


Article

Estimating Linear Dynamic Panels with Recentered Moments

Yong Bao 

Department of Economics, Purdue University, West Lafayette, IN 47907, USA; ybao@purdue.edu

Abstract: This paper proposes estimating linear dynamic panels by explicitly exploiting the endogeneity of lagged dependent variables and expressing the cross moments between the endogenous lagged dependent variables and disturbances in terms of model parameters. These moments, when recentered, form the basis for model estimation. The resulting estimator's asymptotic properties are derived under different asymptotic regimes (large number of cross-sectional units or long time spans), stable conditions (with or without a unit root), and error characteristics (homoskedasticity or heteroskedasticity of different forms). Monte Carlo experiments show that it has very good finite-sample performance.

Keywords: dynamic panels; endogeneity; heteroskedasticity

JEL Classification: C23; C13; C26

1. Introduction

It is well known that the standard fixed-effects or within-group (WG) estimator for linear dynamic panel (DP) models suffers from the issue of bias (Nickell 1981) when the time series dimension T is small. Alternative estimators include those based on instrumental variables (IVs) (Anderson and Hsiao 1981), the generalized method of moments (GMM) (e.g., Arellano and Bond 1991; Holtz-Eakin et al. 1988), the maximum likelihood (ML) approach (e.g., Alvarez and Arellano 2003, 2022; Blundell and Smith 1991; Hsiao et al. 2002; Lancaster 2002), those that correct the score function in the ML framework (e.g., Alvarez and Arellano 2022; Breitung et al. 2022; Dhaene and Jochmans 2016), and those that directly correct the bias of the WG estimator (e.g., Bao and Yu 2023; Bun and Carree 2005; Dhaene and Jochmans 2015; Everaert and Pozzi 2007; Gouriéroux et al. 2010; Hahn and Kuersteiner 2002; Kiviet 1995).¹

This paper proposes estimating linear dynamic panels by recentering the cross moments between the lagged dependent variables, which are endogenous, and the error term in the model by their non-zero expectations.² The resulting estimator is named the recentered method of moments (RMM) estimator accordingly. These recentered moments are functions of model parameters and, together with moment conditions from other exogenous regressors, if any, form the basis for model estimation. Essentially, it is based on the idea that the best “instrument” for any of the endogenous lagged dependent variables is itself. As such, one does not need to search for IVs and can avoid issues of weak instruments and many instruments that exist in the GMM framework. It is closely related to the bias-correction literature, but there is no correction procedure involved. In particular, Appendix B illustrates that the estimator proposed in this paper is numerically equivalent to the indirect inference (II) estimator, as introduced by Bao and Yu (2023). However, they are motivated very differently. The strategy of Bao and Yu (2023) starts with a biased estimator, but this paper does not have a biased estimator to begin with and designs recentered moment conditions directly by expressing the cross moments between the endogenous lagged dependent variables and disturbances in terms of model parameters.³ Furthermore, there are three major contributions in this paper that are absent elsewhere.



Citation: Bao, Yong. 2024. Estimating Linear Dynamic Panels with Recentered Moments. *Econometrics* 12: 3. <https://doi.org/10.3390/econometrics12010003>

Academic Editor: Julie Le Gallo

Received: 20 November 2023

Revised: 8 January 2024

Accepted: 15 January 2024

Published: 17 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

First, it allows for more general assumptions regarding the data-generating process (DGP). Specifically, the most general form of error heteroskedasticity, arising from cross-sectional units, time, or both, is considered. Breitung et al. (2022) focus primarily on the situation of cross-sectional heteroskedasticity.⁴ Alvarez and Arellano (2022) consider only time-series heteroskedasticity. Bao and Yu (2023) discuss a robust version of their II estimator that is valid under both forms of heteroskedasticity, though falling short of deriving its asymptotic distribution. They focus instead on the case when there is only time-series heteroskedasticity. Bun and Carree (2006) derive the asymptotic bias of the WG estimator when both forms of heteroskedasticity are present and propose bias correcting the WG estimator when T is fixed for the first-order DP (DP(1)) model. Juodis (2013) points out that the bias-correction procedure of Bun and Carree (2006) is, in fact, inconsistent and designs a consistent one for a higher-order DP.⁵ Note that both Bun and Carree (2006) and Juodis (2013) estimate the temporal heteroskedasticity parameters so that they can be plugged into the bias expression of the WG estimator for the purpose of bias correction. However, they do not provide insights on how to conduct asymptotic inference on the resulting bias-corrected estimator. In contrast, Alvarez and Arellano (2022) jointly estimate the temporal heteroskedasticity parameters and model parameters and also derive their joint asymptotic distribution.

Second, this paper explicitly includes the case when there is a unit root. If the time span is short, this may not matter, as the inference procedure is under the large- N asymptotic regime, where N is the number of cross-sectional units. Bun and Carree (2006), Juodis (2013), Alvarez and Arellano (2022), and Bao and Yu (2023) all consider short panels. But for long panels, the issue of unit root cannot be simply ignored. Dhaene and Jochmans (2015) do not consider the unit-root case, and their jackknife method is developed under the rectangular-array asymptotic regime (namely, both N and T are large). Breitung et al. (2022) also present some results under the rectangular-array regime, but do not explicitly consider the unit-root case.

Third, asymptotic distribution results are derived that, in general, do not require both N and T to be large. The asymptotic distribution of the proposed RMM estimator under large T resembles the familiar ordinary least squares (OLS) result in traditional regression analysis, and its asymptotic variance achieves the efficiency bound under homoskedasticity. Similar to time series literature, the convergence rate of the estimator of the autoregressive parameters is different when there is a unit root under large T , but the standard t -test procedure carries through when one is conducting hypothesis testing. Under homoskedasticity, Han and Phillips (2010) report a unified asymptotic distribution result for their first difference least squares (FDLS) estimator for DP(1) when the unit-root case is allowed. In their setup, the fixed effects disappear under the unit-root case. Hayakawa (2009) derives the asymptotic properties of the IV estimator for higher-order DP models, with neither heteroskedasticity nor exogenous regressors present, when the panel is dynamically stable under large N and large T , though he remarks that it should also hold under other asymptotic regimes.

The plan of this paper is as follows. The next section introduces the model and notation. Section 3 presents the RMM estimator under the baseline set-up of homoskedasticity, though it is further shown that the estimator is robust to cross-sectional heteroskedasticity as well. An important message here is that when T is large, there is no asymptotic bias, standing in contrast to the consistent WG estimator that may possess an asymptotic bias. Section 4 introduces a robust estimator under cross-sectional and temporal heteroskedasticity. Section 5 illustrates the good finite-sample performance of the proposed estimator by Monte Carlo experiments. The last section concludes and discusses possible future research. Technical details, including lemmas that are used for the proofs of the main results in this paper, together with some extended discussions and additional simulation results, are provided in the appendices. Throughout, matrix/vector dimension subscripts are typically omitted unless confusion may arise. A subscript 0 is used to signify the true value of a parameter that is to be estimated.

2. Model, Notation, and Assumptions

The p -th order linear dynamic panel model, DP(p) for short, is

$$y_{it} = \sum_{\ell=1}^p \phi_{\ell} y_{i,t-\ell} + x'_{it} \beta + \alpha_i + u_{it} = w'_{it} \theta + \alpha_i + u_{it}, \quad (1)$$

$i = 1, \dots, N$, $t = 1, \dots, T$, where the dependent variable y_{it} is related to its lagged values, up to order p , fixed effects α_i , the $k \times 1$ vector of exogenous variables $x_{it} = (x_{it,1}, \dots, x_{it,k})'$, and an idiosyncratic disturbance term u_{it} . The vector $w_{it} = (y_{i,t-1}, \dots, y_{i,t-p}, x'_{it})'$ collects all the right-hand side variables and $\theta = (\phi', \beta')'$, where $\phi = (\phi_1, \dots, \phi_p)'$. For each i , let $y_i = (y_{i1}, \dots, y_{iT})'$, $y_{i(-\ell)} = (y_{i,1-\ell}, \dots, y_{i,T-\ell})'$, $Y_i = (y_{i(-1)}, \dots, y_{i(-p)})'$, $X_i = (x_{i1}, \dots, x_{iT})'$, $W_i = (Y_i, X_i) = (w_{i1}, \dots, w_{iT})'$, and $u_i = (u_{i1}, \dots, u_{iT})'$. Stacking over i , one has $\alpha = (\alpha_1, \dots, \alpha_N)'$, $y = (y'_1, \dots, y'_N)'$, $y_{(-\ell)} = (y'_{1(-\ell)}, \dots, y'_{N(-\ell)})'$, $Y = (Y'_1, \dots, Y'_N) = (y'_{(-1)}, \dots, y'_{(-p)})'$, $X = (X'_1, \dots, X'_N)'$, $W = (Y, X) = (W'_1, \dots, W'_N)'$, and $u = (u'_1, \dots, u'_N)'$. Let \otimes denote the matrix Kronecker product operator and $\mathbf{1} = \mathbf{1}_T$ be a $T \times 1$ column vector of ones. Then, in matrix notation, (1) can be written as

$$y = W\theta + \alpha \otimes \mathbf{1} + u. \quad (2)$$

Suppose there exists a matrix A such that when pre-multiplying it to (2), the fixed effects are wiped out, namely,

$$Ay = AW\theta + Au. \quad (3)$$

The $NT \times NT$ matrix $I_N \otimes M_T$ comes as a natural choice for A , where $M_T = I_T - T^{-1}\mathbf{1}\mathbf{1}' \equiv M$, and I_N (I_T) denotes the identity matrix of size N (T). In this case, applying the OLS procedure to (3) yields the WG estimator. For a given A , the regression model (3) contains endogenous lagged dependent variables $y_{(-1)}, \dots, y_{(-p)}$ as regressors and $E(W'A'Au) \neq 0$. The IV/GMM literature focuses on using various instruments for $y_{(-\ell)}$ (or its various differences). The best “instrument” for $y_{(-\ell)}$ is of course itself, though it violates the definition of instrument. Nevertheless, if one can explicitly analyze how $y_{(-\ell)}$ is correlated with u , subject to the transformation induced by A , then this piece of information can be used to estimate model parameters. This is essentially the idea of this paper. In the sequel, $A = I_N \otimes M$ is used (and thus, $A'A = A$), and the following assumptions are made. Different assumptions about the idiosyncratic error term are deferred into the next two sections when different forms of heteroskedasticity are discussed.

Assumption 1. The series of fixed effects α_i , $i = 1, \dots, N$, is i.i.d. across individuals with finite moments up to the fourth order.

Assumption 2. The error terms u_{it} and fixed effects α_i are independent of each other for any $i = 1, \dots, N$, $t = 1, \dots, T$.

Assumption 3. The regressors X , if present, are either fixed or random and $(NT)^{-1}X'AX$ converges (in probability) to a nonsingular matrix as $NT \rightarrow \infty$. When they are fixed, $x_{it} = O(1)$. When they are random: (i) they are strictly exogenous with respect to error terms; (ii) $x_{it} = O_p(1)$ with finite moments up to the fourth order; (iii) $E(\alpha_i^{r_1} x_{it,s}^{r_2}) = O(1)$ and $\text{Cov}(\alpha_i^{r_1} x_{it,s_i}^{r_2}, \alpha_j^{r_1} x_{jt,s_j}^{r_2}) = 0$, $r_1 + r_2 \leq 4$, $r_1 \geq 0$, $r_2 \geq 0$, $i \neq j$, $i, j = 1, \dots, N$, $s, s_i, s_j = 1, \dots, k$.

Assumption 4. The initial values $y_{i,-s}$, $i = 1, \dots, N$, $s = 0, 1, \dots, p$, are either fixed or random. When they are fixed, $y_{i,-s} = O(1)$. When they are random: (i) $y_{i,-s} = O_p(1)$ with finite moments up to the fourth order; (ii) $E(\alpha_i^{r_1} y_{i,-s}^{r_2}) = O(1)$ and $\text{Cov}(\alpha_i^{r_1} y_{i,-s}^{r_2}, \alpha_j^{r_1} y_{j,-s}^{r_2}) = 0$, $r_1 + r_2 \leq 4$, $r_1 \geq 0$, $r_2 \geq 0$, $i \neq j$, $i, j = 1, \dots, N$, $s = 0, \dots, p$; (iii) $\text{Cov}(y_{i,-s}, u_{it}) = 0$, $s = 0, \dots, p$, $t = 1, \dots, T$, $i = 1, \dots, N$.

Without loss of generality, in what follows, the analysis in this paper is conditioning on X for ease of presentation. No distribution assumption is made, so long as some moment conditions hold. Assumption 3 does not rule out the possible correlation between X and the fixed effects, but rules out correlation among the products of them. For simplicity, the strictly exogenous regressors contained in X are bounded (in probability). For the estimation strategy to be introduced in the next section, the inclusion of explosive exogenous regressors can affect the convergence rate of the associated parameter estimators, but does not affect the convergence rate of those associated with the lagged dependent variables. Assumption 4 does not spell out how the initial values are generated, except that there are restrictions on how they may be correlated with the fixed effects and idiosyncratic errors. The panel under consideration is not restricted to be dynamically stable. The fixed effects in Assumption 1 are treated as randomly generated from some distribution. This assumption itself is not needed for the purpose of estimation, since the recentered moment conditions to be presented wipe out the fixed effects. However, it is used, together with Assumptions 2–4 and Assumption 5 or 6, to ensure that the (scaled) moments and the associated gradient have properly defined probability limits (see, for example, Lemma A5 in the Appendix C), which in turn are used to establish the asymptotic distribution of the RMM estimator.⁶ When there is a unit root, the asymptotic behavior of the RMM estimator to be introduced depends on the fixed effects, so for the sake of tractability, they are assumed to be deterministic by following Hahn and Kuersteiner (2002).⁷

In the next two sections, the asymptotic properties are derived of the RMM estimator under homoskedasticity and heteroskedasticity, without or with a unit root. Under large N and finite T , the asymptotic distribution is of the same form, whereas under large T , subject to the appropriate scaling factors, the asymptotic distribution resembles the OLS result in traditional regression analysis.

Note that, as established in Bao and Yu (2023), the vectors of lagged observations from model (1) (at the true parameter vector θ_0) for each cross-sectional unit can be written as

$$y_{i,(-\ell)} = \Phi_p^{-1}[L^\ell(u_i + \alpha_i \mathbf{1} + X_i \beta_0) + \sum_{s=0}^{\ell-1} \Phi_s L^{\ell-1-s} e_1 y_{i,-s} + \sum_{s=\ell}^{p-1} \Phi_{(s-p)} L^{\ell-1-s} e_1 y_{i,-s}], \quad (4)$$

$\ell = 1, \dots, p$, where L is a $T \times T$ strict lower triangular matrix with 1's on the first sub-diagonals and 0's elsewhere, $e_1 = (1, 0, \dots, 0)'$ is a $T \times 1$ vector, $\Phi_p = \Phi_p(\phi_0)$, $\Phi_p(\phi) = I - \phi_1 L - \dots - \phi_p L^p$, and $\Phi_{(s-p)} = \Phi_s(\phi_0) - \Phi_p(\phi_0)$.⁸ Then, stacking over index i , one has, for $\ell = 1, \dots, p$,

$$\begin{aligned} y_{(-\ell)} &= (I_N \otimes \Phi_p^{-1} L^\ell) u + (I_N \otimes \Phi_p^{-1} L^\ell) X \beta_0 + (I_N \otimes \Phi_p^{-1} L^\ell \mathbf{1}) \alpha \\ &\quad + \sum_{s=0}^{\ell-1} (I_N \otimes \Phi_p^{-1} \Phi_s L^{\ell-1-s} e_1) \mathbf{y}_{(-s)} + \sum_{s=\ell}^{p-1} (I_N \otimes \Phi_p^{-1} \Phi_{(s-p)} L^{\ell-1-s} e_1) \mathbf{y}_{(-s)}, \end{aligned} \quad (5)$$

where $\mathbf{y}_{(-s)} = (y_{1,-s}, \dots, y_{N,-s})'$ is an $N \times 1$ vector collecting initial cross-sectional observations at time $-s = 0, \dots, 1 - p$. This representation of $y_{(-\ell)}$ is in terms of u , X , α , and initial conditions, so $W' A u$ essentially boils down to linear and quadratic forms in the random vector u . This facilitates the derivation of the expectation of $W' A u$, which forms the basis of the moment conditions in this paper for estimating model parameters. When $u_{i,t}$ is independent and identically distributed (i.i.d.) across i and t , moments of quadratic forms in u can be found in Bao and Ullah (2010), and in the presence of heteroskedasticity, results are provided in Appendix A.7 of Ullah (2004).

3. The Baseline Set-Up

This section presents the estimator under the framework when the idiosyncratic errors are homoskedastic across i and t .

Assumption 5. $u_{it}, i = 1, \dots, N, t = 1, \dots, T$, is i.i.d. across i and t , $E(u_{it}) = 0$, $\text{Var}(u_{it}) = \sigma^2$, and has finite moments up to the fourth order.

3.1. Estimation

Given Assumptions 1–5 and using the representation (5), one has, at θ_0 ,

$$E(W' Au) = \begin{pmatrix} N\sigma^2 \text{tr}(M\Phi_p^{-1}L^1) \\ \vdots \\ N\sigma^2 \text{tr}(M\Phi_p^{-1}L^p) \\ \mathbf{0}_k \end{pmatrix} = -\frac{N\sigma^2}{T} \begin{pmatrix} \mathbf{1}'\Phi_p^{-1}L\mathbf{1} \\ \vdots \\ \mathbf{1}'\Phi_p^{-1}L^p\mathbf{1} \\ \mathbf{0}_k \end{pmatrix}, \quad (6)$$

where $\text{tr}(\cdot)$ is the matrix trace operator. The first equality in (6) follows when one substitutes (5) into $W = (\mathbf{y}_{(-1)}, \dots, \mathbf{y}_{(-p)}, X)$ and takes expectation of quadratic forms in \mathbf{u} and the second equality follows from $\text{tr}(M\Phi_p^{-1}L^\ell) = \text{tr}(\Phi_p^{-1}L^\ell) - T^{-1}\mathbf{1}'\Phi_p^{-1}L^\ell\mathbf{1}$, where, since $L^\ell\Phi_p^{-1}$ is strictly lower triangular, $\text{tr}(\Phi_p^{-1}L^\ell) = 0$.

This set of moment conditions additionally involves the variance parameter σ^2 . Since $E(\mathbf{u}'A\mathbf{u}) = N(T-1)\sigma^2$, one can define the following

$$\mathbf{g}_{NT}(\theta) = \frac{1}{NT}W'A(\mathbf{y} - W\theta) + \frac{1}{NT}(\mathbf{y} - W\theta)'A(\mathbf{y} - W\theta)\mathbf{h}(\theta), \quad (7)$$

where $\mathbf{h}(\theta) = [T(T-1)]^{-1}(\mathbf{1}'\Phi_p^{-1}(\theta)L\mathbf{1}, \dots, \mathbf{1}'\Phi_p^{-1}(\theta)L^p\mathbf{1}, \mathbf{0}_k)'$, such that $E(\mathbf{g}_{NT}(\theta_0)) = \mathbf{0}$. Thus, an estimator can be defined as $\hat{\theta} = \arg_{\theta}\{\mathbf{g}_{NT}(\theta) = \mathbf{0}\}$. Appendix B discusses several closely related estimators that are motivated very differently. The traditional method of moments (MM) or GMM is based on moment conditions that have expectations (or the probability limits) equal to zero exactly, where these moments are usually defined as cross moments of exogenous regressors or instruments and disturbances in the model under consideration. Here, the cross moments pertaining to the endogenous lagged dependent variables have non-zero expectations and the moment conditions $\mathbf{g}_{NT}(\theta) = \mathbf{0}$ for estimation are designed by recentering the cross moments by their non-zero expectation parts. For this reason, $\hat{\theta}$ is named the RMM estimator in this paper.

Note that one can equivalently write the set of moment conditions (7) as

$$\begin{aligned} \mathbf{g}_{NT}(\theta) &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [(\mathbf{w}_{it} - \bar{\mathbf{w}}_i)\epsilon_{it}(\theta) + (\epsilon_{it}(\theta) - \bar{\epsilon}_i(\theta))\epsilon_{it}(\theta)\mathbf{h}(\theta)] \\ &\equiv \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [(z_{it}(\theta) - \bar{z}_i(\theta))\epsilon_{it}(\theta)], \end{aligned}$$

where $\epsilon_{it}(\theta) = y_{it} - \mathbf{w}_{it}'\theta$, $z_{it}(\theta) = \mathbf{w}_{it} + \mathbf{h}(\theta)\epsilon_{it}(\theta)$, $\bar{\mathbf{w}}_i = T^{-1}\sum_{t=1}^T \mathbf{w}_{it}$, and $\bar{\epsilon}_i(\theta)$ and $\bar{z}_i(\theta)$ are defined similarly. It turns out that $T^{-1}\sum_{t=1}^T [(z_{it}(\theta) - \bar{z}_i(\theta))\epsilon_{it}(\theta)]$ is the same as $\mathbf{m}_{Ti}(\theta)$ in Breitung et al. (2022).⁹ They show that $\mathbf{m}_{Ti}(\theta)$ has the property of $E(\mathbf{m}_{Ti}(\theta_0)) = \mathbf{0}$ under cross-sectional heteroskedasticity, namely, $E(u_{it}^2) = \sigma_i^2$. Essentially, this is because, in this case, $\hat{\sigma}_{Ti}^2(\theta_0) = (T-1)^{-1}\sum_{t=1}^T (\epsilon_{it}(\theta_0) - \bar{\epsilon}_i(\theta_0))\epsilon_{it}(\theta_0)$ is an unbiased estimator of σ_i^2 and thus $(T-1)^{-1}\sum_{i=1}^N \sum_{t=1}^T (\epsilon_{it}(\theta_0) - \bar{\epsilon}_i(\theta_0))\epsilon_{it}(\theta_0) = (T-1)^{-1}\mathbf{u}'A\mathbf{u}$ is in fact an unbiased estimator of $\sum_{i=1}^N \sigma_i^2$. In other words, if Assumption 5 is modified such that $E(u_{it}^2) = \sigma_i^2$, then it is still the case that $E(\mathbf{g}_{NT}(\theta_0)) = \mathbf{0}$ and the estimation procedure does not change. This is invalid, however, when there is time-series heteroskedasticity ($E(u_{it}^2) = \sigma_t^2$) or both forms of heteroskedasticity ($E(u_{it}^2) = \sigma_{it}^2$). Remark 2 of Breitung et al. (2022) suggests that $\mathbf{m}_{Ti}(\theta)$ may still be considered a valid set of moment conditions when both forms of heteroskedasticity are present, in the sense that the limit of $\mathbf{m}_{Ti}(\theta)$ (as $T \rightarrow \infty$) goes to zero, provided that: (i) there is no unit root and (ii) T is large. In the next section, a robust estimator is introduced such that it is based on a set of moment conditions having exact expectation zero under the most general form of heteroskedasticity without imposing

the two conditions. To fix ideas, various results in this section are presented by assuming homoskedasticity.

Obviously,

$$\begin{aligned} \mathbf{G}_{NT}(\boldsymbol{\theta}) &= \frac{\partial \mathbf{g}_{NT}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &= \frac{1}{NT} [(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})' \mathbf{A}(\mathbf{y} - \mathbf{W}\boldsymbol{\theta}) \mathbf{H}(\boldsymbol{\theta}) - \mathbf{W}' \mathbf{A} \mathbf{W} - 2\mathbf{h}(\boldsymbol{\theta})(\mathbf{y} - \mathbf{W}\boldsymbol{\theta})' \mathbf{A} \mathbf{W}], \end{aligned} \quad (8)$$

where

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial \mathbf{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{1}{T(T-1)} \begin{pmatrix} \mathbf{R}_p(\boldsymbol{\phi}) & \mathbf{O}_{p \times k} \\ \mathbf{O}_{k \times p} & \mathbf{O}_k \end{pmatrix}, \quad (9)$$

and $\mathbf{R}_p(\boldsymbol{\phi})$ is a $p \times p$ matrix with $\mathbf{1}' \boldsymbol{\Phi}_p^{-1}(\boldsymbol{\phi}) \mathbf{L}^s \boldsymbol{\Phi}_p^{-1}(\boldsymbol{\phi}) \mathbf{L}^\ell \mathbf{1}$ in its (ℓ, s) -th position, $\ell, s = 1, \dots, p$. In light of (5), one notices that both $\mathbf{g}_{NT} = \mathbf{g}_{NT}(\boldsymbol{\theta}_0) = (NT)^{-1} [\mathbf{W}' \mathbf{A} \mathbf{u} + \mathbf{u}' \mathbf{A} \mathbf{u} \mathbf{h}]$ and $\mathbf{G}_{NT} = \mathbf{G}_{NT}(\boldsymbol{\theta}_0) = (NT)^{-1} [\mathbf{u}' \mathbf{A} \mathbf{u} \mathbf{H} - \mathbf{W}' \mathbf{A} \mathbf{W} - 2\mathbf{h} \mathbf{u}' \mathbf{A} \mathbf{W}]$ are in terms of linear and quadratic forms in \mathbf{u} , whereas $\mathbf{h} = \mathbf{h}(\boldsymbol{\phi}_0)$ and $\mathbf{H} = \mathbf{H}(\boldsymbol{\phi}_0)$ are purely functions of $\boldsymbol{\phi}_0$. From Lemma A1 in Appendix C, regardless of N and T , under homoskedasticity,

$$\text{Var}(\sqrt{NT} \mathbf{g}_{NT}) = \frac{1}{NT} \text{Var}(\mathbf{W}' \mathbf{A} \mathbf{u}) - \sigma^4 \left[\frac{2(T-1)}{T} + \gamma_2 \frac{(T-1)^2}{T^2} \right] \mathbf{h} \mathbf{h}' \equiv \boldsymbol{\Omega}_{NT}, \quad (10)$$

where γ_2 is the excess kurtosis of u_{it} .¹⁰

At the true parameter vector $\boldsymbol{\theta}_0$, when there is no unit root, $-(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = O_p(1)$, $T\mathbf{h} = O(1)$, $T\mathbf{H} = O(1)$, $N^{-1} \mathbf{u}' \mathbf{A} \mathbf{W} = O(1) + O_p(N^{-1/2} T^{1/2})$, and $(NT)^{-1} \mathbf{u}' \mathbf{A} \mathbf{u} = O_p(1)$, so $\mathbf{G}_{NT} = O_p(1) + [O(T^{-2}) + O_p(N^{-1/2} T^{-3/2})] + O_p(T^{-1})$. When T is large, \mathbf{G} is obviously dominated by the $O_p(1)$ term, namely, $-(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$. (When there is a unit root, the properly scaled \mathbf{G}_{NT} has the scaled $-\mathbf{W}' \mathbf{A} \mathbf{W}$ as its leading term.) From Lemma A6 in Appendix C (or Lemma A14 when there is a unit root), the estimator is locally identified in large samples. When T is finite but N is large, it is still very likely that $-(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$ is relatively large compared with the other two terms of finite order in \mathbf{G}_{NT} . In finite samples, one can always check numerically whether $\mathbf{G}_{NT}(\boldsymbol{\theta})$ is singular against a grid of values of $\boldsymbol{\theta}$. This is similar to the approach adopted by Gospodinov et al. (2017) and Bao and Yu (2023). In this section and the one to follow, it is implicitly assumed that the estimator is (first-order) locally identified. A general statement about sufficient conditions for identification would be desirable but extremely difficult.¹¹

3.2. Inference under Large N

If $N \rightarrow \infty$ and T is finite, by writing $\mathbf{g}_{NT} = (NT)^{-1} \sum_{i=1}^N (\mathbf{W}'_i \mathbf{M} \mathbf{u}_i + \mathbf{u}'_i \mathbf{M} \mathbf{u}_i \mathbf{h}) \equiv (NT)^{-1} \sum_{i=1}^N \mathbf{g}_i$, where \mathbf{g}_i is independent (across i) with mean $\mathbf{0}$ and variance $O(1)$ under Assumptions 1–5, one can apply Lyapunov's central limit theorem to $N^{-1/2} \mathbf{g}_{NT}$. This straightforwardly gives the following result.

Theorem 1. Under Assumptions 1–5, if T is finite and $\mathbf{G}_T = \text{plim}_{N \rightarrow \infty} \mathbf{G}_{NT}$ and $\boldsymbol{\Omega}_T = \lim_{N \rightarrow \infty} \boldsymbol{\Omega}_{NT}$ exist and are nonsingular, then as $N \rightarrow \infty$,

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_T^{-1} \boldsymbol{\Omega}_T \mathbf{G}_T^{-1}). \quad (11)$$

For practical inference, while \mathbf{G}_T may be consistently estimated by $\hat{\mathbf{G}}_T = (NT)^{-1} [(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})' \mathbf{A}(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}) \hat{\mathbf{H}} - \mathbf{W}' \mathbf{A} \mathbf{W} - 2\hat{\mathbf{h}}(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})' \mathbf{A} \mathbf{W}]$ with $\hat{\mathbf{h}} = \mathbf{h}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{H}} = \mathbf{H}(\hat{\boldsymbol{\theta}})$, it is not advisable to use a plug-in approach to estimate $\boldsymbol{\Omega}_T$. Recall $\boldsymbol{\Omega}_{NT}$ (see (10)) contains $\text{Var}(\mathbf{W}' \mathbf{A} \mathbf{u})$, which, as shown in Bao and Yu (2023), has a very complicated expression involving further the skewness and kurtosis of u_{it} , fixed effects, and possible interactions between the fixed effects and initial conditions. Given these complications and the fact that $\mathbf{W}'_i \mathbf{M} \mathbf{u}_i + \mathbf{u}'_i \mathbf{M} \mathbf{u}_i \mathbf{h}$ is independent (across i), a White-type (White 1980) estimator is natural

for estimating Ω_T in this asymptotic regime of large N and finite T . In consequence, a consistent variance estimator of the asymptotic variance of $\sqrt{N}(\hat{\theta} - \theta_0)$ is

$$\hat{V}(\hat{\theta}) = \hat{G}_T^{-1} \left(\frac{1}{N} \sum_{i=1}^N \hat{g}_{Ti} \hat{g}_{Ti}' \right) \hat{G}_T^{-1'}, \quad (12)$$

where $\hat{g}_{Ti} = T^{-1}(W_i' M \hat{u}_i + \hat{u}_i' M \hat{u}_i \hat{h})$ with $M \hat{u}_i = M(y_i - W_i \hat{\theta})$.

One can show that G_T is equal to $\text{plim}_{N \rightarrow \infty} \nabla m_{NT}(\theta_0)$ in Breitung et al. (2022), which, in turn, is equal to $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N E[\nabla m_{Ti}(\theta_0)]$, where $\nabla m_{NT}(\theta) (= \partial m_{NT}(\theta) / \partial \theta')$ and $E[\nabla m_{Ti}(\theta)]$ are given in Appendix A of Breitung et al. (2022). Further, Ω_T is the same as $\text{plim}_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N m_{Ti}(\theta_0) m_{Ti}(\theta_0)'$. Thus, $\hat{V}(\hat{\theta})$ is asymptotically the same as the fixed- T consistent variance matrix estimator of Breitung et al. (2022) (see their equation (6)). Again, despite the discussion in this section being under the set-up of homoskedastic errors, it also applies to the case of cross-sectional heteroskedasticity.

3.3. Inference under Large T

If $T \rightarrow \infty$, regardless of N , upon substituting (5), one can check that the term $y'_{(-\ell)} A u$, $\ell = 1, \dots, p$, in $W' A u$ will be dominated by $u' A (I_N \otimes \Phi_p^{-1} L^\ell) u + u' A (I_N \otimes \Phi_p^{-1} L^\ell) X \beta_0$. When there is a unit root, $u' A (I_N \otimes \Phi_p^{-1} L^1) \alpha$ is also a dominating term. In the following, the stable and unit-root cases are discussed separately, since the leading terms in $y'_{(-\ell)} A u$ are of different orders. The relevant matrices in the quadratic forms are uniformly (in T) bounded in row and column sums under the stable case, but not under the unit-root case. These differences create different distribution results for $\hat{\theta}$ that are mainly in terms of the scaling factors. In what follows, let λ_r , $r = 1, \dots, p$, denote the inverse of roots of $1 - \phi_1 z - \dots - \phi_p z^p = 0$ when $\phi = \phi_0$. The stable case refers to the situation if $|\lambda_r| < 1$, $r = 1, \dots, p$, and without loss of generality, the unit-root case refers to the scenario when $\lambda_1 = 1$ and $|\lambda_r| < 1$, $r = 2, \dots, p$.

