KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.879

"Asymptotic Inference for Dynamic Panel Estimators of Infinite Order Autoregressive Processes"

Yoon-Jin Lee, Ryo Okui and Mototsugu Shintani

October 2013

KYOTO UNIVERSITY

KYOTO, JAPAN

Asymptotic Inference for Dynamic Panel Estimators of Infinite Order Autoregressive Processes *

Yoon-Jin Lee[†], Ryo Okui[‡] and Mototsugu Shintani[§]

This version: October 2013

*The authors thank Yoonseok Lee, Yoshihiko Nishiyama, Tatsushi Oka, Peter Phillips, Laurent Pauwels, Igor Kheifets and seminar and conference participants at the National University of Singapore, the Kansai Econometric Society Meetings in Osaka, the 2011 Meetings of the Midwest Econometrics Group in Chicago, North Carolina State University, Kyoto University, SETA2012, the Asian Meeting of the Econometric Society in Delhi and Otaru University of Commerce for their helpful comments and discussion. Haruo Iwakura provided excellent research assistance. Okui appreciates financial support from the Japan Society of the Promotion of Science under KAKENHI22730176 and KAKENHI22330067. Shintani gratefully acknowledges financial support by the National Science Foundation Grant SES-1030164.

[†]Department of Economics, Indiana University, Bloomington, IN 47405-7104, USA. Email: lee243@indiana.edu

[‡]Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui@kier.kyoto-u.ac.jp

[§]Department of Economics, Vanderbilt University, Box 1819 Station B, Nashville, TN 37235, USA. Email: mototsugu.shintani@vanderbilt.edu.

Abstract

In this paper we consider the estimation of a dynamic panel autoregressive (AR) process of possibly infinite order in the presence of individual effects. We utilize the sieve AR approximation with its lag order increasing with the sample size. We establish the consistency and asymptotic normality of the standard dynamic panel data estimators, including the fixed effects estimator, the generalized methods of moments estimator and Hayakawa's instrumental variables estimator, using double asymptotics under which both the cross-sectional sample size and the length of time series tend to infinity. We also propose a bias-corrected fixed effects estimator based on the asymptotic result. Monte Carlo simulations demonstrate that the estimators perform well and the asymptotic approximation is useful. As an illustration, proposed methods are applied to dynamic panel estimation of the law of one price deviations among US cities.

Key Words: Autoregressive Sieve Estimation; Bias Correction; Double Asymptotics; Fixed Effects Estimator; GMM; Instrumental Variables Estimator.

JEL Classification: C13; C23; C26.

1 Introduction

Because economic relationships are often dynamic in nature, dynamic panel models have been considered very useful in the analysis of microeconomic data. When the time series dimension (T) is limited in the data, relative to the cross-sectional dimension (N), its dynamic properties can be identified only if some parsimonious specifications are employed in dynamic panel models. For this reason, simple autoregressive (AR) models of order one (or of a finite order) are typically employed in practice. Accordingly, various estimation methods for simple dynamic panel models have been proposed, and their theoretical properties have been investigated in many studies under a fixed T and large N asymptotic framework, including Anderson and Hsiao (1981), Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991), to name a few. More recently, however, an increasing number of panel data with longer T have become available in practice. Motivated by the availability of longer panel data, Hahn and Kuersteiner (2002), Alvarez and Arellano (2003) and Hayakawa (2009), among others, have investigated asymptotic properties of various estimators for finite order panel AR models, using an alternative asymptotic approximation when both T and N tend to infinity. We obtain further insight along this line of analysis by showing that the panel data with long T is also useful in allowing a more general time series structure.

In this paper, we consider the estimation of a general dynamic panel structure in the presence of unobserved individual effects. To this end, we employ a sieve approach to approximate a panel AR model of infinite order by a panel AR model of order p that increases with sample size T and N. Our specification of infinite order

AR models covers a very general class of stationary linear processes, which nests standard autoregressive and moving average (ARMA) models of finite order. Therefore, in comparison with those of previous studies in the dynamic panel literature, the estimation results from our approach are less subject to problems caused by possible model misspecifications. Such an idea of the AR sieve approximation in estimating a general linear model has long been used in the literature of time series analysis.[1] In principle, a similar approach should be applicable in the estimation of dynamic panel models as long as a moderately large T is available. However, to the best of our knowledge, this approach has yet to be used in the inference of dynamic panel models. The problem is that a naïve analogy of time series results cannot directly be used, due to several technical issues specific to dynamic panel data analysis under a large T and large N asymptotic framework. It is our intention to fill the gap between the two bodies of literature.[2]

The AR sieve approximation retains the computational simplicity of the finite order AR models, which can be conveniently estimated by a linear regression estimator. In our analysis, we take advantage of the fact that dynamic panel estimators commonly used in practice are all linear estimators originally designed to estimate finite order AR models. In particular, we consider sieve variants of (i) the fixed effects estimator (Hahn and Kuersteiner, 2002), (ii) the generalized methods of moments (GMM) estimator (Holtz-Eakin, Newey and Rosen, 1988, Arellano and Bond, 1991,

[1]For example, see Berk (1974), Lewis and Reinsel (1985), Lütkepohl and Poskitt (1991), Lütkepohl and Saikkonen (1997) and Gonçalves and Kilian (2007).

[2]There are several studies similar in spirit to ours. Phillips and Moon (1999) consider the issue of consistent long-run variance estimation in a panel data model but do not consider the inference of long autoregressions. Lee (2006) examines the asymptotic bias of the fixed effects estimator of a panel AR model of infinite order, but his results do not include the asymptotic distribution of the estimator. Okui (2010) considers the asymptotically unbiased estimation of autocovariances and autocorrelation, which does not require pre-specified dynamic panel models.

and Alvarez and Arellano, 2003) and (iii) an efficient instrumental variables (IV) estimator (Hayakawa, 2009). We show the consistency and asymptotic normality of each estimator. We further construct consistent standard errors for all the estimators and an asymptotically valid automatic lag selection procedure in an AR sieve approximation.

