the parameter θ . This linearity assumption leads us to focus solely on the estimation uncertainty of the plugged-in estimator, $S_T(\hat{\theta}_1)$, which is the primary motivation of the finite-sample corrected GMM in Windmeijer (2005). The formulation of the finite-sample correction in the non-linear case is analogous, and it is provided in

Online Appendix B.1. Because of the linearity, the term $A(\hat{\theta}_2, S_T(\hat{\theta}_1))$ is equal to $G_T' S_T^{-1}(\hat{\theta}_1) G_T$, and the higher-order approximation error term $O_p(T^{-1/2})$ in (4) is dropped, hence we have that

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) = -(G_T' S_T^{-1}(\hat{\theta}_1) G_T)^{-1} G_T' S_T^{-1}(\hat{\theta}_1) \sqrt{T} f_T(\theta_0). \quad (5)$$

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

$$S_T(\hat{\theta}) = S_T(\theta_0) + o_p(1). \quad (6)$$

Using this result, we can approximate (5) as

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) = \underbrace{-(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0)}_{=\sqrt{T}(\tilde{\theta}_2 - \theta_0)} + o_p(1). \quad (7)$$

Any standard approximation of the two-step GMM estimator $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ is based on the first term in (7), 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, $S_T^{-1}(\theta_0)$. Thus $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ 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 the existing asymptotic analysis, including the fixed-smoothing asymptotics in Sun (2014b) and Hwang and Sun (2017).

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 difference in the finite-sample behavior of $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ and $\sqrt{T}(\tilde{\theta}_2 - \theta_0)$. By estimating the term $o_p(1)$ in (7), a finite-sample corrected variance estimate is obtained. Windmeijer (2005) shows that his corrected variance estimate approximates the true finite-sample variance well and leads to a more accurate result in hypothesis testing. Windmeijer (2005) assumes that the moment process $f(v_t, \theta_0)$ is i.i.d., but his idea of a corrected variance estimate can be accommodated to our time-series setup. In doing so, the key step is that instead of eliminating the estimation uncertainty of $\hat{\theta}_1$ in (5), we further approximate $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ in equation (5) by a Taylor expansion, as a function of $\hat{\theta}_1$, in the estimated weight matrix $S_T(\hat{\theta}_1)$ as follows:

$$\begin{aligned} \sqrt{T}(\hat{\theta}_2 - \theta_0) = & -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) \\ & + D(\theta_0, S_T(\theta_0)) \sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \quad (8)$$

where

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

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

$$\begin{aligned} D(\theta_0, S_T(\theta_0))[:, j] = & -(G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} \\ & \times S_T^{-1}(\theta_0) G_T (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0) \\ & + (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} S_T^{-1}(\theta_0) f_T(\theta_0), \end{aligned} \quad (9)$$

where $\partial S_T(\theta)/\partial\theta_j = \Upsilon_j(\theta) + \Upsilon'_j(\theta)$ for $j \in \{1, \dots, d\}$,

$$\Upsilon_j(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T Q_h\left(\frac{t}{T}, \frac{s}{T}\right) \left(g_j(v_s, \theta) - \frac{1}{T} \sum_{s=1}^T g_j(v_s, \theta) \right) \left(f(v_t, \theta_0) - \frac{1}{T} \sum_{s=1}^T f(v_s, \theta) \right)', \quad (10)$$

and $g_j(v_s, \theta) = \partial f(v_s, \theta)/\partial\theta_j$. The expansion in (8) 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, we can improve the approximation of the variance of $\hat{\theta}_2$ in finite samples by taking the correction term into account.

3 Asymptotics for Finite-Sample Corrected Statistics

3.1 Formulation of finite-sample corrected variance

The idea of finite-sample corrected variance estimate starts by assuming that the distribution of $S_T(\theta_0)$ is approximated by the true population counterpart Ω . In time series, the approximation is based on the conventional increasing-smoothing asymptotics which considers $h \rightarrow \infty$ as $T \rightarrow \infty$ such that $h/T \rightarrow 0$, (e.g., Andrews (1991) and Sun (2013, 2014a)). 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 $N(0, G'\Omega^{-1}G)$.

To formulate a finite-sample corrected variance estimator, we use $\overset{a}{\sim}$ to denote a notion of asymptotic equivalence in distribution. That is, $\xi_T \overset{a}{\sim} \eta_T$ indicates that ξ_T and η_T share the same weak limits for two stochastically bounded sequences of random vectors ξ_T and η_T . Keeping Windmeijer's correction term $\tilde{D}(\theta_0, S_T(\theta_0))$ in (8), the asymptotically equivalent representation for (8) is given as

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) \overset{a}{\sim} - \left((G'\Omega^{-1}G)^{-1}, \tilde{D}(\theta_0, S_T(\theta_0))(G'W^{-1}G)^{-1} \right) \begin{pmatrix} G'\Omega^{-1}\Lambda Z \\ G'W^{-1}\Lambda Z \end{pmatrix},$$

where $Z \sim N(0, I_m)$, $\tilde{D}(\theta_0, S_T(\theta_0))$ is a $d \times d$ random matrix which joins the same marginal distribution as that of $D(\theta_0, S_T(\theta_0))$, and $Z \perp \tilde{D}(\theta_0, S_T(\theta_0))$. Conditioning on $\tilde{D}(\theta_0, S_T(\theta_0))$, the sum of two normal distributions can be represented as a normal distribution as well. Thus we can obtain a normal representation, $N(0, \Xi_T)$, of the approximated distribution of $\sqrt{T}(\hat{\theta}_2 - \theta_0)$, whose variance-covariance matrix, $\Xi_T := \Xi_T(\theta_0, S_T(\theta_0))$, is given by

$$\begin{aligned} \Xi_T &= (G'\Omega^{-1}G)^{-1} + \tilde{D}(\theta_0, S_T(\theta_0))(G'\Omega^{-1}G)^{-1} + (G'\Omega^{-1}G)^{-1} \tilde{D}(\theta_0, S_T(\theta_0))' \\ &\quad + \tilde{D}(\theta_0, S_T(\theta_0))(G'W^{-1}G)^{-1} (G'W^{-1}\Omega W^{-1}G) (G'W^{-1}G)^{-1} \tilde{D}(\theta_0, S_T(\theta_0))'. \end{aligned}$$

Motivated by this, the corrected variance estimate is given as $\widehat{var}_c(\hat{\theta}_2) = \hat{\Xi}_T(\hat{\theta}_2, S_T(\hat{\theta}_1))$, where

$$\begin{aligned} \hat{\Xi}_T(\hat{\theta}_2, S_T(\hat{\theta}_1)) &= \widehat{var}(\hat{\theta}_2) + D(\hat{\theta}_2, S_T(\hat{\theta}_1))\widehat{var}(\hat{\theta}_2) + \widehat{var}(\hat{\theta}_2)D(\hat{\theta}_2, S_T(\hat{\theta}_1))' \\ &\quad + D(\hat{\theta}_2, S_T(\hat{\theta}_1))\widehat{var}(\hat{\theta}_1)D(\hat{\theta}_2, S_T(\hat{\theta}_1))', \end{aligned} \quad (11)$$

$\widehat{var}(\hat{\theta}_1)$ and $\widehat{var}(\hat{\theta}_2)$ are the standard GMM variance estimates, that is,

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

and

$$D(\hat{\theta}_2, S_T(\hat{\theta}_1))[:, j] = (G'_T S_T^{-1}(\hat{\theta}_1) G_T)^{-1} G'_T S_T^{-1}(\hat{\theta}_1) \left(\Upsilon_j(\hat{\theta}_1) + \Upsilon'_j(\hat{\theta}_1) \right) S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2).$$

Using the FOC, $G'_T S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2) = 0$, hence the first term of $D(\theta_0, S_T(\theta_0))[:, j]$ on the right-hand side of (9) 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 (11) is in the same spirit as in Windmeijer (2005), where it is assumed to be 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 statistic using the finite-sample corrected variance $\widehat{var}_c(\hat{\theta}_2)$ can be constructed as follows:

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

whereas the uncorrected Wald statistic using the conventional formula, $F(\hat{\theta}_2)$, uses the standard sandwich variance formula $\widehat{var}(\hat{\theta}_2)$.¹ When $p = 1$ and for one-sided alternative hypotheses, one can similarly construct the corresponding finite-sample corrected and uncorrected t statistics as $t_c(\hat{\theta}_2)$ and $t(\hat{\theta}_2)$, respectively. We assume the following.

Assumption 1 (i) For kernel LRV estimators, the kernel function $k(\cdot) \in [-1, 1]$ 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 $\Phi_j(x)$ are piecewise monotonic, continuously differentiable, and orthonormal in $L^2[0, 1]$, and $\int_0^1 \Phi_j(x) dx = 0$.

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

Assumption 3 For any $\hat{\theta} = \theta_0 + o_p(1)$, $G_{[rT]}(\hat{\theta}) = T^{-1} \sum_{t=1}^{[rT]} \frac{\partial f(v_t, \theta)}{\partial \theta'} \Big|_{\theta=\hat{\theta}} = rG + o_p(1)$ uniformly in r , where $G = G(\theta_0)$ has rank d and $G(\theta) = E[\partial f(v_t, \theta)/\partial \theta']$. When $r = 1$, we have that $G_T(\hat{\theta}) = G + O_p(T^{-1/2})$.

Assumption 4 For each $j \in \{1, \dots, d\}$ and any $\hat{\theta} = \theta_0 + o_p(1)$, $H_{[rT],j}(\hat{\theta}) = rH_j + o_p(1)$ uniformly in r , where $H_{[rT],j}(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^{[rT]} \frac{\partial g_j(v_t, \hat{\theta})}{\partial \theta'}$ and $H_j = H_j(\theta_0)$ with $H_j(\theta) = E[\partial g_j(v_t, \theta)/\partial \theta']$. When $r = 1$, we have that $H_{T,j}(\hat{\theta}) = H_j + O_p(T^{-1/2})$.

Assumption 5 For each $j \in \{1, \dots, d\}$, $\{g_j(v_t, \theta_0)\}$ 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 FCLT.

¹In constructing the Wald statistic, we divide it by the number of hypotheses p , but only because we anticipate a more convenient F approximation in the next subsection.

Assumptions 1–3 are standard assumptions in the literature on HAR inference, and they are similar to those in Sun (2014) and Hwang and Sun (2017, 2018). Assumptions 4 and 5 are needed to prove the asymptotic validity of the finite-sample corrected variance formula. Assumptions 4 and 5 are not needed to prove our asymptotic theory if the moment conditions are linear in the parameters, as illustrated in Section 2. For completeness, we keep them to prove Lemma 1 for a more general (non-linear) case in Online Appendix B.1, which includes the linear-moment condition as a particular case.

Lemma 1 *Under Assumptions 1–5, 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 is fixed as $T \rightarrow \infty$, or when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$.

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 the smoothing parameter h being fixed as $T \rightarrow \infty$, or on $h \rightarrow \infty$ such that $h/T \rightarrow 0$. 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 states this result.

Theorem 2 *Under Assumptions 1–5, we have that*

$$(a) \ t_c(\hat{\theta}_2) = t(\hat{\theta}_2) + o_p(1);$$

$$(b) \ F_c(\hat{\theta}_2) = F(\hat{\theta}_2) + o_p(1),$$

where (a) and (b) hold when $h \rightarrow \infty$ such that $h/T \rightarrow 0$, or when h is fixed as $T \rightarrow \infty$.

Under the increasing-smoothing asymptotics, that is, $h \rightarrow \infty$ but $h/T \rightarrow 0$, 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 results in Theorem 2 justify that the distribution of our finite-sample corrected t and Wald statistics can be approximated by

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

respectively.

3.2 Fixed-smoothing asymptotics for finite-sample corrected variance

The conventional increasing-smoothing asymptotics is a key device for proving conventional normal and chi-square approximations of our finite-sample corrected statistics, but it often fails to reflect finite-sample variations of the non-parametric LRV estimation in time-series data. See, for example, Kiefer and Vogelsang (2002 a&b, 2005) and Hwang and Sun (2017). To overcome this problem, we derive a more accurate fixed-smoothing asymptotics for the finite-sample corrected $t_c(\hat{\theta}_2)$ and $F_c(\hat{\theta}_2)$. We consider LRV estimation using the following orthonormal series (OS) weighting function:

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

where the smoothing parameter h in $Q_h(r/T, s/T)$ is now equal to $K \geq m$, which is the number of orthonormal series functions. The OS-LRV estimator is then defined as $S_T(\theta) = K^{-1} \sum_{j=1}^K U_j(\theta) U_j(\theta)'$, where

$$U_j(\theta) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) \left[f(v_t, \theta) - \frac{1}{T} \sum_{\tau=1}^T f(v_\tau, \theta) \right]$$

for each $j \in \{1, \dots, K\}$. The OS-LRV estimator has gained considerable attention in the recent HAR literature, (e.g., Phillips (2005), Müller (2007), Sun (2011, 2014b), Lazarus et al. (2018), and Lazarus et al. (2019)). In this paper, we consider

$$\{\Phi_j(r)\}_{j=1}^K = \{\Phi_{2j-1}(r) = \sqrt{2} \sin(2\pi jr), \Phi_{2j}(r) = \sqrt{2} \cos(2\pi jr), j = 1, 2, \dots, K/2\}, \quad (12)$$

where K is an even number. By construction, the OS-LRV becomes an equal-weighted periodogram (EWP) estimator using the first $K/2$ periodograms (e.g., Sun (2013) and Lazarus et al. (2019)), which is proportional to an estimator of the scaled spectral density point at zero.

The fixed-smoothing approximation of the OS-LRV, $S_T(\hat{\theta}_1)$, captures the finite-sample variability of each periodogram by $U_j(\hat{\theta}_1) \stackrel{a}{\sim} \Lambda \mathbb{U}_j$ for each $j \in \{1, \dots, K\}$, where

$$\mathbb{U}_j = T^{-1/2} \sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) \left(e_t - \frac{1}{T} \sum_{s=1}^T e_s \right)$$

and $e_t \stackrel{\text{i.i.d.}}{\sim} N(0, I_m)$. From the properties of the Fourier basis functions, $\sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) = 0$ and $T^{-1} \sum_{t=1}^T \Phi_i\left(\frac{t}{T}\right) \Phi_j\left(\frac{t}{T}\right) = 1 (i \neq j)$, it is straightforward to show that $\mathbb{U}_j = \frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_j\left(\frac{t}{T}\right) e_t \stackrel{\text{i.i.d.}}{\sim} N(0, I_m)$ for $j \in \{1, \dots, K\}$. Then, holding K fixed, Sun (2013) shows that the OS-LRV is approximated by

$$S_T(\hat{\theta}_1) \stackrel{a}{\sim} \mathbb{S} = \begin{pmatrix} \mathbb{S}_{11} & \mathbb{S}_{12} \\ d \times d & d \times q \\ \mathbb{S}_{21} & \mathbb{S}_{22} \\ q \times d & q \times q \end{pmatrix} := \Lambda \left(\frac{1}{K} \sum_{j=1}^K \mathbb{U}_j \mathbb{U}_j' \right) \Lambda'. \quad (13)$$

Note that the approximated random variable $\mathbb{S} \sim K^{-1} \mathbb{W}_p(K, I_m)$ is a scaled Wishart random matrix with K degrees of freedom. Using this result, Theorem 1 in Sun (2014b) shows that 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

$$t(\hat{\theta}_2) \xrightarrow{d} \mathbb{T} \stackrel{d}{=} \frac{Z_1 - \mathbb{S}_{12} \mathbb{S}_{22}^{-1} Z_2}{\sqrt{\mathbb{S}_{11.2}}} \quad (14)$$

and

$$F(\hat{\theta}_2) \xrightarrow{d} \mathbb{F} \stackrel{d}{=} \frac{1}{p} (Z_1 - \mathbb{S}_{12} \mathbb{S}_{22}^{-1} Z_2)' \mathbb{S}_{11.2}^{-1} (Z_1 - \mathbb{S}_{12} \mathbb{S}_{22}^{-1} Z_2), \quad (15)$$

respectively, where $Z_1 \sim N(0, I_p)$, $Z_2 \sim N(0, I_q)$, $Z_1 \perp Z_2$, and $\mathbb{S}_{11.2} = \mathbb{S}_{11} - \mathbb{S}_{12} \mathbb{S}_{22}^{-1} \mathbb{S}_{21}$.