3.3.1. Stable Case

In this case, the matrix $M \Phi_p^{-1} L^\ell$ in $u' A (I_N \otimes \Phi_p^{-1} L^\ell) u + u' A (I_N \otimes \Phi_p^{-1} L^\ell) X \beta_0 = u' (I_N \otimes M \Phi_p^{-1} L^\ell) u + u' (I_N \otimes M \Phi_p^{-1} L^\ell) X \beta_0$ is uniformly (with respect to T) bounded in row and column sums. This boundedness property also holds for $I_N \otimes M \Phi_p^{-1} L^\ell$ and thus the central limit theorem of Kelejian and Prucha (2010) on linear and quadratic forms can be invoked.¹² That is, as $T \rightarrow \infty$, $\Omega_{1,NT}^{-1/2} \sqrt{NT} g_{NT} \xrightarrow{d} N(0, I)$ (see Lemma A7), in which

$$\Omega_{1,NT} = \frac{1}{NT} \begin{pmatrix} \sigma^4 N \Gamma_1 + \sigma^2 \Gamma_2 & \sigma^2 F_1' \\ \sigma^2 F_1 & \sigma^2 X' A X \end{pmatrix}, \quad (13)$$

where Γ_1 and Γ_2 are $p \times p$ matrices, consisting respectively of elements $\text{tr}[(\Phi_p^{-1} L^\ell)' M \Phi_p^{-1} L^s]$ and $\beta_0' X' [I_N \otimes (\Phi_p^{-1} L^\ell)' M \Phi_p^{-1} L^s] X \beta_0$, $\ell, s = 1, \dots, p$, in their (ℓ, s) -positions, and F_1 is $k \times p$ with $X' (I_N \otimes M \Phi_p^{-1} L^\ell) X \beta_0$, $\ell = 1, \dots, p$, as its ℓ -th column. Note that $\Omega_{1,NT}$ is not the exact variance of $\sqrt{NT} g_{NT}$, which is Ω_{NT} as defined by (10). It is better interpreted as an approximation, namely, $\Omega_{1,NT} = \Omega_{NT} + o(1)$. It follows that $\Omega_{NT}^{-1/2} \sqrt{NT} g_{NT} = \Omega_{1,NT}^{-1/2} \sqrt{NT} g_{NT} + o_p(1)$, regardless of N . In contrast to the exact variance matrix Ω_{NT} , the variance matrix $\Omega_{1,NT}$ involves θ_0 , X , and σ^2 , but not the fixed effects, initial conditions, or the skewness and kurtosis of the error distribution.

Theorem 2. Under Assumptions 1–5 and that $DP(p)$ is dynamically stable, if $\text{plim}_{T \rightarrow \infty} G_{NT}$ and $\lim_{T \rightarrow \infty} \Omega_{1,NT}$ exist and are nonsingular, then as $T \rightarrow \infty$,

$$\sqrt{NT}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left(0, \left(\text{plim}_{T \rightarrow \infty} G_{NT} \right)^{-1} \left(\lim_{T \rightarrow \infty} \Omega_{1,NT} \right) \left(\text{plim}_{T \rightarrow \infty} G_{NT} \right)^{-1'} \right). \quad (14)$$

If N is finite, one could have replaced the convergence rate \sqrt{NT} by \sqrt{T} and denoted $\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}$ by \mathbf{G}_N and $\lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}$ by $\mathbf{\Omega}_{1,N}$. But since there is no restriction in N (except that the probability limit of \mathbf{G}_{NT} and the limit of $\mathbf{\Omega}_{1,NT}$ are well defined when $T \rightarrow \infty$), which may be finite or also diverge as $T \rightarrow \infty$, no attempt is made to implement such replacements here (and also in the next (sub)sections).¹³ Lemma A6 gives $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = \lim_{T \rightarrow \infty} \sigma^{-2} \mathbf{\Omega}_{1,NT}$. Combining all these results, one has the following.

Corollary 1. *Under the conditions in Theorem 2, as $T \rightarrow \infty$,*

$$\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(\mathbf{0}, \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}' \mathbf{A} \mathbf{W}\right)^{-1} \left(\lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}\right) \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}' \mathbf{A} \mathbf{W}\right)^{-1}\right) \quad (15)$$

$$\xrightarrow{d} N\left(\mathbf{0}, \sigma^4 \left(\lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}\right)^{-1}\right) \quad (16)$$

$$\xrightarrow{d} N\left(\mathbf{0}, \sigma^2 \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}' \mathbf{A} \mathbf{W}\right)^{-1}\right). \quad (17)$$

The asymptotic variances in (15)–(17) resemble the variance formula of a consistent OLS estimator. Recall that the WG estimator, which is also named the least squares dummy variable (LSDV) estimator, is essentially an OLS estimator and is consistent under large T , though it may still possess an asymptotic bias, depending on the divergence rates of N and T . The asymptotic distribution result (17) is the same as Theorem 1(ii) of Breitung et al. (2022) that is derived under cross-sectional heteroskedasticity.¹⁴ In view of Theorem 3 in Hahn and Kuersteiner (2002), the asymptotic variance (17) achieves the efficiency bound.

In practice, the asymptotic variance may be estimated by $\hat{\mathbf{G}}_{NT}^{-1} \hat{\mathbf{\Omega}}_{1,NT} \hat{\mathbf{G}}_{NT}^{-1'}$, where $\hat{\mathbf{\Omega}}_{1,NT} = \mathbf{\Omega}_{1,NT}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{G}}_{NT} = \mathbf{G}_{NT}(\hat{\boldsymbol{\theta}})$, with $\hat{\boldsymbol{\theta}}$ replacing $\boldsymbol{\theta}_0$ in $\mathbf{\Omega}_{1,NT}$ and \mathbf{G}_{NT} , respectively, or $(\mathbf{W}' \mathbf{A} \mathbf{W} / NT)^{-1} \hat{\mathbf{\Omega}}_1 (\mathbf{W}' \mathbf{A} \mathbf{W} / NT)^{-1}$, or $\hat{\sigma}^4 \hat{\mathbf{\Omega}}_{1,NT}^{-1}$, or $\hat{\sigma}^2 (\mathbf{W}' \mathbf{A} \mathbf{W} / NT)^{-1}$, where $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})' \mathbf{A} (\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}}) / [N(T - 1)]$. The last choice is perhaps the easiest one to use.

When $k = 0$ and $p = 1$, $\mathbf{\Omega}_{1,NT} / \sigma^4 = \text{tr}[(\boldsymbol{\Phi}_1^{-1} \mathbf{L})' \mathbf{M} \boldsymbol{\Phi}_1^{-1} \mathbf{L}] / T = 1 / (1 - \phi_0^2) + O(T^{-1})$, so

$$\sqrt{NT}(\hat{\phi} - \phi_0) \xrightarrow{d} N(0, 1 - \phi_0^2), \quad (18)$$

as $T \rightarrow \infty$. This is also the asymptotic distribution of the bias-corrected WG estimator in Hahn and Kuersteiner (2002).

3.3.2. Unit-Root Case

When there is a unit root, the leading terms in $\mathbf{W}' \mathbf{A} \mathbf{u}$ are linear forms in \mathbf{u} . The matrices in quadratic forms in \mathbf{u} that appear in $\mathbf{W}' \mathbf{A} \mathbf{u}$ are no longer uniformly (in T) bounded in row and column sums, but they are of lower order compared with the leading linear forms. Thus, define the scaling matrix

$$\mathbf{Y} = \begin{pmatrix} NT^3 \mathbf{I}_p & \mathbf{O}_{p \times k} \\ \mathbf{O}_{k \times p} & NT \mathbf{I}_k \end{pmatrix}.$$

Correspondingly, define $\mathbf{g}_{NT}^* = \mathbf{Y}^{-1} [\mathbf{W}' \mathbf{A} \mathbf{u} - E(\mathbf{W}' \mathbf{A} \mathbf{u})]$ and

$$\mathbf{\Omega}_{1,NT}^*(\boldsymbol{\alpha}) = \mathbf{Y}^{-1/2} \sigma^2 \begin{pmatrix} \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3(\boldsymbol{\alpha}) & \mathbf{F}_1' + \mathbf{F}_2'(\boldsymbol{\alpha}) \\ \mathbf{F}_1 + \mathbf{F}_2(\boldsymbol{\alpha}) & \mathbf{X}' \mathbf{A} \mathbf{X} \end{pmatrix} \mathbf{Y}^{-1/2}, \quad (19)$$

where $\boldsymbol{\Gamma}_2$ and \mathbf{F}_1 are defined in (13), $\boldsymbol{\Gamma}_3(\boldsymbol{\alpha})$ has $\mathbf{1}'(\boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell)' \mathbf{M} \boldsymbol{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} \boldsymbol{\alpha}' \boldsymbol{\alpha}$ in its (ℓ, s) -position, $\ell, s = 1, \dots, p$, and $\mathbf{F}_2(\boldsymbol{\alpha})$ has $\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{M} \boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell) \boldsymbol{\alpha}$ as its ℓ -th column, $\ell = 1, \dots, p$. Note that the matrix $\mathbf{\Omega}_{1,NT}^*(\boldsymbol{\alpha})$ involves the fixed effects. If the fixed effects are deterministic

and of finite magnitudes, then $\Omega_{1,NT}^*(\alpha)$ can be interpreted as the approximate variance of $Y^{1/2}g_{NT}^*$, namely, $\Omega_{1,NT}^*(\alpha) = \text{Var}(Y^{1/2}g_{NT}^*) + o(1)$. Lemma A15 in Appendix C shows that $\Omega_{1,NT}^{*-1/2}(\alpha)Y^{1/2}g_{NT}^* \xrightarrow{d} N(0, I)$ as $T \rightarrow \infty$. Denote $G_{NT}^* = G_{NT}^*(\theta_0)$ with $G_{NT}^*(\theta) = Y^{-1/2}\{\partial[W'A(y - W\theta) - E(W'A((y - W\theta)))]/\partial\theta'\}Y^{-1/2}$.

Theorem 3. Under Assumptions 3–5 and that $DP(p)$ has a unit root, assume that the fixed effects are deterministic and of finite magnitudes. If $\text{plim}_{T \rightarrow \infty} G_{NT}^*$ and $\lim_{T \rightarrow \infty} \Omega_{1,NT}^*(\alpha)$ exist and are nonsingular, then as $T \rightarrow \infty$,

$$Y^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left(\text{plim}_{T \rightarrow \infty} G_{NT}^*\right)^{-1} \left(\lim_{T \rightarrow \infty} \Omega_{1,NT}^*(\alpha)\right) \left(\text{plim}_{T \rightarrow \infty} G_{NT}^*\right)^{-1'}\right). \quad (20)$$

This, together with Lemmas A13 and A14, implies the following.

Corollary 2. Under the conditions in Theorem 3, as $T \rightarrow \infty$,

$$Y^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left(\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AWY^{-1/2}\right)^{-1} \left(\lim_{T \rightarrow \infty} \Omega_{1,NT}^*(\alpha)\right) \left(\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AWY^{-1/2}\right)^{-1'}\right) \quad (21)$$

$$\xrightarrow{d} N\left(0, \sigma^2 \left(\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AWY^{-1/2}\right)^{-1}\right). \quad (22)$$

When there are no fixed effects ($\alpha = 0$), but there are still relevant exogenous regressors contained in X , the asymptotic distribution result still holds as $T \rightarrow \infty$. When there is no X and $p = 1$, the leading term in $W'AW = \sum_{i=1}^N y'_{i,(-1)} M y_{i,(-1)}$ is $\sum_{i=1}^N \alpha_i^2 \mathbf{1}' L \Phi_1^{-1'} M \Phi_1^{-1} L \mathbf{1}$. As $T \rightarrow \infty$, $T^{-3} \mathbf{1}' L \Phi_1^{-1'} M \Phi_1^{-1} L \mathbf{1} \rightarrow 1/12$ and $T^{-3} \text{Var}(W'Au) \rightarrow \sigma^2 \sum_{i=1}^N \alpha_i^2 / 12$.

Corollary 3. Under Assumptions 4 and 5, if further $k = 0$, $p = 1$, $\phi_0 = 1$, the fixed effects are deterministic and of finite magnitudes, and $\alpha \neq 0$, then as $T \rightarrow \infty$,

$$\sqrt{NT^3}(\hat{\phi} - \phi_0) \xrightarrow{d} N\left(0, \frac{12\sigma^2}{N^{-1}\alpha'\alpha}\right). \quad (23)$$

This echoes Theorem 5 of Hahn and Kuersteiner (2002). If $N \rightarrow \infty$ also, then the asymptotic variance in (23) is replaced accordingly with $12\sigma^2 / (\lim_{N \rightarrow \infty} N^{-1}\alpha'\alpha)$ (assuming that the limit exists).¹⁵

It should be pointed out that unit-root situation considered here is different from those in Han and Phillips (2010) and Kruiniger (2018), both focusing on $DP(1)$ and ruling out trends due to the fixed effects and/or exogenous regressors. The specification in Han and Phillips (2010) is $y_{it} = \phi y_{i,t-1} + (1 - \phi)\alpha_i + u_{it}$. Their FDLS estimator has a unified convergence rate of $\sqrt{N(T-1)}$ for $\phi_0 \in (-1, 1]$ and the limiting distribution is Gaussian for any N/T ratio as long as $N(T-1) \rightarrow \infty$. The model in Kruiniger (2018) is $y_{it} = \phi y_{i,t-1} + x'_{it}(\beta - \phi\beta) + (1 - \phi)\alpha_i + u_{it}$. He shows that the modified ML approach (Lancaster 2002) is undesirable in the unit-root case since ϕ_0 is first-order unidentified and is only second-order identified under the large- N -finite- T asymptotic regime. His generalized modified ML estimator, on the other hand, exists in the large- N -finite- T asymptotic regime with probability approaching 1; it is uniquely identified with probability 1 if it exists, and the asymptotic distribution of the estimated autoregressive parameter is non-Gaussian with a convergence rate of $N^{1/4}$ in the unit-root case. In this paper, it is assumed that either there are always fixed effects or when the fixed effects do not exist, there are relevant exogenous regressors. In this framework, the convergence rate of the RMM estimator is different only under large T for the unit-root case and the asymptotic distribution is still normal.

4. Heteroskedasticity

This section considers the situation of heteroskedastic idiosyncratic errors. Assumption 5 is modified accordingly such that both forms of heteroskedasticity are allowed.

Assumption 6. The series of error terms u_{it} , $i = 1, \dots, N$, $t = 1, \dots, T$, is independent across i and t , $E(u_{it}) = 0$, $\text{Var}(u_{it}) = \sigma_{it}^2$, and has finite $(4 + \eta)$ -th moments, $\eta > 0$.

Let $\text{Var}(\mathbf{u}_i) = \mathbf{\Sigma}_i = \text{Dg}(\sigma_{i1}^2, \dots, \sigma_{iT}^2)$, where $\text{Dg}(\cdot)$ denotes a diagonal matrix with its arguments in order on the diagonal. (When $\text{Dg}(\cdot)$ has a matrix argument, it collects, in order, the diagonal elements of the matrix and forms a diagonal matrix.) It is obvious that

$$E(\mathbf{W}'\mathbf{A}\mathbf{u}) = \sum_{i=1}^N \begin{pmatrix} \text{tr}(\mathbf{M}\mathbf{\Phi}_p^{-1}\mathbf{L}^1\mathbf{\Sigma}_i) \\ \vdots \\ \text{tr}(\mathbf{M}\mathbf{\Phi}_p^{-1}\mathbf{L}^p\mathbf{\Sigma}_i) \\ \mathbf{0}_k \end{pmatrix}. \quad (24)$$

Following Bao and Yu (2023), define

$$\mathbf{\Psi}_\ell(\boldsymbol{\phi}) = \frac{T}{T-2} \text{Dg}(\mathbf{M}\mathbf{\Phi}_p^{-1}(\boldsymbol{\phi})\mathbf{L}^\ell) - \frac{\text{tr}(\mathbf{M}\mathbf{\Phi}_p^{-1}(\boldsymbol{\phi})\mathbf{L}^\ell)}{(T-1)(T-2)} \mathbf{I}.$$

Then, using $\text{tr}(\mathbf{M}\mathbf{\Phi}_p^{-1}\mathbf{L}^\ell\mathbf{\Sigma}_i) = E(\mathbf{u}_i'\mathbf{M}\mathbf{\Psi}_\ell\mathbf{M}\mathbf{u}_i)$, where $\mathbf{\Psi}_\ell = \mathbf{\Psi}_\ell(\boldsymbol{\phi}_0)$, one may define the moment vector as

$$\begin{aligned} \mathbf{g}_{NT}(\boldsymbol{\theta}) &= \frac{1}{NT} \sum_{i=1}^N \begin{pmatrix} \mathbf{y}_{i,(-1)}' \mathbf{M}(\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}) - (\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta})' \mathbf{M}\mathbf{\Psi}_1(\boldsymbol{\phi})\mathbf{M}(\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}) \\ \vdots \\ \mathbf{y}_{i,(-p)}' \mathbf{M}(\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}) - (\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta})' \mathbf{M}\mathbf{\Psi}_p(\boldsymbol{\phi})\mathbf{M}(\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}) \\ \mathbf{X}_i' \mathbf{M}(\mathbf{y}_i - \mathbf{W}_i\boldsymbol{\theta}) \end{pmatrix} \\ &\equiv \frac{1}{NT} \sum_{i=1}^N \mathbf{g}_i(\boldsymbol{\theta}) \end{aligned} \quad (25)$$

such that $E(\mathbf{g}_{NT}) = \mathbf{0}$. Note that this set of moment conditions is still valid under Assumption 5. In the previous section under homoskedasticity, σ^2 is treated as a nuisance parameter, and it is replaced with $\mathbf{u}'\mathbf{A}\mathbf{u}/[N(T-1)]$ as if it were estimated. (It has also been emphasized that $\mathbf{u}'\mathbf{A}\mathbf{u}/[N(T-1)]$, in fact, is unbiased for estimating $\sum_{i=1}^N \sigma_i^2$ under cross-sectional heteroskedasticity.) Under the most general form of heteroskedasticity, there is no feasible way of estimating all the variance parameters unbiasedly. So this section does not seek to estimate $\mathbf{\Sigma}_i$. Instead, the set of moment conditions (25) is designed such that it is robust to heteroskedasticity.

Denote $\mathbf{\Omega}_{NT} = \text{Var}(\sqrt{NT}\mathbf{g}_{NT})$ as before, but its expression is different, see (A9) in Appendix C. The gradient (evaluated at $\boldsymbol{\theta}_0$) now becomes

$$\begin{aligned} \mathbf{G}_{NT} &= \frac{1}{NT} \sum_{i=1}^N \begin{pmatrix} 2\mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_1 \mathbf{M}\mathbf{W}_i - \mathbf{y}_{i,(-1)}' \mathbf{M}\mathbf{W}_i \\ \vdots \\ 2\mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_p \mathbf{M}\mathbf{W}_i - \mathbf{y}_{i,(-p)}' \mathbf{M}\mathbf{W}_i \\ -\mathbf{X}_i' \mathbf{M}\mathbf{W}_i \end{pmatrix} \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \begin{pmatrix} \mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_{11} \mathbf{M}\mathbf{u}_i & \cdots & \mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_{1p} \mathbf{M}\mathbf{u}_i & \mathbf{0}_k' \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_{p1} \mathbf{M}\mathbf{u}_i & \cdots & \mathbf{u}_i' \mathbf{M}\mathbf{\Psi}_{pp} \mathbf{M}\mathbf{u}_i & \mathbf{0}_k' \\ \mathbf{0}_k & \cdots & \mathbf{0}_k & \mathbf{O}_k \end{pmatrix}, \end{aligned} \quad (26)$$

where

$$\Psi_{\ell s} = \frac{\partial \Psi_{\ell}(\theta_0)}{\partial \phi_s} = \frac{T}{T-2} \text{Dg}(\mathbf{M}\Phi_p^{-1}\mathbf{L}^s\Phi_p^{-1}\mathbf{L}^{\ell}) - \frac{\text{tr}(\mathbf{M}\Phi_p^{-1}\mathbf{L}^s\Phi_p^{-1}\mathbf{L}^{\ell})}{(T-1)(T-2)}\mathbf{I}. \quad (27)$$

By substituting (4), one can see that $\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_{\ell}\mathbf{M}\mathbf{u}_i$ that populates $\mathbf{g}_i(\theta_0)$ is in terms of linear and quadratic forms in \mathbf{u}_i (see (A10) in the proof of Lemma A1 in Appendix C). This structure is to be exploited for one to study the asymptotic distribution of the resulting RMM estimator from the set of sample moment conditions (25).

4.1. Inference under Large N

Given any finite T , for each $\ell = 1, \dots, p$, one can verify that the Lyapunov condition holds by checking $N^{-1} \sum_{i=1}^N \text{Var}(\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_{\ell}\mathbf{M}\mathbf{u}_i) = O(1)$ (see Lemma A1 in Appendix C for the expression of $\text{Var}(\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_{\ell}\mathbf{M}\mathbf{u}_i)$) and $E(|\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_{\ell}\mathbf{M}\mathbf{u}_i|^{2+\delta}) < \infty$ for some $\delta > 0$. The Lyapunov condition also holds for $\mathbf{u}'_i\mathbf{M}\mathbf{X}_i$, since $N^{-1} \sum_{i=1}^N \text{Var}(\mathbf{u}'_i\mathbf{M}\mathbf{X}_i) = N^{-1} \sum_{i=1}^N \text{tr}(\mathbf{M}\mathbf{X}_i\mathbf{X}'_i\mathbf{M}\Sigma_i) = O(1)$ and $E(|\mathbf{u}'_i\mathbf{M}\mathbf{X}_i|^{2+\delta}) < \infty$. Moreover, for any linear combination of $\mathbf{y}'_{i,(-1)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_1\mathbf{M}\mathbf{u}_i, \dots, \mathbf{y}'_{i,(-p)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_p\mathbf{M}\mathbf{u}_i$, and $\mathbf{u}'_i\mathbf{M}\mathbf{X}_i$, the Lyapunov condition holds as well. Thus, with Ω_{NT} and \mathbf{G}_{NT} updated, one has the same asymptotic distribution result as in Theorem 1, and to save space, it is not listed explicitly as a separate theorem.

4.2. Inference under Large T

As in the previous section, cases without and with a unit root are discussed separately, since the orders of various dominating terms in \mathbf{G}_{NT} and Ω_{NT} are different.

4.2.1. Stable Case

Note that, for any $\ell = 1, \dots, p$, $\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i$ is dominated by $\mathbf{u}'_i\mathbf{M}\Phi_p^{-1}\mathbf{L}^{\ell}(\mathbf{u}_i + \mathbf{X}_i\beta_0)$. $\text{Dg}(\mathbf{M}\Phi_p^{-1}\mathbf{L}^{\ell})$ has diagonal elements of order $O(T^{-1})$ (see Lemma A2 in Appendix C). Therefore, $\mathbf{y}'_{i,(-\ell)}\mathbf{M}\mathbf{u}_i - \mathbf{u}'_i\mathbf{M}\Psi_{\ell}\mathbf{M}\mathbf{u}_i$ is dominated by $\mathbf{u}'_i\mathbf{M}\Phi_p^{-1}\mathbf{L}^{\ell}(\mathbf{u}_i + \mathbf{X}_i\beta_0)$.

Theorem 4. Under Assumptions 1–4 and 6 and that $DP(p)$ is dynamically stable, if $\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}$ and $\lim_{T \rightarrow \infty} \Omega_{1,NT}$ exist and are nonsingular, then as $T \rightarrow \infty$,

$$\sqrt{NT}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \left(\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}\right)^{-1} \left(\lim_{T \rightarrow \infty} \Omega_{1,NT}\right) \left(\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}\right)^{-1'}\right), \quad (28)$$

where

$$\Omega_{1,NT} = \frac{1}{NT} \begin{pmatrix} \Gamma_1 + \Gamma_2 & \mathbf{F}'_1 \\ \mathbf{F}_1 & \sum_{i=1}^N \mathbf{X}'_i\mathbf{M}\Sigma_i\mathbf{M}\mathbf{X}_i \end{pmatrix} \quad (29)$$

in which Γ_1 and Γ_2 are $p \times p$ matrices with, respectively, $\sum_{i=1}^N \text{tr}(\Sigma_i\mathbf{L}^{\ell'}\Phi_p^{-1'}\mathbf{M}\Sigma_i\mathbf{M}\Phi_p^{-1}\mathbf{L}^s)$ and $\sum_{i=1}^N \beta'_0\mathbf{X}'_i\mathbf{L}^{\ell'}\Phi_p^{-1'}\mathbf{M}\Sigma_i\mathbf{M}\Phi_p^{-1}\mathbf{L}^s\mathbf{X}_i\beta_0$ in their (ℓ, s) -th positions, $\ell, s = 1, \dots, p$, and \mathbf{F}_1 is $k \times p$ with $\mathbf{X}'_i\mathbf{M}\Sigma_i\mathbf{M}\Phi_p^{-1}\mathbf{L}^{\ell}\mathbf{X}_i\beta_0$ as its ℓ -th column, $\ell = 1, \dots, p$.

In view of Lemmas A9 and A10 in Appendix C, one also has the following corollary.

Corollary 4. Under the assumptions in Theorem 4, as $T \rightarrow \infty$,

$$\begin{aligned} & \sqrt{NT}(\hat{\theta} - \theta_0) \\ & \xrightarrow{d} N\left(\mathbf{0}, \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}'\mathbf{A}\mathbf{W}\right)^{-1} \left(\lim_{T \rightarrow \infty} \Omega_{1,NT}\right) \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}'\mathbf{A}\mathbf{W}\right)^{-1}\right) \end{aligned} \quad (30)$$

$$\xrightarrow{d} N\left(\mathbf{0}, \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}' \mathbf{A} \mathbf{W}\right)^{-1} \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \boldsymbol{\Sigma}_i \mathbf{M} \mathbf{W}_i\right) \left(\text{plim}_{T \rightarrow \infty} \frac{1}{NT} \mathbf{W}' \mathbf{A} \mathbf{W}\right)^{-1}\right). \quad (31)$$

The asymptotic variance (31) resembles the OLS variance result under heteroskedasticity in cross-sectional regressions. In practice, for meaningful inference, one needs to estimate the meat part of the sandwich form of the asymptotic variance. From the proof of Lemma A8, \mathbf{g}_{NT} is dominated by $(NT)^{-1} \sum_{i=1}^N (\mathbf{y}'_{i(-1)} \mathbf{M} \mathbf{u}_i, \dots, \mathbf{y}'_{i(-p)} \mathbf{M} \mathbf{u}_i, \mathbf{u}'_i \mathbf{M} \mathbf{X}_i)' = (NT)^{-1} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{u}_i$. Therefore, as $T \rightarrow \infty$,

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{u}_i \mathbf{u}'_i \mathbf{M} \mathbf{W}_i &= \lim_{T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N E(\mathbf{W}'_i \mathbf{M} \mathbf{u}_i \mathbf{u}'_i \mathbf{M} \mathbf{W}_i) \\ &= \lim_{T \rightarrow \infty} \text{Var}(\sqrt{NT} \mathbf{g}_{NT}) \\ &= \lim_{T \rightarrow \infty} \boldsymbol{\Omega}_{1,NT}. \end{aligned}$$

But still, \mathbf{u}_i is not observable. Given that $\mathbf{M} \mathbf{u}_i = \mathbf{M}(\mathbf{y}_i - \mathbf{W}_i \boldsymbol{\theta}_0) \equiv \mathbf{M} \mathbf{v}_i$, one may be tempted to use $\mathbf{M} \hat{\mathbf{v}}_i = \mathbf{M}(\mathbf{y}_i - \mathbf{W}_i \hat{\boldsymbol{\theta}}) = \mathbf{M}[\mathbf{u}_i + \mathbf{W}_i(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})]$. By substitution,

$$\begin{aligned} &\frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}'_i \mathbf{M} \mathbf{W}_i \\ &= \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{u}_i \mathbf{u}'_i \mathbf{M} \mathbf{W}_i + \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{W}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{W}'_i \mathbf{M} \mathbf{W}_i \\ &\quad - \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{W}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mathbf{u}'_i \mathbf{M} \mathbf{W}_i - \frac{1}{NT} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{u}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{W}'_i \mathbf{M} \mathbf{W}_i, \end{aligned}$$

where $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(1/\sqrt{NT})$, $\mathbf{W}'_i \mathbf{M} \mathbf{W}_i = O_p(T)$, and $\mathbf{W}'_i \mathbf{M} \mathbf{u}_i = O(1) + O_p(\sqrt{T/N})$. If it is also the case that N diverges, then $(NT)^{-1} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{W}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mathbf{u}'_i \mathbf{M} \mathbf{W}_i = o_p(1)$ and $(NT)^{-1} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{W}_i (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})' \mathbf{W}'_i \mathbf{M} \mathbf{W}_i = o_p(1)$.

Therefore, if $N, T \rightarrow \infty$, a feasible consistent estimator of the asymptotic variance is

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\theta}}) = NT(\mathbf{W}' \mathbf{A} \mathbf{W})^{-1} \left(\sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}'_i \mathbf{M} \mathbf{W}_i \right) (\mathbf{W}' \mathbf{A} \mathbf{W})^{-1}. \quad (32)$$

The asymptotic distribution result (31) with the variance matrix estimated by (32) resembles Theorem 3 of Hansen (2007). Appendix D discusses the asymptotic distribution of $(NT)^{-1} \sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}'_i \mathbf{M} \mathbf{W}_i$ if N is fixed and its implication on asymptotic inference when one is using (32). In particular, the $\sqrt{N/(N-1)} t_{N-1}$ approximation proposed by Hansen (2007), where t_{N-1} is the t distribution with $N-1$ degrees of freedom, may be used to approximate the asymptotic distribution of the t -statistic.

4.2.2. Unit-Root Case

In this case, the diagonal elements of $\mathbf{M} \boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell \boldsymbol{\Sigma}_i = \boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell \boldsymbol{\Sigma}_i - T^{-1} \mathbf{1} \mathbf{1}' \boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell \boldsymbol{\Sigma}_i$ are $O(1)$, since $\boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell \boldsymbol{\Sigma}_i$ is strictly lower triangular and $T^{-1} \mathbf{1} \mathbf{1}' \boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell$ has $O(1)$ elements. Further, $\boldsymbol{\Psi}_\ell$ is dominated by $(T-2)^{-1} T \text{Dg}(\boldsymbol{\Phi}_p^{-1} \mathbf{L}^\ell \boldsymbol{\Sigma}_i)$. As before, the scaling matrix \mathbf{Y} is used such that the set of moment conditions $\mathbf{g}_{NT}^*(\boldsymbol{\theta})$ and gradient matrix $\mathbf{G}_{NT}^*(\boldsymbol{\theta})$ are defined accordingly. Specifically, $\mathbf{g}_{NT}^*(\boldsymbol{\theta})$ is $\mathbf{Y}^{-1} [NT \mathbf{g}_{NT}(\boldsymbol{\theta})]$ with $\mathbf{g}_{NT}(\boldsymbol{\theta})$ given by (25) and $\mathbf{G}_{NT}^*(\boldsymbol{\theta})$ is $\mathbf{Y}^{-1/2} [NT \mathbf{G}_{NT}(\boldsymbol{\theta}) \mathbf{Y}^{-1/2}]$ with $\mathbf{G}_{NT}(\boldsymbol{\theta})$ given by (26).