Our main theoretical results can be summarized as follows. First, when T is only moderately large, a fixed effects estimator suffers from asymptotic bias due to the incidental parameters problem (Neyman and Scott, 1948). We show that a simple bias correction method, which is analogous to that for finite-order AR models, works well for the infinite-order AR model.[3] Second, since the number of lags increases with T, the GMM estimator involves many moment conditions even when T is moderately large.[4] We find that N must be much larger relative to T in order for the GMM estimator to behave well without suffering from the many moments bias. Third, Hayakawa's IV estimator is shown to be consistent and asymptotically normal under a weaker condition on the relative magnitude of N and T than those required for the other estimators. Overall, our theoretical results suggest that the choice among estimators should be based on the relative magnitude of N and T and their finite sample properties.

Our Monte Carlo simulation to evaluate the finite sample properties provides useful guidance for practitioners in choosing among the estimators. Our proposed bias-corrected estimator works well in reducing the bias of the fixed effects estimator

[3] Kiviet (1995), Hahn and Kuersteiner (2002) and Lee (2012a) discuss bias correction in finite order panel AR models. Han, Phillips and Sul (2011) propose an alternative transformation of the data that leads to unbiased estimation.

[4] The problems of the GMM and other IV estimators in the presence of many moment conditions have been investigated in many studies. See, e.g., Kunitomo (1980), Morimune (1983), Bekker (1994), Newey and Smith (2004) and Newey and Windmeijer (2009).

without inflating the dispersion of the estimator. When N becomes larger, however, the GMM estimator has a smaller bias than the bias-corrected fixed effects estimator. The bias of the fixed effects estimator is not negligible even when T is fairly large, which illustrates the importance of bias correction. Among all the estimators, Hayakawa's IV estimator has the smallest bias at the cost of larger dispersion. An automatic lag selection procedure also helps to choose the approximation models that produce precise estimates.

Finally, as an empirical illustration, proposed methods are applied to dynamic panel estimation of the law of one price (LOP) deviations among US cities. The speed of individual good price adjustment is evaluated by the estimated sum of AR coefficients using competing estimators. We find that both the fixed effects estimator and the GMM estimator often provide values less than those provided by the bias-corrected fixed effects estimator and Hayakawa's IV estimator.

The remainder of this paper is organized as follows. Our model is described in Section 2. Our estimators are introduced and their asymptotic properties are investigated in Section 3. The finite sample performance of estimators is examined in Section 4, and our approach is applied to the real data in Section 5. Concluding remarks are made in Section 6. All mathematical proofs are collected in the mathematical appendix.

We use the following notation: For a sequence of vector a_{it}, we let $a_t = (a_{1t}, \ldots, a_{Nt})'$. The same convention applies to a sequence of vector denoted by $a_{it}(p)$ so that $a_t(p) = (a_{1t}(p) \ldots, a_{Nt}(p))'$. A constant C represents an arbitrary constant.

2 The Model

Suppose that we observe panel data $\{\{y_{it}\}_{t=1}^{T}\}_{i=1}^{N}$. We assume that y_{it} is generated from an AR process of possibly infinite order with individual specific effects. Namely, the model is

$$y_{it} = \mu_i + \sum_{k=1}^{\infty} \alpha_k y_{i,t-k} + \epsilon_{it}, \tag{1}$$

where μ_i is an unobservable individual effect and ϵ_{it} is an unobservable innovation with mean zero and variance σ^2. The individual effect, μ_i, which is assumed to be independent of ϵ_{it}, is included in order to capture the heterogeneity across individuals. Controlling for unobserved heterogeneity using individual effects is an important advantage of panel data analysis. The stationarity of y_{it} is imposed throughout the paper. In what follows, we consider the situation under which both the cross-sectional sample size, N, and the length of time series, T, are large.

The specification (1) is quite general and can include various linear stationary time series such as stationary and invertible panel ARMA models with individual effects. This representation is useful as some aspects of the interesting dynamic nature of the model can be examined by using the linear (or nonlinear) transformation of AR parameters $(\alpha_1, \ldots, \alpha_k, \ldots)'$. For example, the first AR coefficient, α_1, which represents the one-period-ahead effect of a shock, and the sum of AR coefficients (SAR),

$$SAR = \sum_{k=1}^{\infty} \alpha_k, \tag{2}$$

which captures the long-run cumulative effect of a shock, are often of interest in practice.

To estimate (1), we follow the time series literature of the AR sieve estimation and utilize its approximated model:

$$y_{it} = \mu_i + \sum_{k=1}^{p} \alpha_k y_{i,t-k} + u_{it,p}, \tag{3}$$

where $u_{it,p} = b_{it,p} + \epsilon_{it}$ with $b_{it,p} = \sum_{k=p+1}^{\infty} \alpha_k y_{i,t-k}$. The term $b_{it,p}$ here represents the error caused by approximating the true infinite order AR model given by (1) using the AR model with a truncated lag, p, given by (3). This approximated model is convenient in maintaining the computational simplicity of the parametric finite order AR model while making the effect of the model misspecification disappear asymptotically.

We make the following assumptions throughout the paper.