A natural question is whether $t_c(\hat{\theta}_2)$ and $F_c(\hat{\theta}_2)$ are asymptotically free of nuisance parameters, including the correction term $D(\hat{\theta}_2, S_T(\hat{\theta}_1))$. 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

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

respectively. The fixed- K limiting distributions in (14) and (15) are nonstandard. To investigate further, we use the following well-known properties of the Wishart distribution: $\mathbb{S}_{11.2} \sim \mathbb{W}_p(K - p - q + 1, I_p)/G$, and $\mathbb{S}_{11.2}$ 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

$$\frac{K - p - q + 1}{K} \mathbb{F} \stackrel{d}{=} \frac{K - p - q + 1}{K} \frac{(Z_1 + \Delta)' \mathbb{S}_{11.2}^{-1} (Z_1 + \Delta)}{p} \stackrel{d}{=} \mathcal{F}_{p, K-p-q+1} \left(\|\Delta\|^2 \right),$$

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 the non-standard limiting distribution \mathbb{F} , and in practice the critical values need to be simulated. With our finite-sample corrected test statistics, from a practical standpoint it would be more convenient to use a standard F critical value after we modify the non-standard source in Δ . The modified t and Wald statistics are

$$\tilde{t}_c(\hat{\theta}_2) = \frac{K - q}{K} \cdot \frac{t_c(\hat{\theta}_2)}{\sqrt{1 + \frac{1}{K} J(\hat{\theta}_2)}}; \quad (16)$$

$$\tilde{F}_c(\hat{\theta}_2) = \frac{K - p - q + 1}{K} \cdot \frac{F_c(\hat{\theta}_2)}{1 + \frac{1}{K} J(\hat{\theta}_2)}, \quad (17)$$

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.

Theorem 3 *Under Assumptions 1–5, for a fixed K as $T \rightarrow \infty$, 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 alter 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 in testing $H_0 : R\theta_0 = r$, 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 statistic. 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. Together with the finite-sample corrected variance formula, the standard t and F limits provide convenient solutions to the efficient GMM inference problem for finite samples.

The Windmeijer correction requires estimating the smaller-order correction term, $D(\theta_0, S_T(\theta_0))$, by $D(\hat{\theta}_2, S_T(\hat{\theta}_1))$, and the estimation cost has the same order of magnitude with the true correction, that is,

$$D(\hat{\theta}_2, S_T(\hat{\theta}_1)) = \underbrace{D(\theta_0, S_T(\theta_0))}_{=O_p\left(\frac{1}{\sqrt{T}}\right)} + O_p\left(\frac{1}{\sqrt{T}}\right).$$

The estimation uncertainty can change the corrected variance estimates in an unexpected way in finite samples, that is, we could obtain $\widehat{var}_c(\hat{\theta}_2) - var(\hat{\theta}_2) < 0$. A smaller variance estimate after

the correction does not necessarily imply that it gets closer to the actual finite-sample variance. In fact, the deflation of the original (uncorrected) variance estimate after Windmeijer's correction can introduce greater risk of committing type-I error in testing problems. Since the motivation of our corrected variance formula and corresponding t and Wald inferences is to reduce the excessive size distortion in finite samples, we want to set the gap between $\widehat{var}_c(\hat{\theta}_2)$ and $\widehat{var}(\hat{\theta}_2)$ to be non-negative. Thus we propose to replace $\widehat{var}_c(\hat{\theta}_2)$ by $\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$, where the adjusted variance correction $\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$ satisfies

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

The adjustment step comes from looking at the matrix $R_T = \widehat{var}_c(\hat{\theta}_2) - \widehat{var}(\hat{\theta}_2)$, which is a small-order difference between $\widehat{var}_c(\hat{\theta}_2)$ and $\widehat{var}(\hat{\theta}_2)$. After calculating the spectral decomposition of $\tilde{R}_T = V_T L_T V_T'$, we 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 $\widehat{var}_c(\hat{\theta}_2)$ can be constructed as

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

Using $\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$, the corrected t and Wald statistics, $\tilde{t}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, are defined similarly as in (16) and (17). By construction, the corrected statistics always satisfy

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

Thus the size distortions of $\tilde{t}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ are no greater than the uncorrected $\tilde{t}(\hat{\theta}_2)$ and $\tilde{F}(\hat{\theta}_2)$, respectively, in finite samples. Also, Theorem 2 indicates that $R_T = o_p(1)$ and $\tilde{R}_T = o_p(1)$. This further implies that the proposed $\tilde{t}_c^{\text{adj}}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ are asymptotically t and F distributed, that is, that

$$t_c^{\text{adj}}(\hat{\theta}_2) \xrightarrow{d} t_{K-q} \text{ and } \tilde{F}_c^{\text{adj}}(\hat{\theta}_2) \xrightarrow{d} \mathcal{F}_{p, K-p-q+1} \quad (18)$$

hold under fixed-smoothing asymptotics.

Our eigenvalue adjustments overcome the limitation in Windmeijer's original method, because our method compares the difference between the corrected and uncorrected variance estimates and adjusts the corrected variance to be at least as large as in the original (uncorrected) formula. By doing so, the resulting test statistic does not produce more size distortion than the uncorrected test statistic does. The eigenvalue adjustments after the Windmeijer correction guarantee that we do not over-correct the original variance estimates when the true variance is larger than the corrected variance estimate, that is, when $\widehat{var}_c(\hat{\theta}_2) - \widehat{var}(\hat{\theta}_2) < 0$. This ensures that our formula becomes immune to the side effect of Windmeijer's formula, which has an additional chance of committing type-I error after the correction.

4 Comparison to Continuously Updating GMM

In an attempt to improve the finite-sample performances of the two-step GMM estimator, Hansen et al. (1996) propose and investigate two alternative GMM estimators: iterated GMM and continuously updating (CU) GMM. In Online Appendix B.2, we show that the finite-sample corrected variance estimate for iterated GMM estimator can be formulated in the same way that we correct the two-step GMM. This section analyzes the finite-sample distribution of the CU-GMM to the same order shown for the two-step GMM and compares the uncorrected CU-GMM inference with the inference using the finite-sample corrected two-step GMM.

The CU-GMM estimator continuously optimizes the entire GMM criterion function, including the inverted LRV estimator:

$$\hat{\theta}_{\text{CU}} = \arg \min_{\theta \in \Theta} M(\theta, S_T(\theta)),$$

where $M(\theta, S_T(\theta)) = f_T(\theta)' S_T^{-1}(\theta) f_T(\theta)$. In HAR inference, Zhang (2016) and Hwang and Sun (2017) show that the CU-GMM Wald tests are first-order equivalent to the two-step GMM tests under fixed-smoothing asymptotics.

Define a $dm \times m$ matrix of $\Upsilon(\theta)$ such that

$$\Upsilon'(\theta) = [\Upsilon'_1(\theta), \dots, \Upsilon'_d(\theta)],$$

where

$$\Upsilon_j(\theta) = \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k\left(\frac{t}{T}\right) \frac{\partial f(v_t, \theta)}{\partial \theta} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \Phi_k\left(\frac{t}{T}\right) f(v_t, \theta) \right)'$$

The FOC of the CU-GMM objective function is

$$Q((\hat{\theta}_{\text{CU}}, S_T(\hat{\theta}_{\text{CU}})) := \frac{1}{2} \frac{\partial M(\theta, S_T(\theta))}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{CU}}} = \tilde{G}_T(\hat{\theta}_{\text{CU}}) [S_T(\hat{\theta}_{\text{CU}})]^{-1} f_T(\hat{\theta}_{\text{CU}}) = 0,$$

where $\tilde{G}_T(\theta) := G_T - \Upsilon'(\theta) \{I_d \otimes (S_T^{-1}(\theta) f_T(\theta))\}$ is an effective Jacobian function for the CU-GMM estimator.² Donald and Newey (2006) show that $\tilde{G}_T(\theta)$ is a key device for CU-GMM, since it provides an appealing higher-order bias property for $\hat{\theta}_{\text{CU}}$ to the two-step GMM point estimator, $\hat{\theta}_2$. Newey and Smith (2004) derive a higher-order bias formula for $\hat{\theta}_{\text{CU}}$ in the i.i.d. setting. Anatolyev (2005) extends the results of Newey and Smith (2004) in the time series setting.

Zhang (2016) proves that under some mild conditions,

$$\sqrt{T}(\hat{\theta}_{\text{CU}} - \theta_0) = - [G_T' S_T^{-1}(\theta_0) G_T]^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) + o_p(1) \quad (19)$$

holds under the first-order fixed-smoothing asymptotics. That is, the CU-GMM estimator is asymptotically equivalent to the (in)feasible two-step GMM estimator. This leads us to construct the J -statistic modification of the CU-Wald statistic as

$$\tilde{F}(\hat{\theta}_{\text{CU}}) = \frac{K - p - q + 1}{K} \cdot \frac{F(\hat{\theta}_{\text{CU}})}{1 + \frac{1}{K} J(\hat{\theta}_{\text{CU}})},$$

where $F(\hat{\theta}_{\text{CU}})$ is the Wald statistic using the uncorrected sandwich variance formula $\widehat{\text{var}}(\hat{\theta}_{\text{CU}})$. Hwang and Sun (2017) show that the J -statistic-modified t and Wald statistics, $\tilde{t}(\hat{\theta}_{\text{CU}})$ and $\tilde{F}(\hat{\theta}_{\text{CU}})$, are asymptotically t and F distributed, respectively, so they are first-order equivalent to the two-step GMM tests considered in subsection 3.2.

However, Guggenberger (2005, 2008) provides numerical evidence of no-moment and heavy-tail problems which are embodied in finite-sample distributions of the CU-GMM estimators, which cannot be thoroughly explained by (19). To provide a better description of $\hat{\theta}_{\text{CU}}$ in finite

²See the proof of Lemma B.4 in Online Appendix B for a derivation of the FOC function $Q(\hat{\theta}_{\text{CU}}, S_T(\hat{\theta}_{\text{CU}}))$.

samples, we derive an expansion of $\hat{\theta}_{CU}$ in a way that is similar to our derivation of the corrected two-step GMM in Section 2. Define a $dm \times d$ matrix

$$\tilde{H}(\theta_0) = \begin{pmatrix} \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_1} \Big|_{\theta=\theta_0} - \Upsilon_1(\theta_0) S_T^{-1}(\theta_0) [\tilde{G}_T(\theta_0)] \\ \vdots \\ \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_d} \Big|_{\theta=\theta_0} - \Upsilon_d(\theta_0) S_T^{-1}(\theta_0) [\tilde{G}_T(\theta_0)] \end{pmatrix},$$

where $\partial \tilde{G}_T(\theta) / \partial \theta_j = \tilde{G}_{j,2}^{(a)}(\theta) + \tilde{G}_{j,2}^{(b)}(\theta) + \tilde{G}_{j,2}^{(c)}(\theta)$ for $j \in \{1, \dots, d\}$,

$$\tilde{G}_{j,2}^{(a)}(\theta) = - \left[\Upsilon_{1j}^{(2)'} S_T^{-1}(\theta) f_T(\theta), \dots, \Upsilon_{dj}^{(2)'} S_T^{-1}(\theta) f_T(\theta) \right],$$

$$G_{j,2}^{(b)}(\theta) = \left[\Upsilon_1'(\theta) S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) f_T(\theta), \dots, \Upsilon_d'(\theta) S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) f_T(\theta) \right],$$

$$G_{j,2}^{(c)}(\theta) = - \left[\Upsilon_1'(\theta) S_T^{-1}(\theta) G_{jT}, \dots, \Upsilon_d'(\theta) S_T^{-1}(\theta) G_{jT} \right],$$

and $\Upsilon_{ji}^{(2)} = \partial \Upsilon_j(\theta) / \partial \theta_i$.

Assumption 6 $\hat{\theta}_{CU} = \theta_0 + o_p(1)$ and satisfies the first-order asymptotics in (19).

Theorem 4 Suppose that Assumptions 1–5 hold with $\hat{\theta}_{CU}$ and that Assumption 6 holds. Let $\tilde{A}(\theta_0, S_T(\theta_0)) = \tilde{A}_1(\theta_0, S_T(\theta_0)) + \tilde{A}_2(\theta_0, S_T(\theta_0))$, where

$$\tilde{A}_1(\theta_0, S_T(\theta_0)) = \left[\tilde{G}_T(\theta_0) \right]' S_T^{-1}(\theta_0) \left[\tilde{G}_T(\theta_0) \right] + \{I_d \otimes (f_T(\theta_0)' S_T^{-1}(\theta_0))\} \tilde{H}(\theta_0),$$

$$\tilde{A}_2(\theta_0, S_T(\theta_0)) = \frac{1}{2} \sum_{j=1}^d \frac{\partial Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_0} \psi_j(\theta_0, S_T(\theta_0)),$$

$$\psi_j(\theta_0, S_T(\theta_0)) = e_j' [G_T' S_T^{-1}(\theta_0) G_T]^{-1} G_T' S_T^{-1}(\theta_0) f_T(\theta_0),$$

and e_j is a d -component vector with j -th component equal to 1, and 0 otherwise. Then, assuming that $\tilde{A}(\theta_0, S_T(\theta_0))$ is invertible, the CU-GMM admits the following expansion:

$$\sqrt{T}(\hat{\theta}_{CU} - \theta_0) = - \left[\tilde{A}(\theta_0, S_T(\theta_0)) \right]^{-1} \left[\tilde{G}_T(\theta_0) \right]' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (20)$$

The proof of Theorem 4 is in Online Appendix B. Theorem 4 provides an analytical expansion of the CU-GMM estimator whose order of magnitude is the same as that of two-step GMM estimator shown in (8). The exact expression for $\partial Q(\theta, S_T(\theta)) / \partial \theta_j \partial \theta'$ is provided in Lemma B.4 in Online Appendix B.4. It is not surprising that the CU scheme does not depend on the initial weighting matrix W_T in $\hat{\theta}_1$. The difference between CU-GMM and two-step GMM is also captured by the terms $\tilde{A}_1(\theta_0, S_T(\theta_0))$, $\tilde{A}_2(\theta_0, S_T(\theta_0))$, and $\tilde{G}_T(\theta_0)$ in (20). In the proof of Theorem 4, which is available in Online Appendix B.4, we show that these terms appear mainly because of the effective Jacobian function, $\tilde{G}_T(\theta)$, which is a non-linear function of θ . This contrasts with the two-step GMM estimator, whose Jacobian term, G_T , does not depend on θ .

The result in Theorem 4 indicates that the uncorrected $\tilde{F}(\hat{\theta}_{CU})$, which is based on the first-order expansion in (19), may not reflect potentially large finite-sample variations embodied in the non-linear CU estimation. In contrast, our proposed finite-sample corrected $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ does not

involve the non-linearity in the CU-GMM estimation and uses the corrected variance formula that explicitly considers the finite-sample uncertainties in the estimated optimal weighting matrix. In the next section, we numerically explore finite-sample distributions of $\hat{\theta}_2$ and $\hat{\theta}_{\text{CU}}$ by investigating the performance of the corrected two-step GMM test, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, and the uncorrected CU-GMM test, $\tilde{F}(\hat{\theta}_{\text{CU}})$.

5 Simulation Results

We follow the simulation design in Sun (2014b) and Hwang and Sun (2017) 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 $z_{0,t}, z_{1,t}, \dots, z_{m-1,t}$, with $z_{0,t} \equiv 1$. The reduced-form equations for $x_{1,t}$, $x_{2,t}$, and $x_{3,t}$ are given by

$$x_{j,t} = z_{j,t} + \sum_{i=d-1}^{m-1} z_{i,t} + \epsilon_{x_{j,t}} \text{ for } j \in \{1, 2, 3\}. \quad (21)$$

We assume that the $z_{i,t}$ for $i \geq 1$ follow an AR(1) process, that is, $z_{i,t} = \rho z_{i,t-1} + \sqrt{1 - \rho^2} e_{z_{i,t}}$ where $(e_{z_{1,t}}^1, \dots, e_{z_{m-1,t}}^{m-1})' \sim N(0, V_e)$. The diagonal elements of V_e are equal to 1 and the off-diagonal elements are equal to ψ . The data-generation process (DGP) for $\epsilon_t = (\epsilon_{y,t}, \epsilon_{x_{1,t}}, \epsilon_{x_{2,t}}, \epsilon_{x_{3,t}})'$ is the same as the DGP for $(z_{1,t}, \dots, z_{m-1,t})'$ except for the dimensionality difference. Thus the parameter $\psi \in [0, 1)$ serves as a degree of endogeneity between the regressor $x_{j,t}$ and $\epsilon_{y,t}$. By construction, the vectors, ϵ_t and $(z_{1,t}, \dots, z_{m-1,t})'$ are independent of each other. We consider the true parameters to be $\theta_0 = (0, 0, 0, 0)'$, $\rho \in \{0.3, 0.5, 0.7, 0.9\}$, and $\psi = 0.5$.

Define $x_t = (x_{1,t}, x_{2,t}, x_{3,t})'$ and $z_t = (z_{0,t}, z_{1,t}, \dots, z_{m-1,t})'$. Then we have that the m moment conditions are given by

$$E[f(v_t, \theta_0)] = E[z_t(y_t - x_t' \theta_0)] \in \mathbb{R}^m. \quad (22)$$

The closed-form expressions for the one-step, two-step, and iterated GMM estimators and the corresponding formulas for the (un)corrected asymptotic variance estimators are shown in Online Appendix B.3.