Theorem 5. Under Assumptions 3, 4 and 6 and that $DP(p)$ has a unit root, assume the fixed effects are deterministic and of finite magnitudes. If $\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}^*$ and $\lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}^*(\alpha)$ exist and are nonsingular, then as $T \rightarrow \infty$,

$$\mathbf{Y}^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{N}\left(\mathbf{0}, \left(\text{plim}_{T \rightarrow \infty} \mathbf{G}^*\right)^{-1} \left(\lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}^*(\alpha)\right) \left(\text{plim}_{T \rightarrow \infty} \mathbf{G}^*\right)^{-1'}\right), \quad (33)$$

where

$$\mathbf{\Omega}_{1,NT}^*(\alpha) = \mathbf{Y}^{-1/2} \begin{pmatrix} \mathbf{\Gamma}_2 + \mathbf{\Gamma}_3(\alpha) & \mathbf{F}'_1 + \mathbf{F}'_2(\alpha) \\ \mathbf{F}_1 + \mathbf{F}_2(\alpha) & \sum_{i=1}^N \mathbf{X}'_i \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{X}_i \end{pmatrix} \mathbf{Y}^{-1/2}, \quad (34)$$

in which $\mathbf{\Gamma}_2$ and \mathbf{F}_1 are defined in (29), $\mathbf{\Gamma}_3(\alpha)$ is $p \times p$ with $\sum_{i=1}^N \mathbf{1}' \mathbf{\Sigma}_i \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} \alpha_i^2$ in its (ℓ, s) -th position, $\ell, s = 1, \dots, p$, and $\mathbf{F}_2(\alpha)$ is $k \times p$ with $\sum_{i=1}^N \mathbf{X}'_i \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^{\ell} \mathbf{1} \alpha_i$ as its ℓ -th column, $\ell = 1, \dots, p$.

Based on Lemmas A17 and A18, one also has the following result.

Corollary 5. Under the assumptions in Theorem 5, as $T \rightarrow \infty$,

$$\mathbf{Y}^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{V}), \quad (35)$$

where $\mathbf{V} = (\text{plim}_{T \rightarrow \infty} \mathbf{Y}^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{W} \mathbf{Y}^{-1/2})^{-1} \text{plim}_{T \rightarrow \infty} \mathbf{Y}^{-1/2} (\sum_{i=1}^N \mathbf{W}'_i \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{W}_i) \mathbf{Y}^{-1/2} (\text{plim}_{T \rightarrow \infty} \mathbf{Y}^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{W} \mathbf{Y}^{-1/2})^{-1}$.

Similar to (32), for practical inference, one needs $N, T \rightarrow \infty$ so that $\mathbf{W}'_i \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{W}_i$ can be replaced with $\mathbf{W}'_i \mathbf{M} \hat{\boldsymbol{\vartheta}}_i \hat{\boldsymbol{\vartheta}}'_i \mathbf{M} \mathbf{W}_i$ in (35) and the probability limits are replaced with the sample analogs to form an estimated asymptotic variance.

5. Monte Carlo Evidence

In this section, Monte Carlo simulations are conducted to assess the finite-sample performance of the proposed estimator. Bao and Yu (2023) provide simulation results for the first- and second-order models when the idiosyncratic errors are either homoskedastic or temporally heteroskedastic. Recall that under the baseline framework, the proposed estimator in this paper is also robust to cross-sectional heteroskedasticity. This section considers a third-order model, and for one to get a comprehensive spectrum of possible heteroskedasticity, four scenarios are included: homoskedasticity (across i and t), cross-sectional heteroskedasticity (across i only), temporal heteroskedasticity (over t only), and double (both cross-sectional and temporal) heteroskedasticity.

The following DP(3) is used:

$$\begin{aligned} y_{it} &= \alpha_i + \phi_1 y_{i,t-1} + \phi_2 y_{i,t-2} + \phi_3 y_{i,t-3} + \beta_1 x_{1,it} + \beta_2 x_{2,it} + u_{it}, \\ x_{1,it} &= \rho_x x_{1,i,t-1} + \xi_{1,it}, \quad \rho_x = 0.8, \\ x_{2,it} &= \rho_i \alpha_i + \xi_{2,it}, \end{aligned} \quad (36)$$

where α_i is i.i.d. (across i) following a standard normal distribution, ρ_i is i.i.d. (across i) uniformly on the interval $[0, 1]$, $\xi_{1,it}$ is i.i.d. (across i and t) following a standard normal distribution, and $\xi_{2,it}$ is similarly defined. The initial value of $x_{1,it}$ is set to be $\xi_{0,i} / \sqrt{1 - \rho_x^2}$, where $\xi_{0,i}$ is i.i.d. (across i) following a standard normal distribution. The initial observations on y_{it} are simulated as $\alpha_i / (1 - \phi_{01} - \phi_{02} - \phi_{03}) + \mathbf{x}'_{it} \boldsymbol{\beta}_0 + u_{it} \sqrt{\text{Var}(\delta_t)}$ if there is no unit root and $\alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta}_0 + u_{it}$ otherwise, where δ_t follows a stationary zero-mean third-order autoregressive (AR(3)) process with coefficients ϕ_{01} , ϕ_{02} , and ϕ_{03} , and its shock term is a unit-variance white noise.

Let e_{it} be i.i.d. (across i and t) with mean 0 and variance 1. The error term u_{it} is simulated as follows in the four different scenarios (homoskedasticity, cross-sectional heteroskedasticity, temporal heteroskedasticity, and double heteroskedasticity):

$$u_{it} = \begin{cases} e_{it} & z_i \sim U[0.5, i] \text{ or } \chi_{10}^2 \text{ if } U[0.5, i] \geq 100 \\ \sqrt{z_i}e_{it} & z_t \sim U[0.5, t^2] \text{ or } \chi_{10}^2 \text{ if } U[0.5, t^2] \geq 100 \\ \sqrt{z_t}e_{it} & z_{it} \sim U[0.5, i] \cdot U[0.5, t^2] \text{ or } \chi_{10}^2 \text{ if } U[0.5, i] \cdot U[0.5, t^2] \geq 100 \end{cases} \quad (37)$$

where, for example, $z_i \sim U[0.5, i] \text{ or } \chi_{10}^2 \text{ if } U[0.5, i] \geq 100$ means that z_i is independently (across i) drawn from a uniform distribution on the interval $[0.5, i]$, and if the realization is no smaller than 100, then it is redrawn from a chi-squared distribution with 10 degrees of freedom. e_{it} is simulated from a normal distribution, but results under non-normal distributions are available upon request and the conclusions in this section are largely consistent across all distributions.¹⁶

In reality, data themselves rarely reveal clearly the lack or presence of heteroskedasticity, so estimation results from the estimator using the recentered moment conditions (7) and the robust one using (25), referred to as RMM and RMM_r, respectively, are presented. When the empirical rejection rates of a 5% two-sided t -test of the relevant parameter equal to its true value are reported, for WG and GMM, also included are their empirical sizes from t -ratios using the White-type robust standard errors (clustered at the individual level, referred to as WG(h) and GMM(h), respectively). There are different choices for the GMM estimator, depending on what and how many instruments are used. [Breitung et al. \(2022\)](#) show that the one-step estimator of [Arellano and Bond \(1991\)](#) is very comparable to other popular choices ([Ahn and Schmidt 1995](#); [Blundell and Bond 1998](#)), so the GMM estimator used in this section is the one-step one.¹⁷ Different combinations of N and T are experimented with: [100 10; 50 20; 50, 50; 25 40; 20 50; 10 100]. Regardless of N and T , one can always use the estimator in this paper, but in reality, in a situation like $N = 25$, $T = 40$, to conduct inference, there is typically no convincing evidence for one to favor one asymptotic regime over the other. So in what follows, in addition to the bias and root mean squared error (RMSE), all out of 10,000 simulations, also reported are the empirical sizes using standard errors constructed under different asymptotics. In particular, RMM(N) and RMM_r(N) denote the empirical sizes from t -ratios using the large- N standard errors (see Sections 3.2 and 4.1) for RMM and RMM_r, respectively. Likewise, RMM(T) refers to the empirical size from the t -ratio when the large- T standard errors from (17) (or (22)) are used for RMM. Furthermore, RMM_r(NT) means when the feasible variance (32) (or its version under unit root) is used for RMM_r, which is valid if both N and T are large. On the other hand, RMM_r(T) does not mean that a different t -ratio is used and instead, it signals that, for the RMM_r t -ratio when the feasible variance is used, the $\sqrt{N/(N-1)}t_{N-1}$ approximation developed by [Hansen \(2007\)](#) under large T is used for conducting hypothesis testing.

Included for comparison are the bias-corrected estimators of [Bun and Carree \(2006\)](#) (BC for short) and [Juodis \(2013\)](#) (BCJ for short).¹⁸ As [Juodis \(2013\)](#) points out, BC is consistent under homoskedasticity and cross-sectional heteroskedasticity, but invalid under temporal heteroskedasticity. BCJ, on the other hand, is robust to both forms of heteroskedasticity. Neither [Bun and Carree \(2006\)](#) nor [Juodis \(2013\)](#) derives the asymptotic distributions of their bias-corrected estimators, so only their bias and RMSE results are reported in this section. Included also is the half-panel jackknife (HPJ) estimator of [Chudik et al. \(2018\)](#), where errors can be heteroskedastic across both i and t .

In the experiments, $\beta_0 = (1, 1)'$ and three sets of parameter configurations of ϕ_0 are used: $(0.3, 0.3, 0.2)'$, $(0.3, -0.2, -0.1)'$, and $(0.3, 0.6, 0.1)'$. The first configurations reflects a situation when the degree of time-series correlation is relatively strong in the sense that the cumulative partial effect of a past shock, measured by powers of $\phi_{01} + \phi_{02} + \phi_{03}$, can be high. The second set corresponds to a case of zero past effect, and the last one is a unit-root case where the past effect never dies out. To save space, only results related to $\phi_{01} + \phi_{02} + \phi_{03}$

under the first set of parameter configuration $(0.3, 0.3, 0.2)'$ are reported, whereas results under the other two parameter configurations are contained in Appendix F.¹⁹

Table 1 reports the bias and RMSE, both multiplied by 100, and empirical rejection rate of the two-sided 5% t -test related to $\phi_{01} + \phi_{02} + \phi_{03}$ when u_{it} is homoskedastic. One sees clearly the superb performance of both RMM and RMM_r, with the smallest bias and lowest RMSE. GMM performs reasonably well, but its bias and RMSE peak at $(N, T) = (50, 20)$. The WG estimator performs the worst when T is small but improves as T gets relatively larger, but is still more biased even when $T = 100$ compared with RMM or RMM_r. BC and BCJ report larger bias and higher RMSE than GMM on many occasions, though there are also cases where they are much better than GMM. Note that BC and BCJ bias-correct the WG estimator, which is in fact consistent under large T . Table A1 in Appendix F reveals that when there is a unit root, BC and BCJ give a much larger bias than WG, even when T is relatively large. So, in this case, the action of bias-correction gives more biased estimates. HPJ is the most biased and possesses the highest RMSE among all the consistent estimators at $(N, T) = (100, 10)$, namely, when the panel is relatively short. Its performance improves quickly as T increases, though still quite a bit below RMM/RMM_r.

Table 1. Simulation results: $\phi_0 = (0.3, 0.3, 0.2)'$ under homoskedasticity.

	(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
Bias ($\times 100$)	WG	−5.75	−2.26	−0.66	−0.91	−0.70	−0.33
	GMM	−1.59	−2.20	−0.66	−0.91	−0.70	−0.75
	BC	0.50	2.15	0.92	1.23	0.89	0.26
	BCJ	0.44	2.15	0.92	1.23	0.89	0.26
	HPJ	3.45	1.10	0.23	0.35	0.21	0.06
	RMM	−0.01	−0.02	−0.01	−0.03	−0.04	−0.04
	RMM _r	−0.01	−0.02	−0.01	−0.03	−0.04	−0.04
RMSE ($\times 100$)	WG	6.06	2.60	0.88	1.34	1.15	0.87
	GMM	2.48	2.55	0.88	1.34	1.15	1.52
	BC	2.33	2.71	1.11	1.66	1.33	0.86
	BCJ	2.31	2.70	1.11	1.66	1.33	0.86
	HPJ	4.65	2.23	0.75	1.32	1.17	0.91
	RMM	1.93	1.31	0.57	0.98	0.92	0.81
	RMM _r	1.95	1.31	0.57	0.98	0.92	0.81
Size(5%)	WG	88.54	44.09	20.95	15.80	12.31	6.98
	WG(h)	86.78	44.01	22.13	18.65	15.52	12.47
	GMM	13.32	42.10	20.93	15.72	12.21	8.64
	GMM(h)	14.37	43.12	22.56	19.39	16.32	16.76
	HPJ	21.36	8.96	4.51	4.68	4.22	3.87
	RMM(N)	5.70	6.45	6.13	7.49	8.17	10.52
	RMM(T)	6.45	6.25	5.50	5.70	5.56	5.34
	RMM _r (N)	5.74	6.50	6.14	7.47	8.13	10.52
	RMM _r (NT)	6.65	6.98	6.62	7.84	8.41	10.52
	RMM _r (T)	6.01	5.98	5.23	5.56	5.53	4.40

Note: WG is the within-group estimator; GMM is the one-step estimator of Arellano and Bond (1991); BC is the bias-corrected estimator of Bun and Carree (2006); BCJ is the bias-corrected estimator of Juodis (2013); HPJ is the half-panel jackknife estimator of Chudik et al. (2018); RMM and RMM_r are, respectively, the recentered method of moments estimator in this paper and its robust version. For the size performance, WG(h) and GMM(h) are based on clustered standard errors at the individual level, RMM(N) and RMM_r(N) are based on the large- N standard errors, RMM(T) is based on the large- T standard error, RMM_r(NT) is based on the standard error that is valid under large N and large T , and RMM_r(T) is based on the $\sqrt{N/(N-1)}t_{N-1}$ approximation. Reported are the bias and RMSE (both scaled up by 100) out of 10,000 simulations of the estimated $\phi_{01} + \phi_{02} + \phi_{03}$ and the empirical size of the 5% two-sided t -test for testing the sum of the autoregressive parameters equal to its true value. The DGP follows (36) and (37).

Now, consider the size performance of the associated two-sided 5% test from the different estimators. From Table 1, one sees severe size distortions of the WG-based inference when T is relatively small, but its empirical size is close to the nominal size when $T = 100$. The GMM-based inference performs really poorly. The t -test based on HPJ severely over-rejects when $(N, T) = (100, 10)$, but its size distortion goes down quickly when the panel has longer spans. The large- N -based inferences from RMM and RMM_r, namely, RMM(N) and RMM_r(N), give empirical sizes close to 5% when N is relatively

large. On the other hand, $RMM(T)$ delivers a reasonably good size performance even when T is relatively small. Results from $RMM_r(NT)$ and $RMM_r(N)$ are mixed: when they work, they may perform slightly worse than $RMM(N)$ and $RMM(T)$, but their performances get worse when N is small and T is large. Finally, the size result from $RMM_r(T)$, namely, using the $\sqrt{N/(N-1)}t_{N-1}$ approximation for the asymptotic distribution of the t -ratio, is good in almost all cases. Recall that the $\sqrt{N/(N-1)}t_{N-1}$ approximation is valid under homoskedasticity. These results are not surprising, given that homoskedastic errors are simulated and that, for the robust estimator, the feasible standard errors based on (32) (or its version under unit root) require both N and T to be large. In fact, when $(N, T) = (50, 50)$, the size distortion from $RMM_r(NT)$ is the smallest compared with other (N, T) combinations with relatively smaller N or T .

Table 2 reports results under cross-sectional heteroskedasticity. Relative performances of these different estimators largely stay the same. Notably, BC and BCJ perform really poorly when $(N, T) = (100, 10)$, though they are designed to bias-correct the WG estimator under large N and fixed T . In contrast to the homoskedastic case, HPJ gives a smaller bias and lower RMSE than RMM/RMM_r when $(N, T) = (100, 10)$, but usually performs worse in other cases. In terms of hypothesis testing, when N is relatively large, both $RMM(N)$ and $RMM_r(N)$ work well; when T is relatively large, both $RMM(T)$ and $RMM_r(T)$ provide empirical rejection rates close to the nominal size. Recall that the RMM estimator designed under homoskedasticity is also valid under cross-sectional heteroskedasticity, and so is the related inference procedure. There is pronounced size distortion from HPJ for the unit root case (see Table A2 in Appendix F), but its size performance is reasonably good under the other two parameter configurations when T is not relatively small.

When there is time-series heteroskedasticity, the robust estimator usually dominates all the other estimators in terms of bias and RMSE, as demonstrated in Table 3 (and Table A3 in Appendix F). In terms of hypothesis testing, the large- N RMM_r t -ratio delivers very good size performance when N is relatively large, whereas the $\sqrt{N/(N-1)}t_{N-1}$ approximation ($RMM_r(T)$) provides good results when T is relatively large. The large- N -large- T RMM_r -based inference reports upward size distortions in almost all cases.

Finally, Table 4 provides results under double heteroskedasticity, namely, when both cross-sectional heteroskedasticity and time-series heteroskedasticity are present. RMM_r is the least biased, except when $(N, T) = (100, 10)$, where HPJ is slightly better. On the other hand, RMM , which ignores temporal heteroskedasticity, reports very small bias on many occasions, especially under the second and third parameter configurations (see Table A4 in Appendix F). In terms of RMSE, BCJ and RMM_r are comparable. In terms of the size performance of the associated t -test, the story is very similar to that reported when there is temporal heteroskedasticity only, namely, under large N , $RMM_r(N)$ is most trustworthy and $RMM_r(T)$ is the one under large T . GMM has the most size distortions in almost all cases, and HPJ gives severe size distortions when there is a unit root.

Summarizing all the simulation results, one can learn the following. (i) When there is no heteroskedasticity, the proposed estimator can be safely used, either the one based on (7) or the robust one based on (25), regardless of N and T . When heteroskedasticity in the time dimension is present, then the robust version is the best. (ii) The presence of a unit root or not has no substantial impact on its performance. (iii) In terms of inference, when N is relatively large or is of comparable size relative to T , the large N -based inference from RMM_r has very good size performance, regardless of heteroskedasticity. When T is relatively large, the large T -based inference from RMM gives reliable inference under homoskedasticity and cross-sectional heteroskedasticity. (iv) When T is large, the $\sqrt{N/(N-1)}t_{N-1}$ approximation for the t -ratio from RMM_r , with the feasible robust variance usually has good size performance, regardless of heteroskedasticity.

Table 2. Simulation results: $\phi_0 = (0.3, 0.3, 0.2)'$ under cross-sectional heteroskedasticity.

	(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
Bias ($\times 100$)	WG	−39.47	−14.10	−4.51	−4.18	−2.78	−0.84
	GMM	−21.83	−14.08	−4.51	−4.18	−2.78	−1.89
	BC	−10.21	−3.01	−1.13	−0.69	−0.35	−0.03
	BCJ	−10.24	−3.01	−1.13	−0.69	−0.35	−0.03
	HPJ	0.69	3.07	1.12	1.26	0.82	0.15
	RMM	1.18	−0.10	−0.10	−0.13	−0.12	−0.09
	RMM _r	1.00	−0.08	−0.10	−0.13	−0.12	−0.09
RMSE ($\times 100$)	WG	39.87	14.57	4.82	4.74	3.41	1.57
	GMM	23.05	14.57	4.82	4.74	3.41	2.89
	BC	12.56	5.09	2.03	2.35	2.00	1.33
	BCJ	12.59	5.09	2.03	2.35	2.00	1.33
	HPJ	9.66	6.70	2.59	3.40	2.77	1.54
	RMM	11.14	4.57	1.77	2.32	2.01	1.34
	RMM _r	11.08	4.61	1.77	2.32	2.01	1.34
Size (5%)	WG	100.00	99.25	84.05	52.52	33.41	9.73
	WG(<i>h</i>)	100.00	97.68	78.30	50.04	34.99	15.40
	GMM	89.51	99.14	84.02	52.28	33.22	14.19
	GMM(<i>h</i>)	85.70	97.55	78.96	51.70	36.68	23.15
	HPJ	8.00	13.17	10.04	8.42	7.02	4.20
	RMM(<i>N</i>)	6.47	5.78	6.84	7.65	8.48	10.35
	RMM(<i>T</i>)	42.66	16.90	10.09	8.36	7.60	6.09
	RMM _r (<i>N</i>)	6.37	5.64	6.83	7.67	8.42	10.38
	RMM _r (<i>NT</i>)	24.67	11.40	8.46	8.35	9.07	10.22
	RMM _r (<i>T</i>)	23.20	9.63	7.09	6.17	5.92	4.27

Note: See Table 1.

Table 3. Simulation results: $\phi_0 = (0.3, 0.3, 0.2)'$ under temporal heteroskedasticity.

	(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
Bias ($\times 100$)	WG	−34.39	−19.57	−5.78	−8.18	−6.01	−2.37
	GMM	−22.39	−20.05	−5.78	−8.18	−6.01	−6.23
	BC	−3.76	−5.74	−2.06	−3.18	−2.27	−0.64
	BCJ	−4.46	−5.71	−2.06	−3.17	−2.27	−0.64
	HPJ	9.37	−3.21	0.74	0.17	0.63	0.43
	RMM	17.84	−2.01	−0.62	−1.16	−0.82	−0.27
	RMM _r	0.22	0.22	−0.11	−0.24	−0.30	−0.20
RMSE ($\times 100$)	WG	34.84	19.92	6.04	8.70	6.63	3.31
	GMM	23.86	20.41	6.04	8.70	6.63	7.41
	BC	7.45	7.56	2.73	4.49	3.63	2.38
	BCJ	7.68	7.55	2.73	4.49	3.63	2.38
	HPJ	13.59	6.98	2.59	4.37	3.98	2.86
	RMM	18.62	5.27	1.94	3.41	3.03	2.36
	RMM _r	6.84	5.39	1.91	3.44	3.05	2.36
Size (5%)	WG	100.00	100.00	94.76	85.08	64.10	18.74
	WG(<i>h</i>)	100.00	99.97	92.51	81.63	62.22	23.52
	GMM	84.78	100.00	94.76	84.94	63.95	45.46
	GMM(<i>h</i>)	81.94	99.96	92.85	82.73	64.03	45.42
	HPJ	24.21	13.71	9.41	8.92	8.36	6.40
	RMM(<i>N</i>)	0.32	10.69	8.00	9.49	8.96	11.01
	RMM(<i>T</i>)	92.02	17.16	9.67	10.68	8.74	7.00
	RMM _r (<i>N</i>)	5.57	4.94	6.36	6.92	7.71	10.93
	RMM _r (<i>NT</i>)	9.26	12.81	9.64	11.75	10.44	11.41
	RMM _r (<i>T</i>)	8.61	11.24	8.18	8.15	7.07	4.90

Note: See Table 1.

Table 4. Simulation Results: $\phi_0 = (0.3, 0.3, 0.2)'$ under double heteroskedasticity.

	(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
Bias ($\times 100$)	WG	−37.33	−16.83	−4.58	−6.95	−5.18	−2.29
	GMM	−17.49	−16.50	−4.58	−6.95	−5.18	−5.89
	BC	−9.98	−4.70	−1.22	−2.31	−1.66	−0.63
	BCJ	−9.81	−4.66	−1.22	−2.30	−1.66	−0.63
	HPJ	−0.37	1.46	1.27	1.47	1.21	0.42
	RMM	−6.25	−2.71	−0.31	−0.88	−0.58	−0.30
	RMM _r	0.91	0.13	−0.08	−0.27	−0.24	−0.25
RMSE ($\times 100$)	WG	37.70	17.24	4.85	7.52	5.83	3.18
	GMM	18.59	16.92	4.85	7.52	5.83	7.04
	BC	11.84	6.48	2.02	3.73	3.15	2.28
	BCJ	11.72	6.47	2.01	3.73	3.15	2.28
	HPJ	8.91	6.35	2.57	4.46	3.92	2.75
	RMM	10.11	5.23	1.68	3.13	2.82	2.26
	RMM _r	10.01	5.60	1.68	3.16	2.84	2.26
Size (5%)	WG	100.00	99.94	87.27	77.04	56.65	18.48
	WG(<i>h</i>)	100.00	99.71	83.81	72.77	55.02	24.20
	GMM	83.66	99.91	87.22	76.84	56.45	42.85
	GMM(<i>h</i>)	81.15	99.61	84.34	73.94	56.53	44.42
	HPJ	9.01	12.02	12.82	11.46	10.67	6.20
	RMM(<i>N</i>)	19.97	11.99	6.61	8.48	8.68	10.59
	RMM(<i>T</i>)	45.32	22.73	8.55	10.64	9.04	6.57
	RMM _r (<i>N</i>)	6.18	4.62	6.28	7.32	8.36	10.53
	RMM _r (<i>NT</i>)	23.10	15.80	8.83	10.76	10.67	10.88
	RMM _r (<i>T</i>)	21.93	14.08	7.32	7.95	7.13	4.70

Note: See Table 1.

6. Conclusions and Directions of Future Research

This paper proposes an estimation strategy that does not rely on instrumental variables. One can view the estimation strategy as using the endogenous lagged dependent variables as their own instruments and then constructing recentered moment conditions by explicitly exploiting the correlation between the endogenous variables and error term in the model. The asymptotic properties of the new estimator are thoroughly investigated under various conditions that relate to the sizes of cross-sectional units and time periods, heterogeneity in the error variance, and the issue of whether a unit root is present. In general, the asymptotic distribution does not require both N and T to be large. Under large T , it resembles the familiar OLS result in traditional regression analysis and its asymptotic variance achieves the efficiency bound under homoskedasticity. Similar to time series autoregressions, the convergence rate of the estimator of the autoregressive parameters is different when there is a unit root under large T , but the standard t -test procedure carries through in hypothesis testing. Monte Carlo simulations demonstrate that it possesses good finite-sample properties in various situations. All the theoretical results and Monte Carlo evidence in this paper suggest two directions for future research.

6.1. Cross-Sectional Correlation

Note that cross-sectional correlation is not explicitly discussed in this paper. If there is a weak cross-sectional correlation (in the sense that the covariance matrix of $(u_{1t}, \dots, u_{Nt})'$ has a bounded norm as $N \rightarrow \infty$), then the estimation strategy proposed in this paper still holds, see discussions in Appendix E. However, both N and T need to be large. This form of cross-sectional correlation may cover situations where correlation arises from spatial contiguity as in the spatial econometrics literature. Nevertheless, when there are dominating units (Pesaran and Yang 2021) whose errors are always correlated with those from other units, the assumption of weak correlation is violated. Another situation of violation is when cross-sectional correlation is due to common factors. Under homoskedasticity, for the case of DP(1), De Vos and Everaert (2021) augment the model by the cross-sectional averages of the right-hand side regressors, which can be interpreted

as proxies for the unknown factors. Then, they derive the asymptotic bias of the common correlated effects pooled (CCEP) estimator of the main parameter vector under large N and fixed T and design an estimator by matching the CCEP estimator with its asymptotic bias.²⁰ One could have directly focused on the endogeneity of the defactored lagged dependent variables, namely, $(I_T - P)w_i$, where P is the projection matrix on the proxies, consisting of cross-sectional averages (including 1_T). The exact expectation of $\sum_{i=1}^N w_i'(I_T - P)u_i$ is not easy to derive, but one can follow De Vos and Everaert (2021) to approximate it and construct the resulting estimator. It is left for future research to extend the robust estimator to cases when there is a strong cross-sectional correlation in addition to heteroskedasticity in higher-order DP with the possibly of a unit root.

6.2. Inference under Heteroskedasticity for Long Panels

It has been observed from Section 5 that all the other estimators in the experiments perform much worse in terms of bias under cross-sectional and temporal heteroskedasticity than the robust estimator when T is not very small. On the other hand, non-negligible size distortions using the feasible variance (32) are also documented and one may need to resort to the approximation of Hansen (2007) for inference purposes, though one has yet to show rigorously that the approximation is still valid in the presence of heteroskedasticity. The re-sampling approach (Kapetanios 2008) may not be applicable under small N and large T . In addition, if one re-samples the data across i (when N is large and T is small) or over t (when T is large and N is small), there is the issue of heteroskedasticity or temporal correlation that one needs to take into account. Designing a practical and reliable inference procedure for the robust estimator in long panels is a second avenue for future research.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Acknowledgments: The author thanks three anonymous referees, Mohitosh Kejriwal, Joon Park, Keli Xu, Xuewen Yu, Xiaoyan Zhou, and seminar participants at Fudan University and Indiana University (Bloomington) for their helpful comments.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Some Preliminary Results

Note that $\Phi_p(\phi)$ is a lower triangular Toeplitz matrix with non-zero diagonals, so it is always invertible and its inverse is also lower triangular Toeplitz. From Linz (1985, p. 172), one has

$$\Phi_p^{-1}(\phi) = \begin{pmatrix} 1 & & & & \\ b_1 & 1 & & & \\ b_2 & b_1 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{T-1} & \cdots & \cdot & b_1 & 1 \end{pmatrix},$$

where b_t can be obtained from the recursion, $b_t = \sum_{s=0}^{t-1} \phi_{t-s} b_s$, $t = 1, \dots, T-1$, $b_0 = 1$ and $\phi_{t-s} = 0$ for $t-s > p$. This defines a p -th order difference equation for the series $\{b_t\}_{t=0}^{T-1}$. The solution of the p -th order homogeneous equation $b_t - \phi_1 b_{t-1} - \dots - \phi_p b_{t-p} = 0$ is

$$b_t = \sum_{r=1}^p a_r \lambda_r^t, \quad (A1)$$

where λ_i 's are the inverse of the roots of $1 - \phi_1 z - \dots - \phi_p z^p = 0$, and a_r 's are constants determined by the first p initial conditions, where, without loss of generality, it is assumed that all the λ_i 's are distinct.²¹

For any $\ell \geq 0$, let the (i, j) -th element of the $T \times T$ matrix $\Phi_p^{-1} L^\ell$ be denoted by $b_{i-j-\ell} = O(1)$ with the understanding that $b_{i-j-\ell} = 0$ whenever $i - j - \ell < 0$. The (i, j) -th element of the $T \times T$ matrix $\Phi_p^{-1} L^\ell \Sigma_k$ is then $b_{i-j-\ell} \sigma_{kj}^2$.