Assumption 1. *(i)* $\{\epsilon_{it}\}$ *is independently and identically distributed (i.i.d.) over time and across individuals with mean zero,* $0 < E(\epsilon_{it}^2) = \sigma^2 < \infty$, *and* $E|\epsilon_{it}|^{2r} \leq C, r > 2$*; (ii)* ϵ_{it} *is independent of* μ_i *for all* i *and* t*; (iii)* $\sum_{k=0}^{\infty} |\alpha_k| < \infty$ *and* $\sum_{k=0}^{\infty} \alpha_k z^k \neq 0$ *for any* $|z| \leq 1$*; (iv)* $y_{i,1-s}$, $s = 0, 1, 2, \ldots,$ *are generated from the stationary distribution; (v)* $p^{1/2} \sum_{k=p+1}^{\infty} |\alpha_k| \to 0$ *as* $p \to \infty$*.*

In Assumption 1(i), we focus on the i.i.d. error $\{\epsilon_{it}\}$ for the sake of simplicity. If somewhat stronger moment conditions are employed, the i.i.d. error assumption can be relaxed to allow for a martingale difference sequence such that $E(\epsilon_{it}|\mathcal{F}_{it-1}) = 0$ *a.s.*, where $\mathcal{F}_{i,t-1} = \sigma(\epsilon_{is}, s \leq t-1)$ is the σ-field generated by $\{\epsilon_{i,t-1}, \epsilon_{i,t-2}, \cdots\}$.[5] Assumption 1(ii) is used for the moving average representation of the model and is also used for the validity of moment conditions for the GMM estimator. Assumption 1(iii) indicates that y_{it} is stationary and can be represented by an infinite order moving average process. Considering cases in which y_{it} is an integrated process is beyond

[5] See, e.g., Gonçalves and Killian (2007).

the scope of the paper. Assumption 1(iv) can be relaxed because the influence of the initial observations is not decisive when T is sufficiently large. However, we keep this assumption to make the mathematical arguments simple. Assumption 1(v) is a commonly used assumption in the literature of the AR sieve estimation, which implies that the approximation error should not be too large.

It is also useful for our purpose to introduce an infinite order moving average representation of (1):

$$y_{it} = \eta_i + \sum_{k=0}^{\infty} \psi_k \epsilon_{i,t-k},$$

where $\psi_0 \equiv 1$, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\eta_i = \mu_i/(1 - \sum_{k=1}^{\infty} \alpha_k)$. This representation is justified by Assumption 1. Let Γ_p be the variance-covariance matrix of the vector $(w_{it}, \ldots, w_{i,t-p+1})'$ where $w_{it} = y_{it} - \eta_i = \sum_{k=0}^{\infty} \psi_k \epsilon_{i,t-k}$. Note that Γ_p does not depend on i.

We assume that Γ_p is positive definite and its eigenvalues do not diverge.

Assumption 2. *There exists $C_1 > 0$ such that the minimum eigenvalue of Γ_p is greater than C_1 for any p. There exists $C_2 < \infty$ such that the the maximum eigenvalue of Γ_p is smaller than C_2 for any p.*

Assumption 2 imposes a restriction on the persistence of the dynamics. This assumption is satisfied, for example, when the y_{it} follows a stationary panel ARMA(p, q) model with finite p and q.

3 Estimation

This section introduces several estimation methods. All the methods estimate parameters in the approximated model (3). We then show the asymptotic properties of the estimators.

3.1 The fixed effects estimator

Let us begin our analysis by considering the conventional fixed effects estimator. To define the estimator, we introduce the vector representation of the approximated model (3) as

$$y_{it} = \mu_i + x_{it}(p)'\alpha(p) + u_{it,p}$$

where $x_{it}(p) = (y_{i,t-1}, \ldots, y_{i,t-p})'$ and $\alpha(p) = (\alpha_1, \ldots, \alpha_p)'$. The first step of the fixed effects estimation is to eliminate the individual effects by subtracting individual averages. Let

$$\begin{aligned} \tilde{y}_{it} &= y_{it} - \frac{1}{T-p}\left(y_{i,p+1} + \cdots + y_{iT}\right), \\ \tilde{x}_{it}(p) &= x_{it}(p) - \frac{1}{T-p}\left(x_{i,p+1}(p) + \cdots + x_{iT}(p)\right) \end{aligned}$$

and $\tilde{u}_{it,p}$ be similarly defined. By rewriting the model (3) in terms of the transformed variables, we have

$$\tilde{y}_{it} = \tilde{x}_{it}(p)'\alpha(p) + \tilde{u}_{it,p}, \tag{4}$$

which does not contain the individual effects. Applying OLS to (4) yields the fixed effects estimator, denoted by $\hat{\alpha}_F(p)$:

$$\hat{\alpha}_F(p) = \left(\sum_{t=p+1}^{T} \tilde{x}_t(p)'\tilde{x}_t(p)\right)^{-1} \sum_{t=p+1}^{T} \tilde{x}_t(p)'\tilde{y}_t.$$

We define consistency as the property that the probability limit of the distance between the estimator and the true value of the parameter converges to zero where we use the Euclidean distance $||a|| = \sqrt{a'a}$ for a vector a.[6] The following theorem shows the consistency of $\hat{\alpha}_F(p)$.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied. Then, if $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/(NT) \to 0$ and $p^3/T^2 \to 0$, we have*

$$||\hat{\alpha}_F(p) - \alpha(p)|| \to_p 0.$$

Next we show the asymptotic normality of a linear combination of the estimated AR parameters. Let ℓ_p be an arbitrary deterministic sequence of $p \times 1$ vectors such that $0 < C_1 \leq ||\ell_p||^2 = \ell_p'\ell_p \leq C_2 < \infty$ for $p = 1, 2, \cdots$ for some C_1 and C_2. Our parameter of interest is $\lim_{p\to\infty} \ell_p'\alpha(p)$. For example, if we are interested in the kth AR coefficient, α_k, for $1 \leq k \leq p$, our choice of ℓ_p would be $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)'$ where e_k is a $p \times 1$ selection vector with the kth element being one and other elements being zero. Instead, if we are interested in the SAR, our choice of ℓ_p would be $\ell_p^* = \iota_p/\sqrt{p} = (1/\sqrt{p}, \ldots, 1/\sqrt{p})'$ where ι_p is a $p \times 1$ vector of ones, and we define its estimator by

$$\widehat{SAR}_F = \sqrt{p}\ell_p^{*\prime}\hat{\alpha}_F(p).$$

The following theorem presents the asymptotic distribution of the estimator $\ell_p'\hat{\alpha}_F(p)$. Let $v_p^2 = \sigma^2\ell_p'\Gamma_p^{-1}\ell_p$, which turns out to be the asymptotic variance of all the estimators considered in this paper.