5.1 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 orthonormal Fourier basis functions introduced in (12). For the choice of K^* in the OS-LRV estimation, we use $K^* = \max\{K_{\text{MSE}}, 8\}$ for $q \in \{1, 3\}$ and $K^* = \max\{K_{\text{MSE}}, 10\}$ for $q = 5$, where K_{MSE} is the AMSE-optimal formula in Phillips (2005) and Sun (2013):

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

$\lceil \cdot \rceil$ is the ceiling function, \mathbb{K}_{mm} is the $m^2 \times m^2$ commutation matrix, and $B^* = -\pi^2/6 \sum_{j=-\infty}^{\infty} j^2 E u_t u_{t-j}'$ with $u_t = f(v_t, \theta_0)$. Similarly, in the case of the kernel LRV estimation, we select the smoothing

parameter b^* according to the asymptotic mean squared error (AMSE)-optimal formula in Andrews (1991).³ The unknown parameters in the AMSE can be either calibrated or data driven using the VAR(1) plug-in approach. Here, we use the data-driven VAR(1) plug-in approach, following Andrews (1991). 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 over-identification are $q = \{1, 3, 5\}$. The number of replications in all of our Monte Carlo simulations is 10,000. We look at the finite-sample performance of the estimated values for β_1 in the GMM point estimators, $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_{\text{iter}}^\infty$, and the corresponding finite-sample variances and median values of the asymptotic variance estimates considered in the paper. Tables 1 and 2 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 Windmeijer (2005). Second, in contrast to the one-step GMM, the conventional sandwich 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 $T = 100$, $q = 3$, and $\rho = 0.30$, Table 1 indicates that the sandwich asymptotic variance estimate, $\widehat{\text{var}}(\hat{\theta}_2)$, is about 35% downward biased from the true finite-sample variance $\text{var}(\hat{\theta}_2)$. When $T = 200$, the downward bias decreases by 13%. The bias becomes larger as the degree of over-identification q increases. Table 2 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 proposed in this paper, $\widehat{\text{var}}_c^{\text{adj}}(\hat{\theta}_2)$, successfully reduces the downward bias of $\widehat{\text{var}}(\hat{\theta}_2)$. For instance, when $T = 100$, $q = 3$, and $\rho = 0.30$, $\widehat{\text{var}}_c^{\text{adj}}(\hat{\theta}_2)$ reduces the bias of $\widehat{\text{var}}(\hat{\theta}_2)$ by 29%. The improvement in the bias correction of our estimator increases as q increases. We find the same quantitative messages using the OS-LRV and the iterated GMM estimator, as shown in Table 2.

Although our corrected variance formula improves the finite-sample behavior 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 yields a large amount of finite-sample variability from the non-parametric LRV estimator. 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 and 2. As illustrated in subsection 3.2, the non-parametric LRV estimate, $\widehat{\text{var}}_c^{\text{adj}}(\hat{\theta}_2)$, converges in distribution to a random matrix under fixed-smoothing asymptotics. In the next subsection, we consider this feature and investigate the finite-sample performances of our proposed F test using $\widehat{\text{var}}_c^{\text{adj}}(\hat{\theta}_2)$ in the Wald statistic.

Lastly, Tables 1 and 2 indicate that the bias in LRV estimation is less severe for the OS estimator using Fourier basis functions than for the Bartlett kernel estimator. This is consistent with previous findings in Lazarus, Lewis, and Stock (2019) in the exactly identified regression

³We note that, as is well known, minimizing the AMSE is equivalent to minimizing bias²+ variance for the LRV estimator, can be different from the paper's primary purpose, namely, tests for the structural parameter vector. This is because minimizing the MSE leads to an underweighting of the bias that we address in this paper. Alternatively, we can choose a larger smoothing parameter based on testing-oriented criteria. For example, Sun and Phillips (2009) show that larger smoothing parameters are justified from the standpoint of reducing size distortions in the kernel LRV estimation. The same caution can be applied to the choice of AMSE-optimal K^* . We thank an anonymous referee for pointing this out.

context.

5.2 Hypothesis testing

We consider the following null hypothesis of interest,

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, \quad (23)$$

where the number of restricted parameters in R is $p = 3$ and the nominal significance level α is 5%. Given its superior performance to Bartlett LRV in point estimations, we consider only test statistics using the OS LRV with Fourier basis functions. Under (23), we study the empirical rejection probability (ERP) of the uncorrected Wald statistic, $F(\hat{\theta}_2)$, and the finite-sample corrected Wald statistic, $F_c^{\text{adj}}(\hat{\theta}_2)$, which use chi-square critical values derived from the conventional increasing-smoothing asymptotics. We examine the uncorrected Wald statistic with the J -statistic modification, $\tilde{F}(\hat{\theta}_2)$, as in Hwang and Sun (2017). We also examine $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, which is a finite-sample corrected version of $\tilde{F}(\hat{\theta}_2)$. Both $\tilde{F}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ use asymptotic F critical values which are derived under the fixed smoothing asymptotics. The same Wald test statistics using the iterated GMM estimator, $F(\hat{\theta}_{\text{iter}}^\infty)$, $F_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$, $\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$, and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$, are considered. The results are reported in Tables 3 and 4.

We first find that two-step GMM tests based on the uncorrected variance estimates and chi-square critical values, $F(\hat{\theta}_2)$, suffer from severe size distortions on finite samples. For example, when $T = 100$ and $\rho = 0.50$, Table 3 reports that the ERPs of $F(\hat{\theta}_2)$ are around 20%–43%, and these size distortions increase up to 58%–87% when ρ increases to 0.90, as shown in Table 4. As we point out, one possible reason for the failure of the chi-square test is the difference in behavior between the asymptotic variance estimate and the actual finite-sample variance of the two-step GMM estimators. That difference can be reflected in the corrected variance estimates that we provide. When the corrected version of the test statistic, $F_c^{\text{adj}}(\hat{\theta}_2)$, is used, it can help reduce the finite-sample size distortions. For example, the ERPs of $F(\hat{\theta}_2)$ when $\rho = 0.50$ and $T = 100$ are reduced to 17%–27%, as shown Table 3. When $\rho = 0.90$ and $T = 100$, the ERPs of $F(\hat{\theta}_2)$ are reduced to 25%–44%, as shown Table 4. Lastly, the results in Tables 3 and 4 also indicate that the size distortions become larger as the degree of over-identification q increases.

Although our simulation results suggest that the finite-sample variance correction can improve the Wald inference, there are still significant size distortions for $F_c^{\text{adj}}(\hat{\theta}_2)$, as shown in Tables 3 and 4, which indicate the limitations of chi-square tests. This is because the chi-square critical value from the increasing-smoothing asymptotics cannot capture the estimation uncertainty in the nonparametric weight matrix $S_T(\theta_0)$. This is why we employ F critical values using the J -statistic-modified Wald statistic driven by the fixed-smoothing asymptotics and obtain a further improvement in the finite-sample inference. The results are provided in Tables 3 and 4. We first observe that the size distortions of all testing procedures are substantially reduced. For example, the ERPs of $\tilde{F}(\hat{\theta}_2)$ are reported to be around 9% when $T = 100$ and $\rho = 0.50$. Thus the F tests clearly reduce the finite-sample size distortion from previous chi-square tests by 23% on average. This agrees with the previous literature, such as Hwang and Sun (2017) and Sun (2011, 2013), which highlights the accuracy of the fixed-smoothing asymptotics using OS LRV.

We find that the two-step GMM test with the corrected variance estimates, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, proposed in this paper can further reduce the finite-sample size distortions, which is shown in Tables 3 and 4. For example, when $T = 100$ and $\rho = 0.50$, the 8%–10% size distortions of $\tilde{F}(\hat{\theta}_2)$ are reduced to 4%–8% after the corrected variance estimates are used in $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$. When $T = 100$

and $\rho = 0.90$, the 12%–21% size distortions of $\tilde{F}(\hat{\theta}_2)$ are reduced to 3%–14%. The other results in Tables 3 and 4 exhibit similar quantitative and qualitative interpretations for the iterated GMM when the value of ρ is small and $q < 5$. When ρ is high or q increases, which induces a smaller choice of the optimal K_{MSE} , we find that the performance of the iterated GMM test, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$, is dominated by that of the two-step GMM test, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$. We note that this finding does not support the first-order equivalence of the two-step and iterated procedures in GMM.

In sum, our numerical findings for two-step GMM are consistent with the theoretical results developed in this paper, which indicate that the $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ procedure is able to further refine the fixed-smoothing asymptotics by capturing the initial estimation uncertainty from the non-parametric LRV estimator. It is interesting to note that the amount of finite-sample improvement using $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ is increasing in the degree of over-identification q . However, we find that using our corrected formula, $\widehat{\text{var}}_c^{\text{adj}}(\hat{\theta}_2)$, can lead to some conservatism for certain parameter constellations, for example when $\rho = 0.9$ and $q = 5$.

Tables 3 and 4 explore finite-sample distributions of $\hat{\theta}_2$ and $\hat{\theta}_{\text{CU}}$ in a numerical dimension by investigating the performances of the corrected two-step GMM test and the uncorrected CU-GMM test. When the degree of persistence is mild, for example when $\rho \in \{0.3, 0.5\}$, the CU-GMM test, $\tilde{F}(\hat{\theta}_{\text{CU}})$, which uses F critical values and the J -statistic modification, behaves similarly to the uncorrected F -test using two-step GMM, $\tilde{F}(\hat{\theta}_2)$. However, $\tilde{F}(\hat{\theta}_{\text{CU}})$ is more size-distorted than $\tilde{F}(\hat{\theta}_2)$ when the degree of over-identification, q , increases to 5. In all of these cases, the finite-sample corrected two-step test $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ performs better than the uncorrected $\tilde{F}(\hat{\theta}_{\text{CU}})$. We also find that the ERPs of $\tilde{F}(\hat{\theta}_{\text{CU}})$ perform its best on finite samples when the degree of over-identification q is equal to 1, that is, when the model is close to being exactly identified.

Lastly, we point out that more than half of reductions in empirical size distortions for $F(\hat{\theta}_2)$ and $F(\hat{\theta}_{\text{iter}}^\infty)$ are driven by using F critical values with the J -statistic modification, as originally shown in Hwang and Sun (2017). However, the result is not unexpected in our asymptotic theory, because under our asymptotics the order of the finite-sample correction is $O_p(T^{-1/2})$. Our finite-sample correction is designed to target the smaller order of error still embodied in the ERPs of the J -statistic-modified F test of Hwang and Sun (2017). Our simulation results in Tables 3 and 4 successfully show that the error of order $O_p(T^{-1/2})$ could be further reduced with the finite-sample corrected statistics proposed in this paper.

5.3 Weak identification

GMM inference often suffers from identification issues as well; see, for example, the July 1996 special issue of the *Journal of Business & Economic Statistics*. Stock and Wright (2000) point out that the weak identification of the GMM parameters can be another source of the poor finite-sample properties of HAR tests. In this subsection, we use Monte Carlo simulations to explore the issue of weak identification and to investigate how our proposed tests fare relative to their alternatives when the identification in (22) is weakened. Namely, we alter our previous DGP by premultiplying the first two summands in the first stage regression (21) by a constant $\pi = 0.10$, that is,

$$x_{j,t} = \pi \left(z_{j,t} + \sum_{i=d-1}^{m-1} z_{i,t} \right) + \epsilon_{x_{j,t}} \text{ for } j \in \{1, 2, 3\}. \quad (24)$$

From simple algebra, we can derive that the value of R -squared is given by $(q+1)/((q+1)+1/\pi^2)$. With our choices of $q \in \{1, 3, 5\}$ and $\pi = 0.10$, the set of R -squared values corresponds to $\{0.0196,$

0.0385, 0.0566}. We consider the degree of endogeneity, ψ , in $\{0.5, 0.9\}$. All other DGPs are the same as in previous sections. To save space, we report results only for $\rho \in \{0.70, 0.90\}$.

The simulation results in Table 5 show that the presence of weak identification affects the performance of all the tests considered. To be specific, with a mild endogeneity degree $\psi = 0.5$, the corrected F test, $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$, becomes more under-sized in most cases, but over-sized when ρ increases to 0.90, $q \in \{1, 3\}$, and $T = 100$. When the degree of endogeneity, ψ , increases to 0.90, however, we find that both the uncorrected and corrected F tests are severely size distorted in finite samples. For example, with $\rho = 0.90$, $q = 3$, and $T = 200$, the ERP of the corrected F tests is about 19%. Nevertheless, we find that our finite-sample correction remains effective in reducing a large amount of the size distortion in the uncorrected F tests. We also find that the uncorrected CU-GMM test with the J -statistic modification and F critical values, $\tilde{F}(\hat{\theta}_{\text{CU}})$, performs similarly to the uncorrected F test in most cases, but it becomes more over-sized as q increases. When ψ is 0.90, the uncorrected CU-GMM is more size-distorted than when $\psi = 0.50$, which is also the case for the uncorrected two-step GMM test $\tilde{F}(\hat{\theta}_2)$.

Summing up, our finite-sample evidence indicates that in cases of both strong and weak identification, the uncorrected CU-GMM performs best when the degree of over-identification is equal to 1. However, the performance of the uncorrected CU-GMM test has no advantage over the uncorrected two-step GMM test. Also, the finite-sample corrected two-step test performs better than the uncorrected CU-GMM test.

6 Conclusion

We develop an improved heteroskedasticity autocorrelated robust (HAR) inference that uses a finite-sample bias-corrected asymptotic variance estimate for the efficient GMM estimator in time series. We extend Windmeijer's (2005) approach to the time-series setting by explicitly considering the non-parametric LRV estimator in the bias-corrected variance formula. We formally show the consistency of the finite-sample corrected variance estimate when the smoothing parameter in the LRV estimator is fixed or is increasing with respect to the sample size. After the correction, our eigenvalue adjustments ensure that the corrected Wald test statistic is immune to the potential side effect of Windmeijer's formula which can introduce an additional type-I error.

With our finite-sample corrected variance estimator, this paper constructs t and Wald tests using the fixed-smoothing asymptotics. The standard t and F limiting distributions provide a convenient solution to the efficient GMM inference problem. We also provide an analytical expansion of the CU-GMM estimator and point out that the standard Wald inference using the CU-GMM cannot reflect potentially significant finite-sample variations embodied in the non-linear CU estimation.

Our Monte Carlo results show that the asymptotic t and F tests developed in this paper reduce the empirical size distortions compared to those in the existing tests in Sun (2014b) and Hwang and Sun (2017). Our findings show that the amount of size improvement increases as the degree of over-identification increases or as the time-series dependence increases. Also, we find that the finite-sample corrected two-step test outperforms the uncorrected CU-GMM test on finite samples.