When the DP(p) is dynamically stable, the i -th element of $\mathbf{1}' \Phi_p^{-1} L^\ell$ is the sum of the i -th column of $\Phi_p^{-1} L^\ell$, given by

$$\sum_{j=i+\ell}^T b_{j-i-\ell} = \sum_{r=1}^p a_r \sum_{j=i+\ell}^T \lambda_r^{j-i-\ell} = \sum_{r=1}^p a_r \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r}, \quad i \leq T + 1 - \ell. \quad (\text{A2})$$

The i -th element of $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_k$ is

$$\sum_{j=i+\ell}^T b_{j-i-\ell} \sigma_{ki}^2 = \sigma_{ki}^2 \sum_{r=1}^p a_r \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r} \quad i \leq T + 1 - \ell.$$

The i -th element of $\Phi_p^{-1} L^\ell \mathbf{1}$ is the sum of the i -th row of $\Phi_p^{-1} L^\ell$, given by

$$\sum_{j=1}^{i-\ell} b_{i-j-\ell} = \sum_{r=1}^p a_r \sum_{j=1}^{i-\ell} \lambda_r^{i-j-\ell} = \sum_{r=1}^p a_r \frac{1 - \lambda_r^{i-\ell}}{1 - \lambda_r}, \quad \ell \leq i. \quad (\text{A3})$$

The i -th element of $\Phi_p^{-1} L^\ell \Sigma_k \mathbf{1}$ is

$$\sum_{j=1}^{i-\ell} b_{i-j-\ell} \sigma_{kj}^2 = \sum_{r=1}^p a_r \sum_{j=1}^{i-\ell} \sigma_{kj}^2 \lambda_r^{i-j-\ell}.$$

The (i, j) -th element of $(\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell$ is

$$\sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{j_1=1}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} = \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} w_{i,j,s,\ell,r_1,r_2}, \quad (\text{A4})$$

where

$$w_{i,j,s,\ell,r_1,r_2} = \begin{cases} \frac{\lambda_{r_2}^{i+s-j-\ell} - \lambda_{r_1}^{T+1-i-s} \lambda_{r_2}^{T+1-j-\ell}}{1 - \lambda_{r_1} \lambda_{r_2}} & j + \ell \leq i + s \leq T + 1 \\ \frac{\lambda_{r_1}^{j+\ell-i-s} - \lambda_{r_1}^{T+1-i-s} \lambda_{r_2}^{T+1-j-\ell}}{1 - \lambda_{r_1} \lambda_{r_2}} & i + s < j + \ell \leq T + 1 \end{cases}.$$

The (i, j) -th element of $(\Phi_p^{-1} L^s \Sigma_k)' \Phi_p^{-1} L^\ell \Sigma_k$ is

$$\sigma_{ki}^2 \sigma_{kj}^2 \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{j_1=1}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} = \sigma_{ki}^2 \sigma_{kj}^2 \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} w_{i,j,s,\ell,r_1,r_2}.$$

When there is a unit root, let $\lambda_1 = 1$ and $|\lambda_r| < 1$ be all distinct, $r = 2, \dots, p$. Now, the (i, j) -th element of $\Phi_p^{-1} L^\ell$ is $b_{i-j-\ell}$ with $b_{i-j-\ell} = a_1 + \sum_{r=2}^p a_r \lambda_r^{i-j-\ell}$. Note that the i -th element of $\mathbf{1}' \Phi_p^{-1} L^\ell$ is

$$\begin{aligned} \sum_{j=i+\ell}^T b_{j-i-\ell} &= \sum_{r=1}^p a_r \sum_{j=i+\ell}^T \lambda_r^{j-i-\ell} \\ &= a_1(T + 1 - i - \ell) + \sum_{r=2}^p a_r \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r}, \quad i \leq T + 1 - \ell. \end{aligned} \quad (\text{A5})$$

The i -th element of $\Phi_p^{-1} L^\ell \mathbf{1}$ is

$$\sum_{j=1}^{i-\ell} b_{i-j-\ell} = \sum_{r=1}^p a_r \sum_{j=1}^{i-\ell} \lambda_r^{i-j-\ell} = a_1(i-\ell) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{i-\ell}}{1-\lambda_r}, \quad \ell \leq i. \quad (\text{A6})$$

The (i, j) -th element of $(\Phi_p^{-1} \mathbf{L}^s)' \Phi_p^{-1} \mathbf{L}^\ell$ is, when $i + s \geq j + \ell$,

$$\begin{aligned} & \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{j_1=i+s}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} \\ &= a_1^2(T+1-i-s) + a_1 \sum_{r_2=2}^p a_{r_2} \sum_{j_1=i+s}^T \lambda_{r_2}^{j_1-j-\ell} \\ &+ a_1 \sum_{r_1=2}^p a_{r_1} \sum_{j_1=i+s}^T \lambda_{r_1}^{j_1-i-s} + \sum_{r_1=2}^p \sum_{r_2=2}^p a_{r_1} a_{r_2} \sum_{j_1=i+s}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} \\ &= a_1^2(T+1-i-s) + a_1 \sum_{r=2}^p a_r \frac{(1+\lambda_r^{i+s-j-\ell})(1-\lambda_r^{T+1-i-s})}{1-\lambda_r} \\ &+ \sum_{r_1=2}^p \sum_{r_2=2}^p a_{r_1} a_{r_2} w_{i,j,s,\ell,r_1,r_2}, \end{aligned} \quad (\text{A7})$$

where w_{i,j,s,ℓ,r_1,r_2} is given by (A4), and when $i + s < j + \ell$,

$$\begin{aligned} & \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{j_1=j+t}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} \\ &= a_1^2(T+1-j-\ell) + a_1 \sum_{r_2=2}^p a_{r_2} \sum_{j_1=j+t}^T \lambda_{r_2}^{j_1-j-\ell} \\ &+ a_1 \sum_{r_1=2}^p a_{r_1} \sum_{j_1=j+t}^T \lambda_{r_1}^{j_1-i-s} + \sum_{r_1=2}^p \sum_{r_2=2}^p a_{r_1} a_{r_2} \sum_{j_1=j+t}^T \lambda_{r_1}^{j_1-i-s} \lambda_{r_2}^{j_1-j-\ell} \\ &= a_1^2(T+1-j-\ell) + a_1 \sum_{r=2}^p a_r \frac{(1+\lambda_r^{j+t-i-s})(1-\lambda_r^{T+1-j-\ell})}{1-\lambda_r} \\ &+ \sum_{r_1=2}^p \sum_{r_2=2}^p a_{r_1} a_{r_2} w_{i,j,s,\ell,r_1,r_2}. \end{aligned} \quad (\text{A8})$$

Appendix B. Discussion of Several Related Estimators

It is worth discussing several closely related estimators. Bao and Yu (2023) motivate their estimator by matching the inconsistent WG estimator $\hat{\theta}_{WG} = (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbf{W}'\mathbf{A}\mathbf{y} = \theta_0 + (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbf{W}'\mathbf{A}\mathbf{u}$, when T is finite, with its approximate analytical expectation. More specifically, their indirect inference estimator is based on solving a random sample binding function, namely, $\hat{\theta}_{II} = \arg_{\theta} \{\hat{\theta}_{WG} = \theta + (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbb{E}[\mathbf{W}'\mathbf{A}\mathbf{u}(\theta)]\}$, where $\mathbf{A}\mathbf{u}(\theta) = \mathbf{A}(\mathbf{y} - \mathbf{W}\theta)$ and $\mathbb{E}[\mathbf{W}'\mathbf{A}\mathbf{u}(\theta)]$ is an analytical function of θ (and σ^2). Since, by definition, $\hat{\theta}_{WG} - \theta_0 = (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbf{W}'\mathbf{A}\mathbf{u}$, their estimation strategy numerically amounts to matching $\mathbf{W}'\mathbf{A}\mathbf{u}$ with $\mathbb{E}(\mathbf{W}'\mathbf{A}\mathbf{u})$, as considered in this paper, provided that $(\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}$ is nonsingular. Under time-series heteroskedasticity, they introduce a robust estimator that matches $\hat{\mathbb{E}}(\mathbf{W}'\mathbf{A}\mathbf{u})$ with $\mathbb{E}(\mathbf{W}'\mathbf{A}\mathbf{u})$, where $\hat{\mathbb{E}}(\mathbf{W}'\mathbf{A}\mathbf{u})$ is a function of θ_0 and \mathbf{W} but not involving the variance parameters. Their estimator is robust in the sense that the moment conditions are valid under time-series heteroskedasticity, and the variance parameters $\sigma_{it}^2 = \text{Var}(u_{it})$ are not estimated explicitly.

Under temporal heteroskedasticity, Alvarez and Arellano (2022) consider explicitly estimating σ_{it}^2 , such that the whole set of moment conditions includes both first-order conditions (of the log-likelihood function) with respect to θ and those with respect to the variance parameters. When there is no temporal heteroskedasticity, the estimator of

Alvarez and Arellano (2022) is essentially based on moment conditions matching $W' Au$ with $E(W' Au)$ (without the variance parameter concentrated out) plus an additional one for the variance. Under cross-sectional heteroskedasticity, namely, $\text{Var}(u_{it}) = \sigma_i^2$, Breitung et al. (2022) construct their estimator by matching the “numerator” of the profile score function (associated with the parameter vector ϕ) with its expectation. This, together with the exogeneity condition on X , constitutes the moment condition for their estimator. As discussed in the main text, it turns out that the estimator arising from matching $W' Au$ with $E(W' Au)$ under homoskedasticity is, in fact, robust to pure cross-sectional heteroskedasticity, and the estimator of Bao and Yu (2023) is equivalent to that of Breitung et al. (2022). However, neither estimator explicitly allows for heteroskedasticity in both cross-sectional and time-series dimensions. Breitung et al. (2022) mention this extension under stringent restrictions, but do not proceed to formally investigate the properties of the resulting estimator.

Bun and Carree (2006) and Juodis (2013) consider both forms of heteroskedasticity. In their bias-correction procedure, the variance parameters are estimated and then plugged into the bias expression to bias-correct the WG estimator. Notwithstanding, they focus on short panels and do not derive the asymptotic distribution of their bias-corrected estimator.

Appendix C. Lemmas and Proofs

For the ease of presentation, let $C_\ell = M\Phi_p^{-1}L^\ell$, $C_{\ell s} = C'_\ell C_s$, $D_\ell = \text{Dg}(C_\ell)$, $B_\ell = C_\ell - M\Psi_\ell M$, $C_{s,\ell-s} = M\Phi_p^{-1}\Phi_s L^{\ell-s-1}$, and $C_{(s-p),\ell-s} = M\Phi_p^{-1}\Phi_{(s-p)} L^{\ell-s-1}$. Recall that $\Phi_p = \Phi_p(\phi_0) = I - \phi_{01}L - \dots - \phi_{0p}L^p$, $\Phi_{(s-p)} = \Phi_s(\phi_0) - \Phi_p(\phi_0)$, $\Psi_\ell = (T-2)^{-1}TD_\ell - (T-2)^{-1}(T-1)^{-1}\text{tr}(D_\ell)I$. Denote $\Sigma_i^{(3)} = \text{Dg}(E(u_{i1}^3), \dots, E(u_{iT}^3))$ and $\Sigma_i^{(4)} = \text{Dg}(E(u_{i1}^4) - 3\sigma_{i1}^4, \dots, E(u_{iT}^4) - 3\sigma_{iT}^4)$ when there is heteroskedasticity. $\text{dg}(\cdot)$ creates a column vector consisting of the diagonal elements in order of its argument, and \odot is the matrix Hadamard (element-by-element) product operator.

Lemma A1. (i) Under Assumption 5 or 6, $u' Au = O_P(NT)$. (ii) Under Assumptions 1–5, $\text{Var}(\sqrt{NT}g_{NT}) = (NT)^{-1}\text{Var}(W' Au) - \sigma^4[2(T-1)/T + \gamma_2(T-1)^2/T^2]hh'$. (iii) Under Assumptions 1–4 and 6,

$$\text{Var}(\sqrt{NT}g_{NT}) = \frac{1}{NT} \sum_{i=1}^N \begin{pmatrix} c_{i,1,1} & \cdots & c_{i,1,p} & d'_{i,1} \\ \vdots & \ddots & \vdots & \vdots \\ c_{i,p,1} & \cdots & c_{i,p,p} & d'_{i,p} \\ d_{i,1} & \cdots & d_{i,p} & X'_i M \Sigma_i M X_i \end{pmatrix}, \quad (\text{A9})$$

where $c_{i,\ell,s} = \text{Cov}(y'_{i,(-\ell)} Mu_i - u'_i M \Psi_\ell Mu_i, y'_{i,(-s)} Mu_i - u'_i M \Psi_s Mu_i)$ is given by the sum of the 15 terms in (A11) and $d_{i,\ell} = \text{Cov}(X'_i Mu_i, y'_{i,(-\ell)} Mu_i - u'_i M \Psi_\ell Mu_i)$ is given by (A12).

Proof. (i): Using moments of quadratic forms (e.g., Bao and Ullah 2010), one has $E(u' Au) = \sigma^2 N \text{tr}(M) = \sigma^2 N(T-1)$ under Assumption 5 and $E(u' Au) = \sum_{i=1}^N \text{tr}(M \Sigma_i)$ under Assumption 6. Note that M is uniformly (in T) bounded in row and column sums and Σ_i has $O(1)$ elements. Thus, using Lemma A2 of Bao et al. (2020), $\text{tr}(M \Sigma_i) = O(T)$, leading to $\sum_{i=1}^N \text{tr}(M \Sigma_i) = O(NT)$. Further, $\text{Var}(u' Au) = \sigma^4 N[2\text{tr}(M) + \gamma_2 \text{tr}(M \odot M)] = \sigma^4 N[2(T-1) + \gamma_2(T-1)^2/T] = O(NT)$ under Assumption 5 and $\text{Var}(u' Au) = \sum_{i=1}^N [\text{tr}(\Sigma_i^{(4)} \odot M \odot M) + 2\text{tr}(\Sigma_i M \Sigma_i M)]$ under Assumption 6, where it is obvious that $\text{tr}(\Sigma_i^{(4)} \odot M \odot M) = O(T)$. Using Lemmas A2 of Bao et al. (2020), again, one can claim $\text{tr}(\Sigma_i M \Sigma_i M) = O(T)$. So in either case, one has $\text{Var}(u' Au) = O(NT)$ and one can claim that $u' Au = O_P(NT)$. (ii): See Supplementary Appendix D of Bao and Yu (2023). (iii): With $g_{NT} = (NT)^{-1} \sum_{i=1}^N (y'_{i,(-1)} Mu_i - u'_i M \Psi_1 Mu_i, \dots, y'_{i,(-p)} Mu_i - u'_i M \Psi_p Mu_i, u'_i M X_i)'$, one can derive $\text{Var}(\sqrt{NT}g_{NT})$ such that its top-left $p \times p$ block contains in its (ℓ, s) -th position $(NT)^{-1} \sum_{i=1}^N \text{Cov}(y'_{i,(-\ell)} Mu_i - u'_i M \Psi_\ell Mu_i, y'_{i,(-s)} Mu_i - u'_i M \Psi_s Mu_i)$, $\ell, s =$

$1, \dots, p$, its lower-left $k \times p$ block consists of $k \times 1$ columns like $(NT)^{-1} \sum_{i=1}^N \text{Cov}(\mathbf{X}_i' \mathbf{M} \mathbf{u}_i, \mathbf{y}_{i,(-\ell)}' \mathbf{M} \mathbf{u}_i - \mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_\ell \mathbf{M} \mathbf{u}_i)$, $\ell = 1, \dots, p$, and its lower-right $k \times k$ block is $(NT)^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{X}_i$. Substituting (4) into (25), one has, for $\ell = 1, \dots, p$,

$$\begin{aligned} \mathbf{y}_{i,(-\ell)}' \mathbf{M} \mathbf{u}_i - \mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_\ell \mathbf{M} \mathbf{u}_i &= \mathbf{u}_i' \mathbf{B}_\ell \mathbf{u}_i + \alpha_i \mathbf{u}_i' \mathbf{C}_\ell \mathbf{1} + \mathbf{u}_i' \mathbf{C}_\ell \mathbf{X}_i \beta_0 \\ &+ \sum_{j=0}^{\ell-1} \mathbf{u}_i' \mathbf{C}_{j,\ell-j} \mathbf{e}_1 \mathbf{y}_{i,-j} + \sum_{j=\ell}^{p-1} \mathbf{u}_i' \mathbf{C}_{(j-p),\ell-j} \mathbf{e}_1 \mathbf{y}_{i,-j}. \end{aligned} \quad (\text{A10})$$

Thus, $\text{Cov}(\mathbf{y}_{i,(-\ell)}' \mathbf{M} \mathbf{u}_i - \mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_\ell \mathbf{M} \mathbf{u}_i, \mathbf{y}_{i,(-s)}' \mathbf{M} \mathbf{u}_i - \mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_s \mathbf{M} \mathbf{u}_i)$ has the following 15 terms,

$$\begin{aligned} &\text{tr}(\mathbf{\Sigma}_i^{(4)} \odot \mathbf{B}_\ell \odot \mathbf{B}_s) + \text{tr}[\mathbf{\Sigma}_i \mathbf{B}_\ell \mathbf{\Sigma}_i (\mathbf{B}_s + \mathbf{B}_s')], \\ &\text{Var}(\alpha_i) \mathbf{1}' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_s \mathbf{1}, \\ &\beta_0' \mathbf{X}_i' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_s \mathbf{X}_i \beta_0, \\ &\sum_{j_1=1}^{\ell-1} \sum_{j_2=1}^{s-1} \text{E}(\mathbf{y}_{i,-j_1} \mathbf{y}_{i,-j_2}) \mathbf{e}_1' \mathbf{C}_{j_1,\ell-j_1}' \mathbf{\Sigma}_i \mathbf{C}_{j_2,s-j_2} \mathbf{e}_1, \\ &\sum_{j_1=\ell}^{p-1} \sum_{j_2=s}^{p-1} \text{E}(\mathbf{y}_{i,-j_1} \mathbf{y}_{i,-j_2}) \mathbf{e}_1' \mathbf{C}_{(j_1-p),\ell-j_1}' \mathbf{\Sigma}_i \mathbf{C}_{(j_2-p),s-j_2} \mathbf{e}_1, \\ &\text{E}(\alpha_i) [\text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_\ell) \mathbf{C}_s \mathbf{1} + \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_s) \mathbf{C}_\ell \mathbf{1}], \\ &\text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_\ell) \mathbf{C}_s \mathbf{X}_i \beta_0 + \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_s) \mathbf{C}_\ell \mathbf{X}_i \beta_0, \\ &\sum_{j=0}^{s-1} \text{E}(\mathbf{y}_{i,-j}) \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_\ell) \mathbf{u}_i' \mathbf{C}_{j,s-j} \mathbf{e}_1 + \sum_{j=0}^{\ell-1} \text{E}(\mathbf{y}_{i,-j}) \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_s) \mathbf{u}_i' \mathbf{C}_{j,\ell-j} \mathbf{e}_1, \\ &\sum_{j=s}^{p-1} \text{E}(\mathbf{y}_{i,-j}) \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_\ell) \mathbf{u}_i' \mathbf{C}_{(j-p),s-j} \mathbf{e}_1 + \sum_{j=\ell}^{p-1} \text{E}(\mathbf{y}_{i,-j}) \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_s) \mathbf{u}_i' \mathbf{C}_{(j-p),\ell-j} \mathbf{e}_1, \\ &\text{E}(\alpha_i) (\mathbf{1}' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_s \mathbf{X}_i \beta_0 + \mathbf{1}' \mathbf{C}_s' \mathbf{\Sigma}_i \mathbf{C}_\ell \mathbf{X}_i \beta_0), \\ &\sum_{j=0}^{s-1} \text{E}(\mathbf{y}_{i,-j} \alpha_i) \mathbf{1}' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_{j,s-j} \mathbf{e}_1 + \sum_{j=0}^{\ell-1} \text{E}(\mathbf{y}_{i,-j} \alpha_i) \mathbf{1}' \mathbf{C}_s' \mathbf{\Sigma}_i \mathbf{C}_{j,\ell-j} \mathbf{e}_1, \\ &\sum_{j=s}^{p-1} \text{E}(\mathbf{y}_{i,-j} \alpha_i) \mathbf{1}' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_{(j-p),s-j} \mathbf{e}_1 + \sum_{j=\ell}^{p-1} \text{E}(\mathbf{y}_{i,-j} \alpha_i) \mathbf{1}' \mathbf{C}_s' \mathbf{\Sigma}_i \mathbf{C}_{(j-p),\ell-j} \mathbf{e}_1, \\ &\sum_{j=0}^{s-1} \text{E}(\mathbf{y}_{i,-j}) \beta_0' \mathbf{X}_i' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_{j,s-j} \mathbf{e}_1 + \sum_{j=0}^{\ell-1} \text{E}(\mathbf{y}_{i,-j}) \beta_0' \mathbf{X}_i' \mathbf{C}_s' \mathbf{\Sigma}_i \mathbf{C}_{j,\ell-j} \mathbf{e}_1, \\ &\sum_{j=s}^{p-1} \text{E}(\mathbf{y}_{i,-j}) \beta_0' \mathbf{X}_i' \mathbf{C}_\ell' \mathbf{\Sigma}_i \mathbf{C}_{(j-p),s-j} \mathbf{e}_1 + \sum_{j=\ell}^{p-1} \text{E}(\mathbf{y}_{i,-j}) \beta_0' \mathbf{X}_i' \mathbf{C}_s' \mathbf{\Sigma}_i \mathbf{C}_{(j-p),\ell-j} \mathbf{e}_1, \\ &\sum_{j_1=0}^{\ell-1} \sum_{j_2=s}^{p-1} \text{E}(\mathbf{y}_{i,-j_1} \mathbf{y}_{i,-j_2}) \mathbf{e}_1' \mathbf{C}_{j_1,\ell-j_1}' \mathbf{\Sigma}_i \mathbf{C}_{(j_2-p),s-j_2} \mathbf{e}_1 \\ &+ \sum_{j_1=0}^{s-1} \sum_{j_2=\ell}^{p-1} \text{E}(\mathbf{y}_{i,-j_1} \mathbf{y}_{i,-j_2}) \mathbf{e}_1' \mathbf{C}_{j_1,s-j_1}' \mathbf{\Sigma}_i \mathbf{C}_{(j_2-p),\ell-j_2} \mathbf{e}_1, \end{aligned} \quad (\text{A11})$$

and

$$\begin{aligned}
& \text{Cov}(\mathbf{X}_i' \mathbf{M} \mathbf{u}_i, \mathbf{y}_{i(-\ell)}' \mathbf{M} \mathbf{u}_i - \mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_\ell \mathbf{M} \mathbf{u}_i) \\
&= \mathbf{X}_i' \mathbf{M} \text{dg}(\mathbf{\Sigma}_i^{(3)} \odot \mathbf{B}_\ell) + \text{E}(\alpha_i) \mathbf{X}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{C}_\ell \mathbf{1} + \mathbf{X}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{C}_\ell \mathbf{X}_i \beta_0 \\
&+ \sum_{j=0}^{\ell-1} \text{E}(y_{i,-j}) \mathbf{X}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{C}_{j,\ell-j} \mathbf{e}_1 + \sum_{j=\ell}^{p-1} \text{E}(y_{i,-j}) \mathbf{X}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{C}_{(j-p),\ell-j} \mathbf{e}_1
\end{aligned} \tag{A12}$$

in view of results on moments of quadratic forms (Bao and Ullah 2010; Ullah 2004). \square

Lemma A2. For the stable case, namely, $|\lambda_r| < 1$ for all $r = 1, \dots, p$, the following hold for $\ell, s = 1, \dots, T$: (i) $\mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{1} = O(T)$; (ii) $\mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} = O(T)$; (iii) $\mathbf{1}' \mathbf{C}_{\ell s} \mathbf{1} = O(1)$ and $\mathbf{d}' \mathbf{C}_\ell \mathbf{1} = O(1)$, where $\mathbf{d} = (d_1, \dots, d_T)' \neq \mathbf{0}$ consists of $O(1)$ elements; (iv) $\text{tr}(\mathbf{C}_\ell \mathbf{C}_s) = O(1)$; (v) $\text{tr}(\mathbf{C}_\ell \odot \mathbf{C}_s) = O(T^{-1})$; (vi) $\text{tr}(\mathbf{C}_{\ell s}) = O(T)$; (vii) all the non-zero elements of $\text{dg}(\mathbf{I}_N \otimes \mathbf{C}_\ell)$ are $O(T^{-1})$; (viii) $\mathbf{1}' \mathbf{C}_{\ell s} \mathbf{C}_{s\ell} \mathbf{1} = O(1)$; (ix) $\text{tr}(\mathbf{C}_{\ell s} \odot \mathbf{C}_{\ell s}) = O(T)$; (x) $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{s\ell}) = O(T)$; (xi) $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{\ell s}) = O(T)$.

Proof. These results follow by substituting (A2), (A3), and (A4) into the various terms involved.

(i): Using (A2), one has

$$\begin{aligned}
\mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{1} &= \sum_{i=1}^{T+1-\ell} \sum_{r=1}^p a_r \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r} = \sum_{r=1}^p a_r \frac{(T - \ell)(1 - \lambda_r) - \lambda_r(1 - \lambda_r^{T-\ell})}{(1 - \lambda_r)^2} \\
&= T \sum_{r=1}^p \frac{a_r}{1 - \lambda_r} + O(1).
\end{aligned}$$

(ii): Using (A2) and (A3), one has

$$\begin{aligned}
\mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} &= \sum_{i=s}^{T+1-\ell} \left(\sum_{r=1}^p a_r \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r} \right) \left(\sum_{r=1}^p a_r \frac{1 - \lambda_r^{i-s}}{1 - \lambda_r} \right) \\
&= \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{i=s}^{T+1-\ell} \frac{1 - \lambda_{r_1}^{T+1-i-\ell}}{1 - \lambda_{r_1}} \frac{1 - \lambda_{r_2}^{i-s}}{1 - \lambda_{r_2}} \\
&= T \sum_{r_1=1}^p \sum_{r_2=1}^p \frac{a_{r_1} a_{r_2}}{(1 - \lambda_{r_1})(1 - \lambda_{r_2})} + O(1).
\end{aligned}$$

(iii): Without loss of generality, assume $\ell \geq s$. First consider $\mathbf{d} = \mathbf{1}$. Using (A3), one has

$$\begin{aligned}
\mathbf{1}' (\mathbf{\Phi}_p^{-1} \mathbf{L}^\ell)' \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} &= \sum_{i=\ell}^T \left(\sum_{r=1}^p a_r \frac{1 - \lambda_r^{i-\ell}}{1 - \lambda_r} \right) \left(\sum_{r=1}^p a_r \frac{1 - \lambda_r^{i-s}}{1 - \lambda_r} \right) \\
&= \sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \sum_{i=\ell}^T \frac{1 - \lambda_{r_1}^{i-\ell}}{1 - \lambda_{r_1}} \frac{1 - \lambda_{r_2}^{i-s}}{1 - \lambda_{r_2}} \\
&= T \sum_{r_1=1}^p \sum_{r_2=1}^p \frac{a_{r_1} a_{r_2}}{(1 - \lambda_{r_1})(1 - \lambda_{r_2})} + O(1).
\end{aligned}$$

Further,

$$\begin{aligned}
\frac{1}{T} \mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1} \cdot \mathbf{1}' \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{1} &= \frac{1}{T} \left(T \sum_{r=1}^p \frac{a_r}{1 - \lambda_r} + O(1) \right) \left(T \sum_{r=1}^p \frac{a_r}{1 - \lambda_r} + O(1) \right) \\
&= T \sum_{r_1=1}^p \sum_{r_2=1}^p \frac{a_{r_1} a_{r_2}}{(1 - \lambda_{r_1})(1 - \lambda_{r_2})} + O(1).
\end{aligned}$$

It follows that

$$\mathbf{1}'C_{\ell s}\mathbf{1} = \mathbf{1}'(\Phi_p^{-1}L^\ell)'M\Phi_p^{-1}L^s\mathbf{1} - \frac{1}{T}\mathbf{1}'\Phi_p^{-1}L^s\mathbf{1} \cdot \mathbf{1}'\Phi_p^{-1}L^\ell\mathbf{1} = O(1).$$

For $\ell = s$, $\mathbf{1}'C_{\ell\ell}\mathbf{1} = \mathbf{1}'(\Phi_p^{-1}L^\ell)'M\Phi_p^{-1}L^\ell\mathbf{1}$ is the sum of square of elements of $C_\ell\mathbf{1}$. This implies that, at most, some finite number of elements of $C_\ell\mathbf{1}$ can be $O(1)$, and all other elements are $O(T^{-1})$. Thus, for any non-zero $T \times 1$ vector \mathbf{d} of $O(1)$ elements, $\mathbf{d}'C_\ell\mathbf{1} = O(1)$.

- (iv): In light of the facts that $T^{-1}\mathbf{1}'\Phi_p^{-1}L^s\Phi_p^{-1}L^\ell\mathbf{1} = O(1)$ and $T^{-1}\mathbf{1}'\Phi_p^{-1}L^\ell\Phi_p^{-1}L^s\mathbf{1} = O(1)$ from (ii), $T^{-2}\mathbf{1}'\Phi_p^{-1}L^s\mathbf{1} \cdot \mathbf{1}'\Phi_p^{-1}L^\ell\mathbf{1} = O(T^{-1})$ from (iii), and $\text{tr}(\Phi_p^{-1}L^\ell\Phi_p^{-1}L^s) = 0$ since both $\Phi_p^{-1}L^\ell$ and $\Phi_p^{-1}L^s$ are strictly lower triangular,

$$\begin{aligned}\text{tr}(C_\ell C_s) &= \text{tr}(\Phi_p^{-1}L^\ell M \Phi_p^{-1}L^s) - \frac{1}{T}\mathbf{1}'\Phi_p^{-1}L^\ell M \Phi_p^{-1}L^s\mathbf{1} \\ &= \text{tr}(\Phi_p^{-1}L^\ell\Phi_p^{-1}L^s) - \frac{1}{T}\mathbf{1}'\Phi_p^{-1}L^s\Phi_p^{-1}L^\ell\mathbf{1} \\ &\quad - \frac{1}{T}\mathbf{1}'\Phi_p^{-1}L^\ell\Phi_p^{-1}L^s\mathbf{1} + \frac{1}{T^2}\mathbf{1}'\Phi_p^{-1}L^s\mathbf{1} \cdot \mathbf{1}'\Phi_p^{-1}L^\ell\mathbf{1} \\ &= O(1).\end{aligned}$$

- (v): Note that both $\Phi_p^{-1}L^\ell$ and $\Phi_p^{-1}L^s$ are strictly lower triangular. It then follows that $\text{tr}(C_\ell \odot C_s) = T^{-2}\text{tr}(\mathbf{1}\mathbf{1}'\Phi_p^{-1}L^\ell \odot \mathbf{1}\mathbf{1}'\Phi_p^{-1}L^s)$, where the (i, i) -th element of $\mathbf{1}\mathbf{1}'\Phi_p^{-1}L^\ell$ is the i -th element of $\mathbf{1}'\Phi_p^{-1}L^\ell$. Therefore, using (A2) and assuming $\ell \geq s$,

$$\begin{aligned}\text{tr}(C_\ell \odot C_s) &= \frac{1}{T^2}\mathbf{1}'\Phi_p^{-1}L^\ell(\Phi_p^{-1}L^s)'\mathbf{1} \\ &= \frac{1}{T^2}\sum_{r_1=1}^p\sum_{r_2=1}^p a_{r_1}a_{r_2}\sum_{i=1}^{T+1-\ell}\frac{1-\lambda_{r_1}^{T+1-i-\ell}}{1-\lambda_{r_1}}\frac{1-\lambda_{r_2}^{T+1-i-s}}{1-\lambda_{r_2}} \\ &= \frac{1}{T^2}\left[T\sum_{r_1=1}^p\sum_{r_2=1}^p\frac{a_{r_1}a_{r_2}}{(1-\lambda_{r_1})(1-\lambda_{r_2})} + O(1)\right] \\ &= O(T^{-1}).\end{aligned}$$

- (vi): Note that $\text{tr}(C_{\ell s}) = \text{tr}((\Phi_p^{-1}L^\ell)'M\Phi_p^{-1}L^s) - T^{-1}\mathbf{1}'\Phi_p^{-1}L^s(\Phi_p^{-1}L^\ell)'\mathbf{1}$, where, from (v), $T^{-1}\mathbf{1}'\Phi_p^{-1}L^s(\Phi_p^{-1}L^\ell)'\mathbf{1} = O(1)$. Thus, using (A4) and assuming $\ell \geq s$,

$$\begin{aligned}\text{tr}(C_{\ell s}) &= \sum_{r_1=1}^p\sum_{r_2=1}^p a_{r_1}a_{r_2}\sum_{i=1}^T w_{i,i,\ell,s,r_1,r_2} + O(1) \\ &= \sum_{r_1=1}^p\sum_{r_2=1}^p a_{r_1}a_{r_2}\sum_{i=1}^T \frac{\lambda_{r_2}^{\ell-s} - \lambda_{r_1}^{T+1-i-\ell}\lambda_{r_2}^{T+1-i-s}}{1-\lambda_{r_1}\lambda_{r_2}} + O(1) \\ &= T\sum_{r_1=1}^p\sum_{r_2=1}^p \frac{a_{r_1}a_{r_2}\lambda_{r_2}^{\ell-s}}{(1-\lambda_{r_1}\lambda_{r_2})} + O(1).\end{aligned}$$

- (vii): Since $\Phi_p^{-1}L^\ell$ is strictly lower triangular, $\text{dg}(I_N \otimes C_\ell) = -T^{-1}\text{dg}(I_N \otimes \mathbf{1}\mathbf{1}'\Phi_p^{-1}L^\ell)$. From (A2), the diagonal elements of $\mathbf{1}\mathbf{1}'\Phi_p^{-1}L^\ell$ are all $O(1)$. Thus, one can claim that all the non-zero elements of $\text{dg}(I_N \otimes C_\ell)$ are $O(T^{-1})$.