Theorem 2. *Suppose that Assumptions 1 and 2 are satisfied. Then, if $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $\sqrt{NTp}\sum_{k=p+1}^{\infty} |\alpha_k| \to 0$, $p^3/(NT) \to 0$, $p^2/T \to 0$ and*

[6] An alternative way of defining the consistency of $\hat{\alpha}(p)$ is $\sup_k |\hat{\alpha}_k - \alpha_k| \to_p 0$, where we set $\hat{\alpha}_k = 0$ for $k > p$. Note that our definition of consistency is actually stronger than this alternative definition. This is because $\sup_k |\hat{\alpha}_k - \alpha_k| \leq ||\hat{\alpha}(p) - \alpha(p)|| + \sup_{k>p} |\alpha_k| \to_p 0$ if $||\hat{\alpha}(p) - \alpha(p)|| \to_p 0$ and $p \to \infty$.

$p^4N/T^3 \to 0$, *we have*

$$\sqrt{NT}\left[\ell_p'\hat{\alpha}_F(p) - \ell_p'\alpha(p) + \ell_p'\Gamma_p^{-1}B\right]/v_p \xrightarrow{d} N(0,1),$$

where

$$B = E\left(\frac{T-p}{T}\left(\frac{1}{T-p}\sum_{k=p}^{T-1} w_{ik}(p)\right)\frac{1}{T-p}\sum_{k=p+1}^{T}\epsilon_{ik}\right) = \frac{1}{T(T-p)}\sum_{t=p+1}^{T}\sum_{m=p+1}^{t-1}\sigma^2\psi_{t-1-m}^{-}(p),$$

$w_{ik}(p) = (w_{i,k}, \dots, w_{i,k-p+1})'$ *and* $\psi_{t-1-m}^{-}(p) = (\psi_{t-1-m}, \psi_{t-m}, \dots, \psi_{t-2-m+p})$.

The theorem shows that $\ell_p'\hat{\alpha}_F(p)$ is asymptotically normal, but also asymptotically biased. We see as well that the convergence rate of $\ell_p'\hat{\alpha}_F(p)$ is $\sqrt{NT}$ when we ignore the bias term. Note that since the convergence rate of $\ell_p^{*\prime}\hat{\alpha}_F(p)$ is also $\sqrt{NT}$, the convergence rate of $\widehat{SAR}_F$ is $\sqrt{NT/p}$, which is slower than $\sqrt{NT}$.

To better understand the structure of the bias in our analysis, we can utilize a convenient decomposition formula. Since $u_{it,p} = b_{it,p} + \epsilon_{it}$, the transformed error $\tilde{u}_{it,p}$ is the sum of $\tilde{b}_{it,p} = b_{it,p} - \sum_{t'=p+1}^{T} b_{it',p}/(T-p)$ and $\tilde{\epsilon}_{it} = \epsilon_{it} - \sum_{t'=p+1}^{T}\epsilon_{it'}/(T-p)$. For this reason, the total bias can be decomposed as

$$\begin{aligned}
& E\left(\hat{\alpha}_F(p) - \alpha(p)\right) \\
= \; & E\left((\hat{\Gamma}_p^F)^{-1}\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{u}_{t,p}\right) \\
= \; & \underbrace{E\left((\hat{\Gamma}_p^F)^{-1}\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{b}_{t,p}\right)}_{\text{truncation bias}} + \underbrace{E\left((\hat{\Gamma}_p^F)^{-1}\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{\epsilon}_t\right)}_{\text{fundamental bias}}
\end{aligned}$$

where

$$\hat{\Gamma}_p^F = \frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{x}_t(p).$$

The first term is the bias that arises because we estimate the AR model with a truncated lag length, not the true infinite order AR model. Throughout the paper, we refer this bias to 'truncation bias.' Similarly, we refer the second term to 'fundamental bias' since this part of bias is present even if we estimate the true finite order AR model with the correct lag length. While the truncation bias may not be negligible in finite samples, it vanishes in our asymptotic analysis due to our assumption $\sqrt{NTp}\sum_{k=p+1}^{\infty}|\alpha_k| \to 0$. This assumption implies that $\sup_{k>p+1}|\alpha_k| = o(\sqrt{NTp})$ is needed. If p increases very slowly, the approximation error does not vanish fast enough and a bias of the estimator appears. For example, if w_{it} follows a finite order stationary and invertible ARMA process, α_k decays exponentially and p must grow at a rate faster than $\log(NT)$ (i.e., $\log(NT)/p \to 0$ is needed).

It is the second term, namely the fundamental bias, that appears in the theorem. The order of this bias term is $\sqrt{p}/T$. This result corresponds to the well-known outcome that the fixed effects estimator is asymptotically biased in dynamic panel data models (see, e.g., Nickell, 1981, Kiviet, 1995 and Hahn and Kuersteiner, 2002). The formula in the theorem includes only the fundamental bias term partly because it is estimable and can be corrected.