Table 1: Finite-sample performance of GMM estimators and asymptotic variance estimates using Bartlett LRV with AR(1) coefficient $\rho \in \{0.30, 0.50, 0.70, 0.90\}$

	Point estimations using Bartlett LRV											
	AR(1) coefficient $\rho=0.30$				AR(1) coefficient $\rho=0.50$				AR(1) coefficient $\rho=0.90$			
	q=1	q=3	q=5	q=1	q=3	q=5	q=1	q=3	q=5	q=1	q=3	q=5
$\hat{\theta}_1$	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
	0.0011	0.0015	0.0012	0.0015	0.0009	-0.0004	0.0015	0.0018	0.0016	0.0016	0.0003	0.0001
$var(\hat{\theta}_1)$	0.0174	0.0088	0.0168	0.0081	0.0162	0.0081	0.0244	0.0122	0.0231	0.0113	0.0221	0.0111
$\widehat{var}(\hat{\theta}_2)$	0.0141	0.0075	0.0135	0.0072	0.0131	0.0070	0.0170	0.0094	0.0160	0.0089	0.0153	0.0087
$\hat{\theta}_2$	0.0013	0.0018	0.0018	0.0020	0.0008	-0.0006	0.0014	0.0019	0.0025	0.0019	-0.0008	-0.0002
$var(\hat{\theta}_2)$	0.0181	0.0089	0.0185	0.0086	0.0193	0.0089	0.0254	0.0125	0.0258	0.0122	0.0269	0.0125
$\widehat{var}(\hat{\theta}_2)$	0.0137	0.0073	0.0120	0.0067	0.0107	0.0063	0.0162	0.0091	0.0136	0.0081	0.0117	0.0074
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0146	0.0076	0.0139	0.0073	0.0133	0.0071	0.0179	0.0096	0.0169	0.0091	0.0161	0.0088
$\hat{\theta}_{\text{iter}}^\infty$	0.0015	0.0017	0.0014	0.0019	0.0014	-0.0007	0.0018	0.0017	0.0019	0.0016	-0.0005	0.0001
$var(\hat{\theta}_{\text{iter}}^\infty)$	0.0181	0.0089	0.0188	0.0086	0.0200	0.0090	0.0255	0.0125	0.0264	0.0123	0.0284	0.0128
$\widehat{var}(\hat{\theta}_{\text{iter}}^\infty)$	0.0137	0.0073	0.0121	0.0067	0.0107	0.0063	0.0162	0.0091	0.0137	0.0081	0.0117	0.0074
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.0147	0.0076	0.0142	0.0073	0.0139	0.0072	0.0181	0.0097	0.0173	0.0093	0.0167	0.0090
$\hat{\theta}_1$	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005	0.0090	0.0037	-0.0026	0.0038	0.0049	0.0007
$var(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194	0.2309	0.0747	0.1384	0.0624	0.0981	0.0577
$\widehat{var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127	0.0370	0.0284	0.0295	0.0245	0.0255	0.0225
$\hat{\theta}_2$	0.0039	0.0030	-0.0012	0.0016	0.0004	0.0014	0.0081	0.0065	-0.0004	0.0002	0.0069	-0.0004
$var(\hat{\theta}_2)$	0.0463	0.0222	0.0453	0.0223	0.0465	0.0226	0.2434	0.0790	0.1543	0.0732	0.1222	0.0719
$\widehat{var}(\hat{\theta}_2)$	0.0214	0.0133	0.0166	0.0111	0.0132	0.0096	0.0322	0.0252	0.0199	0.0174	0.0138	0.0136
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0254	0.0147	0.0241	0.0138	0.0229	0.0130	0.0474	0.0327	0.0444	0.0305	0.0445	0.0305
$\theta_{\text{it}}^\infty$	0.0028	0.0031	-0.0010	0.0011	0.0012	0.0006	0.0098	0.0058	-0.0019	-0.0000	0.0114	0.0014
$var(\hat{\theta}_{\text{iter}}^\infty)$	0.0469	0.0223	0.0473	0.0227	0.0508	0.0236	0.2469	0.0806	0.1584	0.0765	0.1290	0.0772
$\widehat{var}(\hat{\theta}_{\text{iter}}^\infty)$	0.0214	0.0133	0.0165	0.0111	0.0131	0.0096	0.0316	0.0252	0.0189	0.0172	0.0126	0.0127
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.0258	0.0148	0.0243	0.0141	0.0218	0.0136	0.0451	0.0329	0.0335	0.0282	0.0246	0.0242

Note: $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_{\text{iter}}^\infty$ indicate median values of one-step, two-step, and iterated GMM estimates, respectively, for the first component of the regression coefficient β_1 in $\theta = (\alpha, \beta_1, \beta_2, \beta_3)'$. The true value of β_1 is 0. $var(\hat{\theta}_1)$, $var(\hat{\theta}_2)$, and $var(\hat{\theta}_{\text{iter}}^\infty)$ indicate finite-sample variances of the corresponding GMM estimators for β_1 when $T \in \{100, 200\}$. $\widehat{var}(\hat{\theta}_1)$, $\widehat{var}(\hat{\theta}_2)$, and $\widehat{var}(\hat{\theta}_{\text{iter}}^\infty)$ are median values of the uncorrected sandwich variance estimates for the corresponding GMM estimators. $\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$ and $\widehat{var}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$ are median values of finite-sample corrected variance estimates for corresponding GMM estimators. The number of replications is 10,000.

Table 2: Finite-sample performance of GMM estimators and asymptotic variance estimates using OS LRV with AR(1) coefficient $\rho \in \{0.30, 0.50, 0.70, 0.90\}$

Point estimations using OS LRV												
	AR(1) coefficient $\rho=0.30$				AR(1) coefficient $\rho=0.50$				AR(1) coefficient $\rho=0.90$			
	q=1		q=3		q=5		q=1		q=3		q=5	
	T=100	T=200	T=100	T=200	T=100	T=200	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	0.0015	0.0018	0.0016	0.0016	0.0003	0.0001
$\widehat{var}(\hat{\theta}_1)$	0.0174	0.0088	0.0168	0.0081	0.0162	0.0081	0.0244	0.0122	0.0231	0.0113	0.0221	0.0111
$\widehat{var}(\hat{\theta}_1)$	0.0141	0.0075	0.0135	0.0072	0.0131	0.0070	0.0170	0.0094	0.0160	0.0089	0.0153	0.0087
$\hat{\theta}_2$	0.0016	0.0017	0.0023	0.0019	0.0019	-0.0005	0.0018	0.0018	0.0012	0.0026	-0.0008	-0.0010
$\widehat{var}(\hat{\theta}_2)$	0.0183	0.0089	0.0192	0.0087	0.0209	0.0091	0.0259	0.0125	0.0277	0.0125	0.0302	0.0132
$\widehat{var}(\hat{\theta}_2)$	0.0142	0.0075	0.0122	0.0069	0.0105	0.0064	0.0173	0.0097	0.0139	0.0084	0.0113	0.0076
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0152	0.0078	0.0142	0.0074	0.0132	0.0071	0.0192	0.0103	0.0176	0.0095	0.0161	0.0090
$\hat{\theta}_{\text{iter}}^\infty$	0.0016	0.0017	0.0024	0.0020	0.0029	-0.0005	0.0027	0.0018	0.0011	0.0028	-0.0000	-0.0008
$\widehat{var}(\hat{\theta}_{\text{iter}}^\infty)$	0.0183	0.0089	0.0196	0.0088	0.0219	0.0092	0.0261	0.0125	0.0286	0.0126	0.0321	0.0135
$\widehat{var}(\hat{\theta}_{\text{iter}}^\infty)$	0.0143	0.0075	0.0123	0.0069	0.0106	0.0064	0.0174	0.0097	0.0140	0.0085	0.0114	0.0076
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.0154	0.0078	0.0144	0.0075	0.0135	0.0072	0.0194	0.0103	0.0177	0.0096	0.0158	0.0091
AR(1) coefficient $\rho=0.70$												
	q=1		q=3		q=5		q=1		q=3		q=5	
	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
	0.0048	0.0030	-0.0007	-0.0003	0.0029	-0.0005	0.0090	0.0037	-0.0026	0.0038	0.0049	0.0007
$\widehat{var}(\hat{\theta}_1)$	0.0440	0.0213	0.0398	0.0199	0.0372	0.0194	0.2309	0.0747	0.1384	0.0624	0.0981	0.0577
$\widehat{var}(\hat{\theta}_1)$	0.0231	0.0140	0.0212	0.0132	0.0196	0.0127	0.0370	0.0284	0.0295	0.0245	0.0255	0.0225
$\hat{\theta}_2$	0.0029	0.0023	-0.0011	0.0024	-0.0013	0.0031	0.0068	0.0073	-0.0015	0.0025	0.0076	-0.0012
$\widehat{var}(\hat{\theta}_2)$	0.0478	0.0226	0.0511	0.0236	0.0593	0.0249	0.2396	0.0819	0.1684	0.0853	0.1446	0.0889
$\widehat{var}(\hat{\theta}_2)$	0.0237	0.0146	0.0168	0.0117	0.0119	0.0098	0.0394	0.0290	0.0207	0.0180	0.0134	0.0126
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_2)$	0.0283	0.0161	0.0257	0.0146	0.0238	0.0134	0.0555	0.0379	0.0498	0.0338	0.0496	0.0341
$\hat{\theta}_{\text{it}}^\infty$	0.0020	0.0020	0.0004	0.0019	0.0014	0.0011	0.0103	0.0059	-0.0029	0.0001	0.0084	-0.0006
$\widehat{var}(\hat{\theta}_{\text{it}}^\infty)$	0.0484	0.0226	0.0530	0.0242	0.0608	0.0261	0.2423	0.0831	0.1717	0.0852	0.1421	0.0908
$\widehat{var}(\hat{\theta}_{\text{it}}^\infty)$	0.0237	0.0146	0.0168	0.0118	0.0119	0.0099	0.0386	0.0290	0.0195	0.0176	0.0120	0.0118
$\widehat{var}_c^{\text{adj}}(\hat{\theta}_{\text{it}}^\infty)$	0.0286	0.0162	0.0241	0.0147	0.0182	0.0134	0.0506	0.0368	0.0300	0.0266	0.0186	0.0189

See note in Table 1.

Table 3: Empirical rejection probabilities of two-step, iterated, and continuously updating (CU) GMM tests using OS LRV at nominal level $\alpha = 0.05$ with AR(1) coefficient $\rho \in \{0.30, 0.50\}$

HAR inferences using OS LRV at $\alpha = 0.05$																
Increasing-smoothing asymptotics with chi-square critical values																
	AR(1) coefficient $\rho=0.30$				AR(1) coefficient $\rho=0.50$											
	q=1			q=3	q=5			q=1			q=3			q=5		
	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$F(\hat{\theta}_2)$	0.1437	0.1013	0.2098	0.1253	0.3025	0.1645	0.2092	0.1361	0.3097	0.1724	0.4243	0.2379				
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.1206	0.0915	0.1549	0.1013	0.2035	0.1227	0.1765	0.1191	0.2195	0.1318	0.2684	0.1731				
$F(\hat{\theta}_{\text{iter}}^\infty)$	0.1439	0.1015	0.2131	0.1263	0.3165	0.1679	0.2093	0.1369	0.3191	0.1745	0.4415	0.2444				
$F_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.1209	0.0908	0.1571	0.1005	0.2181	0.1228	0.1769	0.1184	0.2347	0.1322	0.3120	0.1766				
$F(\hat{\theta}_{\text{cu}})$	0.1409	0.0996	0.2017	0.1204	0.2992	0.1573	0.2067	0.1365	0.3089	0.1681	0.4355	0.2297				

Fixed-smoothing asymptotics with F-critical values and J-statistic modifications																
AR(1) coefficient $\rho=0.30$																
	q=1			q=3	q=5			q=1			q=3			q=5		
	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
	$\tilde{F}(\hat{\theta}_2)$	0.0736	0.0695	0.0764	0.0665	0.0798	0.0716	0.0959	0.0817	0.0953	0.0746	0.0926	0.0855			
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.0636	0.0629	0.0521	0.0507	0.0424	0.0484	0.0769	0.0711	0.0604	0.0538	0.0417	0.0509				
$\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$	0.0807	0.0733	0.1054	0.0807	0.1373	0.0985	0.1088	0.0894	0.1429	0.0975	0.1762	0.1276				
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.0676	0.0664	0.0712	0.0615	0.0828	0.0680	0.0861	0.0770	0.0924	0.0680	0.1020	0.0807				
$\tilde{F}(\hat{\theta}_{\text{cu}})$	0.0798	0.0729	0.0998	0.0779	0.1288	0.0914	0.1082	0.0888	0.1396	0.0932	0.1757	0.1197				

Note: $F(\hat{\theta}_2)$ and $F_c^{\text{adj}}(\hat{\theta}_2)$ indicate two-step GMM Wald tests for testing the null hypothesis in (23) with conventional chi-square critical values using uncorrected and corrected variance estimates, respectively. $\tilde{F}(\hat{\theta}_2)$ and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$ indicate two-step GMM Wald tests with alternative F critical values and J -statistic modifications using uncorrected and corrected variance estimates, respectively. $F(\hat{\theta}_{\text{iter}}^\infty)$, $\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$, and $\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$ are defined similarly using the iterated GMM estimator. $F(\hat{\theta}_{\text{cu}})$ and $\tilde{F}(\hat{\theta}_{\text{cu}})$ indicate uncorrected continuously updating GMM Wald tests with conventional chi-square critical values and alternative F critical values with J -statistic modifications, respectively. The number of replications is 10,000.

Table 4: Empirical rejection probabilities of two-step, iterated, and continuously updating (CU) GMM tests using OS LRV at nominal level $\alpha = 0.05$ with $\text{AR}(1)$ coefficient $\rho \in \{0.70, 0.90\}$

HAR inferences using OS LRV at $\alpha = 0.05$																	
Increasing-smoothing asymptotics with chi-square critical values																	
AR(1) coefficient $\rho=0.70$																	
q=1			q=3			q=5			q=1			q=3			q=5		
T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$F(\hat{\theta}_2)$	0.3287	0.2036	0.4943	0.2931	0.6365	0.4043	0.5816	0.4467	0.7827	0.6427	0.8677	0.7697					
$F_c^{\text{adj}}(\hat{\theta}_2)$	0.2622	0.1762	0.3139	0.2120	0.3070	0.2673	0.4379	0.3457	0.3592	0.3392	0.2538	0.2850					
$F(\hat{\theta}_{\text{iter}}^\infty)$	0.3324	0.2045	0.5082	0.2986	0.6543	0.4118	0.5889	0.4532	0.8015	0.6529	0.8865	0.7888					
$F_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.2729	0.1762	0.3905	0.2211	0.5162	0.2962	0.5099	0.3794	0.6989	0.5274	0.8031	0.6638					
$F(\hat{\theta}_{\text{cu}})$	0.3310	0.2027	0.5098	0.2926	0.6706	0.4054	0.5641	0.4539	0.7778	0.6578	0.8643	0.7972					

Fixed-smoothing asymptotics with F-critical values and J-statistic modifications																	
AR(1) coefficient $\rho=0.70$																	
q=1			q=3			q=5			q=1			q=3			q=5		
T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$\tilde{F}(\hat{\theta}_2)$	0.1221	0.1029	0.1255	0.0956	0.1068	0.0993	0.2073	0.1548	0.1742	0.1411	0.1224	0.1178					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.0924	0.0828	0.0611	0.0617	0.0359	0.0509	0.1408	0.1078	0.0684	0.0591	0.0250	0.0298					
$\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$	0.1420	0.1132	0.1943	0.1337	0.2308	0.1769	0.2598	0.1844	0.3306	0.2412	0.3586	0.2852					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.1089	0.0933	0.1249	0.0892	0.1364	0.1059	0.2058	0.1394	0.2277	0.1593	0.2496	0.1800					
$\tilde{F}(\hat{\theta}_{\text{cu}})$	0.1414	0.1119	0.1913	0.1334	0.2274	0.1734	0.2323	0.1784	0.2648	0.2173	0.2720	0.2500					

See note in Table 3.

Table 5: Empirical rejection probabilities of two-step, iterated, and continuously updating (CU) GMM tests using OS LRV at nominal level $\alpha = 0.05$ with AR(1) coefficient $\rho \in \{0.70, 0.90\}$ under weak identification, with $\pi = 0.10$, and degree of endogeneity $\psi \in \{0.50, 0.90\}$

HAR inferences using OS LRV at $\alpha = 0.05$ under weak identification ($\pi=0.10$)																	
Fixed-smoothing asymptotics with F-critical values and J-statistic modifications																	
Degree of endogeneity $\psi=0.50$																	
AR(1) coefficient $\rho=0.70$																	
q=1			q=3			q=5			q=1			q=3			q=5		
T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$\tilde{F}(\hat{\theta}_2)$	0.0746	0.0496	0.1196	0.0731	0.1269	0.0916	0.1757	0.1062	0.2219	0.1611	0.1892	0.1536					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.0430	0.0327	0.0533	0.0361	0.0361	0.0364	0.1111	0.0652	0.0882	0.0621	0.0489	0.0409					
$\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$	0.0833	0.0540	0.1777	0.1001	0.2393	0.1511	0.2064	0.1247	0.3576	0.2464	0.4340	0.3201					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.0567	0.0374	0.1091	0.0567	0.1400	0.0855	0.1598	0.0868	0.2562	0.1572	0.3179	0.2060					
$\tilde{F}(\hat{\theta}_{\text{cu}})$	0.0837	0.0588	0.1547	0.1080	0.2013	0.1537	0.1876	0.1150	0.2774	0.2020	0.3196	0.2496					

Fixed-smoothing asymptotics with F-critical values and J-statistic modifications																	
Degree of endogeneity $\psi=0.90$																	
AR(1) coefficient $\rho=0.90$																	
q=1			q=3			q=5			q=1			q=3			q=5		
T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$\tilde{F}(\hat{\theta}_2)$	0.3021	0.2030	0.3196	0.1997	0.2725	0.1851	0.4992	0.3922	0.5071	0.3893	0.4032	0.3207					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_2)$	0.2120	0.1456	0.1579	0.1078	0.0973	0.0802	0.3669	0.2824	0.2617	0.1874	0.1507	0.1085					
$\tilde{F}(\hat{\theta}_{\text{iter}}^\infty)$	0.3222	0.2147	0.3949	0.2360	0.4164	0.2568	0.5340	0.4172	0.6488	0.4976	0.6595	0.5169					
$\tilde{F}_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$	0.2549	0.1650	0.2758	0.1558	0.2753	0.1562	0.4553	0.3421	0.5277	0.3714	0.5398	0.3746					
$\tilde{F}(\hat{\theta}_{\text{cu}})$	0.2878	0.2110	0.3443	0.2741	0.3924	0.3296	0.4582	0.3643	0.5174	0.4109	0.5485	0.4450					

See note in Table 3.