- (viii): From (iii), at most, some finite number of elements of $C_\ell\mathbf{1}$ can be $O(1)$ and all other elements are $O(T^{-1})$. Also, all the elements of $\Phi_p^{-1}L^s(\Phi_p^{-1}L^s)'$ are $O(1)$. Thus, $\mathbf{1}'C_{\ell s}C_{s\ell}\mathbf{1} = \mathbf{1}'(\Phi_p^{-1}L^\ell)'M\Phi_p^{-1}L^s(\Phi_p^{-1}L^s)'M\Phi_p^{-1}L^\ell\mathbf{1} = O(1)$.

- (ix): The (i, j) -th element of $\mathbf{C}_{\ell s} = (\Phi_p^{-1} \mathbf{L}^\ell)' \Phi_p^{-1} \mathbf{L}^s - T^{-1} (\Phi_p^{-1} \mathbf{L}^\ell)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} \mathbf{L}^s$, in light of (A2) and (A4), is

$$\sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} \left[w_{i,j,\ell,s,r_1,r_2} - \frac{1}{T} \left(\frac{1 - \lambda_{r_1}^{T-i-\ell+1}}{1 - \lambda_{r_1}} \right) \left(\frac{1 - \lambda_{r_2}^{T-j-s+1}}{1 - \lambda_{r_2}} \right) \right] = O(1).$$

Therefore, $\text{tr}(\mathbf{C}_{\ell s} \odot \mathbf{C}_{\ell s}) = O(T)$.

- (x): From the previous part, the (i, j) -th element of $\mathbf{C}_{\ell s}$ is $\sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} w_{i,j,\ell,s,r_1,r_2} + O(T^{-1})$. Thus, the i -th diagonal element of $\mathbf{C}_{\ell s} \mathbf{C}_{s\ell}$ is

$$\begin{aligned} & \sum_{j=1}^T \left(\sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} w_{i,j,\ell,s,r_1,r_2} \right) \left(\sum_{r_1=1}^p \sum_{r_2=1}^p a_{r_1} a_{r_2} w_{i,j,\ell,s,r_1,r_2} \right) + O(T^{-1}) \\ &= \sum_{r_1=1}^p \sum_{r_2=1}^p \sum_{r_3=1}^p \sum_{r_4=1}^p a_{r_1} a_{r_2} a_{r_3} a_{r_4} \left(\sum_{j=1}^T w_{i,j,\ell,s,r_1,r_2} w_{i,j,\ell,s,r_3,r_4} \right) + O(T^{-1}) \\ &= O(1). \end{aligned}$$

Therefore, $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{s\ell}) = O(T)$.

- (xi): By similar reasoning as in (x), one can show that the i -th diagonal element of $\mathbf{C}_{\ell s} \mathbf{C}_{\ell s}$ is $O(1)$, and then $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{\ell s}) = O(T)$. \square

Lemma A3. For the unit-root case, namely, $\lambda_1 = 1$ and $|\lambda_r| < 1$, $r = 2, \dots, p$, the following hold for $\ell, s = 1, \dots, T$: (i) $\mathbf{1}' \Phi_p^{-1} \mathbf{L}^\ell \mathbf{1} = O(T^2)$ and $\mathbf{d}' \mathbf{C}_\ell \mathbf{1} = O(T^2)$, where $\mathbf{d} = (d_1, \dots, d_T)' \neq \mathbf{0}$ consists of $O(1)$ elements; (ii) $\mathbf{1}' \Phi_p^{-1} \mathbf{L}^\ell \Phi_p^{-1} \mathbf{L}^s \mathbf{1} = O(T^3)$; (iii) $\mathbf{1}' \mathbf{C}_{\ell s} \mathbf{1} = O(T^3)$; (iv) $\text{tr}(\mathbf{C}_\ell \mathbf{C}_s) = O(T^2)$; (v) $\text{tr}(\mathbf{C}_\ell \odot \mathbf{C}_s) = O(T)$; (vi) $\text{tr}(\mathbf{C}_{\ell s}) = O(T^2)$; (vii) all the non-zero elements of $\text{dg}(\mathbf{I}_N \otimes \mathbf{C}_\ell)$ are $O(1)$. (viii) $\mathbf{1}' \mathbf{C}_{\ell s} \mathbf{C}_{s\ell} \mathbf{1} = O(T^5)$; (ix) $\text{tr}(\mathbf{C}_{\ell s} \odot \mathbf{C}_{\ell s}) = O(T^3)$; (x) $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{s\ell}) = O(T^4)$; (xi) $\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{\ell s}) = O(T^4)$.

Proof. These results follow by substituting (A5), (A6), (A7), and (A8) into the various terms involved.

- (i): Using (A5), one has

$$\begin{aligned} \mathbf{1}' \Phi_p^{-1} \mathbf{L}^\ell \mathbf{1} &= a_1 \sum_{i=1}^{T+1-\ell} (T+1-i-\ell) + \sum_{r=2}^p a_r \sum_{i=1}^{T+1-\ell} \frac{1 - \lambda_r^{T+1-i-\ell}}{1 - \lambda_r} \\ &= \frac{a_1}{2} (T-\ell)(T-\ell+1) + \sum_{r=2}^p a_r \frac{(T-\ell)(1-\lambda_r) - \lambda_r(1-\lambda_r^{T-\ell})}{(1-\lambda_r)^2} \\ &= \frac{a_1}{2} T^2 + \frac{a_1}{2} (1-2\ell)T + T \sum_{r=2}^p \frac{a_r}{1-\lambda_r} + O(1). \end{aligned}$$

Similarly, using (A6)

$$\begin{aligned} \mathbf{d}' \Phi_p^{-1} \mathbf{L}^\ell &= a_1 \sum_{i=\ell}^T d_i (i-\ell) + \sum_{r=2}^p a_r \sum_{i=\ell}^T d_i \frac{1 - \lambda_r^{i-\ell}}{1 - \lambda_r} \\ &= a_1 \sum_{i=1}^T i d_i - \ell a_1 \sum_{i=1}^T d_i + O(T) \\ &= a_1 \sum_{i=1}^T i d_i + O(T) = O(T^2) \end{aligned}$$

and thus, by denoting $\bar{\mathbf{d}} = T^{-1} \sum_{i=1}^T d_i$,

$$\mathbf{d}'\mathbf{M}\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} = \mathbf{d}'\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} - \bar{\mathbf{d}}'\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} + O(T) = a_1 \sum_{i=1}^T i(d_i - \bar{d}) = O(T^2),$$

which may retain a term of order $O(T^2)$, unless all the elements of \mathbf{d} are the same. So, in general, $\mathbf{d}'\mathbf{M}\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} = O(T^2)$.

(ii): Using (A5) and (A6),

$$\begin{aligned} & \mathbf{1}'\Phi_p^{-1}\mathbf{L}^\ell\Phi_p^{-1}\mathbf{L}^s\mathbf{1} \\ &= \sum_{i=s}^{T+1-\ell} \left(a_1(T+1-i-\ell) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{T+1-i-\ell}}{1-\lambda_r} \right) \left(a_1(i-s) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{i-s}}{1-\lambda_r} \right) \\ &= a_1^2 \sum_{i=s}^{T+1-\ell} (T+1-i-\ell)(i-s) + a_1 \sum_{r=2}^p a_r \sum_{i=s}^{T+1-\ell} (i-s) \frac{1-\lambda_r^{T+1-i-\ell}}{1-\lambda_r} \\ &\quad + a_1 \sum_{r=2}^p a_r \sum_{i=s}^{T+1-\ell} (T+1-i-\ell) \frac{1-\lambda_r^{i-s}}{1-\lambda_r} + T \sum_{r_1=2}^p \sum_{r_2=2}^p \frac{a_{r_1}a_{r_2}}{(1-\lambda_{r_1})(1-\lambda_{r_2})} + O(1) \\ &= \frac{a_1^2}{6} (T-\ell-s)(T-\ell-s+1)(T-\ell-s+2) \\ &\quad + a_1 \sum_{r=2}^p a_r \left(\frac{T^2}{1-\lambda_r} + \frac{T[1-3\lambda_r-2(1-\lambda_r)(\ell+s)]}{(1-\lambda_r)^2} \right) \\ &\quad + T \sum_{r_1=2}^p \sum_{r_2=2}^p \frac{a_{r_1}a_{r_2}}{(1-\lambda_{r_1})(1-\lambda_{r_2})} + O(1) \\ &= \frac{a_1^2}{6} T^3 + \frac{a_1^2(1-\ell-s)}{2} T^2 + a_1 T^2 \sum_{r=2}^p \frac{a_r}{1-\lambda_r} + O(T). \end{aligned}$$

(iii): Without loss of generality, assume $\ell \geq s$. First, note that

$$\begin{aligned} & \mathbf{1}'(\Phi_p^{-1}\mathbf{L}^\ell)'\Phi_p^{-1}\mathbf{L}^s\mathbf{1} \\ &= \sum_{i=\ell}^T \left(a_1(i-\ell) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{i-\ell}}{1-\lambda_r} \right) \left(a_1(i-s) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{i-s}}{1-\lambda_r} \right) \\ &= a_1^2 \sum_{i=\ell}^T (i-\ell)(i-s) + a_1 \sum_{r=2}^p a_r \sum_{i=\ell}^T \left[(i-\ell) \frac{1-\lambda_r^{i-s}}{1-\lambda_r} + (i-s) \frac{1-\lambda_r^{i-\ell}}{1-\lambda_r} \right] + O(T) \\ &= \frac{a_1^2}{3} T^3 + \frac{a_1^2(1-\ell-s)}{2} T^2 + a_1 T^2 \sum_{r=2}^p \frac{a_r}{1-\lambda_r} + O(T). \end{aligned}$$

Further, from (i),

$$\begin{aligned} & \frac{1}{T} \mathbf{1}'\Phi_p^{-1}\mathbf{L}^s\mathbf{1} \cdot \mathbf{1}'\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} \\ &= \frac{1}{T} \left[\frac{a_1}{2} T^2 + \frac{a_1}{2} (1-2s)T + T \sum_{r=2}^p \frac{a_r}{1-\lambda_r} \right] \left[\frac{a_1}{2} T^2 + \frac{a_1}{2} (1-2\ell)T + T \sum_{r=2}^p \frac{a_r}{1-\lambda_r} \right] + O(T) \\ &= \frac{a_1^2}{4} T^3 + \frac{a_1^2(1-\ell-s)}{2} T^2 + a_1 T^2 \sum_{r=2}^p \frac{a_r}{1-\lambda_r} + O(T). \end{aligned}$$

It follows that

$$\mathbf{1}'\mathbf{C}_{\ell s}\mathbf{1} = \mathbf{1}'(\Phi_p^{-1}\mathbf{L}^\ell)'\Phi_p^{-1}\mathbf{L}^s\mathbf{1} - \frac{1}{T} \mathbf{1}'\Phi_p^{-1}\mathbf{L}^s\mathbf{1} \cdot \mathbf{1}'\Phi_p^{-1}\mathbf{L}^\ell\mathbf{1} = \frac{a_1^2}{12} T^3 + O(T).$$

(iv): Using results from (ii) and (iii),

$$\begin{aligned}
& \text{tr}(\mathbf{C}_\ell \mathbf{C}_s) \\
&= \frac{1}{T^2} \mathbf{1}' \Phi_p^{-1} L^s \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^\ell \mathbf{1} - \frac{1}{T} \mathbf{1}' \Phi_p^{-1} L^s \Phi_p^{-1} L^\ell \mathbf{1} - \frac{1}{T} \mathbf{1}' \Phi_p^{-1} L^\ell \Phi_p^{-1} L^s \mathbf{1} \\
&= \frac{a_1^2}{4} T^2 + \frac{a_1^2}{2} T(1 - \ell - s) + a_1 T \sum_{r=2}^p \frac{a_r}{1 - \lambda_r} \\
&\quad - \left[\frac{a_1^2}{3} T^2 + a_1^2 T(1 - \ell - s) + 2a_1 T \sum_{r=2}^p \frac{a_r}{1 - \lambda_r} \right] + O(1) \\
&= -\frac{a_1^2}{12} T^2 - \frac{a_1^2(1 - \ell - s)}{2} T - a_1 T \sum_{r=2}^p \frac{a_r}{1 - \lambda_r} + O(1).
\end{aligned}$$

(v): Assuming $\ell \geq s$,

$$\begin{aligned}
& \text{tr}(\mathbf{C}_\ell \odot \mathbf{C}_s) \\
&= \frac{1}{T^2} \mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^s)' \mathbf{1} \\
&= \frac{1}{T^2} \sum_{i=1}^{T+1-s} [a_1(T+1-i-\ell) + O(1)] [a_1(T+1-i-s) + O(1)] \\
&= \frac{a_1^2}{3} T + O(1).
\end{aligned}$$

(vi): Assuming $\ell \geq s$,

$$\begin{aligned}
\text{tr}(\mathbf{C}_{\ell s}) &= \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s) - \frac{1}{T} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \mathbf{1} \\
&= \sum_{i=1}^{T+1-\ell} \left[a_1^2(T+1-i-\ell) + a_1 \sum_{r=2}^p \frac{a_r(1 + \lambda_r^{\ell-s})(1 - \lambda_r^{T+1-i-\ell})}{1 - \lambda_r} + \sum_{r_1=2}^p \sum_{r_2=2}^p a_{r_1} a_{r_2} w_{i,j,\ell,s,r_1,r_2} \right] \\
&\quad - \frac{1}{T} \sum_{i=1}^{T+1-\ell} \left[a_1(T+1-i-s) + \sum_{r=2}^p \frac{a_r(1 - \lambda_r^{T+1-i-s})}{1 - \lambda_r} \right] \left[a_1(T+1-i-\ell) \right. \\
&\quad \left. + \sum_{r=2}^p \frac{a_r(1 - \lambda_r^{T+1-i-\ell})}{1 - \lambda_r} \right] \\
&= \frac{a_1^2}{2} T^2 - \frac{a_1^2}{3} T^2 + O(T) \\
&= \frac{a_1^2}{6} T^2 + O(T).
\end{aligned}$$

(vii): From (A5), the diagonal elements of $\mathbf{1}' \Phi_p^{-1} L^\ell$ are all $O(T)$. Thus, one can claim that all the non-zero elements of $\text{dg}(\mathbf{I}_N \otimes \mathbf{C}_\ell) = -T^{-1} \text{dg}(\mathbf{I}_N \otimes \mathbf{1}' \Phi_p^{-1} L^\ell)$ are $O(1)$.

(viii): By substitution,

$$\begin{aligned}
& \mathbf{1}' \mathbf{C}_{\ell s} \mathbf{C}_s \mathbf{1} \\
&= \mathbf{1}' (\Phi_p^{-1} L^\ell)' \mathbf{M} \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{M} \Phi_p^{-1} L^\ell \mathbf{1} \\
&= \mathbf{1}' (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell \mathbf{1} - \frac{1}{T} \mathbf{1}' (\Phi_p^{-1} L^\ell)' \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell \mathbf{1} \\
&\quad - \frac{1}{T} \mathbf{1}' (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^\ell \mathbf{1} \\
&\quad + \frac{1}{T^2} \mathbf{1}' (\Phi_p^{-1} L^\ell)' \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^\ell \mathbf{1} \\
&= \mathbf{1}' (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell \mathbf{1} - \frac{2}{T} \mathbf{1}' \Phi_p^{-1} L^\ell \mathbf{1} \cdot \mathbf{1}' (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \\
&\quad + \frac{1}{T^2} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \cdot (\mathbf{1}' \Phi_p^{-1} L^\ell \mathbf{1})^2.
\end{aligned}$$

Assume $\ell \geq s$. Using (A7) and (A8),

$$\begin{aligned} & \mathbf{1}'(\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s(\Phi_p^{-1}L^s)' \Phi_p^{-1}L^\ell \mathbf{1} \\ &= \sum_{i=1}^T \left[\sum_{j=1}^{i+\ell-s} a_1^2(T+1-i-\ell) + \sum_{j=i+\ell-s+1}^{T+1-j} a_1^2(T+1-j-s) + O(1) \right]^2 \\ &= \frac{2a_1^4}{15}T^5 + O(T^4). \end{aligned}$$

Using (i), (A5), (A7), and (A8),

$$\begin{aligned} & \frac{2}{T} \mathbf{1}' \Phi_p^{-1}L^\ell \mathbf{1} \cdot \mathbf{1}'(\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s(\Phi_p^{-1}L^s)' \mathbf{1} \\ &= \frac{2}{T} \left(\frac{a_1}{2}T^2 + O(T) \right) \\ & \quad \times \sum_{i=1}^{T+1-\ell} \left[\sum_{j=1}^{i+\ell-s} a_1^2(T+1-i-\ell) + \sum_{j=i+\ell-s+1}^{T+1-j} a_1^2(T+1-j-s) + O(1) \right] \\ & \quad \times [a_1(T+1-i-s) + O(1)] \\ &= \frac{5a_1^4}{24}T^5 + O(T^4). \end{aligned}$$

Using (i) and (v),

$$\begin{aligned} \frac{1}{T^2} \mathbf{1}' \Phi_p^{-1}L^s(\Phi_p^{-1}L^s)' \mathbf{1} \cdot (\mathbf{1}' \Phi_p^{-1}L^\ell \mathbf{1})^2 &= \left(\frac{a_1^2}{3}T + O(1) \right) \left(\frac{a_1}{2}T^2 + O(T) \right)^2 \\ &= \frac{a_1^4}{12}T^5 + O(T^4). \end{aligned}$$

Therefore,

$$\mathbf{1}'C_{\ell s}C_{s\ell}\mathbf{1} = \frac{2a_1^4}{15}T^5 - \frac{5a_1^4}{24}T^5 + \frac{a_1^4}{12}T^5 + O(T^4) = \frac{a_1^4}{120}T^5 + O(T^4).$$

(ix): By substituting $C_{\ell s} = (\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s - T^{-1}(\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s$,

$$\begin{aligned} \text{tr}(C_{\ell s} \odot C_{\ell s}) &= \text{tr}((\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s \odot (\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s) \\ & \quad - \frac{2}{T} \text{tr}((\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s \odot (\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s) \\ & \quad + \frac{1}{T^2} \text{tr}((\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s \odot (\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s). \end{aligned}$$

Assume $\ell \geq s$. Using (A7),

$$\text{tr}((\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s \odot (\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s) = \sum_{i=1}^{T+1-\ell} \left[a_1^2(T+1-i-\ell) + O(1) \right]^2 = \frac{a_1^4}{3}T^3 + O(T^2).$$

Using (A5), the i -th diagonal element of $(\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s$ is, $i \leq T+1-\ell$,

$$\left[a_1(T+1-i-\ell) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{T+1-i-\ell}}{1-\lambda_r} \right] \left[a_1(T+1-i-s) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{T+1-i-s}}{1-\lambda_r} \right].$$

Thus,

$$\frac{2}{T} \text{tr}((\Phi_p^{-1}L^\ell)' \Phi_p^{-1}L^s \odot (\Phi_p^{-1}L^\ell)' \mathbf{1}\mathbf{1}' \Phi_p^{-1}L^s)$$

$$\begin{aligned}
&= \frac{2}{T} \sum_{i=1}^{T+1-\ell} \left[a_1^2(T+1-i-\ell) + O(1) \right] \\
&\quad \times [a_1(T+1-i-\ell) + O(1)][a_1(T+1-i-s) + O(1)] \\
&= \frac{a_1^4}{2} T^3 + O(T^2),
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{T^2} \text{tr}((\Phi_p^{-1} L^\ell)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^s \odot (\Phi_p^{-1} L^\ell)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^s) \\
&= \frac{1}{T^2} \sum_{i=1}^{T+1-\ell} \{ [a_1(T+1-i-\ell) + O(1)][a_1(T+1-i-s) + O(1)] \}^2 \\
&= \frac{a_1^4}{5} T^3 + O(T^2).
\end{aligned}$$

Therefore,

$$\text{tr}(\mathbf{C}_{\ell s} \odot \mathbf{C}_{\ell s}) = \frac{a_1^4}{3} T^3 - \frac{a_1^4}{2} T^3 + \frac{a_1^4}{5} T^3 + O(T^2) = \frac{a_1^4}{30} T^3 + O(T^2).$$

(x): By substitution,

$$\begin{aligned}
\text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{s\ell}) &= \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell) \\
&\quad - \frac{1}{T} \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^\ell) \\
&\quad - \frac{1}{T} \text{tr}((\Phi_p^{-1} L^\ell)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell) \\
&\quad + \frac{1}{T^2} \text{tr}((\Phi_p^{-1} L^\ell)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \mathbf{1}' \Phi_p^{-1} L^\ell) \\
&= \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell) \\
&\quad - \frac{2}{T} \mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \\
&\quad + \frac{1}{T^2} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \cdot \mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^\ell)' \mathbf{1}.
\end{aligned}$$

Assume $\ell \geq s$. Using (A7),

$$\begin{aligned}
&\text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \Phi_p^{-1} L^\ell) \\
&= \sum_{i=1}^T \left\{ \sum_{j=1}^{i+\ell-s} \left[a_1^2(T+1-i-\ell) + O(1) \right]^2 + \sum_{j=i+\ell-s+1}^T \left[a_1^2(T+1-j-s) + O(1) \right]^2 \right\} \\
&= \frac{a_1^4}{6} T^4 + O(T^3).
\end{aligned}$$

Using (v),

$$\begin{aligned}
&\frac{1}{T} (\mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1}) \cdot \frac{1}{T} \mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^\ell)' \mathbf{1} \\
&= \left[\frac{a_1^2}{3} T^2 + O(T) \right] \left[\frac{a_1^2}{3} T^2 + O(T) \right] \\
&= \frac{a_1^4}{9} T^4 + O(T^3).
\end{aligned}$$

Using (A5), (A7), and (A8),

$$\begin{aligned}
& \frac{2}{T} \mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^s)' \mathbf{1} \\
&= \frac{2}{T} \sum_{i=1}^{T+1-\ell} \sum_{j=1}^{i+\ell-s} (a_1(T+1-i-\ell)) (a_1^2(T+1-i-\ell)) (a_1(T+1-j-s)) \\
&+ \frac{2}{T} \sum_{i=1}^{T+1-\ell} \sum_{j=i+\ell-s+1}^{T+1-s} [a_1(T+1-i-\ell)] [a_1^2(T+1-j-s)] [a_1(T+1-j-s)] \\
&+ O(T^3) \\
&= \frac{4a_1^4}{15} T^4 + O(T^3).
\end{aligned}$$

Therefore,

$$\text{tr}(C_{\ell s} C_{s\ell}) = \frac{a_1^4}{6} T^4 + \frac{a_1^4}{9} T^4 - \frac{4a_1^4}{15} T^4 + O(T^3) = \frac{a_1^4}{90} T^4 + O(T^3).$$

(xi): Similarly,

$$\begin{aligned}
\text{tr}(C_{\ell s} C_{\ell s}) &= \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s) \\
&- \frac{2}{T} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \mathbf{1} \\
&+ \frac{1}{T^2} (\mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^s)' \mathbf{1})^2.
\end{aligned}$$

Assume $\ell \geq s$. If $\ell = s$, from (x), $\text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s) = a_1^4 T^4 / 6 + O(T^3)$, in view of (A7) and (A8). Assume $\ell = s + 1$. Note that the (i, j) -th element of $(\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s$ has leading term $a_1^2(T+1-i-\ell)$ if $i \geq j-1$ and $T+1-j-s = T+2-j-\ell$ if $i < j-1$. The (j, i) -th element of $(\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s$ has leading term $a_1^2(T+1-j-\ell)$ if $j \geq i-1$ and $T+1-i-s = T+2-i-\ell$ if $j < i-1$. Then,

$$\begin{aligned}
& \text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s) \\
&= \sum_{i=1}^{T+1-\ell} \sum_{j=1}^{i-2} a_1^2(T+1-i-\ell) a_1^2(T+2-i-\ell) \\
&+ \sum_{j=i+2}^{T+1-\ell} \sum_{i=1}^{j-2} a_1^2(T+2-j-\ell) a_1^2(T+1-j-\ell) + O(T^3) \\
&= \frac{a_1^4}{6} T^4 + O(T^3).
\end{aligned}$$

Similarly, for any $\ell \geq s$, $\text{tr}((\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s) = a_1^4 T^4 / 6 + O(T^3)$. Using (v), one has

$$\frac{1}{T^2} (\mathbf{1}' \Phi_p^{-1} L^\ell (\Phi_p^{-1} L^s)' \mathbf{1})^2 = \left[\frac{a_1^2}{3} T^2 + O(T) \right]^2 = \frac{a_1^4}{9} T^4 + O(T^3).$$

Using (A5), (A7), (A8), and following similarly the proof in the previous part,

$$\frac{2}{T} \mathbf{1}' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \Phi_p^{-1} L^s (\Phi_p^{-1} L^\ell)' \mathbf{1} = \frac{4a_1^4}{15} T^4 + O(T^3).$$

Therefore,

$$\text{tr}(C_{\ell s} C_{\ell s}) = \frac{a_1^4}{6} T^4 + \frac{a_1^4}{9} T^4 - \frac{4a_1^4}{15} T^4 + O(T^3) = \frac{a_1^4}{90} T^4 + O(T^3). \quad \square$$

Lemma A4. Under the conditions of Theorem 2,

$$\text{Var}((NT)^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{u}) = \frac{1}{NT} \begin{pmatrix} \sigma^4 N \mathbf{\Gamma}_1 + \sigma^2 \mathbf{\Gamma}_2 & \sigma^2 \mathbf{F}'_1 \\ \sigma^2 \mathbf{F}_1 & \sigma^2 \mathbf{X}' \mathbf{A} \mathbf{X} \end{pmatrix} + O(T^{-1}),$$

where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ have $\text{tr}(\mathbf{C}_{\ell s})$ and $\beta'_0 \mathbf{X}'(\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{X} \beta_0$, respectively, in their (ℓ, s) -positions, $\ell, s = 1, \dots, p$, and \mathbf{F}_1 has $\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{C}_{\ell}) \mathbf{X} \beta_0$ as its ℓ -th column, $\ell = 1, \dots, p$.

Proof. Following Bao and Yu (2023), write $\text{Var}(\mathbf{W}' \mathbf{A} \mathbf{u}) = \text{Var}(\overline{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Var}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Cov}(\overline{\mathbf{W}}' \mathbf{A} \mathbf{u}, \widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Cov}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}, \overline{\mathbf{W}}' \mathbf{A} \mathbf{u})$, where $\overline{\mathbf{W}} = \mathbf{E}(\mathbf{W})$ and $\widetilde{\mathbf{W}} = \mathbf{W} - \overline{\mathbf{W}}$ (and this kind of notation to denote the fixed and random components of a random term is used henceforth). $\text{Var}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u})$ has a non-zero top-left $p \times p$ block with $\text{Cov}(\tilde{\mathbf{y}}'_{(-\ell)} \mathbf{A} \mathbf{u}, \tilde{\mathbf{y}}'_{(-s)} \mathbf{A} \mathbf{u})$ populating its (ℓ, s) -th position, $\ell, s = 1, \dots, p$. In particular, $\text{Cov}(\tilde{\mathbf{y}}'_{(-\ell)} \mathbf{A} \mathbf{u}, \tilde{\mathbf{y}}'_{(-s)} \mathbf{A} \mathbf{u})$ is equal to

$$\begin{aligned} & \sigma^4 N [\text{tr}(\mathbf{C}_{\ell} \mathbf{C}_s) + \text{tr}(\mathbf{C}_{\ell s}) + \gamma_2 \text{tr}(\mathbf{C}_{\ell} \odot \mathbf{C}_s)] + \sigma^2 \mathbf{1}' \mathbf{C}_{\ell s} \mathbf{1} \mathbf{E}(\tilde{\alpha}' \tilde{\alpha}) \\ & + \sigma^2 \sum_{j_1=0}^{s-1} \sum_{j_2=0}^{\ell-1} \mathbf{e}'_1 \Phi_{(j_1, j_2)}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\mathbf{y}}'_{-j_1} \tilde{\mathbf{y}}_{-j_2}) \\ & + \sigma^2 \sum_{j_1=s}^{p-1} \sum_{j_2=\ell}^{p-1} \mathbf{e}'_1 \Phi_{((j_1-p), (j_2-p))}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\mathbf{y}}'_{-j_1} \tilde{\mathbf{y}}_{-j_2}) \\ & + \sigma^2 \sum_{j_2=0}^{\ell-1} \mathbf{1}' \Phi_{(0, j_2)}^{(s, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\alpha}' \tilde{\mathbf{y}}_{-j_2}) + \sigma^2 \sum_{j_2=\ell}^{p-1} \mathbf{1}' \Phi_{(0, (j_2-p))}^{(s, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\alpha}' \tilde{\mathbf{y}}_{-j_2}) \\ & + \sigma^2 \sum_{j_1=0}^{s-1} \mathbf{1}' \Phi_{(0, j_1)}^{(\ell, s-1-j_1)} \mathbf{e}_1 \mathbf{E}(\tilde{\alpha}' \tilde{\mathbf{y}}_{-j_1}) + \sigma^2 \sum_{j_1=0}^{s-1} \sum_{j_2=\ell}^{p-1} \mathbf{e}'_1 \Phi_{(j_1, (j_2-p))}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\mathbf{y}}'_{-j_1} \tilde{\mathbf{y}}_{-j_2}) \\ & + \sigma^2 \sum_{j_1=s}^{p-1} \mathbf{1}' \Phi_{(0, (j_1-p))}^{(\ell, s-1-j_1)} \mathbf{e}_1 \mathbf{E}(\tilde{\alpha}' \tilde{\mathbf{y}}_{-j_1}) + \sigma^2 \sum_{j_1=s}^{p-1} \sum_{j_2=0}^{\ell-1} \mathbf{e}'_1 \Phi_{((j_1-p), (j_2-p))}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1 \mathbf{E}(\tilde{\mathbf{y}}'_{-j_1} \tilde{\mathbf{y}}_{-j_2}), \end{aligned}$$

where $\Phi_{(r, j)}^{(\ell, s)} = \Phi'_r \mathbf{L}^{\ell'} \Phi_p^{-1'} \mathbf{M} \Phi_p^{-1} \mathbf{L}^s \Phi_j$. (Recall that $\Phi_{(\ell-p)} = \Phi_{\ell} - \Phi_p$. Thus, for example, $\Phi_{(r, (j-p))}^{(\ell, s)} = \Phi'_r \mathbf{L}^{\ell'} \Phi_p^{-1'} \mathbf{M} \Phi_p^{-1} \mathbf{L}^s \Phi_{(j-p)}$.) From Lemma A2, in the first term, $\text{tr}(\mathbf{C}_{\ell} \mathbf{C}_s) = O(1)$, $\text{tr}(\mathbf{C}_{\ell s}) = O(T)$, and $\text{tr}(\mathbf{C}_{\ell} \odot \mathbf{C}_s) = O(1)$. The second term $\sigma^2 \mathbf{1}' \mathbf{C}_{\ell s} \mathbf{1} \mathbf{E}(\tilde{\alpha}' \tilde{\alpha})$ is $O(N)$. Terms like $\mathbf{e}'_1 \Phi_{(j_1, j_2)}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1$, $\mathbf{e}'_1 \Phi_{(j_1, (j_2-p))}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1$, and $\mathbf{e}'_1 \Phi_{((j_1-p), (j_2-p))}^{(s-1-j_1, \ell-1-j_2)} \mathbf{e}_1$ pick up just one element of the relevant matrices in the quadratic forms in \mathbf{e}_1 and thus are $O(1)$, whereas terms like $\mathbf{1}' \Phi_{(0, j_2)}^{(s, \ell-1-j_2)} \mathbf{e}_1$, $\mathbf{1}' \Phi_{(0, (j_2-p))}^{(s, \ell-1-j_2)} \mathbf{e}_1$, $\mathbf{1}' \Phi_{(0, j_1)}^{(\ell, s-1-j_1)} \mathbf{e}_1$, and $\mathbf{1}' \Phi_{(0, (j_1-p))}^{(\ell, s-1-j_1)} \mathbf{e}_1$ pick up the sum of the first column of the relevant matrices in the quadratic forms in \mathbf{e}_1 , which are again $O(1)$. Thus, $\text{Var}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u})$ is dominated by elements like $\sigma^4 N \text{tr}(\mathbf{C}_{\ell s}) = O(NT)$. Further,

$$\text{Cov}(\overline{\mathbf{W}}' \mathbf{A} \mathbf{u}, \widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}) = \mathbf{E} \begin{pmatrix} \overline{\mathbf{y}}'_{(-1)} \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-1)} & \cdots & \overline{\mathbf{y}}'_{(-1)} \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-p)} & \mathbf{0}'_k \\ \vdots & \ddots & \vdots & \vdots \\ \overline{\mathbf{y}}'_{(-p)} \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-1)} & \cdots & \overline{\mathbf{y}}'_{(-p)} \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-p)} & \mathbf{0}'_k \\ \mathbf{X}' \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-1)} & \cdots & \mathbf{X}' \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-p)} & \mathbf{O}_k \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{E}(\overline{\mathbf{y}}'_{(-\ell)} \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-s)}) &= \gamma_1 \sigma^3 \overline{\mathbf{y}}'_{(-\ell)} \text{Adg}(\mathbf{I}_N \otimes \mathbf{C}_s) = O(N), \\ \mathbf{E}(\mathbf{X}' \mathbf{A} \mathbf{u} \mathbf{u}' \mathbf{A} \tilde{\mathbf{y}}_{(-\ell)}) &= \gamma_1 \sigma^3 \mathbf{X}' \text{Adg}(\mathbf{I}_N \otimes \mathbf{C}_{\ell}) = O(N). \end{aligned}$$

with γ_1 denoting the skewness coefficient of the distribution of u_{it} . Thus,

$$\begin{aligned}
& \text{Var}((NT)^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{u}) \\
&= \frac{1}{NT} [\text{Var}(\overline{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Var}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Cov}(\overline{\mathbf{W}}' \mathbf{A} \mathbf{u}, \widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}) + \text{Cov}(\widetilde{\mathbf{W}}' \mathbf{A} \mathbf{u}, \overline{\mathbf{W}}' \mathbf{A} \mathbf{u})] \\
&= \sigma^2 \frac{1}{NT} \overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} + \sigma^4 \frac{1}{T} \begin{pmatrix} \text{tr}(\mathbf{C}_{11}) & \cdots & \text{tr}(\mathbf{C}_{1p}) & \mathbf{0}'_k \\ \vdots & \ddots & \vdots & \vdots \\ \text{tr}(\mathbf{C}_{p1}) & \cdots & \text{tr}(\mathbf{C}_{pp}) & \mathbf{0}'_k \\ \mathbf{0}_k & \cdots & \mathbf{0}_k & \mathbf{O}_k \end{pmatrix} + O(T^{-1}),
\end{aligned}$$

where $\text{tr}(\mathbf{C}_{\ell s}) = O(T)$ from Lemma A2, and

$$\frac{1}{NT} \overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} = \frac{1}{NT} \begin{pmatrix} \overline{\mathbf{y}}'_{(-1)} \mathbf{A} \overline{\mathbf{y}}_{(-1)} & \cdots & \overline{\mathbf{y}}'_{(-1)} \mathbf{A} \overline{\mathbf{y}}_{(-p)} & \overline{\mathbf{y}}'_{(-1)} \mathbf{A} \mathbf{X} \\ \vdots & \ddots & \vdots & \vdots \\ \overline{\mathbf{y}}'_{(-p)} \mathbf{A} \overline{\mathbf{y}}_{(-1)} & \cdots & \overline{\mathbf{y}}'_{(-p)} \mathbf{A} \overline{\mathbf{y}}_{(-p)} & \overline{\mathbf{y}}'_{(-p)} \mathbf{A} \mathbf{X} \\ \mathbf{X}' \mathbf{A} \overline{\mathbf{y}}_{(-1)} & \cdots & \mathbf{X}' \mathbf{A} \overline{\mathbf{y}}_{(-p)} & \mathbf{X}' \mathbf{A} \mathbf{X} \end{pmatrix}.$$