We now consider a bias correction. Let $\hat{B}$ be an estimator of B. Let ν_{NTp} be the inverse of the convergence rate of $\hat{B}$ such that $\left\|\hat{B} - B\right\| = O_p(\nu_{NTp})$. A bias corrected estimator is given by

$$\hat{\alpha}_{BF}(p) = \hat{\alpha}_F(p) + (\hat{\Gamma}_p^F)^{-1}\hat{B}.$$

We note that $B \approx \sigma^2 \sum_{k=0}^{\infty}\psi_k \iota_p/T$, where ι_p is the $p \times 1$ vector of ones. Since $\sum_{k=0}^{\infty}\psi_k = 1/(1-\sum_{k=1}^{\infty}\alpha_k)$, a possible choice of $\hat{B}$ is $(\hat{\sigma}^2/(1-\sum_{k=1}^{p}\tilde{\alpha}_k))\iota_p/T$, where $\tilde{\alpha}_k$'s are some preliminary estimates for α_k's. If we are interested in SAR, i.e., when $\ell_p = \ell_p^*$, a simple bias correction method can be employed. In this case, we have $\ell_p^{*\prime}\Gamma_p^{-1}\iota_p \approx p^{1/2}(1-\sum_{k=1}^{p}\alpha_k)^2/\sigma^2$ so that the bias is approximated by $p^{1/2}(1-\iota_p'\alpha(p))/T$. As a result, we can simply replace $\ell_p^{*\prime}\hat{\alpha}_F(p)$ by

$$\frac{T-p}{T}\ell_p^{*\prime}\hat{\alpha}_F(p)+\frac{p^{1/2}}{T}=p^{-1/2}\iota_p'\hat{\alpha}_F(p)+\frac{p^{1/2}}{T}(1-\iota_p'\hat{\alpha}_F(p))$$

and use

$$\widehat{SAR}_{BF}=\frac{T-p}{T}\widehat{SAR}_F+\frac{p}{T}$$

as a bias corrected estimator of SAR.[7]

The following theorem gives the consistency of the bias corrected fixed effects estimator.

Theorem 3. *Suppose that Assumptions 1 and 2 are satisfied. Suppose that $\nu_{NTp}\to 0$. Then, if $N\to\infty$, $T\to\infty$ and $p\to\infty$ with $p^2/(NT)\to 0$ and $p^3/T^2\to 0$, we have*

$$||\hat{\alpha}_{BF}(p)-\alpha(p)||\to_p 0.$$

The asymptotic normality result is provided below.

Theorem 4. *Suppose that Assumptions 1 and 2 are satisfied. Then, if $N\to\infty$, $T\to\infty$ and $p\to\infty$ with $\sqrt{NTp}\sum_{k=p+1}^{\infty}|\alpha_k|\to 0$, $p^3/(NT)\to 0$, $p^2/T\to 0$ and $\sqrt{NT}\nu_{NTp}\to 0$, we have*

$$\sqrt{NT}\left[\ell_p'\hat{\alpha}_{BF}(p)-\ell_p'\alpha(p)\right]/v_p\xrightarrow{d}N(0,1).$$

This theorem shows that our bias corrected estimator can effectively eliminate the asymptotic bias.

Note that Theorems 3 and 4 require the condition on the convergence rate of ν_{NTp}. When the bias is estimated by $\hat{B}=(\hat{\sigma}^2/(1-\sum_{k=1}^{p}\tilde{\alpha}_k))\iota_p/T$, ν_{NTp} would be

[7] A remarkable observation is that this bias correction for the estimator of SAR reduces not only the bias but also the variance. Since $p^{1/2}/T$ is nonrandom and $(T-p)/T<1$, the variance of the bias corrected estimator must be smaller than that of the fixed effects estimator.

$(p/(\sqrt{N}T^{3/2}) + p/T^2)$ because the convergence rate of $\sum_{k=1}^{p} \tilde{\alpha}_k$ is $\sqrt{p/(NT)} + \sqrt{p}/T$. In this case, the condition $\nu_{NTp} \to 0$ is always satisfied because $p^2/T \to 0$ is imposed. For the asymptotic normality result, $\sqrt{NT}\nu_{NTp} \to 0$ requires an additional condition that $p^2 N/T^3 \to 0$.

3.2 The GMM estimator

In this section, we consider the GMM estimator based on Holtz-Eakin, Newey and Rosen (1988) and Arellano and Bond (1991).

For the GMM estimator, we apply the forward filter to the variables to eliminate the individual effects.[8] Let

$$y_{it}^* = \sqrt{\frac{T-t}{T-t+1}} \left(y_{it} - \frac{1}{T-t}(y_{i,t+1} + \cdots + y_{iT}) \right),$$

and $x_{it}^*(p)$ and $u_{it,p}^*$ be similarly defined.[9] In this paper, a variable with $*$ superscript is a variable transformed by the forward filter even when it is not explicitly mentioned. The transformed variables satisfy the following relationship:

$$y_{it}^* = x_{it}^*(p)'\alpha(p) + u_{it,p}^* \tag{5}$$

so that the individual effect μ_i is eliminated.

The GMM estimator exploits the following moment conditions.

$$E[y_{i,t-s}\epsilon_{it}^*] = 0 \text{ for } s \geq 1.$$

[8] We use the forward filtering because it provides a simple representation of the GMM estimator. Note that the GMM estimator considered here is numerically equivalent to the efficient GMM estimator based on the equations in first differences (Arellano and Bover, 1995).

[9] Note that $y_{i,t-k}^*$ and the forward filtered lagged dependent variable in $x_{it}^*(p)$ are different. For example, the kth element of $x_{it}^*(p)$ is

$$\sqrt{\frac{T-t}{T-t+1}} \left(y_{i,t-k} - \frac{1}{T-t}(y_{i,t-k+1} + \cdots + y_{i,T-k}) \right).$$

To obtain the kth element of $x_{it}^*(p)$, we subtract the average of $(y_{i,t-k+1}, \ldots, y_{i,T-k})$ instead of the average of $(y_{i,t-k+1}, \ldots, y_{i,T})$ as for $y_{i,t-k}^*$.

We note that there are $T - p - 1$ equations (one equation for each time period) to be estimated and the equation for y_{it}^* has $t - 1$ instruments. Therefore, there are $\sum_{t=p+1}^{T-1}(t-1) = (T-2)(T-1)/2 - (p-1)p/2$ moment conditions in total, and the number of moment conditions can be very large even when T is moderately large.