Appendix A: Proofs of Main Results

Proof of Lemma 1. We prove the result for the general non-linear case. For each $j \in \{1, \dots, d\}$, we have

$$\begin{aligned} D(\hat{\theta}_2, S_T(\hat{\theta}_1))[\cdot, j] &= (A(\hat{\theta}_2, S_T(\hat{\theta}_1))^{-1} G_T(\hat{\theta}_2)' S_T^{-1}(\hat{\theta}_1) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_1} S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2) \\ &= (G_T'(\theta_0) S_T^{-1}(\theta_0) G_T(\theta_0))^{-1} G_T'(\theta_0) S_T^{-1}(\theta_0) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_1} S_T^{-1}(\theta_0) f_T(\hat{\theta}_2) (1 + o_p(1)), \end{aligned}$$

where the second equality holds by Assumptions 3 and 4, together with (6). Using the first-order Taylor expansion of $f_T(\hat{\theta}_2)$, we have

$$\begin{aligned} D(\hat{\theta}_2, S_T(\hat{\theta}_1))[\cdot, j] &= \{G_T'(\theta_0) S_T^{-1}(\theta_0) G_T(\theta_0)\}^{-1} G_T(\theta_0)' S_T^{-1}(\theta_0) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_1} S_T^{-1}(\theta_0) f_T(\theta_0) (1 + o_p(1)) \\ &\quad - \left\{ \{G_T'(\theta_0) S_T^{-1}(\theta_0) G_T(\theta_0)\}^{-1} G_T(\theta_0)' S_T^{-1}(\theta_0) \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_1} S_T^{-1}(\theta_0) G_T(\theta_0) \right. \\ &\quad \left. \times \{G_T'(\theta_0) S_T^{-1}(\theta_0) G_T(\theta_0)\}^{-1} G_T(\theta_0)' S_T^{-1}(\theta_0) f_T(\theta_0) \right\} (1 + o_p(1)) \end{aligned}$$

for each $j \in \{1, \dots, d\}$. For the term $\partial S_T(\theta)/\partial \theta_j|_{\theta=\hat{\theta}_1}$, recall that it is equal to $\Upsilon_j(\hat{\theta}_1) + \Upsilon_j'(\hat{\theta}_1)$. We want to show that $\Upsilon_j(\hat{\theta}_1) = \Upsilon_j(\theta_0)(1 + o_p(1))$. 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 $Q_h^*(r, s) = Q_h(r, s) - \int_0^1 Q_h(\tau_1, s) d\tau_1 - \int_0^1 Q_h(r, \tau_2) d\tau_2 + \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2$. We first consider the case in which h is fixed as $T \rightarrow \infty$. For some $\bar{\theta}_{T,j}^*$ for $j \in \{1, \dots, d\}$, and $\check{\theta}_T^*$, both of which are between $\hat{\theta}_1$ and θ_0 , we can write

$$\begin{aligned} \Upsilon_j^*(\hat{\theta}_1) &= \frac{1}{T} \sum_{s,t=1}^T Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) g_j(v_t, \theta_0) f(v_s, \theta_0)' \tag{25} \\ &\quad + \frac{1}{T} \sum_{s,t=1}^T Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right) \left(\frac{\partial g_j(v_t, \bar{\theta}_{T,j}^*)}{\partial \theta'} (\hat{\theta}_1 - \theta_0) \right) f(v_s, \theta_0)' \\ &\quad + \frac{1}{T} \sum_{s,t=1}^T Q_h^* \left(\frac{t}{T}, \frac{s}{T} \right)' \left(\frac{\partial f(v_t, \check{\theta}_T^*)}{\partial \theta'} (\hat{\theta}_1 - \theta_0) \right) g_j(v_s, \theta_0)' \\ &\quad + \frac{1}{T} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \left(\frac{\partial g_j(v_t, \bar{\theta}_{T,j}^*)}{\partial \theta'} (\hat{\theta}_1 - \theta_0) \right) \left[\frac{\partial f(v_\tau, \check{\theta}_T^*)}{\partial \theta'} (\hat{\theta}_1 - \theta_0) \right]' \\ &:= \Upsilon_j^*(\theta_0) + I_1 + I_2 + I_3, \end{aligned}$$

for each $j \in \{1, \dots, d\}$. Let us define $\epsilon_{j,t} = H_{t,j}(\bar{\theta}_{T,j}^*) - (t/T)H_j$ for each $j \in \{1, \dots, d\}$. By Assumption 4 and the definition of initial estimator $\hat{\theta}_1$, $\epsilon_{j,t}$ does not depend on h and satisfies

$\sup_{1 \leq t \leq T} \|\epsilon_{j,t}\| = o_p(1)$ for each $j \in \{1, \dots, d\}$. Setting $A_t = \frac{\partial g_j(v_t, \bar{\theta}_{T,j}^*)}{\partial \theta'}(\hat{\theta}_1 - \theta_0)$ and $B_\tau = u_\tau := f(v_\tau, \theta_0)$, we apply Lemma B.3 in Online Appendix B.4 and re-write I_1 as

$$\begin{aligned}
I_1 &= H_j(\hat{\theta}_1 - \theta_0) \frac{1}{T} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) u'_\tau + T Q_h^*(1, 1) \epsilon_{j,T} (\hat{\theta}_1 - \theta_0) \mathcal{S}'_T(u) \\
&\quad + T \sum_{\tau=1}^{T-1} \left[Q_h^* \left(1, \frac{\tau}{T} \right) - Q_h^* \left(1, \frac{\tau+1}{T} \right) \right] \epsilon_{j,T} (\hat{\theta}_1 - \theta_0) \mathcal{S}'_\tau(u) \\
&\quad + T \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \epsilon_{j,t} (\hat{\theta}_1 - \theta_0) \mathcal{S}'_T(u) \\
&\quad + T \sum_{\tau=1}^{T-1} \sum_{t=1}^{T-1} \nabla Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \epsilon_{j,t} (\hat{\theta}_1 - \theta_0) \mathcal{S}'_\tau(u) \\
&:= I_{11} + I_{12} + I_{13} + I_{14} + I_{15},
\end{aligned}$$

where $\mathcal{S}_\tau(u) = T^{-1} \sum_{s=1}^\tau u_s$ for $\tau = 1, \dots, T$. We want to show that I_{1i} for each $i \in \{1, \dots, 5\}$ is $o_p(1)$ as $T \rightarrow \infty$ holding h fixed. For I_{11} , there exists a finite $M > 0$ which does not depend on t such that

$$\begin{aligned}
\|I_{11}\| &= \left\| H_j \sqrt{T} (\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^T \left[\int_0^1 Q_h^* \left(r, \frac{\tau}{T} \right) dr + O \left(\frac{1}{T} \right) \right] u'_\tau \right\| \\
&\leq \frac{M}{T} \left\| \left(H_j \sqrt{T} (\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^T u'_\tau \right) \right\| = O_p \left(\frac{1}{T} \right) = o_p(1),
\end{aligned} \tag{26}$$

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

$$\begin{aligned}
I_{13} &= \epsilon_{j,T} \sqrt{T} (\hat{\theta}_1 - \theta_0) \left[\frac{1}{\sqrt{T}} \sum_{\tau=1}^T Q_h^* \left(1, \frac{\tau}{T} \right) u_\tau \right] - \epsilon_{j,T} \sqrt{T} (\hat{\theta}_1 - \theta_0) Q_h^*(1, 1) \mathcal{S}'_T(u) \\
&= o_p(1),
\end{aligned} \tag{27}$$

where the last equality follows from the boundedness of the function $Q_h^*(\cdot, \cdot)$, and that fact that $\epsilon_{j,T} = o_p(1)$. For I_{14} , we have that

$$\begin{aligned}
\|I_{14}\| &\leq \left\| \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \epsilon_{j,t} \right\| \left\| \sqrt{T} (\hat{\theta}_1 - \theta_0) \right\| \left\| \sqrt{T} \mathcal{S}'_T(u) \right\| \\
&\leq \sup_{1 \leq t \leq T} \|\epsilon_{j,t}\| \left\| \sum_{t=1}^{T-1} Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right\| \times O_p(1) \\
&= o_p(1) \left| Q_h^* \left(\frac{1}{T}, 1 \right) - Q_h^*(1, 1) \right| = o_p(1).
\end{aligned} \tag{28}$$

Lastly, for I_{15} , we have that

$$\|I_{15}\| \leq \sup_{1 \leq t \leq T} \|\epsilon_{j,t}\| \left\| \sum_{\tau=1}^{T-1} \left[\sum_{t=1}^{T-1} \nabla Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] \sqrt{T} (\hat{\theta}_1 - \theta_0) \sqrt{T} \mathcal{S}'_\tau(u) \right\|. \tag{29}$$

For each $\tau \in \{1, \dots, T\}$,

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

Using this result, we re-express the upper bound on $\|I_{15}\|$ in (29) by

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

where the first equality follows from the FCLT and the continuous mapping theorem. We have therefore shown 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. Using these results, we can conclude that

$$\Upsilon_j^*(\hat{\theta}_1) = \Upsilon_j^*(\theta_0)(1 + o_p(1)) \quad (30)$$

for each $j \in \{1, \dots, d\}$. Using the result in Lemma B.2 in Online Appendix B.4, we obtain

$$\Upsilon_j^*(\hat{\theta}_1) = \Upsilon_j(\hat{\theta}_1)(1 + o_p(1)) \text{ and } \Upsilon_j^*(\theta_0) = \Upsilon_j(\theta_0)(1 + o_p(1)), \quad (31)$$

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

$$D(\hat{\theta}_2, S_T(\hat{\theta}_1)) = D(\theta_0, S_T(\theta_0))(1 + o_p(1)), \quad (32)$$

which is the desired result.

Now, we consider the case where $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. We first want to check that the terms I_1 – I_3 are still $o_p(1)$ when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. For I_{11} , we have that

$$\begin{aligned} \|I_{11}\| &= \left\| H_j \sqrt{T}(\hat{\theta}_1 - \theta_0) \frac{1}{\sqrt{T}} \sum_{\tau=1}^T \left[\frac{1}{T} \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] u'_\tau \right\| \\ &= o(1) \left\| H_j \sqrt{T}(\hat{\theta}_1 - \theta_0) \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^T u'_\tau \right) \right\| = o_p(1), \end{aligned}$$

where the second equality follows from part (c) of Lemma B.1 in Online Appendix B.4. Also, we check that $I_{12} = o_p(1)$ since $\epsilon_{j,T} = o_p(1)$ does not depend on h . A careful inspection of (27)–(29) indicates that the terms I_{13} – I_{15} are also $o_p(1)$ because $Q_h^*(\cdot, \cdot)$ is uniformly bounded in h under Assumption 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. This concludes that the result in (30) also holds when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. From the result in Lemma B.2 in Online Appendix B.4, we also obtain that (31) also holds when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$, which leads to the desired result in (32). ■

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

$$\begin{aligned}\widehat{var}_c^{\text{inf}}(\hat{\theta}_2) &= \widehat{var}(\hat{\theta}_2) + D(\theta_0, S_T(\theta_0))\widehat{var}(\hat{\theta}_2) \\ &\quad + \widehat{var}(\hat{\theta}_2)D(\theta_0, S_T(\theta_0))' + D(\theta_0, S_T(\theta_0))\widehat{var}(\hat{\theta}_1)D(\theta_0, S_T(\theta_0))',\end{aligned}$$

with corresponding statistic $F_c^{\text{inf}}(\hat{\theta}_2) = (R\hat{\theta}_2 - r)'[R\widehat{var}_c^{\text{inf}}(\hat{\theta}_2)R']^{-1}(R\hat{\theta}_2 - r)/p$. Then (32) implies that $\widehat{var}_c(\hat{\theta}_2) = \widehat{var}_c^{\text{inf}}(\hat{\theta}_2)(1 + o_p(1))$. Thus the corresponding infeasible Wald statistic satisfies $F_c(\hat{\theta}_2) = F_c^{\text{inf}}(\hat{\theta}_2)(1 + o_p(1))$. Also, $D_{\theta_0, S_T(\theta_0)} = o_p(1)$ implies that $\widehat{var}_c^{\text{inf}}(\hat{\theta}_2) = \widehat{var}(\hat{\theta}_2)(1 + o_p(1))$, and this leads us to get

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

as desired. ■

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

$$\tilde{F}(\hat{\theta}_2) := \frac{K - p - q + 1}{K} \cdot \frac{F(\hat{\theta}_2)}{1 + \frac{1}{K}J(\hat{\theta}_2)},$$

and similarly define $\tilde{t}(\hat{\theta}_2)$ using $t(\hat{\theta}_2)$. Under Assumptions 2–5, we can apply Theorem 1 in Hwang and Sun (2017), and we obtain that $\tilde{t}(\hat{\theta}_2) \xrightarrow{d} t_{p, K-q}$ and $\tilde{F}(\hat{\theta}_2) \xrightarrow{d} \mathcal{F}_{p, K-p-q+1}$ for a fixed K and $T \rightarrow \infty$. By Theorem 2, we have that

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

and this gives us the desired results

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

■

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] 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.
- [4] 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.
- [5] 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.

- [6] Guggenberger, P. (2005). Monte-carlo evidence suggesting a no moment problem of the continuous updating estimator. *Economics Bulletin*, 3(13), 1–6.
- [7] Guggenberger, P. (2008). Finite sample evidence suggesting a heavy tail problem of the generalized empirical likelihood estimator. *Econometric Reviews*, 27(4–6), 526–541.
- [8] Guggenberger, P., & Smith, R. J. (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory*, 667–709.
- [9] Hall, P., & Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 891–916.
- [10] 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.
- [11] Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 1269–1286.
- [12] Hwang, J. (2021). Simple and trustworthy cluster-robust GMM inference. *Journal of Econometrics*, 222(2), 993–1023.
- [13] Hwang, J., Kang, B., & Lee, S. (2020). A doubly corrected robust variance estimator for linear GMM. Working paper, Department of Economics, University of Connecticut.
- [14] Hwang, J., & Sun, Y. (2017). Asymptotic F and t tests in an efficient GMM setting. *Journal of Econometrics*, 198(2), 277–295.
- [15] Kiefer, N. M., & Vogelsang, T. J. (2002). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70(5), 2093–2095.
- [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] Martínez-Iriarte, J., Sun, Y., & Wang, X. (2020). Asymptotic F tests under possibly weak identification. *Journal of Econometrics*, 218(1), 140–177.
- [21] Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, 141(2), 1331–1352.
- [22] Müller, U. K., & Watson, M. W. (2018). Long-run covariability. *Econometrica*, 86(3), 775–804.
- [23] Newey, W. K., & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1), 219–255.

- [24] 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.
- [25] Phillips, P. C. (2005). HAC estimation by automated regression. *Econometric Theory*, 116–142.
- [26] Pötscher, B. M., & Preinerstorfer, D. (2018). Controlling the size of autocorrelation robust tests. *Journal of Econometrics*, 207(2):406–431.
- [27] Pötscher, B. M., & Preinerstorfer, D. (2019). Further results on size and power of heteroskedasticity and autocorrelation robust tests, with an application to trend testing. *Electronic Journal of Statistics*, 13(2):3893–3942.
- [28] Preinerstorfer, D., & Pötscher, B. M. (2016). On size and power of heteroskedasticity and autocorrelation robust tests. *Econometric Theory*, 32(2):261–358.
- [29] Roodman, D. (2009). How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal*, 9(1), 86–136.
- [30] Stock, J. H., & Wright, J. H. (2000). GMM with weak identification. *Econometrica*, 68(5), 1055–1096.
- [31] Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, 164(2), 345–366.
- [32] Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using an orthonormal series variance estimator. *The Econometrics Journal*, 16(1), 1–26.
- [33] 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.
- [34] Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica*, 82(6), 2327–2370.
- [35] Sun, Y., & Phillips, P. C. B. (2009). Bandwidth choice for interval estimation in GMM regression. Working Paper. Department of Economics, UC San Diego.
- [36] Sun, Y., Phillips, P. C., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity–autocorrelation robust testing. *Econometrica*, 76(1), 175–194.
- [37] Windmeijer, F. (2000). A finite sample correction for the variance of linear efficient two-step GMM estimators. IFS Working paper W00/19, Institute for Fiscal Studies.
- [38] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1), 25–51.
- [39] Zhang, X. (2016). Fixed-smoothing asymptotics in the generalized empirical likelihood estimation framework. *Journal of Econometrics*, 193(1), 123–146.

Online Appendix B: “Finite-sample Corrected Inference for Two-Step GMM in Time Series”

Jungbin Hwang
University of Connecticut
Department of Economics

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

B.1 Finite-sample correction formula for non-linear moments

In the non-linear moment case, recall that the Taylor expansion of the FOC in (3), $\sqrt{T}(\hat{\theta}_2 - \theta_0)$, is expanded as:

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

where the matrix $A(\theta_0, S_T(\hat{\theta}_1))$ of second-order derivatives is given as

$$A(\theta_0, S_T(\hat{\theta}_1)) = G_T(\theta_0)' S_T^{-1}(\hat{\theta}_1) G_T(\theta_0) + H_T(\theta_0)' (I_d \otimes S_T^{-1}(\hat{\theta}_1) f_T(\theta_0))$$

and

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

Then we can formulate an estimator for the asymptotic variance of $\hat{\theta}_2$ as

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

Note that the form of $\widehat{\text{var}}(\hat{\theta}_2)$ in (4) is different from the standard asymptotic variance estimate

$(G_T'(\hat{\theta}_2) S_T^{-1}(\hat{\theta}_1) G_T(\hat{\theta}_2))^{-1}$. Keeping this term, however, could potentially improve the finite-sample performance of $\widehat{\text{var}}(\hat{\theta}_2)$, because the second term in the expression for $A(\hat{\theta}_2, S_T(\hat{\theta}_1))$, $H_T(\hat{\theta}_2)' (I_d \otimes S_T^{-1}(\hat{\theta}_1) f_T(\hat{\theta}_2))$, which is of stochastic order $O_p(T^{-1/2}) = o_p(1)$, is non-zero in a finite sample in over-identified GMM, (e.g., Windmeijer (2005)).