By substituting (5) and using Lemma A2 again, one has

$$\begin{aligned}
\frac{1}{NT} \overline{\mathbf{y}}'_{(-\ell)} \mathbf{A} \overline{\mathbf{y}}_{(-s)} &= \frac{1}{NT} \beta'_0 \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{X} \beta_0 + O(T^{-1}), \\
\frac{1}{NT} \mathbf{X}' \mathbf{A} \overline{\mathbf{y}}_{(-\ell)} &= \frac{1}{NT} \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_\ell) \mathbf{X} \beta_0 + O(T^{-1}).
\end{aligned}$$

Substituting these terms into $\text{Var}((NT)^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{u})$ yields the result. \square

Lemma A5. Under the conditions of Theorem 2, $(NT)^{-1/2} [\mathbf{W}' \mathbf{A} \mathbf{u} - E(\mathbf{W}' \mathbf{A} \mathbf{u})] = O_p(1)$ with $E(\mathbf{W}' \mathbf{A} \mathbf{u}) = O(N)$ and $(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = O_p(1)$. Further, $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$ exists.

Proof. Using Lemmas A2 and A4, one has $E(\mathbf{W}' \mathbf{A} \mathbf{u}) = O(N)$ and $\text{Var}((NT)^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{u}) = O(1) + O(T^{-1})$. Write $\mathbf{W}' \mathbf{A} \mathbf{W} = \overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} + \overline{\mathbf{W}}' \mathbf{A} \widetilde{\mathbf{W}} + \widetilde{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} + \widetilde{\mathbf{W}}' \mathbf{A} \widetilde{\mathbf{W}}$. From Lemma A4, $\overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} = O(NT)$. By substitution, $\widetilde{\mathbf{W}}' \mathbf{A} \widetilde{\mathbf{W}}$ has $\widetilde{\mathbf{y}}'_{(-\ell)} \mathbf{A} \widetilde{\mathbf{y}}_{(-s)}$, $\ell, s = 1, \dots, p$, populating its top-left $p \times p$ block and zero elsewhere, where, by some tedious algebra, $E(\widetilde{\mathbf{y}}'_{(-\ell)} \mathbf{A} \widetilde{\mathbf{y}}_{(-s)})$ has its leading term $N\sigma^2 \text{tr}(\mathbf{C}_{\ell s}) = O(NT)$ and $\text{Var}(\widetilde{\mathbf{y}}'_{(-\ell)} \mathbf{A} \widetilde{\mathbf{y}}_{(-s)})$ has its leading term $\sigma^4 [N \text{tr}(\mathbf{C}_{\ell s} \mathbf{C}_{\ell s} + \mathbf{C}_{\ell s} \mathbf{C}'_{\ell s}) + N\gamma_2 \text{tr}(\mathbf{C}_{\ell s} \odot \mathbf{C}_{\ell s})] = O(NT)$ by using Lemma A2. So, one can claim that $(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = O_p(1)$. The lower-right block of $(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$ is $(NT)^{-1} \mathbf{X}' \mathbf{A} \mathbf{X}$, which, by Assumption 3, has a probability limit. The top-left block of $(NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$ has elements $(NT)^{-1} \mathbf{y}'_{(-\ell)} \mathbf{A} \mathbf{y}_{(-s)}$, $\ell, s = 1, \dots, p$, which are dominated by $(NT)^{-1} [(I_N \otimes \Phi_p^{-1} \mathbf{L}^\ell) \mathbf{u} + (I_N \otimes \Phi_p^{-1} \mathbf{L}^\ell) \mathbf{X} \beta_0]' \mathbf{A} [(I_N \otimes \Phi_p^{-1} \mathbf{L}^s) \mathbf{u} + (I_N \otimes \Phi_p^{-1} \mathbf{L}^s) \mathbf{X} \beta_0]$ as $T \rightarrow \infty$. Note that they, in turn, consist of terms like $(NT)^{-1} \mathbf{u}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{u}$, $(NT)^{-1} \beta'_0 \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{u}$, and $(NT)^{-1} \beta'_0 \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{X} \beta_0$. From Lemma A2, \mathbf{C}_ℓ is uniformly (in T) bounded in row and column sums, and so are $\mathbf{C}_{\ell s} = \mathbf{C}'_\ell \mathbf{C}_s$ and $\mathbf{I}_N \otimes \mathbf{C}_{\ell s}$. Immediately, as $T \rightarrow \infty$, $(NT)^{-1} \beta'_0 \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{X} \beta_0$ exists, so long as $(NT)^{-1} \mathbf{X}' \mathbf{A} \mathbf{X}$ exists. Theorem A.1 of Kelejian and Prucha (2010) implies that $(NT)^{-1} \mathbf{u}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{u}$ and $(NT)^{-1} \beta'_0 \mathbf{X}' (\mathbf{I}_N \otimes \mathbf{C}_{\ell s}) \mathbf{u}$ are asymptotically normal as $T \rightarrow \infty$ (and converge in probability to their respective means). Thus, one can claim that $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W}$ exists. \square

Lemma A6. Under the conditions of Theorem 2, $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = \lim_{T \rightarrow \infty} \sigma^{-2} \Omega_{1,NT}$, where $\Omega_{1,NT}$ is given by (13).

Proof. $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \mathbf{W}' \mathbf{A} \mathbf{W} = \lim_{T \rightarrow \infty} (NT)^{-1} \overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}} + \text{plim}_{T \rightarrow \infty} (NT)^{-1} \widetilde{\mathbf{W}}' \mathbf{A} \widetilde{\mathbf{W}}$, in which $(NT)^{-1} \overline{\mathbf{W}}' \mathbf{A} \overline{\mathbf{W}}$ is given in the proof of Lemma A4 and $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \widetilde{\mathbf{W}}' \mathbf{A} \widetilde{\mathbf{W}}$ has non-zero elements $\lim_{T \rightarrow \infty} \sigma^2 T^{-1} \text{tr}(\mathbf{C}_{\ell s})$, $\ell, s = 1, \dots, p$, populating its top-left $p \times p$ block. \square

Lemma A7. Under the conditions of Theorem 2, as $T \rightarrow \infty$, $\Omega_{1,NT}^{-1/2} \sqrt{NT} \mathbf{g}_{NT} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$, where $\mathbf{g}_{NT} = \mathbf{g}_{NT}(\theta_0)$ and $\Omega_{1,NT}$ are given by (7) and (13), respectively.

Proof. Recall that $E(W' Au) = -E(u' Au)h$, where $h = O(T^{-1})$. Also, $\text{Var}(u' Au) = \sigma^4 N(T-1)(2 + \gamma_2(T-1)/T)$, so $E(W' Au) + u' Auh = O_p(\sqrt{N/T})$. From Lemma A5, $W' Au - E(W' Au) = O_p(\sqrt{NT})$. Thus, one can write $(\sqrt{NT} \mathbf{g}_{NT} = (NT)^{-1/2} [W' Au - E(W' Au) + E(W' Au) + u' Auh] = (NT)^{-1/2} [W' Au - E(W' Au)] + o_p(1)$ as $T \rightarrow \infty$. By substituting (5) into $W' Au$, one has the top $p \times 1$ block of $W' Au$ consisting of, $\ell = 1, \dots, p$,

$$u'(I_N \otimes C_\ell)u + u'(I_N \otimes C_\ell)X\beta_0 + u'(I_N \otimes C_\ell)\mathbf{1}\alpha \\ + \sum_{s=0}^{\ell-1} u'(I_N \otimes C_{s,\ell-s}e_1)\mathbf{y}_{-s} + \sum_{s=\ell}^{p-1} u'(I_N \otimes C_{(s-p),\ell-s}e_1)\mathbf{y}_{(-s)},$$

where $C_{s,\ell-s} = M\Phi_p^{-1}\Phi_s L^{\ell-s-1}$ and $C_{(s-p),\ell-s} = M\Phi_p^{-1}\Phi_{(s-p)} L^{\ell-s-1}$. From Lemma A2, $u'(I_N \otimes C_\ell)\mathbf{1}\alpha$, $\sum_{s=0}^{\ell-1} u'(I_N \otimes C_{s,\ell-s}e_1)\mathbf{y}_{(-s)}$, and $\sum_{s=\ell}^{p-1} u'(I_N \otimes C_{(s-p),\ell-s}e_1)\mathbf{y}_{(-s)}$ are all of order $O_p(\sqrt{N})$, $u'(I_N \otimes C_\ell)X\beta_0 = O_p(\sqrt{NT})$, and $u'(I_N \otimes C_\ell)u - E(u'(I_N \otimes C_p)u) = O_p(\sqrt{NT})$. Therefore, as $T \rightarrow \infty$,

$$\sqrt{NT} \mathbf{g}_{NT} = \frac{1}{\sqrt{NT}} \begin{pmatrix} u'(I_N \otimes C_1)u - E(u'(I_N \otimes C_1)u) + u'(I_N \otimes C_1)X\beta_0 \\ \vdots \\ u'(I_N \otimes C_p)u - E(u'(I_N \otimes C_p)u) + u'(I_N \otimes C_p)X\beta_0 \\ X' Au \end{pmatrix} + o_p(1).$$

From Lemma A2, C_t is uniformly (in T) bounded in row and column sums. It follows that $(I_N \otimes C_\ell)$ are uniformly (in NT) bounded in row and column sums. Obviously, $A = I_N \otimes M$ also has this property. Thus, if $T \rightarrow \infty$, regardless of N , one can invoke Theorem A.1 of Kelejian and Prucha (2010) to have the asymptotic distribution result. \square

Lemma A8. Under the conditions of Theorem 4, as $T \rightarrow \infty$, $\Omega_{1,NT}^{-1/2} \sqrt{NT} \mathbf{g}_{NT} \xrightarrow{d} N(\mathbf{0}, \mathbf{I})$, where $\mathbf{g}_{NT} = \mathbf{g}_{NT}(\theta_0)$ and $\Omega_{1,NT}$ are given by (25) and (29), respectively.

Proof. Recall that Ψ_ℓ is diagonal with $O(T^{-1})$ elements. Since $M\Sigma_i M$ is uniformly (in T) bounded in row and column sums, one has $E(u'_i M\Psi_\ell M u_i) = \text{tr}(M\Sigma_i M\Psi_\ell) = O(1)$ from Lemma A2 of Bao et al. (2020). Likewise, $\text{tr}(\Sigma_i M\Psi_\ell M\Sigma_i M\Psi_\ell M) = \text{tr}(M\Sigma_i M\Psi_\ell M\Sigma_i M\Psi_\ell)$ is $O(T^{-1})$ and each element of $M\Psi_\ell M$ is $O(T^{-1})$ from a similar proof as in Lemma A3 of Bao et al. (2020). Then it follows that $\text{Var}(u'_i M\Psi_\ell M u_i) = \text{tr}(\Sigma_i^{(4)} \odot M\Psi_\ell M \odot M\Psi_\ell M) + 2\text{tr}(\Sigma_i M\Psi_\ell M\Sigma_i M\Psi_\ell M) = O(T^{-1})$ and $u'_i M\Psi_\ell M u_i = O(1) + O_p(T^{-1/2})$. Next, $y'_{i,(-\ell)} M u_i = u'_i C_\ell u_i + \alpha_i u'_i C_\ell \mathbf{1} + u'_i C_\ell X_i \beta_0 + \sum_{j=0}^{\ell-1} u'_i C_{j,\ell-j} e_1 y_{i,-j} + \sum_{j=\ell}^{p-1} u'_i C_{(j-p),\ell-j} e_1 y_{i,-j}$. Term by term, $E(u'_i C_\ell u_i) = \text{tr}(\Sigma_i C_\ell) = \text{tr}(\Sigma_i D_\ell) = O(1)$ since the diagonal elements of D_ℓ are $O(T^{-1})$. Following similar steps as in the proof of Lemma A2, since pre- or post-multiplying $\Phi_p^{-1} L^\ell$ by Σ_i does not change the uniform boundedness of $\Phi_p^{-1} L^\ell$, one can claim $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_i \mathbf{1}$, $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i \mathbf{1}$, $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_i (\Phi_p^{-1} L^s)' \Sigma_i \mathbf{1}$, $\mathbf{1}' \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i (\Phi_p^{-1} L^s)' \mathbf{1}$, and $\text{tr}((\Phi_p^{-1} L^s)' \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i)$ are $O(T)$. Then, by substituting $C_s = M\Phi_p^{-1} L^s$, $C_\ell = M\Phi_p^{-1} L^\ell$, and $M = I - T^{-1} \mathbf{1}\mathbf{1}'$, one has $\text{tr}(C_s \Sigma_i C_\ell \Sigma_i) = O(1)$ and $\text{tr}(C'_s \Sigma_i C_\ell \Sigma_i) = \text{tr}((\Phi_p^{-1} L^s)' \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i) + O(1) = O(T)$. This leads to $\text{Var}(u'_i C_\ell u_i) = \text{tr}(\Sigma_i^{(4)} \odot C_\ell \odot C_\ell) + \text{tr}[(\Sigma_i C_\ell \Sigma_i (C_\ell + C'_\ell))] = O(1) + O(T)$. So, one can claim that $u'_i C_\ell u_i = O(1) + O_p(T^{1/2})$. From the proof of (iii) in Lemma A2, $M\Phi_p^{-1} L^\ell \mathbf{1}$ has, at most, a finite number of $O(1)$ elements, and all remaining elements are $O(T^{-1})$. It is obvious that $\Sigma_i^{1/2} M\Phi_p^{-1} L^\ell \mathbf{1}$ shares the same properties. Then, $\alpha_i u'_i C_\ell \mathbf{1} = O_p(1)$ and $\mathbf{1}' C'_s \Sigma_i C_\ell \mathbf{1} = O(1)$. Note that $\text{Cov}(u'_i C_\ell X_i \beta_0, u'_i C_s X_i \beta_0) = \beta'_0 X'_i C'_\ell \Sigma_i C_s X_i \beta_0 = \beta'_0 X'_i (\Phi_p^{-1} L^\ell)' M \Sigma_i^{1/2} \Sigma_i^{1/2} M \Phi_p^{-1} L^s X_i \beta_0$

and it has the same magnitude as $\beta_0' X_i' (\Phi_p^{-1} L^\ell)' M \Phi_p^{-1} L^s X_i \beta_0$ in view of Lemma A2 of Bao et al. (2020). So, $u_i' C_\ell X_i \beta_0 = O_P(T^{1/2})$. With all these results, one can claim that the variance of $y_{i,(-\ell)}' M u_i - u_i' M \Psi_\ell M u_i$ is dominated by that of $u_i' C_\ell u_i + u_i' C_\ell X_i \beta_0$. The asymptotic distribution result then follows from Theorem A.1 of Kelejian and Prucha (2010). \square

Lemma A9. Under the conditions of Theorem 4, $\text{plim}_{T \rightarrow \infty} (NT)^{-1} G_{NT} = -\text{plim}_{T \rightarrow \infty} (NT)^{-1} W' A W$, where G_{NT} is given by (26)

Proof. Both Ψ_ℓ and $\Psi_{\ell s}$ are diagonal with $O(T^{-1})$ elements, so the terms in $(NT)^{-1} G_{NT}$ are dominated by $-(NT)^{-1} y_{i,(-\ell)}' M W_i$, $\ell = 1, \dots, p$, and $-(NT)^{-1} X_i' M W_i$. \square

Lemma A10. Under the conditions of Theorem 4, $\text{plim}_{T \rightarrow \infty} (NT)^{-1} \sum_{i=1}^N W_i' M \Sigma_i M W_i = \lim_{T \rightarrow \infty} \Omega_{1,NT}$, where $\Omega_{1,NT}$ is given by (29).

Proof. This follows similarly from the proof of Lemma A6. \square

Lemma A11. Under Assumptions 1–5 and that $DP(p)$ has a unit root,

$$\text{Var}(Y^{-1/2} W' A u) = Y^{-1/2} \sigma^2 \begin{pmatrix} \Gamma_2 + \Gamma_3 & F_1' + F_2' \\ F_1 + F_2 & X' A X \end{pmatrix} Y^{-1/2} + o(1),$$

where Γ_2 and Γ_3 have $\beta_0' X' (I_N \otimes C_{\ell s}) X \beta_0$ and $1' C_{\ell s} 1 E(\alpha' \alpha)$, respectively, in their (ℓ, s) -positions, $\ell, s = 1, \dots, p$, and F_1 and F_2 have $X' (I_N \otimes C_\ell) X \beta_0$ and $X' (I_N \otimes C_\ell 1) E(\alpha)$, respectively, as their ℓ -th columns, $\ell = 1, \dots, p$.

Proof. Following the proof of Lemma A4 and using Lemma A3, one sees that now the leading term appearing in $\text{Var}(\widetilde{W}' A u)$ is $\sigma^2 1' C_{\ell s} 1 E(\tilde{\alpha}' \tilde{\alpha}) = O(NT^3)$ and the leading terms in $\widetilde{W}' A \widetilde{W}$ are $\beta_0' X' (I_N \otimes C_{\ell s}) X \beta_0 = O(NT^3)$ and $1' C_{\ell s} 1 \tilde{\alpha}' \tilde{\alpha} = O(NT^3)$. The remaining terms (e.g., $X' (I_N \otimes C_\ell) X \beta_0$ and $X' (I_N \otimes C_\ell 1) \tilde{\alpha}$) are, at most, $O(NT^2)$. Thus, the leading terms in $\text{Var}(W' A u)$ are $\beta_0' X' (I_N \otimes C_{\ell s}) X \beta_0 + 1' C_{\ell s} 1 [\tilde{\alpha}' \tilde{\alpha} + E(\tilde{\alpha}' \tilde{\alpha})] = \beta_0' X' (I_N \otimes C_{\ell s}) X \beta_0 + 1' C_{\ell s} 1 E(\alpha' \alpha) = O(NT^3)$ that appear in its top-left block. \square

Lemma A12. Under Assumptions 1–5 and that $DP(p)$ has a unit root, $Y^{-1/2} [W' A u - E(W' A u)] = O_p(1)$ and $Y^{-1/2} W' A W Y^{-1/2} = O_p(1)$.

Proof. The orders of magnitudes of terms in $Y^{-1/2} [W' A u - E(W' A u)]$ are obvious, given the expressions of $E(W' A u)$ in (6) and $\text{Var}(Y^{-1/2} W' A u)$ from Lemma A11. Use again $W' A W = \widetilde{W}' A \widetilde{W} + \widetilde{W}' A \widetilde{W} + \widetilde{W}' A \widetilde{W} + \widetilde{W}' A \widetilde{W}$. The top-left, lower-left, and lower-right blocks of $\widetilde{W}' A \widetilde{W}$ are $O(NT^3)$, $O(NT^2)$, and $O(NT)$, respectively. The top-left block of $\widetilde{W}' A \widetilde{W}$ consists of $\tilde{y}_{(-\ell)}' A \tilde{y}_{(-s)}$ with mean 0 and variance $\tilde{y}_{(-\ell)}' A \text{Var}(\tilde{y}_{(-s)}) A \tilde{y}_{(-\ell)}$, whose leading term is

$$\begin{aligned} & \left[(I_N \otimes \Phi_p^{-1} L^\ell) X \beta_0 + (I_N \otimes \Phi_p^{-1} L^\ell 1) \tilde{\alpha} \right]' A \\ & \times \left[(I_N \otimes \Phi_p^{-1} L^s) E(uu') (I_N \otimes L^s \Phi_p^{-1}) + (I_N \otimes \Phi_p^{-1} L^s 1) E(\tilde{\alpha} \tilde{\alpha}') (I_N \otimes 1' L^s \Phi_p^{-1}) \right] \\ & \times A \left[(I_N \otimes \Phi_p^{-1} L^s) X \beta_0 + (I_N \otimes \Phi_p^{-1} L^s 1) \tilde{\alpha} \right]. \end{aligned}$$

Using Lemma A3, one can verify the leading term (involving $(I_N \otimes \Phi_p^{-1} L^s 1) E(\tilde{\alpha} \tilde{\alpha}') (I_N \otimes 1' L^s \Phi_p^{-1})$) in the expansion of the above product is $O(NT^6)$. So the top-left block of $\widetilde{W}' A \widetilde{W}$ is $O_p(\sqrt{NT^6})$. The lower-left block of $\widetilde{W}' A \widetilde{W}$ consists of $X' A \tilde{y}_{(-\ell)}$ with mean 0_k and variance $X' A \text{Var}(\tilde{y}_{(-\ell)}) A X$, whose leading term is $X' (I_N \otimes C_\ell C_\ell') X + X' (I_N \otimes$

$C_\ell \mathbf{1})E(\tilde{\alpha}\tilde{\alpha}')(I_N \otimes \mathbf{1}'C'_\ell)X$, where, in view of Lemma A3, $X'(I_N \otimes C_\ell C'_\ell)X = O(NT^3)$ and $X'(I_N \otimes C_\ell \mathbf{1})E(\tilde{\alpha}\tilde{\alpha}')(I_N \otimes \mathbf{1}'C'_\ell)X = O(N^2T^4)$. So the lower-left block of $\widetilde{W}'A\widetilde{W}$ is $O_p(NT^2)$. The top-left block of $\widetilde{W}'A\widetilde{W}$ consists of elements like $\tilde{y}'_{(-\ell)}M\tilde{y}_{(-s)}$, where $E(\tilde{y}'_{(-\ell)}M\tilde{y}_{(-s)})$ has its leading term $\mathbf{1}'C_{\ell s}\mathbf{1}E(\tilde{\alpha}'\tilde{\alpha}) = O(NT^3)$ and $\text{Var}(\tilde{y}'_{(-\ell)}M\tilde{y}_{(-s)})$ has its leading terms $\sigma^2\mathbf{1}'C_{\ell s}C_{s\ell}\mathbf{1}E(\tilde{\alpha}'\tilde{\alpha}) = O(NT^5)$ and $\sigma^2\mathbf{1}'C_{s\ell}C_{\ell s}\mathbf{1}E(\tilde{\alpha}'\tilde{\alpha}) = O(NT^5)$. So the top-left block of $\widetilde{W}'A\widetilde{W}$ is $O(NT^3) + O_p(\sqrt{NT^5})$. \square

Lemma A13. Under the conditions of Theorem 3, $\text{plim}_{T \rightarrow \infty} G_{NT}^* = -\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AW Y^{-1/2}$, where $G_{NT}^* = G_{NT}^*(\theta_0)$, $G_{NT}^*(\theta) = Y^{-1/2}\{\partial[W'A(y - W\theta) - E(W'A((y - W\theta)))]/\partial\theta'\}Y^{-1/2}$.

Proof. Lemma A12, together with Lemma A3, gives

$$\begin{aligned} \text{plim}_{T \rightarrow \infty} G_{NT}^* &= -\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AW Y^{-1/2} \\ &\quad - 2\text{plim}_{T \rightarrow \infty} Y^{-1/2} \begin{pmatrix} O(1) \\ \mathbf{0}_k \end{pmatrix} \begin{pmatrix} O(NT) + O_p(\sqrt{NT^3}) \\ O_p(\sqrt{NT}) \end{pmatrix}' Y^{-1/2} \\ &\quad + \text{plim}_{T \rightarrow \infty} Y^{-1/2} O_p(NT) \begin{pmatrix} O(T) & \mathbf{0}_{p \times k} \\ \mathbf{0}_{k \times p} & \mathbf{0}_k \end{pmatrix} Y^{-1/2} \\ &= -\text{plim}_{T \rightarrow \infty} Y^{-1/2}W'AW Y^{-1/2}. \quad \square \end{aligned}$$

Lemma A14. Under the conditions of Theorem 3, $\text{plim}_{T \rightarrow \infty} \sigma^2 Y^{-1/2}W'AW Y^{-1/2} = \lim_{T \rightarrow \infty} \Omega_{1,NT}^*(\alpha)$, where $\Omega_{1,NT}^*(\alpha)$ is given by (19).

Proof. From the proofs of Lemma A12, $Y^{-1/2}W'AW Y^{-1/2}$ is dominated by $Y^{-1/2}(\widetilde{W}'A\widetilde{W} + \widetilde{W}'A\widetilde{W})Y^{-1/2}$ when the fixed effects are deterministic. Its top-left block consists of leading terms $(NT^3)^{-1}\sigma^2[\beta_0'X'(I_N \otimes C_{\ell s})X\beta_0 + \mathbf{1}'C_{\ell s}\mathbf{1}\alpha'\alpha]$, the lower-left (or top-right) block consists of leading terms $(NT^2)^{-1}\sigma^2[X'(I_N \otimes C_\ell)X\beta_0 + X'(I_N \otimes C_\ell \mathbf{1})\alpha]$ (or its transpose), and the lower-right block is $(NT)^{-1}\sigma^2(X'AX)$. Under the conditions of Theorem 3, in view of Lemma A3, for a given N , all these terms have well defined limits as $T \rightarrow \infty$. If N also diverges as $T \rightarrow \infty$, by writing $Y^{-1/2}W'AW Y^{-1/2} = N^{-1} \sum_{i=1}^N W_i^* M W_i^*$, where W_i^* is the properly scaled version of W_i . (In particular, the first p columns of W_i are multiplied by $T^{-3/2}$ and the last k columns are multiplied by $T^{-1/2}$.) One can check that (the dominating part of) each element of the $(p+k) \times (p+k)$ summand is uniformly (in T) bounded for each i and Theorem 1 of Phillips and Moon (1999) applies, namely, the sequential convergence and joint convergence are the same, provided that the limit is defined when $N \rightarrow \infty$. With the matrix $\Omega_{1,NT}^*(\alpha)$ defined accordingly, the result follows. \square

Lemma A15. Under the conditions of Theorem 3, $\Omega_{1,NT}^{*-1/2}(\alpha)Y^{1/2}g_{NT}^* \xrightarrow{d} N(\mathbf{0}, I)$ as $T \rightarrow \infty$, where $g_{NT}^* = Y^{-1}[W'Au - E(W'Au)]$ and $\Omega_{1,NT}^*(\alpha)$ is given by (19).