We now define the GMM estimator. Let $Z_{it} = (y_{it-1}, \ldots, y_{i1})'$ be the set of instrumental variables for the equation with y_{it}^* as the dependent variable. Let $M_t = Z_t(Z_t'Z_t)^{-1}Z_t'$. The GMM estimator, $\hat{\alpha}_G(p)$, is

$$\hat{\alpha}_G(p) = \left(\sum_{t=p+1}^{T-1} x_t^*(p)' M_t x_t^*(p)\right)^{-1} \sum_{t=p+1}^{T-1} x_t^*(p)' M_t y_t^*.$$

We need the following two additional assumptions to investigate the asymptotic properties of the GMM estimator.

Assumption 3. *$E(\mu_i^4)$ is finite.*

Assumption 3 is not needed for the fixed effects estimator because the fixed effects estimator is based solely on the transformed variables and is thus free from individual effects. On the other hand, the GMM estimator depends on individual effects because the instruments are affected by individual effects.

Assumption 4. $\sum_{k=1}^{\infty} k|\psi_k| < \infty$.

Assumption 4 imposes a stronger restriction on the dependence of the variable than that needed for the fixed effects estimator. However, this assumption is satisfied, for example, when the true process is a finite order ARMA model. This assumption is employed in order to evaluate the bias caused by the presence of many moment conditions.

The following theorems provide the consistency of $\hat{\alpha}_G(p)$ and the asymptotic normality of $\ell_p'\hat{\alpha}_G(p)$.

Theorem 5. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. Then, if* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$ *with* $T/N \to 0$ *and* $p^2/T \to 0$, *we have*

$$||\hat{\alpha}_G(p) - \alpha(p)|| \to_p 0.$$

Theorem 6. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. Then, if* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$ *with* $\sqrt{NTp}\sum_{k=p+1}^{\infty}|\alpha_k| \to 0$, $p^2T/N \to 0$ *and* $p^3 \log T/T \to 0$, *we have*

$$\sqrt{NT}\left[\ell_p'\hat{\alpha}_G(p) - \ell_p'\alpha(p)\right]/v_p \xrightarrow{d} N(0,1).$$

Note that the convergence rate and asymptotic variance of the GMM estimator is identical to those of the fixed effects estimator and the bias corrected fixed effects estimator. Similarly, the GMM estimator of SAR defined by $\widehat{SAR}_G = \sqrt{p}\ell_p^{*\prime}\hat{\alpha}_G(p)$ also has a nonparametric convergence rate of $\sqrt{NT/p}$.

However, it should be noted that our asymptotic normality results of the GMM estimator are derived under the assumption that N grows at a rate faster than T. In general, the GMM estimator suffers from the bias caused by the presence of many moment conditions, but our assumption on N and T allows us to ignore the "many moments bias." We may be able to relax the condition on the relative magnitude of N and T and derive the asymptotic distribution that explicitly evaluates the many moment bias term as done in Alvarez and Arellano (2003). However, evaluating the many moments bias in the current setting is very difficult and should be considered as a separate project. Instead, in the next subsection, we consider an alternative IV estimator, developed by Hayakawa (2009), that is free from the many moments bias.

3.3 Hayakawa's efficient IV estimator

The concept of Hayakawa's (2009) IV estimator is similar to that of the GMM estimator except for the choice of instruments. The instruments are constructed by subtracting the average of past realizations from the regressors. Let $z_{it}(p) = (y_{i,t-1}, \ldots, y_{i,t-p})$ and

$$h_{it}(p) = \sqrt{\frac{T-t}{T-t+1}} \left(z_{it}(p) - \frac{1}{t-p-1}(z_{i,t-1}(p) - \cdots - z_{i,p+1}(p)) \right).$$

The choice of $h_{it}(p)$ as instruments in equation (5) can be motivated by the following observation. In finite order AR models, the optimal instruments for $x^*_{it}(p)$ is $w_{it-1}(p) = (x_{it}(p) - \eta_i)$. The instrument $h_{it}(p)$ may be regarded as an approximation to this optimal instrument, using the average of past realizations of y_{it} in place of η_i. We use only the past realizations so that the moment conditions become valid even when T is small.

An IV estimator can be constructed as

$$\hat{\alpha}_H(p) = \left(\sum_{t=p+2}^{T-1} h_t(p)' x^*_t(p) \right)^{-1} \sum_{t=p+2}^{T-1} h_t(p)' y^*_t.$$

Hayakawa (2009) shows that for finite order AR models, this estimator is consistent regardless of the relative magnitude of N and T and is efficient when $T \to \infty$ under Gaussianity.

The following theorems demonstrate the asymptotic properties of $\hat{\alpha}_H(p)$.

Theorem 7. *Suppose that Assumptions 1 and 2 are satisfied. Then, if $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/T \to 0$, we have*

$$||\hat{\alpha}_H(p) - \alpha(p)|| \to_p 0.$$

Theorem 8. *Suppose that Assumptions 1 and 2 are satisfied. Then, if* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$ *with* $\sqrt{NTp}\sum_{k=p+1}^{\infty}|\alpha_k| \to 0$ *and* $p^3/T \to 0$, *we have*

$$\sqrt{NT}\left[\ell_p'\hat{\alpha}_H(p) - \ell_p'\alpha(p)\right]/v_p \xrightarrow{d} N(0,1).$$

Note that the convergence rate and asymptotic variance of Hayakawa's IV estimator are identical to those of other estimators. Therefore, the estimator of SAR defined by $\widehat{SAR}_H = \sqrt{p}\ell_p^{*\prime}\hat{\alpha}_G(H)$ also has the same convergence rate and limiting distribution.

The most notable fact is that, in contrast to other estimators, the asymptotic normality of Hayakawa's IV estimator does not require any condition on the relative rate for N and T. Nonetheless, this IV estimator possesses the same asymptotic variance as that of the other estimators. This result indicates that Hayakawa's IV estimator may behave better than the GMM estimator when T is relatively large. On the other hand, compared with that of the bias corrected fixed effects estimator, the condition on p is stronger for Hayakawa's IV estimator (i.e., $p^3/T \to 0$ for Hayakawa's IV estimator while $p^3/(NT) \to 0$ and $p^2/T \to 0$ for the bias corrected fixed effects estimator). However, we view this restriction as a relatively minor issue because p can be chosen appropriately by researchers.