Similarly to the expansion in (8) and taking a further expansion of (B.1), we have that

$$\begin{aligned} \sqrt{T}(\hat{\theta}_2 - \theta_0) &= -(A(\theta_0, S_T(\theta_0)))^{-1} G_T(\theta_0)' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) \\ &\quad + D(\theta_0, S_T(\theta_0)) \sqrt{T}(\hat{\theta}_1 - \theta_0) + O_p \left(\frac{1}{\sqrt{T}} \right), \end{aligned} \quad (\text{B.2})$$

where

$$D(\theta_0, S_T(\theta_0)) = \left. \frac{\partial \left[-(A(\theta_0, S_T(\theta)))^{-1} G_T(\theta_0)' S_T^{-1}(\theta) f_T(\theta_0) \right]}{\partial \theta'} \right|_{\theta=\theta_0}$$

is a $d \times d$ matrix. The j -th column of $D(\theta_0, S_T(\theta_0))$ is expressed as

$$D(\theta_0, S_T(\theta_0))[., j] = (A(\theta_0, S_T(\theta_0)))^{-1} \left. \frac{\partial A(\theta_0, S_T(\theta))}{\partial \theta_j} \right|_{\theta=\theta_0} (A(\theta_0, S_T(\theta_0)))^{-1} \quad (\text{B.3})$$

$$\times G_T(\theta_0)' S_T^{-1}(\theta_0) f_T(\theta_0) + (A(\theta_0, S_T(\theta_0)))^{-1} G_T(\theta_0)' S_T^{-1}(\theta_0) \quad (\text{B.4})$$

$$\times \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} S_T^{-1}(\theta_0) f_T(\theta_0). \quad (\text{B.5})$$

From the FOC, the first term in (B.3) is always estimated as zero, so the feasible estimator for $D(\theta_0, S_T(\theta_0))[., j]$ is

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

where the formula for $\partial S_T(\theta)/\partial \theta_j|_{\theta=\hat{\theta}_1}$ is provided in (10). The finite-sample corrected formula is then given by

$$\widehat{\text{var}}_c(\hat{\theta}_2) = \widehat{\text{var}}(\hat{\theta}_2) + D(\hat{\theta}_2, S_T(\hat{\theta}_1)) \widehat{\text{var}}(\hat{\theta}_2) \\ + \widehat{\text{var}}(\hat{\theta}_2) D(\hat{\theta}_2, S_T(\hat{\theta}_1))' + D(\hat{\theta}_2, S_T(\hat{\theta}_1)) \widehat{\text{var}}(\hat{\theta}_1) D(\hat{\theta}_2, S_T(\hat{\theta}_1))',$$

where

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

Unlike the linear moment condition illustrated in Section 2, we note that there still exists a remainder term of order $O_p(T^{-1/2})$ in (B.2) which arises from the Jacobian function $G_T(\theta)$. The remainder term is of the same order as our correction term, $D(\theta_0, S_T(\theta_0))$. This implies that the effect of finite-sample improvement after taking account of $D(\hat{\theta}_2, S_T(\hat{\theta}_1))$ in our corrected formula can depend on its magnitude relative to that of the remainder term in (B.2). The same points are mentioned in Windmeijer (2005) and Hwang et al. (2020).

B.2 Iterated GMM

Let $\hat{\theta}_{\text{iter}}^0$ be the two-step estimator $\hat{\theta}_2$. For $j \geq 1$, the j -th iterated GMM estimator $\hat{\theta}_{\text{IE}}^j$ is defined as the solution to the following minimization problem:

$$\hat{\theta}_{\text{iter}}^j = \arg \min_{\theta \in \Theta} M \left(\theta, S_T \left(\hat{\theta}_{\text{iter}}^{j-1} \right) \right), \quad (\text{B.6})$$

where $M(\theta, S_T(\hat{\theta}_{\text{iter}}^{j-1})) = f_T(\theta)' S_T^{-1}(\hat{\theta}_{\text{iter}}^{j-1}) f_T(\theta)$. The asymptotic distribution of $\sqrt{T}(\hat{\theta}_{\text{iter}}^1 - \theta_0)$ can be represented as follows:

$$\sqrt{T}(\hat{\theta}_{\text{iter}}^1 - \theta_0) = - (G_T' S_T^{-1}(\theta_0) G_T)^{-1} G_T' S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) \quad (\text{B.7})$$

$$+ D_{\theta_0, S_T(\theta_0)} \sqrt{T}(\hat{\theta}_2 - \theta_0) + o_p \left(\frac{1}{\sqrt{T}} \right). \quad (\text{B.8})$$

Substituting the expansion in (8) into (B.8), we can represent the first iteration estimator as

$$\begin{aligned}\sqrt{T}(\hat{\theta}_{\text{iter}}^1 - \theta_0) &= - (I_d + D(\theta_0, S_T(\theta_0))) (G'_T S_T^{-1}(\theta_0) G_T)^{-1} G'_T S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) \\ &\quad + [D(\theta_0, S_T(\theta_0))]^2 \sqrt{T}(\hat{\theta}_1 - \theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

The leading term in $\sqrt{T}(\hat{\theta}_{\text{iter}}^1 - \theta_0)$ consists of an asymptotic normal distribution, part of which is scaled by $I_d + D(\theta_0, S_T(\theta_0))$. Also, 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:

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

When the number of iterations j goes to infinity, $\hat{\theta}_{\text{iter}}^j$ is expected to converge to a random variable $\hat{\theta}_{\text{iter}}^\infty$.¹ Then the impact of $\sqrt{T}(\hat{\theta}_1 - \theta_0)$ on $\sqrt{T}(\hat{\theta}_{\text{iter}}^j - \theta_0)$ through $[D(\theta_0, S_T(\theta_0))]^{j+1} = O_p(T^{-(j+1)/2})$ can be perfectly removed, and we have that

$$\sqrt{T}(\hat{\theta}_{\text{iter}}^\infty - \theta_0) = - (I_d - D(\theta_0, S_T(\theta_0)))^{-1} (G'_T S_T^{-1}(\theta_0) G_T)^{-1} G'_T S_T^{-1}(\theta_0) \sqrt{T} f_T(\theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right),$$

assuming that $I_d - D(\theta_0, S_T(\theta_0))$, which is $I_d + o_p(1)$, is invertible. The corrected variance estimate for $\hat{\theta}_{\text{iter}}^\infty$ is constructed as follows :

$$\widehat{\text{var}}_c(\hat{\theta}_{\text{iter}}^\infty) = \left(I_d - D(\hat{\theta}_{\text{iter}}^\infty, S_T(\hat{\theta}_{\text{iter}}^\infty)) \right)^{-1} \widehat{\text{var}}(\hat{\theta}_{\text{iter}}^\infty) \left(I_d - D(\hat{\theta}_{\text{iter}}^\infty, S_T(\hat{\theta}_{\text{iter}}^\infty)) \right)'^{-1},$$

where $\widehat{\text{var}}(\hat{\theta}_{\text{iter}}^\infty) = T^{-1} (G'_T S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) G_T)^{-1}$ is the standard sandwich variance formula. The corrected formula for $\hat{\theta}_{\text{iter}}^\infty$ extends that of Windmeijer (2000), which is formulated in an i.i.d. setting. The corrected Wald statistic is

$$F_c(\hat{\theta}_{\text{iter}}^\infty) = \frac{1}{p} \left(R \hat{\theta}_{\text{iter}}^\infty - r \right)' \left(R \widehat{\text{var}}_c(\hat{\theta}_{\text{iter}}^\infty) R' \right)^{-1} \left(R \hat{\theta}_{\text{iter}}^\infty - r \right).$$

Similarly, one can construct the corrected t statistic when $p = 1$. The asymptotic distribution of $F_c(\hat{\theta}_{\text{iter}}^\infty)$ can be characterized as

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

¹Hansen and Lee (2020) provide some regular conditions that guarantee that the loop of the iteration sequence, $\hat{\theta}_{\text{iter}}^j$ for $j = 1, 2, \dots$, is a contraction mapping, which implies that the iteration estimator $\hat{\theta}_{\text{iter}}^\infty$ is the fixed point.

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

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

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 $\mathbb{F}_{\text{iter}} = \mathbb{F} + o_p(1)$. Thus instead of approximating $F_c(\hat{\theta}_{\text{iter}}^\infty)$ by a conventional χ_d^2/p distribution, the standard t and F distributions shown in Theorem 3 and (18), together with the corrected variance estimate $\widehat{\text{var}}_c(\hat{\theta}_{\text{iter}}^\infty)$, the J -statistic modification, and the finite-sample adjustments in subsection 3.2, can be used to obtain asymptotic critical values for $t_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$ and $F_c^{\text{adj}}(\hat{\theta}_{\text{iter}}^\infty)$, respectively.

B.3 Finite-sample corrected formula for linear-IV model

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

$$\hat{\theta}_1 = (X'ZW_T^{-1}Z'X)^{-1} (X'ZW_T^{-1}Z'y),$$

and the corresponding asymptotic variance estimator is

$$\widehat{\text{var}}(\hat{\theta}_1) = T (X'ZW_T^{-1}Z'X)^{-1} (X'ZW_T^{-1}S_T(\hat{\theta}_1)W_T^{-1}Z'X) (X'ZW_T^{-1}Z'X)^{-1},$$

where

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

and $\epsilon_{y,t}(\hat{\theta}_1) = y_t - x_t' \hat{\theta}_1$. The efficient two-step GMM estimator is

$$\hat{\theta}_2 = \left(X'ZS_T^{-1}(\hat{\theta}_1)Z'X \right)^{-1} X'ZS_T^{-1}(\hat{\theta}_1)Z'y,$$

and the corresponding uncorrected sandwich variance estimator and the corrected variance estimator are

$$\widehat{\text{var}}(\hat{\theta}_2) = T \left(X'ZS_T^{-1}(\hat{\theta}_1)Z'X \right)^{-1}$$

and

$$\widehat{\text{var}}_c(\hat{\theta}_2) = \widehat{\text{var}}(\hat{\theta}_2) + D(\hat{\theta}_2, S_T(\hat{\theta}_1)) \widehat{\text{var}}(\hat{\theta}_2) + \widehat{\text{var}}(\hat{\theta}_2) D(\hat{\theta}_2, S_T(\hat{\theta}_1))' \\ + D(\hat{\theta}_2, S_T(\hat{\theta}_1)) \widehat{\text{var}}(\hat{\theta}_1) D(\hat{\theta}_2, S_T(\hat{\theta}_1))',$$

respectively, where the j -th column of $D(\hat{\theta}_2, S_T(\hat{\theta}_1))$ is given by

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

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

and $\epsilon_y(\hat{\theta}_1) = (\epsilon_{y,1}(\hat{\theta}_1), \dots, \epsilon_{y,T}(\hat{\theta}_1))'$. The formula for $\Upsilon_j(\hat{\theta}_1)$ is

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

The iterated GMM estimator, $\hat{\theta}_{\text{iter}}^\infty$, which repeats the loop of the iteration sequence in (B.6), is given by

$$\hat{\theta}_{\text{iter}}^\infty = \left(X' Z S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) Z' X \right)^{-1} X' Z S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) Z' y,$$

and the corresponding uncorrected sandwich variance estimator and the corrected variance estimator are

$$\widehat{\text{var}}_c \left(\hat{\theta}_{\text{iter}}^\infty \right) = \left(I_d - D \left(\hat{\theta}_{\text{iter}}^\infty, S_T(\hat{\theta}_{\text{iter}}^\infty) \right) \right)^{-1} \widehat{\text{var}} \left(\hat{\theta}_{\text{iter}}^\infty \right) \left(I_d - D \left(\hat{\theta}_{\text{iter}}^\infty, S_T(\hat{\theta}_{\text{iter}}^\infty) \right) \right)'^{-1}$$

and

$$\widehat{\text{var}} \left(\hat{\theta}_{\text{iter}}^\infty \right) = T \left(X' Z S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) Z' X \right)^{-1},$$

where

$$D \left(\hat{\theta}_{\text{iter}}^\infty, S_T(\hat{\theta}_{\text{iter}}^\infty) \right)[:, j] = - \left(X' Z S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) Z' X \right)^{-1} X' Z S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty)$$

$$\times \left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_{\text{iter}}^\infty} S_T^{-1}(\hat{\theta}_{\text{iter}}^\infty) Z' \epsilon_y(\hat{\theta}_{\text{iter}}^\infty),$$

$$\left. \frac{\partial S_T(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}_{\text{iter}}^\infty} = \Upsilon_j(\hat{\theta}_{\text{iter}}^\infty) + \Upsilon_j'(\hat{\theta}_{\text{iter}}^\infty).$$

B.4 Technical lemmas and proof of Theorem 4

Lemma B.1 *Under Assumption 1, together with $h \rightarrow \infty$ and $T \rightarrow \infty$ such that $h/T \rightarrow 0$, the following hold:*

- (a) $T^{-2} \sum_{\tau_1=1}^T \sum_{\tau_2=1}^T Q_h \left(\frac{\tau_1}{T}, \frac{\tau_2}{T} \right) - \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2 = o(1)$;
- (b) $\sup_{1 \leq t \leq T} \left\{ T^{-1} \sum_{\tau=1}^T Q_h \left(\frac{t}{T}, \frac{\tau}{T} \right) - \int_0^1 Q_h \left(\frac{t}{T}, \tau \right) d\tau \right\} = o(1)$;
- (c) $\sup_{1 \leq \tau \leq T} T^{-1} \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) = o(1)$, where

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

Proof of Lemma B.1. We start by proving the results for the case when $Q_h(r, s) = k((r-s)/b)$, where $h = 1/b$. Denote bT by B_T . For part (a),

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

since $B_T \rightarrow \infty$ such that $B_T/T \rightarrow 0$. By Assumption 1, $k((\tau_1 - \tau_2)/b) \rightarrow 0$ for almost all τ_1 and τ_2 , and this enables us to apply the 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

$$\begin{aligned}
&\sup_{1 \leq t \leq T} \frac{1}{T} \sum_{\tau=1}^T Q_h\left(\frac{t}{T}, \frac{\tau}{T}\right) - \int_0^1 Q_h\left(\frac{t}{T}, \tau_2\right) d\tau_2 \\
&= \sup_{1 \leq t \leq T} \left\{ \frac{1}{T} \sum_{\tau=1}^T k\left(\frac{t - \tau}{B_T}\right) - \int_0^1 k\left(\frac{t/T - \tau_2}{b}\right) d\tau_2 \right\} \\
&\leq \sup_{1 \leq t \leq T} \left(\frac{B_T}{T} \times \frac{1}{B_T} \sum_{\tau=1}^T k\left(\frac{t - \tau}{B_T}\right) \right) + \sup_{1 \leq t \leq T} \left| \int_0^1 k\left(\frac{t/T - \tau_2}{b}\right) d\tau_2 \right| \\
&\leq \frac{B_T}{T} \times \frac{1}{B_T} \sum_{j=-\infty}^{\infty} \left| k\left(\frac{j}{B_T}\right) \right| + \int_0^1 \sup_{0 \leq \tau_1 \leq 1} \left| k\left(\frac{\tau_1 - \tau_2}{b}\right) \right| d\tau_2 \tag{B.9} \\
&= o(1),
\end{aligned}$$

because the first term on the right-hand side of the first equality (B.9) is $O(B_T/T) = o(1)$. Also, Assumption 1 implies that $\sup_{0 \leq \tau_1 \leq 1} |k((\tau_1 - \tau_2)/b)| \rightarrow 0$ for almost all τ_2 , and this enables us to apply the dominated convergence theorem to get the second term on the right-hand side of the first equality in (B.9) to be $o(1)$ when $b = B_T/T \rightarrow 0$.