Proof. By substituting (5) and using Lemma A3,

$$Y^{-1/2}[W'Au - E(W'Au)] = \begin{pmatrix} (NT^3)^{-1/2}[u'(I_N \otimes C_1)X\beta_0 + u'(I_N \otimes C_1 \mathbf{1})\alpha] + O_p(T^{-1/2}) \\ \vdots \\ (NT^3)^{-1/2}[u'(I_N \otimes C_p)X\beta_0 + u'(I_N \otimes C_p \mathbf{1})\alpha] + O_p(T^{-1/2}) \\ (NT)^{-1/2}X'Au \end{pmatrix},$$

where the $O_p(T^{-1/2})$ terms come from $(NT^3)^{-1/2}[u'(I_N \otimes C_{\ell s})u' - E(u'(I_N \otimes C_{\ell s})u')]$, since $\text{Var}(u'(I_N \otimes C_\ell)u') = O(NT^2)$, $\ell = 1, \dots, p$. Consider a term like $(NT^3)^{-1/2}u'(I_N \otimes$

$C_\ell)X\beta_0 = (NT^3)^{-1/2} \sum_{i=1}^N u_i' C_\ell X_i \beta_0$, $\ell = 1, \dots, p$. Let $\zeta_{\ell,i} = C_\ell X_i \beta_0 \neq 0$. Then, $T^{-3/2} u_i' C_\ell X_i \beta_0$ is the sum of T random variables, namely, $T^{-3/2} \sum_{t=1}^T u_{it} \zeta_{\ell,it}$. Note that $u_{it} \zeta_{\ell,it}$'s are independent across t and $\text{Var}(T^{-3/2} \sum_{t=1}^T u_{it} \zeta_{\ell,it}) = T^{-3} \sigma^2 \beta_0' X' C_{\ell\ell} X \beta_0 = O(1) > 0$.²² With Assumption 5, Lyapunov's central limit theorem can be used, and it can be claimed that for each i , $T^{-3/2} u_i' C_\ell X_i \beta_0$ is asymptotically normal as $T \rightarrow \infty$. When summing over i , by the same logic, for any N , one can claim that $(NT^3)^{-1/2} [u'(I_N \otimes C_\ell)X\beta_0]$ is asymptotically normal with $\text{Var}((NT^3)^{-1/2} [u'(I_N \otimes C_\ell)X\beta_0]) = (NT^{-3}) \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0 = O(1) > 0$. If N also diverges, the sequential asymptotic distribution follows straightforwardly, provided that $(NT^{-3}) \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0$ is well defined in the limit. For the joint asymptotic distribution (when T and N diverge at the same time, denoted by $(T, N \rightarrow \infty)$), define $\xi_{i,N,T} = (T^{-3/2} u_i' C_\ell X_i \beta_0) / \sqrt{\sum_{i=1}^N \text{Var}(T^{-3/2} u_i' C_\ell X_i \beta_0)}$, where $\sum_{i=1}^N \text{Var}(T^{-3/2} u_i' C_\ell X_i \beta_0)$ is obviously $T^{-3} \sigma^2 \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0$. Let $\mathbb{I}(\cdot)$ be the indicator function. Then, for any $\varepsilon > 0$,

$$\begin{aligned} & \sum_{i=1}^N \mathbb{E} \left[\xi_{i,N,T}^2 \mathbb{I}(|\xi_{i,N,T}| > \varepsilon) \right] \\ & \leq \sum_{i=1}^N \mathbb{E} \left[\xi_{i,N,T}^2 \mathbb{I} \left(\beta_0' X' C_\ell u_i u_i' C_\ell X_i \beta_0 > \varepsilon^2 \sigma^2 \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0 \right) \right] \\ & \leq \frac{NT^3}{\sigma^2 \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0} \\ & \quad \times \max_i \mathbb{E} \left[\frac{\beta_0' X' C_\ell u_i u_i' C_\ell X_i \beta_0}{T^3} \mathbb{I} \left(\frac{\beta_0' X' C_\ell u_i u_i' C_\ell X_i \beta_0}{T^3} > \varepsilon^2 \sigma^2 \frac{\beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0}{T^3} \right) \right], \end{aligned}$$

where $NT^3 / [\sigma^2 \beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0] = O(1)$, $\beta_0' X' C_\ell u_i u_i' C_\ell X_i \beta_0 / T^3$ is uniformly (in T) integrable, and $\beta_0' X' (I_N \otimes C_{\ell\ell}) X \beta_0 / T^3 \rightarrow \infty$ when $(T, N \rightarrow \infty)$. In view of Theorem 2 of Phillips and Moon (1999), the joint asymptotic distribution is the same. So for any N , one can claim that $(NT^3)^{-1/2} [u'(I_N \otimes C_\ell)X\beta_0]$ is asymptotically normal as $T \rightarrow \infty$, provided that the asymptotic variance matrix is always defined. By the same reasoning, $(NT^3)^{-1/2} u'(I_N \otimes C_\ell) \alpha$ and $(NT)^{-1/2} X' A u$, as well as their linear combinations, have similar properties. So in the end one can claim $Y^{-1/2} [W' A u - E(W' A u)]$ is asymptotically normal. The variance of $Y^{-1/2} W' A u$ is $Y^{-1/2} \text{Var}(W' A u) Y^{-1/2} = Y^{-1/2} \Omega_{1,NT}^*(\alpha) Y^{-1/2} + o(1)$ in view of Lemma A11. Thus, one has

$$[Y^{-1/2} \Omega_{1,NT}^*(\alpha) Y^{-1/2}]^{-1/2} Y^{-1/2} [W' A u - E(W' A u)] \xrightarrow{d} N(0, I).$$

Recall that $E(W' A u) = -E(u' A u)h$, where h has $O(1)$ elements in its top p positions and 0's in its lower k positions, and $\text{Var}(u' A u) = \sigma^4 N(T-1)(2 + \gamma_2(T-1)/T)$. So, $E(W' A u) + u' A u h$ has its top $p \times 1$ block consisting of $O_p(\sqrt{NT})$ terms and bottom $k \times 1$ block of zeros. From Lemma A12, $W' A u - E(W' A u)$ has $O_p(\sqrt{NT^3})$ elements in its top p positions and $O_p(\sqrt{NT})$ in its bottom k positions. Thus, one can write $Y^{1/2} g_{NT}^* = Y^{-1/2} [W' A u - E(W' A u) + E(W' A u) + u' A u h] = Y^{-1/2} [W' A u - E(W' A u)] + o_p(1)$. Combining all the results, one has the asymptotic distribution of $Y^{1/2} g_{NT}^*$. \square

Lemma A16. Under the conditions of Theorem 5,

$$\Omega_{1,NT}^{*-1/2}(\alpha) Y^{1/2} g_{NT}^* \xrightarrow{d} N(0, I),$$

where $g_{NT}^* = g_{NT}^*(\theta_0)$, $g_{NT}^*(\theta)$ is $Y^{-1} [NT g_{NT}(\theta)]$ with $g_{NT}(\theta)$ given by (25), and $\Omega_{1,NT}^*(\alpha)$ is given by (34).

Proof. Note that Ψ_ℓ is diagonal with $O(1)$ elements. Since $M\Sigma_i M$ is uniformly (in T) bounded in row and column sums, one has $E(u_i' M \Psi_\ell M u_i) = \text{tr}(M \Sigma_i M \Psi_\ell) = O(T)$ in view of Lemma A2 of Bao et al. (2020). Similarly, $\text{tr}(\Sigma_i M \Psi_\ell M \Sigma_i M \Psi_\ell M) = \text{tr}(M \Sigma_i M \Psi_\ell M \Sigma_i M \Psi_\ell)$ is $O(T)$ and each elements of $M \Psi_\ell M$ is $O(1)$ from a similar proof as in Lemma A3 of Bao et al. (2020). Then, it follows that $\text{Var}(u_i' M \Psi_\ell M u_i) = \text{tr}(\Sigma_i^{(4)} \odot M \Psi_\ell M \odot M \Psi_\ell M) + 2\text{tr}(\Sigma M \Psi_\ell M \Sigma M \Psi_\ell M) = O(T)$ and $u_i' M \Psi_\ell M u_i = O(T) + O_p(T^{1/2})$. On the other hand, $y'_{i,(-\ell)} M u_i = u_i' C_\ell u_i + \alpha_i u_i' C_\ell \mathbf{1} + u_i' C_\ell X_i \beta_0 + \sum_{j=0}^{t-1} u_i' C_{j,\ell-j} e_1 y_{i,-j} + \sum_{j=\ell}^{p-1} u_i' C_{(j-p),\ell-j} e_1 y_{i,-j}$. Term by term, $E(u_i' C_\ell u_i) = \text{tr}(\Sigma_i C_\ell) = \text{tr}(\Sigma_i D_\ell) = O(T)$ since the diagonal elements of D_ℓ are $O(1)$. One can claim that $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_i \mathbf{1}$ and $\text{tr}((\Phi_p^{-1} L^s)' \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i)$ are $O(T^2)$ by following similar steps as in the proof of Lemma A3. Also, $\mathbf{1}' \Phi_p^{-1} L^s \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i \mathbf{1}$, $\mathbf{1}' \Phi_p^{-1} L^\ell \Sigma_i (\Phi_p^{-1} L^s)' \Sigma_i \mathbf{1}$, and $\mathbf{1}' \Sigma_i \Phi_p^{-1} L^\ell \Sigma_i (\Phi_p^{-1} L^s)' \mathbf{1}$ are $O(T^3)$. This leads to $\text{Var}(u_i' C_\ell u_i) = \text{tr}(\Sigma_i^{(4)} \odot C_\ell \odot C_\ell) + \text{tr}[(\Sigma_i C_\ell \Sigma_i (C_\ell + C_\ell')] = O(T) + O(T^2)$. So one can claim that $u_i' C_\ell u_i = O(T) + O_p(T)$. From the proof of (iii) in Lemma A3, $M \Phi_p^{-1} L^\ell \mathbf{1}$ has, at most, a finite number of $O(T^3)$ elements, and all remaining elements are $O(T^2)$. Then, it is obvious that $\Sigma_i^{1/2} M \Phi_p^{-1} L^\ell \mathbf{1}$ shares the same properties. Therefore, $\mathbf{1}' C_s' \Sigma_i C_\ell \mathbf{1} = \mathbf{1}' (\Phi_p^{-1} L^s)' M \Sigma_i^{1/2} \Sigma_i^{1/2} M \Phi_p^{-1} L^\ell \mathbf{1} = O(T^3)$ and one can claim that $\alpha_i u_i' C_\ell \mathbf{1} = O_p(T^{3/2})$. Note that $\text{Cov}(u_i' C_\ell X_i \beta_0, u_i' C_s X_i \beta_0) = \beta_0' X_i' C_\ell' \Sigma_i C_s X_i \beta_0 = \beta_0' X_i' (\Phi_p^{-1} L^\ell)' M \Sigma_i^{1/2} \Sigma_i^{1/2} M \Phi_p^{-1} L^s X_i \beta_0$, and, in view of Lemma A2 of Bao et al. (2020), it has the same magnitude as $\beta_0' X_i' (\Phi_p^{-1} L^\ell)' M \Phi_p^{-1} L^s X_i \beta_0$. So, $u_i' C_\ell X_i \beta_0 = O_p(T^{3/2})$. With all these results, one can claim that $y'_{i,(-\ell)} M u_i - u_i' M \Psi_\ell M u_i$ is dominated by $\alpha_i u_i' C_\ell \mathbf{1} + u_i' C_\ell X_i \beta_0$. Thus,

$$\mathbf{Y}^{1/2} \mathbf{g}_{NT}^* = \begin{pmatrix} (NT^3)^{-1/2} [u'(I_N \otimes C_1) X \beta_0 + u'(I_N \otimes C_1 \mathbf{1}) \alpha] + O_p(T^{-1/2}) \\ \vdots \\ (NT^3)^{-1/2} [u'(I_N \otimes C_p) X \beta_0 + u'(I_N \otimes C_p \mathbf{1}) \alpha] + O_p(T^{-1/2}) \\ (NT)^{-1/2} X' A u \end{pmatrix}.$$

Following similar steps in the proof of Lemma A15, one can arrive at the desired asymptotic distribution. \square

Lemma A17. Under the conditions of Theorem 5,

$$\text{plim}_{T \rightarrow \infty} \mathbf{G}_{NT}^* = \text{plim}_{T \rightarrow \infty} \mathbf{Y}^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{W} \mathbf{Y}^{-1/2},$$

where $\mathbf{G}_{NT}^* = \mathbf{G}_{NT}^*(\theta_0)$ and $\mathbf{G}_{NT}^*(\theta)$ is $\mathbf{Y}^{-1/2} [NT \mathbf{G}_{NT}(\theta) \mathbf{Y}^{-1/2}]$ with $\mathbf{G}_{NT}(\theta)$ given by (26).

Proof. Given that Ψ_ℓ is diagonal with $O(1)$ elements, it follows that $M \Psi_\ell M$ is uniformly (in T) bounded in row and column sums. Recall $\mathbf{W}_i = (y_{i,(-1)}, \dots, y_{i,(-p)}, X_i)$. Substituting (4), one has the first three leading terms in $u_i' M \Psi_\ell M y_{i,(-s)}$, namely, $u_i' M \Psi_\ell M \Phi_p^{-1} L^\ell u_i$, $\alpha_i u_i' M \Psi_\ell M \Phi_p^{-1} \mathbf{1}$, and $u_i' M \Psi_\ell M \Phi_p^{-1} X_i \beta_0$. One can show that $u_i' M \Psi_\ell M \Phi_p^{-1} L^\ell u_i = O_p(T)$, $\alpha_i u_i' M \Psi_\ell M \Phi_p^{-1} \mathbf{1} = O_p(T^{3/2})$, and $u_i' M \Psi_\ell M \Phi_p^{-1} X_i \beta_0 = O_p(T^{3/2})$ by following the steps in the proof of Lemma A16. Also, $u_i' M \Psi_\ell M X_i = O_p(T^{1/2})$. So, $u_i' M \Psi_\ell M \mathbf{W}_i = O_p(T^{3/2})$. Now consider $y'_{i,(-\ell)} M \mathbf{W}_i = (y'_{i,(-\ell)} M y_{i,(-1)}, \dots, y'_{i,(-\ell)} M y_{i,(-p)}, y'_{i,(-\ell)} M X_i)$. Following the steps in the proof of Lemma A12, one can claim that the leading terms in $y'_{i,(-\ell)} M \mathbf{W}_i$ are $O_p(T^3)$, which dominate $u_i' M \Psi_\ell M \mathbf{W}_i$. Next, consider terms like $u_i' M \Psi_\ell M u_i$ that also appear in (26). The diagonal elements of $M \Phi_p^{-1} L^\ell \Phi_p^{-1} L^s$ are $-T^{-1} \Phi_p^{-1} L^\ell \Phi_p^{-1} L^s \mathbf{1}$. In view of Lemma A3. (ii), all the elements of the second part in (27) are $O(1)$. The i -th element of $\Phi_p^{-1} L^\ell \Phi_p^{-1} L^s \mathbf{1}$ is

$$\sum_{j=s}^{i+\ell} (a_1 + \sum_{r=2}^p a_r \lambda_r^{i-j-\ell}) \left[a_1(j-s) + \sum_{r=2}^p a_r \frac{1-\lambda_r^{j-s}}{1-\lambda_r} \right].$$

So all the elements of the first part in (27) are, at most, $O(T)$. Then, $\mathbf{u}_i' \mathbf{M} \mathbf{\Psi}_{\ell s} \mathbf{M} \mathbf{u}_i$ is dominated by $\mathbf{u}_i' \mathbf{M} \mathbf{D} \mathbf{g}(-T^{-1} \mathbf{\Phi}_p^{-1} \mathbf{L}^\ell \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{1}) \mathbf{M} \mathbf{u}_i$, which has $O(T^2)$ mean and $O(T^3)$ variance. Given all these results, one can claim that the top $p \times (p+k)$ block of \mathbf{G}_{NT}^* is dominated by terms like $(NT^3)^{-1} \sum_{i=1}^N (-\mathbf{y}_{i,(-\ell)}' \mathbf{M} \mathbf{W}_i) = O_p(1)$. The bottom $k \times (p+k)$ block of \mathbf{G}_{NT}^* is $(NT)^{-1} \sum_{i=1}^N (-\mathbf{X}_i' \mathbf{M} \mathbf{W}_i)$. Again, by substituting (4), one can verify that $\mathbf{X}_i' \mathbf{M} \mathbf{W}_i = O_p(T)$. In summary,

$$\mathbf{G}_{NT}^* = \mathbf{Y}^{-1/2} \left[\sum_{i=1}^N \begin{pmatrix} -\mathbf{y}_{i,(-1)}' \mathbf{M} \mathbf{W}_i \\ \vdots \\ -\mathbf{y}_{i,(-p)}' \mathbf{M} \mathbf{W}_i \\ -\mathbf{X}_i' \mathbf{M} \mathbf{W}_i \end{pmatrix} \right] \mathbf{Y}^{-1/2} + o_p(1)$$

and the result follows immediately. \square

Lemma A18. Under the conditions of Theorem 5,

$$\text{plim}_{T \rightarrow \infty} \mathbf{Y}^{-1/2} \left(\sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \mathbf{\Sigma}_i \mathbf{M} \mathbf{W}_i \right) \mathbf{Y}^{-1/2} = \lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,NT}^*(\boldsymbol{\alpha}),$$

where $\mathbf{\Omega}_{1,NT}^*(\boldsymbol{\alpha})$ is given by (34).

Proof. This follows similarly from the proof of Lemma A14. \square

Proof of Theorem 2. This follows from Lemmas A4 and A7. \square

Proof of Theorem 3. This follows from Lemmas A11 and A15. \square

Proof of Theorem 4. This follows from Lemma A8. \square

Proof of Theorem 5. This follows from Lemma A16. \square

Appendix D. Asymptotic Distribution of $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \mathbf{M} \mathbf{W}_i$ When N Is Fixed

If N is fixed, all four terms in the expansion of $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \mathbf{M} \mathbf{W}_i$ are $O_p(1)$. From the proof of Lemma A8 in Appendix C, the top $p \times 1$ block of $\mathbf{W}_i' \mathbf{M} \mathbf{u}_i$ is dominated by $\mathbf{u}_i' \mathbf{C}_\ell \mathbf{u}_i + \mathbf{u}_i' \mathbf{C}_\ell \mathbf{X}_i \boldsymbol{\beta}_0$, $\ell = 1, \dots, p$. For $\mathbf{\Omega}_{1,NT}$ defined by (29), suppose one writes $\mathbf{\Omega}_{1,NT} = N^{-1} \sum_{i=1}^N \mathbf{\Omega}_{1,Ti}$, then as $T \rightarrow \infty$, $T^{-1/2} \mathbf{W}_i' \mathbf{M} \mathbf{u}_i \xrightarrow{d} N(\mathbf{0}, \lim_{T \rightarrow \infty} \mathbf{\Omega}_{1,Ti})$. So, $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \mathbf{u}_i \mathbf{u}_i' \mathbf{M} \mathbf{W}_i \xrightarrow{p} N^{-1} \sum_{i=1}^N \mathbf{\Omega}_{1,Ti}^{1/2} \mathbf{b}_i \mathbf{b}_i' \mathbf{\Omega}_{1,Ti}^{1/2}$, where $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{I})$ is a $(p+k)$ -dimensional normal random vector and $E(\mathbf{b}_i \mathbf{b}_j') = \mathbf{0}$ for $i \neq j$. The lower-right block of $\mathbf{W}_i' \mathbf{M} \mathbf{W}_i$ is $\mathbf{X}_i' \mathbf{M} \mathbf{X}_i$ and its top-left $p \times p$ block is dominated by $\mathbf{u}_i' \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{u}_i + 2\mathbf{u}_i' \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{X}_i \boldsymbol{\beta}_0 + \boldsymbol{\beta}_0' \mathbf{X}_i' \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{X}_i \boldsymbol{\beta}_0$, $\ell, s = 1, \dots, p$ from the proofs of Lemmas A5 and A9.

Assuming that $T^{-1} \mathbf{u}_i' \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{u}_i$ and $T^{-1} \boldsymbol{\beta}_0' \mathbf{X}_i' \mathbf{L}^{\ell'} \mathbf{\Phi}_p^{-1'} \mathbf{M} \mathbf{\Phi}_p^{-1} \mathbf{L}^s \mathbf{X}_i \boldsymbol{\beta}_0$ both have properly defined probability limits as $T \rightarrow \infty$, then one may write $T^{-1} \mathbf{W}_i' \mathbf{M} \mathbf{W}_i \xrightarrow{p} \mathbf{Q}_i$. Further, $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \mathbf{W}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \mathbf{u}_i' \mathbf{M} \mathbf{W}_i \xrightarrow{p} N^{-1} \sum_{i=1}^N \mathbf{Q}_i \mathbf{Q}_i^{-1} \mathbf{b} \mathbf{b}' \mathbf{\Omega}_{1,Ti}^{1/2}$, where $\mathbf{Q} = \sum_{j=1}^N \mathbf{Q}_j$ and $\mathbf{b} = \sum_{j=1}^N \mathbf{\Omega}_{1,Tj}^{1/2} \mathbf{b}_j$, in light of the proof of Theorem 4 in Hansen (2007). Similarly, one also has $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \mathbf{W}_i (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{W}_i' \mathbf{M} \mathbf{W}_i \xrightarrow{p} N^{-1} \sum_{i=1}^N \mathbf{Q}_i \mathbf{Q}_i^{-1} \mathbf{b} \mathbf{b}' \mathbf{Q}_i^{-1} \mathbf{Q}_i$. Thus, $(NT)^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{M} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \mathbf{M} \mathbf{W}_i$ is not consistent for estimating $\mathbf{\Omega}_{1,NT}$, though it has a limiting distribution as $T \rightarrow \infty$ with N fixed.

In the special case when Q_i and $\Omega_{1,Ti}$ stay the same across i , then the limiting distribution of $(NT)^{-1} \sum_{i=1}^N W_i' M \hat{v}_i \hat{v}_i' M W_i$ is proportional to $\Omega_{1,NT}$, and Corollary 4.1 of Hansen (2007) suggests that the t -statistic converges to $\sqrt{N/(N-1)} t_{N-1}$, where t_{N-1} is the t distribution with $N-1$ degrees of freedom.²³ In the presence of (unconditional) cross-sectional heteroskedasticity, it is unlikely that Q_i and Ω_{1i} do not change with i . But under homoskedasticity (with possible conditional temporal heteroskedasticity) or temporal heteroskedasticity, the moment condition (25) and the asymptotic distribution (31) are still valid, and thus, one can use the $\sqrt{N/(N-1)} t_{N-1}$ approximation to conduct the t -test. As Hansen (2007) points out, if $N \rightarrow \infty$, then $\sqrt{N/(N-1)} t_{N-1}$ converges to the standard normal distribution, so the $\sqrt{N/(N-1)} t_{N-1}$ approximation can also be used under large N .

Appendix E. Cross-Sectional Correlation

Now consider the more general case when $\text{Cov}(u_i, u_j) = \Sigma_{ij} = \text{Dg}(\sigma_{ij,1}^2, \dots, \sigma_{ij,T}^2)$. (When $i = j$, $\Sigma_{ii} = \Sigma_i$.) That is, cross-sectional correlation (and heteroskedasticity) may exist, but temporal correlation is ruled out. For this situation, one needs both N and T to be large.

When there is no unit root, from Lemma A8, one sees that, for θ in a neighborhood of θ_0 , $g_i(\theta)$ is dominated by $W_i' M u_i$. (The top $p \times 1$ block of $g_i(\theta)$ in this neighborhood is dominated by $u_i' C_\ell u_i + u_i' C_\ell X_i \beta$.) So,

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T g_{it}(\theta) &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T w_{it}(u_{it} - \bar{u}_i) + o_p(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{\sqrt{N}} \sum_{i=1}^N w_{it} u_{it} - \frac{1}{\sqrt{N}} \sum_{i=1}^N \bar{u}_i \frac{1}{\sqrt{T}} \sum_{t=1}^T w_{it} + o_p(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{\sqrt{N}} \sum_{i=1}^N w_{it} u_{it} + o_p(1) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T N^{-1/2} W_t' u_t + o_p(1), \end{aligned}$$

where $W_t = (w_{1t}, \dots, w_{Nt})'$ and $u_t = (u_{1t}, \dots, u_{Nt})'$. This second last line follows because $T^{-1/2} \sum_{t=1}^T w_{it} = O_p(1)$, and $\bar{u}_i = O_p(T^{-1/2})$. Note that $\{N^{-1/2} W_t' u_t\}_{t=1}^T$ forms a martingale difference sequence, so does $\{N^{-1} W_t' u_t u_t' W_t - N^{-1} W_t' \Sigma_t W_t\}_{t=1}^T$, where $\Sigma_t = E(u_t u_t') = \text{Var}(u_t)$. Therefore,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(N^{-1} W_t' u_t u_t' W_t) = \text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T W_t' \Sigma_t W_t$$

provided that $\text{plim}_{N, T \rightarrow \infty} (NT)^{-1} \sum_{t=1}^T W_t' \Sigma_t W_t$ exists and is positive definite. A sufficient condition is that the positive definite $N \times N$ matrix Σ_t is bounded in the norm as $N \rightarrow \infty$. Then,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T g_{it} \xrightarrow{d} N \left(0, \text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T W_t' \Sigma_t W_t \right)$$

and accordingly, in view of the proof of Lemma A9, one has

$$\begin{aligned} &\sqrt{NT}(\hat{\theta} - \theta_0) \\ &\xrightarrow{d} N \left(0, \left(\text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} W' A W \right)^{-1} \left(\text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^T W_t' \Sigma_t W_t \right) \left(\text{plim}_{N, T \rightarrow \infty} \frac{1}{NT} W' A W \right)^{-1} \right). \end{aligned}$$

In practice, $\text{plim}_{N,T \rightarrow \infty} (NT)^{-1} \sum_{t=1}^T \mathbf{W}_t' \Sigma_t \mathbf{W}_t$ can be estimated by

$$\frac{1}{NT} \sum_{t=1}^T \mathbf{W}_t' (\hat{\mathbf{e}}_t - \tilde{\mathbf{e}}) (\hat{\mathbf{e}}_t - \tilde{\mathbf{e}})' \mathbf{W}_t$$

where $\tilde{\mathbf{e}} = T^{-1} \sum_{t=1}^T \hat{\mathbf{e}}_t$ and $\hat{\mathbf{e}}_t = \mathbf{y}_t - \mathbf{W}_t \hat{\boldsymbol{\theta}}$.

When there is a unit root, from Lemma A16, $g_i(\boldsymbol{\theta})$ is dominated by $\mathbf{W}_i' \mathbf{M} \mathbf{u}_i$, where its top $p \times 1$ block is dominated by elements like $\alpha_i \mathbf{u}_i' \mathbf{C}_\ell \mathbf{1} + \mathbf{u}_i' \mathbf{C}_\ell \mathbf{X}_i \boldsymbol{\beta}$, $\ell = 1, \dots, p$, which are $O_p(T^{3/2})$, and its lower $k \times 1$ block is $\mathbf{X}_i \mathbf{M} \mathbf{u}_i$, which is $O_p(T^{1/2})$. Proceeding similarly, one can write

$$\mathbf{Y}^{-1/2} \sum_{i=1}^N \sum_{t=1}^T g_{it}(\boldsymbol{\theta}) = \mathbf{Y}^{-1/2} \sum_{t=1}^T \mathbf{W}_t' \mathbf{u}_t + o_p(1)$$

and

$$\lim_{T \rightarrow \infty} \mathbf{Y}^{-1/2} \sum_{t=1}^T E(\mathbf{W}_t' \mathbf{u}_t \mathbf{u}_t' \mathbf{W}_t) \mathbf{Y}^{-1/2} = \text{plim}_{N,T \rightarrow \infty} \mathbf{Y}^{-1/2} \left(\sum_{t=1}^T \mathbf{W}_t' \Sigma_t \mathbf{W}_t \right) \mathbf{Y}^{-1/2}.$$

Finally, in view of the proof of Lemma A17, one has

$$\mathbf{Y}^{1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V} = (\text{plim}_{N,T \rightarrow \infty} \mathbf{Y}^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{W} \mathbf{Y}^{-1/2})^{-1} (\text{plim}_{N,T \rightarrow \infty} \mathbf{Y}^{-1/2} (\sum_{t=1}^T \mathbf{W}_t' \Sigma_t \mathbf{W}_t) \mathbf{Y}^{-1/2}) \times (\text{plim}_{N,T \rightarrow \infty} \mathbf{Y}^{-1/2} \mathbf{W}' \mathbf{A} \mathbf{W} \mathbf{Y}^{-1/2})^{-1}$.

Appendix F. Additional Simulation Results

This section provides simulation results when $\boldsymbol{\phi}_0 = (0.3, -0.2, -0.1)'$, $(0.3, 0.6, 0.1)'$ for the DP(3) model, as specified by (36) with the error term (37). The experimental design is exactly the same as that in Section 5 in the main text. Recall all the results are related to the sum of the autoregressive parameters, namely, $\phi_{01} + \phi_{02} + \phi_{03}$.

Table A1. Additional simulation results under homoskedasticity.

$\boldsymbol{\phi}_0$		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
$(0.3, -0.2, -0.1)'$	Bias ($\times 100$)	WG	-5.54	-2.56	-0.98	-1.28	-1.05	-0.52
		GMM	-1.72	-2.57	-0.98	-1.28	-1.05	-1.08
		BC	-0.79	-0.28	-0.07	-0.15	-0.14	-0.06
		BCJ	-0.77	-0.28	-0.07	-0.15	-0.14	-0.06
		HPJ	2.94	0.46	0.06	0.08	0.03	0.03
		RMM	0.03	-0.01	0.00	-0.05	-0.07	-0.04
		RMM _r	0.03	-0.01	0.00	-0.05	-0.07	-0.04
	RMSE ($\times 100$)	WG	6.13	3.47	1.72	2.61	2.47	2.28
		GMM	3.30	3.49	1.72	2.61	2.47	3.37
		BC	2.78	2.40	1.42	2.29	2.25	2.22
		BCJ	2.78	2.40	1.42	2.29	2.25	2.22
		HPJ	4.58	2.69	1.47	2.41	2.34	2.27
		RMM	2.67	2.37	1.42	2.29	2.25	2.22
		RMM _r	2.68	2.37	1.42	2.29	2.25	2.22
	Size (5%)	WG	57.32	19.30	10.39	8.80	7.44	5.68
		WG(h)	57.70	20.91	11.33	11.19	9.92	10.31
		GMM	9.25	19.14	10.37	8.73	7.39	6.16
		GMM(h)	10.05	21.64	11.73	12.00	10.81	12.63
		HPJ	14.31	4.83	4.51	4.85	4.77	4.77
		RMM(N)	6.04	6.03	5.63	6.98	7.08	9.83
		RMM(T)	6.07	5.36	4.96	5.20	5.03	4.96
		RMM _r (N)	6.07	6.00	5.63	7.00	7.10	9.83
		RMM _r (NT)	6.68	6.37	5.76	7.08	7.11	9.88
		RMM _r (T)	6.10	5.17	4.78	5.03	4.44	4.20

Table A1. Cont.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
(0.3, 0.6, 0.1)'	Bias ($\times 100$)	WG	−1.88	−0.61	−0.13	−0.20	−0.14	−0.05
		GMM	−0.46	−0.58	−0.13	−0.20	−0.14	−0.15
		BC	10.76	9.60	4.44	5.47	4.44	2.23
		BCJ	10.77	9.60	4.44	5.47	4.44	2.23
		HPJ	1.99	0.72	0.16	0.25	0.18	0.06
		RMM	0.03	0.01	0.00	0.00	0.00	0.00
		RMM _r	0.03	0.01	0.00	0.00	0.00	0.00
	RMSE ($\times 100$)	WG	2.13	0.81	0.20	0.36	0.28	0.14
		GMM	1.11	0.79	0.20	0.36	0.28	0.39
		BC	11.07	9.72	4.45	5.51	4.46	2.24
		BCJ	11.09	9.72	4.45	5.51	4.46	2.24
		HPJ	2.77	1.22	0.32	0.59	0.48	0.26
		RMM	1.00	0.53	0.14	0.29	0.24	0.13
		RMM _r	1.00	0.53	0.14	0.29	0.24	0.13
	Size(5%)	WG	48.44	22.53	15.58	11.65	9.62	6.56
		WG(h)	47.04	23.68	17.82	16.22	15.05	17.45
		GMM	7.46	20.72	15.54	11.53	9.54	6.76
		GMM(h)	8.42	22.74	18.40	17.11	15.97	18.56
		HPJ	23.33	14.43	11.96	9.48	9.07	7.54
		RMM(N)	6.40	7.23	7.01	9.58	10.84	15.17
		RMM(T)	5.57	5.41	5.13	5.67	5.55	5.03
		RMM _r (N)	6.31	7.11	6.90	9.54	10.87	15.10
		RMM _r (NT)	6.32	7.05	6.92	9.51	10.71	14.94
		RMM _r (T)	5.78	5.88	5.95	7.07	7.18	7.95

Note: See Table 1 in the main text.