3.4 Comparisons of the four estimators

So far, we have considered four estimators. All the estimators are consistent and asymptotically normal with a common asymptotic variance. However, the fixed effects estimator is biased and the GMM estimator requires that N grow faster than T to avoid the bias. On the other hand, the bias corrected fixed effects estimator may require $p^2N/T^3 \to 0$. Hayakawa's estimator does not restrict the relative magnitude of N and T. These theoretical results lead us to recommend the bias corrected fixed effects estimator and Hayakawa's estimator. When N is large, the GMM estimator may also be considered. We also investigate the finite sample properties of these estimators through simulations of which the results are presented in the next section.

3.5 Standard errors

Computing the standard errors of the estimators requires the consistent estimation of $v_p^2 = \sigma^2 \ell_p' \Gamma_p^{-1} \ell_p$. Natural estimators of v_p^2 may be

$$
\begin{aligned}
\hat{v}_{p,F}^2 &= \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\tilde{y}_{it} - \tilde{x}_{it}(p)' \hat{\alpha}_F(p))^2 \right) \ell_p' (\hat{\Gamma}_p^F)^{-1} \ell_p, \\
\hat{v}_{p,BF}^2 &= \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+1}^{T} (\tilde{y}_{it} - \tilde{x}_{it}(p)' \hat{\alpha}_{BF}(p))^2 \right) \ell_p' (\hat{\Gamma}_p^F)^{-1} \ell_p, \\
\hat{v}_{p,G}^2 &= \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+1}^{T-1} (y_{it}^* - x_{it}^*(p)' \hat{\alpha}_G(p))^2 \right) \ell_p' (\hat{\Gamma}_p^G)^{-1} \ell_p \\
\text{and } \hat{v}_{p,H}^2 &= \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+2}^{T-1} (y_{it}^* - x_{it}^*(p)' \hat{\alpha}_H(p))^2 \right) \ell_p' (\hat{\Gamma}_p^H)^{-1} \ell_p
\end{aligned}
$$

where

$$
\hat{\Gamma}_p^G = \frac{1}{NT} \sum_{t=p+1}^{T-1} x_t^*(p)' M_t x_t^*(p) \text{ and } \hat{\Gamma}_p^H = \frac{1}{NT} \sum_{t=p+2}^{T-1} h_t(p)' x_t^*(p).
$$

It would be sensible to use $\hat{v}_{p,F}^2$ for $\hat{\alpha}_F(p)$, $\hat{v}_{p,BF}^2$ for $\hat{\alpha}_{BF}(p)$, $\hat{v}_{p,G}^2$ for $\hat{\alpha}_G(p)$ and $\hat{v}_{p,H}^2$ for $\hat{\alpha}_H(p)$.

The following theorem shows the consistency of these four estimators.

Theorem 9. *Suppose that Assumptions 1 and 2 are satisfied.*

1. *If $N, T, p \to \infty$ with $p^2/(NT) \to 0$ and $p^3/T^2 \to 0$, then $\hat{v}_{p,F}^2 - v_p^2 \to_p 0$.*

2. *If* $N, T, p \to \infty$ *with* $p^2/(NT) \to 0$*,* $p^3/T^2 \to 0$ *and* $\nu_{NTp} \to 0$*, then* $\hat{v}^2_{p,BF} - v^2_p \to_p 0$*.*

3. *Suppose, in addition, that Assumptions 3 and 4 are satisfied. If* $N, T, p \to \infty$ *with* $T/N \to 0$ *and* $p^2/T \to 0$*, then* $\hat{v}^2_{p,G} - v^2_p \to_p 0$*.*

4. *If* $N, T, p \to \infty$ *with* $p^2/T \to 0$*, then* $\hat{v}^2_{p,H} - v^2_p \to_p 0$*.*

Each variance estimator is consistent when the corresponding estimator for $\alpha(p)$ is consistent.

3.6 Lag selection

The estimation procedures require the lag order of the approximated model, p, to be chosen by researchers. In choosing p, we consider the following general-to-specific rule. This automatic rule follows the procedure similar to the one considered in Ng and Perron (1995) which tests for the significance of coefficients on lags.

Each step of the general-to-specific rules uses the t-statistic for the coefficient on the highest lag in the model. Let e_p be the $p \times 1$ vector whose pth element is 1 and other elements are zero. Let

$$t_p(\hat{\alpha}(p)) = \sqrt{NT} e_p' \hat{\alpha}(p) / \hat{v}_p,$$

where $\hat{\alpha}(p)$ and $\hat{v}_p$ are estimators of $\alpha(p)$ and v_p with $\ell_p = e_p$, respectively. The statistics $t_p(\hat{\alpha}(p))$ is the t-test statistics for the null hypothesis $\alpha_p = 0$ based on estimator $\hat{\alpha}(p)$.

The general-to-specific procedure is the following. We a priori set the maximum possible value of p, denoted $p_{\max}$. Let $\hat{p}$ be the maximum value of p such that

$|t_p(\hat{\alpha}(p))| > z_{0.5\alpha}$, where $z_{0.5\alpha}$ is the upper 0.5α quantile of the standard normal distribution, for $p = 1, 2, \ldots, p_{\max}$. This $\hat{p}$ is the lag length chosen by this general-to-specific procedure. An alternative explanation of the rule is the following. We keep the pth-lag if its coefficient is statistically significant in AR(p) specification. Otherwise, we drop the pth lag, estimate the AR($p-1$) model and test the significance of the coefficient of the $(p-1)$th lag. We repeat this process until the coefficient becomes statistically significant or p reaches zero.

The following theorem gives the rate of $\hat{p}$.