For part (c),

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

where the last equality follows from the proof of part (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 $K \rightarrow \infty$ such that $K/T \rightarrow 0$. Then the result of part (a) follows from

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

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

$$\begin{aligned}
& \sup_{1 \leq t \leq T} \left\{ \frac{1}{T} \sum_{\tau=1}^T Q_h \left(\frac{t}{T}, \frac{\tau}{T} \right) - \int_0^1 Q_h \left(\frac{t}{T}, \tau_2 \right) d\tau_2 \right\} \\
&= \sup_{1 \leq t \leq T} \left\{ \frac{1}{T} \sum_{\tau=1}^T \frac{1}{K} \sum_{j=1}^K \Phi_j \left(\frac{t}{T} \right) \Phi_j \left(\frac{\tau}{T} \right) - \int_0^1 \frac{1}{K} \sum_{j=1}^K \Phi_j \left(\frac{t}{T} \right) \Phi_j(\tau_2) d\tau_2 \right\} \\
&= \sup_{1 \leq t \leq T} \left\{ \frac{1}{K} \sum_{j=1}^K \Phi_j \left(\frac{t}{T} \right) \left(\frac{1}{T} \sum_{\tau=1}^T \Phi_j \left(\frac{\tau}{T} \right) \right) - \frac{1}{K} \sum_{j=1}^K \Phi_j \left(\frac{t}{T} \right) \underbrace{\int_0^1 \Phi_j(\tau_2) d\tau_2}_{=0} \right\} \\
&= \left\{ \sup_{1 \leq t \leq T} \frac{1}{K} \sum_{j=1}^K \Phi_j \left(\frac{t}{T} \right) \right\} \left(\int_0^1 \Phi_j(r) dr + O \left(\frac{1}{T} \right) \right) \\
&= O \left(\frac{1}{T} \right) = o(1).
\end{aligned}$$

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

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

■

Lemma B.2 *Let us define*

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

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

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

when h is fixed as $T \rightarrow \infty$, or when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$.

Proof of Lemma B.2. For each $j \in \{1, \dots, d\}$, we have that

$$\begin{aligned}
& \left\| \Upsilon_j^*(\hat{\theta}) - \Upsilon_j(\hat{\theta}) \right\| \tag{B.10} \\
&= \left\| \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e_T(s) g_j(v_t, \hat{\theta}) f(v_s, \hat{\theta})' \right] + \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e_T(t) g_j(v_t, \hat{\theta}) f(v_s, \hat{\theta})' \right] \right. \\
&\quad \left. - \left[\frac{1}{T^2} \sum_{\tau_1=1}^T \sum_{\tau_2=1}^T Q_h \left(\frac{\tau_1}{T}, \frac{\tau_2}{T} \right) - \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2 \right] \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T g_j(v_t, \hat{\theta}) f(v_s, \hat{\theta})' \right] \right\| \\
&\leq \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T g_j(v_t, \hat{\theta}) \right\|}_{:=A} \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T e_T(s) f(v_s, \hat{\theta}) \right\|}_{:=B} + \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T f(v_s, \hat{\theta}) \right\|}_{:=C} \underbrace{\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T e_T(t) g_j(v_t, \hat{\theta}) \right\|}_{:=D} \\
&\quad + \left\| \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T Q_h \left(\frac{t}{T}, \frac{s}{T} \right) - \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2 \right\| \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T g_j(v_t, \hat{\theta}) \right\| \left\| \frac{1}{\sqrt{T}} \sum_{s=1}^T f(v_s, \hat{\theta})' \right\|,
\end{aligned}$$

where

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

We first consider the case that h is fixed where $T \rightarrow \infty$. Note that by Assumption 1, $e_T(t) = O(1/T) = o(1)$ uniformly over t for fixed h . From the proof of Lemma 1-(a) in Sun (2014), we obtain

$$\frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T Q_h \left(\frac{t}{T}, \frac{s}{T} \right) - \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2 = O \left(\frac{1}{T} \right),$$

and

$$B = O_p \left(\frac{1}{T} \right) \text{ and } C = O_p(1), \tag{B.11}$$

as $T \rightarrow \infty$ such that h is fixed. For each $j \in \{1, \dots, d\}$, we apply the mean value theorem to term A (above) and obtain

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g_j(v_t, \check{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^T g_j(v_t, \theta_0) + H_{T,j}(\bar{\theta}_{T,j}^*) \sqrt{T}(\check{\theta}_T - \theta_0) = O_p(1) \tag{B.12}$$

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

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^T e_T(t) g_j(v_t, \hat{\theta}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^T e_T(t) g_j(v_t, \theta_0) + \frac{1}{T} \sum_{t=1}^T e_T(t) \frac{\partial g_j(v_t, \bar{\theta}_{T,j}^*)}{\partial \theta'} \sqrt{T}(\hat{\theta} - \theta_0) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^T e_T(t) g_j(v_t, \theta_0) + \frac{1}{T} \sum_{t=1}^{T-1} [e_T(t) - e_T(t+1)] \times H_{t,j}(\bar{\theta}_{T,j}^*) \sqrt{T}(\hat{\theta} - \theta_0) \\
&\quad + e_T(T) H_{T,j}(\bar{\theta}_{T,j}^*) \sqrt{T}(\hat{\theta} - \theta_0),
\end{aligned} \tag{B.13}$$

where the second equality follows from summation by parts. For the last term on the right-hand side of the second equality in (B.13),

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

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

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

by Assumption 5. Together with Markov's 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 $j \in \{1, \dots, d\}$. By Assumption 4, $\epsilon_{j,t}$ satisfies $\sup_{1 \leq t \leq T} \|\epsilon_{j,t}\| = o_p(1)$. The second term on the right-hand side of the second equality in (B.13) can be written as

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^{T-1} [e_T(t) - e_T(t+1)] H_{t,j}(\bar{\theta}_{T,j}^*) \sqrt{T}(\hat{\theta} - \theta_0) \\ &= \frac{1}{T} \sum_{t=1}^{T-1} [e_T(t) - e_T(t+1)] \epsilon_{j,t} \sqrt{T}(\hat{\theta} - \theta_0) + \frac{1}{T} \sum_{t=1}^{T-1} [e_T(t) - e_T(t+1)] \left(\frac{t}{T}\right) H_j \sqrt{T}(\hat{\theta} - \theta_0) \\ &= \frac{1}{T} \sum_{t=1}^{T-1} [e_T(t) - e_T(t+1)] \epsilon_{j,t} \sqrt{T}(\hat{\theta} - \theta_0) + \frac{1}{T^2} \sum_{t=1}^{T-1} e_T(t) H_j \sqrt{T}(\hat{\theta} - \theta_0) - \frac{e_T(T)}{T} H_j \sqrt{T}(\hat{\theta} - \theta_0) \\ &= O_p\left(\frac{1}{T}\right), \end{aligned}$$

where the last equality follows from $\sqrt{T}(\hat{\theta} - \theta_0) = O_p(1)$, $\sup_{1 \leq t \leq T} \epsilon_{j,t} = o_p(1)$, and $\sup_{1 \leq t \leq T} e_T(t) = O(1/T)$, which leads us to $D = O_p(1/T)$. Incorporating this, together with (B.11) and (B.12), into (B.10), we obtain

$$\begin{aligned} \Upsilon_j(\hat{\theta}) &= \Upsilon_j^*(\hat{\theta}) + O_p\left(\frac{1}{T}\right) \\ &= \Upsilon_j^*(\hat{\theta}) + o_p(1), \end{aligned} \tag{B.14}$$

which is the desired result.

Now we consider the case when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. Using parts (a) and (b) in Lemma B.1, we obtain

$$\frac{1}{T^2} \sum_{\tau_1=1}^T \sum_{\tau_2=1}^T Q_h \left(\frac{\tau_1}{T}, \frac{\tau_2}{T} \right) - \int_0^1 \int_0^1 Q_h(\tau_1, \tau_2) d\tau_1 d\tau_2 = o(1) \quad (\text{B.15})$$

$$\sup_{1 \leq t \leq T} e_T(t) = o(1) \quad (\text{B.16})$$

when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. Note that, compared to the case when h is fixed, the order of the convergence on the right-hand sides of (B.15) and (B.16) becomes $o(1)$ instead of $O(1/T)$. Using (B.15), one can show that the last term in (B.10) is $o_p(1)$ when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. Also, a careful inspection of the proof when h is fixed as $T \rightarrow \infty$ indicates (B.16) leads to the conclusion that

$$A = O_p(1) \text{ and } D = o_p(1)$$

hold when $h \rightarrow \infty$ such that $h/T \rightarrow 0$. Similarly, we can prove that $B = o_p(1)$ and $C = O_p(1)$ also hold when $h \rightarrow \infty$ such that $h/T \rightarrow 0$. Incorporating these results into (B.10), we can conclude that $\Upsilon_j(\hat{\theta}) = \Upsilon_j^*(\hat{\theta}) + o_p(1)$ also holds when $(h, T) \rightarrow \infty$ such that $h/T \rightarrow 0$. ■

Lemma B.3 For a generic sequence of matrices $\{C_t\}$, define $\mathcal{S}_0(C) = 0$ and $\mathcal{S}_t(C) = T^{-1} \sum_{s=1}^t C_s$ for $t \in \{0, 1, \dots, T\}$. Then for any two sequences of matrices $\{A_t\}$ and $\{B_t\}$,

$$\begin{aligned} & \frac{1}{T^2} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) A_t B'_\tau \\ &= \sum_{\tau=1}^{T-1} \sum_{t=1}^{T-1} \nabla Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \mathcal{S}_t(A) \mathcal{S}'_\tau(B) + \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \mathcal{S}_t(A) \mathcal{S}'_T(B)' \\ &+ \sum_{\tau=1}^{T-1} \left[Q_h^* \left(1, \frac{\tau}{T} \right) - Q_h^* \left(1, \frac{\tau+1}{T} \right) \right] \mathcal{S}_T(A) \mathcal{S}'_\tau(B) + Q_h^*(1, 1) \mathcal{S}_T(A) \mathcal{S}'_T(B), \end{aligned}$$

where

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

Proof of Lemma B.3. We use the formula for summation by parts:

$$\frac{1}{T} \sum_{t=1}^T a_t b'_t = \frac{1}{T} a_T C'_T - \frac{1}{T} \sum_{t=1}^{T-1} (a_{t+1} - a_t) C'_t, \text{ where } C_t = \sum_{s=1}^t b_s \quad (\text{B.17})$$

for any conformable vectors a_t and b_t . Consider

$$\frac{1}{T^2} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) A_t B'_\tau = \frac{1}{T} \sum_{\tau=1}^T \left(\frac{1}{T} \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) A_t \right) B'_\tau. \quad (\text{B.18})$$

We first apply the formula in (B.17) to the expression inside the parentheses by setting $a_t = Q_h^*(t/T, \tau/T)$, $b_t = A_t$, and $C_t = \sum_{s=1}^t A_s$. Then

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

for each $\tau \in \{1, \dots, T\}$. Incorporating this into (B.18), we have

$$\begin{aligned} & \frac{1}{T^2} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) A_t B'_\tau \\ &= \frac{1}{T} \sum_{\tau=1}^T Q_h^* \left(1, \frac{\tau}{T} \right) \mathcal{S}_T(A) B'_\tau - \sum_{t=1}^{T-1} \mathcal{S}_t(A) \left[\frac{1}{T} \sum_{\tau=1}^T \left\{ Q_h^* \left(\frac{t+1}{T}, \frac{\tau}{T} \right) - Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right\} B'_\tau \right]. \end{aligned} \quad (\text{B.19})$$

We repeatedly apply the formula in (B.17) to the terms on the right-hand side. For the first term, setting $a_\tau = Q_h^* \left(1, \frac{\tau}{T} \right) \mathcal{S}_T(A)$, $b_\tau = B'_\tau$, and $C_\tau = \sum_{s=1}^\tau B'_s$, we obtain

$$\begin{aligned} & \frac{1}{T} \sum_{\tau=1}^T Q_h^* \left(1, \frac{\tau}{T} \right) \mathcal{S}_T(A) B'_\tau \\ &= Q_h^*(1, 1) \mathcal{S}_T(A) \mathcal{S}_T(B)' - \sum_{\tau=1}^{T-1} Q_h^* \left(1, \frac{\tau+1}{T} \right) - Q_h^* \left(1, \frac{\tau}{T} \right) \mathcal{S}_T(A) \mathcal{S}_\tau(B)' \end{aligned}$$

Inside the braces in the second term, setting $a_\tau = Q_h^* \left(\frac{t+1}{T}, \frac{\tau}{T} \right) - Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right)$, $b_\tau = B'_\tau$, and $C_\tau = \sum_{s=1}^\tau B'_s$, we get

$$\begin{aligned} & \frac{1}{T} \sum_{\tau=1}^T \left\{ Q_h^* \left(\frac{t+1}{T}, \frac{\tau}{T} \right) - Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right\} B'_\tau \\ &= \left[Q_h^* \left(\frac{t+1}{T}, 1 \right) - Q_h^* \left(\frac{t}{T}, 1 \right) \right] \mathcal{S}_T(B)' - \sum_{\tau=1}^{T-1} \left[Q_h^* \left(\frac{t+1}{T}, \frac{\tau+1}{T} \right) - Q_h^* \left(\frac{t}{T}, \frac{\tau+1}{T} \right) \right. \\ & \quad \left. - Q_h^* \left(\frac{t+1}{T}, \frac{\tau}{T} \right) + Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] \mathcal{S}_\tau(B)'. \end{aligned}$$

Incorporating these results into (B.19), we obtain

$$\begin{aligned} & \frac{1}{T^2} \sum_{\tau=1}^T \sum_{t=1}^T Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) A_t B'_\tau \\ &= Q_h^*(1, 1) \mathcal{S}_T(A) \mathcal{S}_T(B)' - \sum_{\tau=1}^{T-1} \left[Q_h^* \left(1, \frac{\tau+1}{T} \right) - Q_h^* \left(1, \frac{\tau}{T} \right) \right] \mathcal{S}_T(A) \mathcal{S}_\tau(B)' \\ & \quad - \sum_{t=1}^{T-1} \mathcal{S}_t(A) \left[Q_h^* \left(\frac{t+1}{T}, 1 \right) - Q_h^* \left(\frac{t}{T}, 1 \right) \right] \mathcal{S}_T(B)' + \sum_{t=1}^{T-1} \mathcal{S}_t(A) \sum_{\tau=1}^{T-1} \left[Q_h^* \left(\frac{t+1}{T}, \frac{\tau+1}{T} \right) - Q_h^* \left(\frac{t}{T}, \frac{\tau+1}{T} \right) \right. \\ & \quad \left. - Q_h^* \left(\frac{t+1}{T}, \frac{\tau}{T} \right) + Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \right] \mathcal{S}_\tau(B)' \\ &= \sum_{t=1}^{T-1} \sum_{\tau=1}^{T-1} \nabla Q_h^* \left(\frac{t}{T}, \frac{\tau}{T} \right) \mathcal{S}_t(A) \mathcal{S}_\tau(B)' + \sum_{t=1}^{T-1} \left[Q_h^* \left(\frac{t}{T}, 1 \right) - Q_h^* \left(\frac{t+1}{T}, 1 \right) \right] \mathcal{S}_t(A) \mathcal{S}_T(B)' \\ & \quad + \sum_{\tau=1}^{T-1} \left[Q_h^* \left(1, \frac{\tau}{T} \right) - Q_h^* \left(1, \frac{\tau+1}{T} \right) \right] \mathcal{S}_T(A) \mathcal{S}_\tau(B)' + Q_h^*(1, 1) \mathcal{S}_T(A) \mathcal{S}_T(B)', \end{aligned}$$

as desired. ■

Lemma B.4 *Let Assumptions 1–3 hold with $\hat{\theta}_{CU}$. Then the CU-FOC function $Q_{\theta, S_T(\theta)}$ satisfies*

$$\left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_T^*} = \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_0} + o_p(1),$$

where θ_T^* is located between θ_0 and $\hat{\theta}_{CU}$.