Table A2. Additional simulation results under cross-sectional heteroskedasticity.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
(0.3, −0.2, −0.1)'	Bias ($\times 100$)	WG	−21.40	−9.10	−3.51	−3.74	−2.73	−1.03
		GMM	−8.69	−9.16	−3.51	−3.74	−2.73	−2.17
		BC	−2.55	−0.62	−0.11	−0.18	−0.10	−0.04
		BCJ	−2.57	−0.62	−0.11	−0.18	−0.10	−0.04
		HPJ	5.93	0.96	0.14	0.21	0.18	0.06
		RMM	−0.20	−0.12	−0.03	−0.10	−0.05	−0.03
		RMM _r	−0.23	−0.12	−0.03	−0.10	−0.05	−0.03
	RMSE ($\times 100$)	WG	22.14	10.36	4.61	5.64	4.88	3.49
		GMM	10.99	10.43	4.61	5.64	4.88	5.29
		BC	6.73	5.17	3.04	4.29	4.10	3.35
		BCJ	6.74	5.17	3.04	4.29	4.10	3.35
		HPJ	10.38	6.02	3.20	4.57	4.32	3.44
		RMM	6.47	5.20	3.04	4.30	4.10	3.35
		RMM _r	6.49	5.20	3.04	4.30	4.10	3.35
	Size (5%)	WG	98.46	53.54	27.99	18.00	13.32	7.25
		WG(h)	96.99	48.94	23.76	18.29	14.45	11.71
		GMM	31.94	53.17	27.93	17.87	13.16	8.14
		GMM(h)	27.36	49.58	24.47	18.94	15.54	14.90
		HPJ	15.22	6.96	5.58	5.27	5.75	5.15
		RMM(N)	6.34	7.02	6.49	7.66	8.19	10.44
		RMM(T)	13.04	9.60	8.49	7.54	7.64	6.21
		RMM _r (N)	6.35	7.03	6.44	7.63	8.19	10.46
		RMM _r (NT)	9.26	7.83	6.83	7.90	8.39	10.57
		RMM _r (T)	8.54	6.70	5.69	5.69	5.81	4.28
	Bias ($\times 100$)	WG	−26.46	−6.50	−1.52	−1.28	−0.74	−0.14
		GMM	−12.14	−6.35	−1.52	−1.28	−0.74	−0.46
		BC	−8.34	−0.64	−0.08	0.51	0.72	1.13
		BCJ	−8.36	−0.64	−0.08	0.51	0.72	1.13
		HPJ	8.07	5.02	1.58	1.39	0.87	0.19
		RMM	0.83	0.10	0.01	0.01	0.00	0.00
		RMM _r	0.92	0.10	0.01	0.01	0.00	0.00

Table A2. Cont.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
	RMSE ($\times 100$)	WG	26.82	6.80	1.62	1.50	0.94	0.27
		GMM	13.06	6.66	1.62	1.50	0.94	0.80
		BC	10.11	2.21	0.61	1.00	0.98	1.19
		BCJ	10.13	2.21	0.61	1.00	0.99	1.19
		HPJ	10.44	5.91	1.83	1.93	1.33	0.47
		RMM	7.33	2.14	0.55	0.77	0.57	0.23
		RMM _r	7.54	2.17	0.55	0.77	0.57	0.23
	Size (5%)	WG	100.00	96.90	87.55	44.27	28.21	9.49
		WG(<i>h</i>)	100.00	93.41	81.93	43.97	31.50	18.86
		GMM	78.51	95.93	87.51	44.06	28.05	11.75
		GMM(<i>h</i>)	74.54	92.19	82.46	45.32	33.42	23.42
		HPJ	21.06	38.18	45.46	21.51	16.91	9.83
		RMM(<i>N</i>)	5.09	6.69	6.91	9.00	9.73	14.41
		RMM(<i>T</i>)	31.87	12.73	8.57	7.25	6.68	5.60
		RMM _r (<i>N</i>)	5.23	6.63	6.87	8.84	9.65	14.43
		RMM _r (<i>NT</i>)	18.50	8.12	6.84	8.32	9.05	13.66
		RMM _r (<i>T</i>)	17.54	6.89	5.73	5.81	6.04	6.79

Note: See Table 1 in the main text.

Table A3. Additional simulation results under temporal heteroskedasticity.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
$(0.3, -0.2, -0.1)'$	Bias ($\times 100$)	WG	-27.50	-10.98	-3.88	-5.11	-4.06	-1.90
		GMM	-12.93	-11.09	-3.88	-5.11	-4.06	-4.14
		BC	-3.32	-0.94	-0.17	-0.39	-0.35	-0.14
		BCJ	-3.38	-0.92	-0.16	-0.38	-0.35	-0.14
		HPJ	7.66	1.11	-0.01	-0.07	-0.10	0.01
		RMM	1.16	-0.41	-0.11	-0.29	-0.29	-0.12
		RMM _r	-0.13	-0.07	-0.06	-0.20	-0.24	-0.11
	RMSE ($\times 100$)	WG	28.15	12.05	4.90	6.99	6.24	4.95
		GMM	15.14	12.17	4.90	6.99	6.24	7.87
		BC	7.50	5.24	3.04	4.88	4.83	4.60
		BCJ	7.51	5.24	3.04	4.88	4.83	4.60
		HPJ	12.82	6.03	3.47	5.54	5.46	5.07
		RMM	7.51	5.25	3.04	4.89	4.83	4.61
		RMM _r	6.87	5.26	3.04	4.89	4.83	4.61
	Size (5%)	WG	99.73	62.19	29.44	21.88	16.53	8.85
		WG(<i>h</i>)	99.63	62.05	27.23	22.19	17.38	12.23
		GMM	41.75	62.62	29.34	21.71	16.39	15.73
		GMM(<i>h</i>)	38.22	63.06	27.86	23.54	18.14	17.61
		HPJ	20.65	6.23	7.50	7.20	7.21	7.14
		RMM(<i>N</i>)	5.92	6.69	5.92	7.57	7.53	9.54
		RMM(<i>T</i>)	12.64	7.52	7.27	7.22	7.27	7.12
		RMM _r (<i>N</i>)	6.32	6.64	5.92	7.51	7.49	9.56
		RMM _r (<i>NT</i>)	8.25	8.04	6.29	8.02	7.89	9.62
		RMM _r (<i>T</i>)	7.66	6.69	5.29	5.53	5.21	4.26
	Bias ($\times 100$)	WG	-21.28	-10.40	-2.46	-3.77	-2.55	-0.73
		GMM	-10.85	-10.54	-2.46	-3.77	-2.55	-2.75
		BC	-2.65	-3.11	-0.54	-1.45	-1.03	-0.38
		BCJ	-3.07	-3.09	-0.53	-1.45	-1.03	-0.38
		HPJ	13.02	3.82	1.29	1.55	1.36	0.68
		RMM	16.75	-0.38	-0.49	-0.69	-0.47	-0.12
		RMM _r	0.13	0.28	0.04	0.18	0.15	0.03
	RMSE ($\times 100$)	WG	21.70	10.64	2.55	4.00	2.78	0.94
		GMM	11.97	10.80	2.55	4.00	2.78	3.22
		BC	5.29	4.24	1.00	2.21	1.68	0.78
		BCJ	5.36	4.23	1.01	2.21	1.68	0.78
		HPJ	14.69	5.40	1.70	2.70	2.28	1.18
		RMM	17.83	2.76	0.83	1.53	1.22	0.61
		RMM _r	4.43	2.82	0.77	1.69	1.37	0.66

Table A3. Cont.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
Size (5%)	WG	WG	99.99	99.99	99.27	92.78	79.90	33.50
		WG(h)	99.93	99.87	97.76	88.17	73.30	35.91
		GMM	63.43	99.98	99.26	92.75	79.81	60.53
		GMM(h)	62.19	99.86	97.90	89.19	75.31	57.29
		HPJ	51.16	21.70	27.82	15.92	16.39	14.10
		RMM(N)	1.91	7.99	15.13	14.12	13.71	15.72
		RMM(T)	91.57	13.60	19.76	15.29	14.46	11.30
		RMM _r (N)	5.78	5.65	6.46	7.07	8.21	14.10
		RMM _r (NT)	6.73	8.68	9.03	10.52	11.09	13.70
		RMM _r (T)	6.22	7.22	7.59	7.74	7.47	6.15

Note: See Table 1 in the main text.

Table A4. Additional simulation results under double heteroskedasticity.

ϕ_0		(N, T)	(100, 10)	(50, 20)	(50, 50)	(25, 40)	(20, 50)	(10, 100)
(0.3, −0.2, −0.1)′	Bias (×100)	WG	−18.86	−9.24	−3.45	−4.69	−3.66	−1.94
		GMM	−6.89	−9.11	−3.45	−4.69	−3.66	−4.05
		BC	−1.98	−0.68	−0.12	−0.36	−0.19	−0.22
		BCJ	−1.93	−0.67	−0.12	−0.36	−0.19	−0.22
		HPJ	6.07	0.84	0.13	0.05	0.14	−0.05
		RMM	−0.94	−0.36	−0.06	−0.27	−0.13	−0.21
		RMM _r	−0.12	−0.12	−0.04	−0.21	−0.10	−0.20
	RMSE (×100)	WG	19.45	10.31	4.43	6.53	5.81	4.84
		GMM	8.80	10.20	4.43	6.53	5.81	7.59
		BC	5.48	4.77	2.83	4.64	4.59	4.47
		BCJ	5.47	4.77	2.83	4.64	4.59	4.47
		HPJ	9.41	5.68	3.07	5.19	5.09	4.81
		RMM	5.29	4.77	2.83	4.65	4.59	4.47
		RMM _r	5.29	4.78	2.83	4.65	4.59	4.47
	Size (5%)	WG	98.41	55.29	26.25	20.08	14.82	8.54
		WG(h)	98.01	54.78	25.58	21.80	16.32	13.04
		GMM	24.19	53.47	26.17	19.93	14.72	14.17
		GMM(h)	24.61	53.78	26.07	22.97	17.41	17.73
		HPJ	18.04	7.34	6.19	7.18	6.85	6.44
		RMM(N)	6.52	6.64	6.35	7.50	7.45	10.04
		RMM(T)	8.87	7.36	6.71	7.06	6.88	6.61
		RMM _r (N)	6.11	6.69	6.36	7.53	7.42	10.08
		RMM _r (NT)	7.80	7.75	6.68	7.97	7.76	10.22
		RMM _r (T)	7.23	6.64	5.56	5.57	5.04	4.53
	Bias (×100)	WG	−23.86	−9.12	−1.80	−3.13	−2.14	−0.66
		GMM	−9.27	−8.78	−1.80	−3.13	−2.14	−2.54
		BC	−6.98	−2.36	−0.19	−0.93	−0.66	−0.30
		BCJ	−6.90	−2.34	−0.18	−0.93	−0.66	−0.30
		HPJ	6.55	4.26	1.56	2.05	1.62	0.69
		RMM	−3.84	−1.65	−0.27	−0.58	−0.37	−0.10
		RMM _r	0.93	0.37	0.01	0.10	0.09	0.02
	RMSE (×100)	WG	24.19	9.40	1.90	3.38	2.38	0.87
		GMM	10.09	9.07	1.90	3.38	2.38	3.02
		BC	8.54	3.61	0.72	1.75	1.37	0.70
		BCJ	8.52	3.62	0.72	1.75	1.37	0.70
		HPJ	8.88	5.57	1.82	2.88	2.34	1.15
		RMM	6.29	2.97	0.65	1.41	1.13	0.57
		RMM _r	6.69	3.21	0.63	1.52	1.23	0.61
	Size (5%)	WG	100.00	99.82	94.38	86.12	70.85	30.99
		WG(h)	100.00	99.24	89.24	77.05	61.59	33.59
		GMM	68.36	99.60	94.38	86.02	70.68	56.04
		GMM(h)	65.89	98.79	89.51	78.65	63.37	51.80
		HPJ	18.99	26.15	41.93	21.92	21.22	14.66
		RMM(N)	18.24	14.35	10.19	12.04	12.00	15.49
		RMM(T)	37.89	25.46	14.43	16.25	14.44	10.97
		RMM _r (N)	4.64	4.49	6.85	7.35	8.27	14.36
		RMM _r (NT)	17.18	12.61	8.11	10.66	10.85	13.78
		RMM _r (T)	16.05	10.80	6.87	7.72	7.03	6.30

Note: See Table 1 in the main text.

Notes

- 1 No attempt is made here to provide an exhaustive list of published papers in this literature. Readers can refer to [Okui \(2021\)](#) for a comprehensive literature review.
- 2 As one referee points out, many papers in the literature distinguish between exogeneity and endogeneity depending on whether a regressor is correlated with the idiosyncratic error by assuming it is correlated with the individual fixed effects. As such, the (random) regressors X in Assumption 3 to be introduced are exogenous. The word “endogeneity” or “endogenous” in this paper exclusively refers to the lagged dependent variables.
- 3 One referee suggests that, in view of the numerical equivalence, the estimator in this paper should be named the implicit indirect inference estimator. Nonetheless, given the different motivations, the suggested name is not adopted in this paper.
- 4 They mention the possibility of both cross-sectional and temporal heteroskedasticity under more stringent conditions, but they do not rigorously derive the properties of the resulting estimator.
- 5 The higher-order DP in [Juodis \(2013\)](#) is presented as a panel first-order vector autoregression (VAR(1)).
- 6 Of course, if one assumes that α_i are fixed constants of bounded magnitudes, then Assumption 1 should be dropped. When they are random, the i.i.d. assumption could be relaxed, so long as the relevant probability limits pertaining to the recentered moments and gradient are well defined.
- 7 In this case, Assumptions 1 and 2 do not apply. One could also follow [Han and Phillips \(2010\)](#) to assume that the fixed effects disappear when there is a unit root, but an extension along this line is not pursued in this paper. In practice, the fixed effects are not estimated anyway, so treating them as random or not does not affect the estimation strategy to be presented in this paper.
- 8 The derivation of (4) is analogous to rewriting lagged time-series vectors from an autoregressive process of order p : $y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t$, $t = 1, \dots, T$, where $y_0, y_{-1}, \dots, y_{t-p+1}$ are given. With obvious notation, $\mathbf{y} = \phi_1 \mathbf{y}_{(-1)} + \dots + \phi_p \mathbf{y}_{(-p)} + \mathbf{u}$. Note that $\mathbf{y}_{(-\ell)} = \mathbf{L} \mathbf{y}_{(-\ell+1)} + \mathbf{e}_1 y_{-(\ell-1)}$, $\ell = 1, \dots, p$, where $\mathbf{y}_{(0)} = \mathbf{y}$. Substituting $\mathbf{y}_{(-1)} = \mathbf{L} \mathbf{y} + \mathbf{e}_1 y_0$, $\mathbf{y}_{(-2)} = \mathbf{L} \mathbf{y}_{(-1)} + \mathbf{e}_1 y_{-1} = \mathbf{L}^2 \mathbf{y} + \mathbf{L} \mathbf{e}_1 y_0 + \mathbf{e}_1 y_{-1}$, and so on to the right-hand side of $\mathbf{y} = \phi_1 \mathbf{y}_{(-1)} + \dots + \phi_p \mathbf{y}_{(-p)} + \mathbf{u}$, one can solve for \mathbf{y} and hence $\mathbf{y}_{(-1)}$, given by $\mathbf{y}_{(-1)} = \Phi_p^{-1}(\mathbf{L} \mathbf{u} + \mathbf{e}_1 y_0 + \sum_{s=1}^{p-1} \Phi_{(s-p)} \mathbf{L}^{-s} \mathbf{e}_1 y_{-s})$ at the true parameter vector. With $\mathbf{y}_{(-1)}$, the expression of $\mathbf{y}_{(-2)}$ follows from $\mathbf{y}_{(-2)} = \mathbf{L} \mathbf{y}_{(-1)} + \mathbf{e}_1 y_{-1}$. By successive substitutions, $\mathbf{y}_{(-\ell)} = \Phi_p^{-1}(\mathbf{L}^\ell \mathbf{u} + \sum_{s=0}^{\ell-1} \Phi_s \mathbf{L}^{\ell-1-s} \mathbf{e}_1 y_{-s} + \sum_{s=\ell}^{p-1} \Phi_{(s-p)} \mathbf{L}^{\ell-1-s} \mathbf{e}_1 y_{-s})$.
- 9 At the time of writing, the author was not aware of the work of [Breitung et al. \(2022\)](#). In addition to differences in the allowable heteroskedasticity, unit root, and asymptotic regimes, their approach is motivated by correcting the profile score from a normal likelihood function, but the estimator in this paper explicitly uses the endogeneity of the within-group transformed lagged dependent variables to construct the recentered moment conditions $E(g_{NT}(\theta_0)) = \mathbf{0}$ in this and the next sections.
- 10 $\text{Var}(\sqrt{NT} \mathbf{g}_{NT}) = [\text{Var}(\mathbf{W}' \mathbf{A} \mathbf{u}) + \mathbf{h} \mathbf{h}' \text{Var}(\mathbf{u}' \mathbf{A} \mathbf{u}) + \text{Cov}(\mathbf{W}' \mathbf{A} \mathbf{u}, \mathbf{u}' \mathbf{A} \mathbf{u}) \mathbf{h}' + \mathbf{h} \text{Cov}(\mathbf{u}' \mathbf{A} \mathbf{u}, \mathbf{W}' \mathbf{A} \mathbf{u})] / NT$. Further, the (i, j) -th element of $\mathbf{h} \mathbf{h}' \text{Var}(\mathbf{u}' \mathbf{A} \mathbf{u})$ is equal to the negative of the (i, j) -th element of $\text{Cov}(\mathbf{W}' \mathbf{A} \mathbf{u}, \mathbf{u}' \mathbf{A} \mathbf{u}) \mathbf{h}'$ and $\text{Var}(\mathbf{u}' \mathbf{A} \mathbf{u}) / NT = \sigma^4 [2(T-1)/T + \gamma_2(T-1)^2/T^2]$. These results lead to the variance expression (10). See [Bao and Yu \(2023\)](#) for the detailed derivation.
- 11 [Dovonon et al. \(2020\)](#) point out that there may exist situations where global identification holds but first-order local identification fails. They provide such an example based on the special case of a unit-root DP(1) with no exogenous covariates, namely, $p = 1$, $k = 0$, and $\phi_0 = 1$, where a GMM estimator is used. When $T = 4$ and $E[(\alpha_i - (1 - \phi_0)y_{i1})^2] = 0$, they show that the Jacobian is a null vector, and thus, the GMM is not able to first-order identify the parameter, though global identification and second-order local identification still hold. For the estimator proposed in this paper, it can be shown that when $\phi_0 = 1$, the (unscaled) moment condition at $\phi_0 = 1$ becomes $\mathbf{y}'_{(-1)} \mathbf{A} \mathbf{u} + \mathbf{u}' \mathbf{A} \mathbf{u} / 2$ and its derivative at $\phi_0 = 1$ becomes $-\mathbf{y}'_{(-1)} \mathbf{A} \mathbf{y}_{(-1)} - \mathbf{y}'_{(-1)} \mathbf{A} \mathbf{u} + (T-2) \mathbf{u}' \mathbf{A} \mathbf{u} / 6$, where $E(\mathbf{y}'_{(-1)} \mathbf{A} \mathbf{y}_{(-1)}) = N(T^2 - 1)\sigma^2 / 6 + (T^3 - T)E(\alpha' \alpha) / 12$, $E(\mathbf{y}'_{(-1)} \mathbf{A} \mathbf{u}) = N\sigma^2(1 - T)/2$, and $E(\mathbf{u}' \mathbf{A} \mathbf{u}) = N(T-1)\sigma^2$. If further $\alpha = \mathbf{0}$, then the Jacobian is equal to 0, and thus, the first-order local identification condition fails in this special case. This is also recognized by [Dhaene and Jochmans \(2016\)](#) (see their Corollary 4.1) when they design their adjusted profile likelihood estimator. For a general DP(p), under some rare circumstances, it may happen that there are multiple zeros when one solves the adjusted profile score function and for local identification, [Dhaene and Jochmans \(2016\)](#) recommend numerical search starting from the WG estimator. Results from numerical grid search in [Dhaene and Jochmans \(2016\)](#) and [Bao and Yu \(2023\)](#) suggest that the issue of multiple zeros may not be of practical concern.
- 12 In [Kelejian and Prucha \(2010\)](#), the linear form in \mathbf{u} involves a vector of non-stochastic elements. Here, X may be random. Checking their proof, which relies on results from their earlier work ([Kelejian and Prucha 2001](#), Theorem 1), one can see that as long as X is strictly exogenous, then the sigma-field that defines the martingale difference array in the proof of Theorem 1 in [Kelejian and Prucha \(2001\)](#) can be extended and the result continues to hold. In the case of random X , one can replace $X'AX$ with $E(X'AX)$ in various variance expressions. Further note that, in view of footnote 13 in [Kelejian and Prucha \(2001\)](#), one can think of their k_n as NT and their n as T .
- 13 If it is also the case that $N \rightarrow \infty$, the sequential $((T, N \rightarrow \infty)_{\text{seq}})$ and joint $((T, N \rightarrow \infty))$ asymptotic distributions may be different ([Phillips and Moon 1999](#)). Under the assumptions in this paper, for the stable case, Theorem 1 of [Kelejian and Prucha \(2001\)](#) essentially states no difference under the two asymptotic regimes. For the unit-root case, Appendix C (Lemma A15) shows that

the two asymptotic regimes deliver the same asymptotic distribution. Recall that $G_{NT} = O_p(1) + [O(T^{-2}) + O_p(N^{-1/2}T^{-3/2})] + O_p(T^{-1})$, where the $O_p(1)$ term is $-(NT)^{-1}W'AW$. Further,

σ^2 is defined as $\lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \sigma_i^2$ in Breitung et al. (2022).

Recall from Note 11, that if further $\alpha = 0$, the method in this paper cannot identify ϕ_0 . This is also in line with Theorem 4 of Hahn and Kuersteiner (2002) that shows that their bias correction is not expected to work under this special case. Checking their proof (see their Lemma 12), one can interpret their recentered WG estimator in this special case as arising from some (unscaled) moment condition, which is $y'_{-1}Au + 3u'(I_N \otimes L'\Phi^{-1}M\Phi^{-1}L)u/(T+1)$ at the true parameter value and has exact expectation 0 when $\phi_0 = 1$. From Note 11, for the RMM estimator in this special case, the (unscaled) moment condition at $\phi_0 = 1$ becomes $y'_{(-1)}Au + u'Au/2$. So, if one designs a (unscaled) moment condition $y'_{(-1)}A(y - \phi y_{(-1)}) + (y - \phi y_{(-1)})'A(y - \phi y_{(-1)})/2$, which is valid at $\phi_0 = 1$ but not valid at $\phi_0 \neq 1$, then its derivative, evaluated at $\phi_0 = 1$, is dominated by $-y'_{(-1)}Ay_{(-1)}$. Using results from Appendix C, one can show $\text{plim}_{N,T \rightarrow \infty} y'_{(-1)}Ay_{(-1)}/(NT^2) = \sigma^2/6$, which is the same as Lemma 11 of Hahn and Kuersteiner (2002). Further, following similarly the proof of Lemma 12 of Hahn and Kuersteiner (2002), one has $(NT^2)^{-1/2}(y'_{(-1)}Au + u'Au/2) \xrightarrow{d} N(0, \sigma^4/12)$ as $N, T \rightarrow \infty$. Correspondingly, $\sqrt{NT^2}(\hat{\phi} - 1) \xrightarrow{d} N(0, 3)$ as $N, T \rightarrow \infty$. Recall that $\hat{\phi}$ here is solved from $y'_{(-1)}A(y - \phi y_{(-1)}) + (y - \phi y_{(-1)})'A(y - \phi y_{(-1)})/2 = 0$. Theorem 4 of Hahn and Kuersteiner (2002) indicates that if one were using the true parameter value in this case, one would have recentered $\hat{\theta} - \theta_0$ by $-3/(T+1)$ instead of the general bias formula $-(1 + \hat{\theta})/T$ (in their notation, where $\hat{\theta}$ is the WG estimator). Similarly here, one would have used the moment condition $y'_{(-1)}A(y - \phi y_{(-1)}) + (y - \phi y_{(-1)})'A(y - \phi y_{(-1)})/2$ instead of the general one $y'_{(-1)}Au(\phi) + u'(\phi)Au(\phi)1'\Phi^{-1}(\phi)L1/[T(T-1)]$.

In particular, additional simulation results are available under four non-normal distributions that are also considered in Bao and Yu (2023): uniform on $[0, 1]$, student- t distribution with five degrees of freedom, log-normal distribution $\ln N(0, 1)$, and mixture of $N(-3, 1)$ and $N(3, 1)$ with half probability each.

The number of instruments for the GMM estimator of Arellano and Bond (1991) is of order $O(T^2)$. To prevent instrument proliferation, the total number of instruments from lagged y is capped at $T_0(T_0 - 1)/2$, where $T_0 = \min(50, T)$, such that when $T > 50$, only the first $q = T_0(T_0 - 1)/2$ columns in the matrix of instruments are retained.

Bun and Carree (2006) consider DP(1) only. The BC estimator is based on the panel VAR(1) representation of DP(p) in Juodis (2013).

The complete results for each single element of θ_0 (including β_0) under each parameter configuration are available upon request and they lead to similar conclusions as reported in this section.

Even though the estimator itself is consistent under large N and fixed T , De Vos and Everaert (2021) assume both to be large to derive its asymptotic distribution. For practical inference, they suggest re-sampling the cross-sectional units and then using the empirical distribution of their estimates from the bootstrapped samples to approximate the asymptotic distribution.

If there are repeated roots, the exact expressions of the various terms in the lemmas to follow are different, but their orders of magnitude stay the same. This is because, for instance, for $|\lambda| < 1$, $\sum_{t=1}^T \lambda^t$, and $\sum_{t=1}^T t^j \lambda^t$, where j is a positive finite integer, are of the same magnitude as $T \rightarrow \infty$.

$T^{-3}\sigma^2\beta_0'X_i'C_{\ell\ell}X_i\beta_0$ is essentially a quadratic form in the idempotent matrix M , which is positive unless $\Phi_p^{-1}L^\ell X_i\beta_0$ is a constant vector.

The matrix $\sum_{i=1}^N W_i'M\hat{\nu}_i\hat{\nu}_i'MW_i$, in this case, may be adjusted by $N/(N-1)$. Further, the F statistic for testing q linear restrictions based on (31) with the variance matrix estimated by (32) (and $\sum_{i=1}^N W_i'M\hat{\nu}_i\hat{\nu}_i'MW_i$ possibly adjusted) converges to $[Nq/(N-q)]F_{q,N-q}$, where $F_{q,N-q}$ denotes an F distribution with q numerator and $N-q$ denominator degrees of freedom.

References

- Ahn, Seung C., and Peter Schmidt. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68: 527. [CrossRef]
- Alvarez, Javier, and Manuel Arellano. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71: 1121–59. [CrossRef]
- Alvarez, Javier, and Manuel Arellano. 2022. Robust likelihood estimation of dynamic panel data models. *Journal of Econometrics* 226: 21–61. [CrossRef]
- Anderson, Theodore W., and Cheng Hsiao. 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76: 598–606. [CrossRef]
- Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58: 277–97. [CrossRef]
- Bao, Yong, and Aman Ullah. 2010. Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference* 140: 1193–205. [CrossRef]
- Bao, Yong, and Xuewen Yu. 2023. Indirect inference estimation of dynamic panel data models. *Journal of Econometrics* 235: 1027–53. [CrossRef]

- Bao, Yong, Xiaotian Liu, and Lihong Yang. 2020. Indirect inference estimation of spatial autoregressions. *Econometrics* 8: 34. [\[CrossRef\]](#)
- Blundell, Richard, and Richard J. Smith. 1991. Initial conditions and efficient estimation in dynamic panel data models. *Annals of Economics and Statistics* 20/21: 109–23.
- Blundell, Richard, and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–43. [\[CrossRef\]](#)
- Breitung, Jörg, Sebastian Kripfganz, and Kazuhiko Hayakawa. 2022. Bias-corrected method of moments estimators for dynamic panel data models. *Econometrics and Statistics* 24: 116–32. [\[CrossRef\]](#)
- Bun, Maurice J. G., and Martin A. Carree. 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics* 23: 200–10. [\[CrossRef\]](#)
- Bun, Maurice J. G., and Martin A. Carree. 2006. Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters* 92: 220–27. [\[CrossRef\]](#)
- Chudik, Alexander, M. Hashem Pesaran, and Jui-Chung Yang. 2018. Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics* 33: 816–36. [. : 10.1002/jae.2623 \[CrossRef\]](#)
- De Vos, Ignace, and Gerdie Everaert. 2021. Bias-corrected common correlated effects pooled estimation in dynamic panels. *Journal of Business & Economic Statistics* 39: 294–306. [\[CrossRef\]](#)
- Dhaene, Geert, and Koen Jochmans. 2015. Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies* 82: 991–1030. [\[CrossRef\]](#)
- Dhaene, Geert, and Koen Jochmans. 2016. Likelihood inference in an autoregression with fixed effects. *Econometric Theory* 32: 1178–215. [\[CrossRef\]](#)
- Dovonon, Prosper, Alastair R. Hall, and Frank Kleibergen. 2020. Inference in second-order identified models. *Journal of Econometrics* 218: 346–72. [. : 10.1016/j.jeconom.2020.04.020 \[CrossRef\]](#)
- Everaert, Gerdie, and Lorenzo Pozzi. 2007. Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control* 31: 1160–84. [\[CrossRef\]](#)
- Gospodinov, Nikolay, Ivana Komunjer, and Serena Ng. 2017. Simulated minimum distance estimation of dynamic models with errors-in-variables. *Journal of Econometrics* 200: 181–93. [\[CrossRef\]](#)
- Gouriéroux, Christian, Peter C. B. Phillips, and Jun Yu. 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* 157: 68–77. [\[CrossRef\]](#)
- Hahn, Jinyong, and Guido Kuersteiner. 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70: 1639–57. [\[CrossRef\]](#)
- Han, Chirok, and Peter C. B. Phillips. 2010. GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory* 26: 119–51. [\[CrossRef\]](#)
- Hansen, Christian B. 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141: 597–620. [\[CrossRef\]](#)
- Hayakawa, Kazuhiko. 2009. A simple efficient instrumental variable estimator for panel AR(p) models when both N and N are large. *Econometric Theory* 25: 873–90. [\[CrossRef\]](#)
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey Rosen. 1988. Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–95. [\[CrossRef\]](#)
- Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu. 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109: 107–50. [\[CrossRef\]](#)
- Juodis, Artūras. 2013. A note on bias-corrected estimation in dynamic panel data models. *Economics Letters* 118: 435–38. [\[CrossRef\]](#)
- Kapetanios, G. 2008. A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal* 11: 377–95. [\[CrossRef\]](#)
- Kelejian, Harry H., and Ingmar R. Prucha. 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104: 219–57. [\[CrossRef\]](#)
- Kelejian, Harry H., and Ingmar R. Prucha. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157: 53–67. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kiviet, Jan F. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68: 53–78. [\[CrossRef\]](#)
- Kruiniger, Hugo. 2018. *A Further Look at Modified ML Estimation of the Panel AR(1) Model with Fixed Effects and Arbitrary Initial Conditions*. MPRA Paper 88623. Munich: University Library of Munich.
- Lancaster, Tony. 2002. Orthogonal parameters and panel data. *The Review of Economic Studies* 69: 647–66. [\[CrossRef\]](#)
- Linz, Peter. 1985. *Analytical and Numerical Methods for Volterra Equations*. Philadelphia: SIAM. [\[CrossRef\]](#)
- Nickell, Stephen. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–26. [\[CrossRef\]](#)
- Okui, Ryo. 2021. Linear dynamic panel data models. In *Handbook of Research Methods and Applications in Empirical Microeconomics*. Edited by Nigar Hashimzade and Michael A. Thornton. Cheltenham: Edward Elgar Publishing, pp. 2–22.
- Pesaran, M. Hashem, and Cynthia Fan Yang. 2021. Estimation and inference in spatial models with dominant units. *Journal of Econometrics* 221: 591–615. [\[CrossRef\]](#)
- Phillips, Peter C. B., and Hyungsik R. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67: 1057–111. [\[CrossRef\]](#)

Ullah, Aman. 2004. *Finite Sample Econometrics*. New York: Oxford University Press.

White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817–38. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.