Proof of Lemma B.4. We start by showing that $\partial Q(\theta, S_T(\theta))/\partial \theta'$ coincides with the definition of $\tilde{A}_1(\theta, S_T(\theta))$. We have that

$$\begin{aligned} \frac{\partial Q(\theta, S_T(\theta))}{\partial \theta'} &= \frac{\partial}{\partial \theta} \left(f_T(\theta)' S_T^{-1}(\theta) \tilde{G}_T(\theta) \right) \\ &= G_T' S_T^{-1}(\theta) \tilde{G}_T(\theta) - \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_1} \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_d} \end{bmatrix} S_T^{-1}(\theta) \tilde{G}_T(\theta) \\ &\quad + \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_1} \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_d} \end{bmatrix}. \end{aligned}$$

The right-hand side of the last equality can be re-written as

$$\begin{aligned} &\left(G_T'(\theta) - \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) (\Upsilon_1(\theta) + \Upsilon_1'(\theta)) \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) (\Upsilon_d(\theta) + \Upsilon_d'(\theta)) \end{bmatrix} \right) S_T^{-1}(\theta) \tilde{G}_T(\theta) \\ &+ \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) \left\{ \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_1} - \Upsilon_1(\theta) S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \right\} \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) \left\{ \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_d} - \Upsilon_d(\theta) S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \right\} \end{bmatrix} \\ &= [\tilde{G}_T(\theta_0)]' S_T^{-1}(\theta_0) [\tilde{G}_T(\theta_0)] + \{I_d \otimes (f_T(\theta_0)' S_T^{-1}(\theta_0))\} \tilde{H}(\theta_0) \\ &= \tilde{A}_1(\theta, S_T(\theta)). \end{aligned}$$

For each $i \in \{1, \dots, d\}$, we have that

$$\begin{aligned} \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} (G_T - \Upsilon'(\theta) \{I_d \otimes (S_T^{-1}(\theta) f_T(\theta))\}) \\ &= -\frac{\partial}{\partial \theta_i} [\Upsilon_1'(\theta) S_T^{-1}(\theta) f_T(\theta), \quad \dots, \quad \Upsilon_d'(\theta) S_T^{-1}(\theta) f_T(\theta)]. \end{aligned}$$

For any $(i, j) \in \{1, \dots, d\} \times \{1, \dots, d\}$, we have that

$$\begin{aligned} -\frac{\partial}{\partial \theta_i} (\Upsilon_j'(\theta) S_T^{-1}(\theta) f_T(\theta)) &= -\left(\frac{\partial}{\partial \theta_i} \Upsilon_j(\theta) \right)' S_T^{-1}(\theta) f_T(\theta) + \Upsilon_j'(\theta) S_T^{-1}(\theta) \left(\frac{\partial}{\partial \theta_i} S_T(\theta) \right) S_T^{-1}(\theta) f_T(\theta) \\ &\quad - \Upsilon_j'(\theta) S_T^{-1}(\theta) \left(\frac{\partial}{\partial \theta_i} f_T(\theta) \right), \end{aligned}$$

and using the linearity of $f(v_t, \theta)$ in θ , we obtain

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \Upsilon_j(\theta) &= \frac{\partial}{\partial \theta_i} \left(\frac{1}{K} \sum_{k=1}^K \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left(\frac{t}{T} \right) \frac{\partial f(v_t, \theta)}{\partial \theta_j} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left(\frac{s}{T} \right) f(v_s, \theta) \right) \right)' \\ &= \frac{1}{K} \sum_{k=1}^K \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \phi_k \left(\frac{t}{T} \right) \frac{\partial f(v_t, \theta)}{\partial \theta_j} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^T \phi_k \left(\frac{s}{T} \right) \frac{\partial f(v_s, \theta)}{\partial \theta_i} \right)' \\ &:= \Upsilon_{ji}^{(2)}, \end{aligned}$$

where, because of the linearity of the moment process, $\Upsilon_{ji}^{(2)}$ does not depend on θ . We have that

$$\begin{aligned} -\frac{\partial}{\partial \theta_i} (\Upsilon_j'(\theta) S_T^{-1}(\theta) f_T(\theta)) &= -\left(\frac{\partial}{\partial \theta_i} \Upsilon_j(\theta) \right)' S_T^{-1}(\theta) f_T(\theta) + \Upsilon_j'(\theta) S_T^{-1}(\theta) \left(\frac{\partial}{\partial \theta_i} S_T(\theta) \right) S_T^{-1}(\theta) f_T(\theta) \\ &\quad - \Upsilon_j'(\theta) S_T^{-1}(\theta) \left(\frac{\partial}{\partial \theta_i} f_T(\theta) \right) \\ &= -\left[\Upsilon_{ji}^{(2)} \right]' S_T^{-1}(\theta) f_T(\theta) + \Upsilon_j'(\theta) S_T^{-1}(\theta) (\Upsilon_i(\theta) + \Upsilon_i(\theta)') S_T^{-1}(\theta) f_T(\theta) \\ &\quad - \Upsilon_j'(\theta) S_T^{-1}(\theta) G_{iT}, \end{aligned}$$

and thus

$$\begin{aligned} \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_j} &= -\frac{\partial}{\partial \theta_j} \left[\Upsilon_1'(\theta) S_T^{-1}(\theta) f_T(\theta), \dots, \Upsilon_d'(\theta) S_T^{-1}(\theta) f_T(\theta) \right] \\ &= -\left[\Upsilon_{1j}^{(2)'} S_T^{-1}(\theta) f_T(\theta), \dots, \Upsilon_{dj}^{(2)'} S_T^{-1}(\theta) f_T(\theta) \right] \\ &\quad + \left[\Upsilon_1'(\theta) S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) f_T(\theta), \dots, \Upsilon_d'(\theta) S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) f_T(\theta) \right] \\ &\quad - \left[\Upsilon_1'(\theta) S_T^{-1}(\theta) G_{jT}, \dots, \Upsilon_d'(\theta) S_T^{-1}(\theta) G_{jT} \right] \\ &:= \tilde{G}_{j,2}^{(a)}(\theta) + \tilde{G}_{j,2}^{(b)}(\theta) + \tilde{G}_{j,2}^{(c)}(\theta), \end{aligned}$$

as desired. For $\partial^2 Q(\theta, S_T(\theta)) / \partial \theta_j \partial \theta'$ with $j \in \{1, \dots, d\}$, we have that

$$\frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} = A_j(\theta) + B_j(\theta) + C_j(\theta),$$

where

$$\begin{aligned} A_j(\theta) &= \frac{\partial}{\partial \theta_j} \left(\tilde{G}_T'(\theta) S_T^{-1}(\theta) \tilde{G}_T(\theta) \right), \\ B_j(\theta) &= \frac{\partial}{\partial \theta_j} \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_1} \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_d} \end{bmatrix}, \\ C_j(\theta) &= -\frac{\partial}{\partial \theta_j} \begin{bmatrix} f_T(\theta)' S_T^{-1}(\theta) \Upsilon_1(\theta) S_T^{-1}(\theta) \left[\tilde{G}_T(\theta) \right] \\ \vdots \\ f_T(\theta)' S_T^{-1}(\theta) \Upsilon_d(\theta) S_T^{-1}(\theta) \left[\tilde{G}_T(\theta) \right] \end{bmatrix}. \end{aligned}$$

For $A_j(\theta)$, it is straightforward to obtain

$$\begin{aligned}
A_j(\theta) &= \left(\frac{\partial}{\partial \theta_j} \tilde{G}'_T(\theta) \right) S_T^{-1}(\theta) \tilde{G}_T(\theta) - \tilde{G}'_T(\theta) S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_j} S_T^{-1}(\theta) \tilde{G}_T(\theta) \\
&\quad + [\tilde{G}_T(\theta)]' S_T^{-1}(\theta) \left(\frac{\partial}{\partial \theta_j} \tilde{G}_T(\theta) \right) \\
&= \left(\tilde{G}'_{j,2}{}^{(a)}(\theta) + \tilde{G}'_{j,2}{}^{(b)}(\theta) + \tilde{G}'_{j,2}{}^{(c)}(\theta) \right)' S_T^{-1}(\theta) \tilde{G}_T(\theta) \\
&\quad - \tilde{G}'_T(\theta) S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) \tilde{G}_T(\theta) \\
&\quad + [\tilde{G}_T(\theta)]' S_T^{-1}(\theta) \left(\tilde{G}_{j,2}{}^{(a)}(\theta) + \tilde{G}_{j,2}{}^{(b)}(\theta) + \tilde{G}_{j,2}{}^{(c)}(\theta) \right).
\end{aligned}$$

For $B_j(\theta)$, its i -th row is expressed as

$$\begin{aligned}
&\frac{\partial}{\partial \theta_j} \left(f_T(\theta)' S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_i} \right) \\
&= \left(\frac{\partial}{\partial \theta_j} f_T(\theta)' \right) S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_i} - f_T(\theta)' S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_j} S_T^{-1}(\theta) \frac{\partial \tilde{G}_T(\theta)}{\partial \theta_i} \\
&\quad + f_T(\theta)' S_T^{-1}(\theta) \frac{\partial^2 \tilde{G}_T(\theta)}{\partial \theta_j \partial \theta_i} \\
&= G'_{jT} S_T^{-1}(\theta) \left(\tilde{G}_{i,2}{}^{(a)}(\theta) + \tilde{G}_{i,2}{}^{(b)}(\theta) + \tilde{G}_{i,2}{}^{(c)}(\theta) \right) \\
&\quad - f_T(\theta)' S_T^{-1}(\theta) (\Upsilon_j(\theta) + \Upsilon_j(\theta)') S_T^{-1}(\theta) \left(\tilde{G}_{i,2}{}^{(a)}(\theta) + \tilde{G}_{i,2}{}^{(b)}(\theta) + \tilde{G}_{i,2}{}^{(c)}(\theta) \right) \\
&\quad + f_T(\theta)' S_T^{-1}(\theta) \left(\frac{\partial \tilde{G}_{i,2}{}^{(a)}(\theta)}{\partial \theta_j} + \frac{\partial \tilde{G}_{i,2}{}^{(b)}(\theta)}{\partial \theta_j} + \frac{\partial \tilde{G}_{i,2}{}^{(c)}(\theta)}{\partial \theta_j} \right).
\end{aligned}$$

Lastly, the i -th row of $C_j(\theta)$ is expressed as

$$\begin{aligned}
&\frac{\partial}{\partial \theta_j} \left(f_T(\theta)' S_T^{-1}(\theta) \Upsilon_i(\theta) S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \right) \\
&= G'_{jT} S_T^{-1}(\theta) \Upsilon_i(\theta) S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \\
&\quad - f_T(\theta)' S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_j} S_T^{-1}(\theta) \Upsilon_i(\theta) S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \\
&\quad + f_T(\theta)' S_T^{-1}(\theta) \Upsilon_{ij}^{(2)} S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \\
&\quad - f_T(\theta)' S_T^{-1}(\theta) \Upsilon_i(\theta) S_T^{-1}(\theta) \frac{\partial S_T(\theta)}{\partial \theta_j} S_T^{-1}(\theta) [\tilde{G}_T(\theta)] \\
&\quad + f_T(\theta)' S_T^{-1}(\theta) \Upsilon_i(\theta) S_T^{-1}(\theta) \left(\frac{\partial \tilde{G}_{i,2}{}^{(a)}(\theta)}{\partial \theta_j} + \frac{\partial \tilde{G}_{i,2}{}^{(b)}(\theta)}{\partial \theta_j} + \frac{\partial \tilde{G}_{i,2}{}^{(c)}(\theta)}{\partial \theta_j} \right).
\end{aligned}$$

From the proof of Lemma 1, it is straightforward to show that

$$\begin{aligned} f_T(\theta_T^*) &= f_T(\theta_0) + o_p(1), \\ S_T(\theta_T^*) &= S_T(\theta_0) + o_p(1), \\ \Upsilon_i(\theta_T^*) &= \Upsilon_i(\theta_0) + o_p(1), \\ \tilde{G}_{j,2}^{(k)}(\theta_T^*) &= \tilde{G}_{j,2}^{(k)}(\theta_0) + o_p(1) \end{aligned}$$

for any $\theta_T^* = \theta_0 + O_p(T^{-1/2})$, $k \in \{a, b, c, \}$, and $j \in \{1, \dots, d\}$. Similarly, it is not difficult to check that

$$\left. \frac{\partial \tilde{G}_{i,2}^{(k)}(\theta)}{\partial \theta_j} \right|_{\theta=\theta_T^*} = \left. \frac{\partial \tilde{G}_{i,2}^{(k)}(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} + o_p(1) \text{ for } k \in \{a, b, c\}.$$

Therefore, we have that

$$\begin{aligned} \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_T^*} &= A_j(\theta_0) + B_j(\theta_0) + C_j(\theta_0) + o_p(1) \\ &= \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_0} + o_p(1), \end{aligned}$$

as desired. ■

Proof of Theorem 4. From the second-order Taylor expansion of the FOC, we have that

$$\begin{aligned} 0 &= Q(\hat{\theta}_{\text{CU}}, S_T(\hat{\theta}_{\text{CU}})) = \tilde{G}_T(\hat{\theta}_{\text{CU}}) [S_T(\hat{\theta}_{\text{CU}})]^{-1} f_T(\hat{\theta}_{\text{CU}}) \\ &= \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} f_T(\theta_0) + \left. \left(\frac{\partial Q_{\theta, S_T(\theta)}}{\partial \theta'} \right) \right|_{\theta=\theta_0} (\hat{\theta}_{\text{CU}} - \theta_0) \\ &\quad + \frac{1}{2} \sum_{j=1}^d \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_T^*} (\hat{\theta}_{\text{CU},j} - \theta_{0,j}) (\hat{\theta}_{\text{CU}} - \theta_0) \\ &= \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} f_T(\theta_0) \\ &\quad + \left\{ \left. \left(\frac{\partial Q(\theta, S_T(\theta))}{\partial \theta'} \right) \right|_{\theta=\theta_0} + \frac{1}{2} \sum_{j=1}^d \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_T^*} (\hat{\theta}_{\text{CU},j} - \theta_{0,j}) \right\} (\hat{\theta}_{\text{CU}} - \theta_0), \end{aligned}$$

where each component of θ_T^* is located between the corresponding components θ_0 and $\hat{\theta}_{\text{CU}}$. By Assumption (6), we obtain

$$\hat{\theta}_{\text{CU},j} - \theta_{0,j} = \underbrace{e'_j [G'_T S_T^{-1}(\theta_0) G_T]^{-1} G'_T S_T^{-1}(\theta_0) f_T(\theta_0)}_{=\psi_j(\theta_0, S_T(\theta_0))} + o_p\left(\frac{1}{T}\right),$$

where the first term on the right-hand side is $O_p(T^{-1/2})$. Lemma B.4 provides that

$$\left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_T^*} = \left. \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \right|_{\theta=\theta_0} + o_p(1).$$

Therefore, we have that

$$\begin{aligned}
& \frac{1}{2} \sum_{j=1}^d \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_T^*} (\hat{\theta}_{\text{CU},j} - \theta_{0,j}) (\hat{\theta}_{\text{CU}} - \theta_0) \\
&= \frac{1}{2} \sum_{j=1}^d \left(\frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_0} + o_p(1) \right) \times \left(\psi_j(\theta_0, S_T(\theta_0)) + o_p\left(\frac{1}{T}\right) \right) (\hat{\theta}_{\text{CU}} - \theta_0) \\
&= \frac{1}{2} \sum_{j=1}^d \left(\frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_0} \right) \psi_j(\theta_0, S_T(\theta_0)) (\hat{\theta}_{\text{CU}} - \theta_0) + o_p\left(\frac{1}{T}\right).
\end{aligned}$$

Combining these results into the second-order Taylor expansion of the FOC, we obtain

$$\begin{aligned}
0 &= Q(\hat{\theta}_{\text{CU}}, S_T(\hat{\theta}_{\text{CU}})) = \tilde{G}_T(\hat{\theta}_{\text{CU}}) [S_T(\hat{\theta}_{\text{CU}})]^{-1} f_T(\hat{\theta}_{\text{CU}}) \\
&= \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} f_T(\theta_0) \\
&+ \left(\frac{\partial Q(\theta, S_T(\theta))}{\partial \theta'} \Big|_{\theta=\theta_0} \right) (\hat{\theta}_{\text{CU}} - \theta_0) + \frac{1}{2} \sum_{j=1}^d \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_T^*} (\hat{\theta}_{\text{CU},j} - \theta_{0,j}) (\hat{\theta}_{\text{CU}} - \theta_0) \\
&= \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} f_T(\theta_0) \\
&+ \left\{ \left(\frac{\partial Q(\theta, S_T(\theta))}{\partial \theta'} \Big|_{\theta=\theta_0} \right) + \frac{1}{2} \sum_{j=1}^d \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_T^*} (\hat{\theta}_{\text{CU},j} - \theta_{0,j}) \right\} (\hat{\theta}_{\text{CU}} - \theta_0) \\
&= \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} f_T(\theta_0) \\
&+ \left\{ \underbrace{\left(\frac{\partial Q(\theta, S_T(\theta))}{\partial \theta'} \Big|_{\theta=\theta_0} + \frac{1}{2} \sum_{j=1}^d \frac{\partial^2 Q(\theta, S_T(\theta))}{\partial \theta_j \partial \theta'} \Big|_{\theta=\theta_0} \psi_j(\theta_0, S_T(\theta_0)) \right)}_{=\tilde{A}(\theta_0, S_T(\theta_0))} \right\} (\hat{\theta}_{\text{CU}} - \theta_0) + o_p\left(\frac{1}{T}\right),
\end{aligned}$$

which leads to the desired result,

$$\sqrt{T} (\hat{\theta}_{\text{CU}} - \theta_0) = - [\tilde{A}(\theta_0, S_T(\theta_0))]^{-1} \tilde{G}_T(\theta_0) [S_T(\theta_0)]^{-1} \sqrt{T} f_T(\theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

■

References

- [1] Hansen, B. E., & Lee, S. (2019). Inference for iterated GMM under misspecification. Working paper.
- [2] Hwang, J., Kang, B., & Lee, S. (2020). A doubly corrected robust variance estimator for linear GMM. Working paper, Department of Economics, University of Connecticut.
- [3] Sun, Y. (2014). Fixed-smoothing asymptotics in a two-step generalized method of moments framework. *Econometrica*, 82(6), 2327–2370.

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