Theorem 10. *Suppose that $\hat{\alpha}(p)$ is $\hat{\alpha}_F(p)$, $\hat{\alpha}_{BF}(p)$, $\hat{\alpha}_G(p)$ or $\hat{\alpha}_H(p)$. Suppose also that Assumptions 1 and 2 are satisfied and that, if $\hat{\alpha}(p) = \hat{\alpha}_G(p)$, Assumptions 3 and 4 are also satisfied. If N, T and $p = p_{\max}$ satisfy conditions for that $\sqrt{NT}\ell_p'(\hat{\alpha}(p) - \alpha(p))/v_p \to N(0,1)$ and $\hat{v}_p \to v_p$, then $\hat{p}$ increases at the same rate as $p_{\max}$.*

This theorem implies that we can choose p by the general-to-specific procedure such that it satisfies the requirement for the asymptotic normality of an estimator by appropriately setting the rate of $p_{\max}$. In the simulations presented below, we set $p_{\max} = O(T^{1/4})$ under which all the conditions for the theoretical analysis hold.

4 Monte Carlo Experiments

In this section, we conduct Monte Carlo simulations to evaluate the accuracy of our asymptotic approximations of distribution of various dynamic panel estimators in finite samples. We would also like to see the effects of different choices of lag orders in approximated models.

We generate samples from the ARMA(1,1) model of the following form:

$$y_{it} = \eta_i + \phi y_{i,t-1} + \epsilon_{it} + \theta \epsilon_{i,t-1}$$

where $\phi = \{0, 0.5, 0.97\}$, $\theta = 0.4$ and $\eta_i \sim N(0,1)$ independent across i, $\epsilon_{it} \sim N(0,1)$ independent across i and t. The individual effect η_i and idiosyncratic error ϵ_{it} are also independently drawn. With a fixed MA parameter set at $\theta = 0.4$, we control the persistence of the process by changing the AR parameter. We estimate the first AR coefficient α_1 and sum of the AR coefficients (SAR) $\sum_{k=1}^{\infty} \alpha_k$ using various estimators. When $\phi = 0$ (DGP1), α_1 and SAR are 0.4 and 0.286, respectively. When $\phi = 0.5$ (DGP2), α_1 is 0.9 and the process becomes more persistent with the SAR being 0.643. The most persistent process with $\phi = 0.97$ (DGP3) has a hump-shaped impulse response with α_1 being 0.9 and a near unit SAR of 0.979.[10] For each process, y_{i0}'s are generated from the (conditional) stationary distribution:

$$y_{i0} | \eta_i \sim N\left(\frac{\eta_i}{1-\phi}, \frac{1+\theta^2+2\phi\theta}{1-\phi^2}\right).$$

The effective sample sizes we consider are $N = \{50, 100, 500, 1000\}$ and $T = \{25, 50, 75\}$.

We consider two rules of thumb and an automatic selection rule to choose the AR lag p. For the rules of thumb, we follow a convention from the time series literature and use $p = [c(T/100)^{1/4}]$ with $c = 4$ and 12, where $[x]$ is integer part of x. The fixed rules with $c = \{4, 12\}$ provide $p = \{2, 8\}$ for $T = 25$, $p = \{3, 10\}$ for $T = 50$ and $p = \{3, 11\}$ for $T = 75$, respectively. The automatic rule follows the general-to-specific procedure described in Section 3.6. The maximum lag is selected as $p_{\max} = [12(T/100)^{1/4}]$, which corresponds to the fixed rule with a choice of $c = 12$.

[10] This third choice of parameters is very close to the value of the ARMA(1,1) process estimated in the literature of purchasing power parity.

Note that both the rule of thumb and automatic lag selection procedures satisfy the required conditions in the theoretical analysis.[11]

We evaluate the following four estimators; the fixed effects estimator, the bias-corrected fixed effects estimator, the GMM estimator and the Hayakawa's IV estimator. We compare the finite sample performance of the four estimators in terms of: (i) median absolute deviation (med abs dev); (ii) median bias; (iii) interquartile range (iqr); and (iv) coverage probability of an asymptotic 95% confidence interval.[12] [13]

The results based on 1000 Monte Carlo runs are provided in Tables 1 to 4. Tables 1, 2 and 3 show the results from DGP1, DGP2 and DGP3, respectively, when $N = \{50, 100, 500, 1000\}$ and $T = 25$. Table 4 shows the results when $N = 100$ and $T = \{50, 75\}$.

[11]We have also tried information criteria in choosing the AR lag. See Lee (2012b) on how information criteria should be modified for dynamic panel data analysis. However, the simulation results are similar to those reported here and we do not report them. Another problem is that in time series literature, the rate of the lag length chosen by an information criterion is founded to be of order $\log(T)$. See, e.g., Hannan and Deistler (1988, Section 6.6) and Ng and Perron (1995). This fact leads us to conjecture that the rate is of order $\log(NT)$ in the case of dynamic panel models. However, such a rate does not satisfy the condition for the truncation bias to be negligible in the asymptotic distribution.

[12]Median absolute deviation is defined as the median of the absolute value of the difference between the estimator and the true value.

[13]We use these robust measures because of the concern about the existence of the moments of the instrumental variables estimators. It is well-known in the literature on instrumental variables estimation that just-identified instrumental variables estimators do not possess any moment (see, e.g., Hayashi 2000, chapter 8, page 542). In fact, we obtain unreasonably high values of root mean squared error or bias for the Hayakawa estimator when $\phi = 0.97$ and N is small. The use of robust measures is thus common in the literature on instrumental variables estimation and panel data. For example, Alvarez and Arellano (2003) and Okui (2009) report robust measures.

Table 1: Finite sample performance of estimators when $T = 25$ (DGP1: $\phi = 0$)

			FE			bias corrected FE			GMM			Hayakawa IV		
	N		fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto
(i) α_1	50	med abs dev	0.053	0.085	0.079	0.022	0.031	0.023	0.042	0.079	0.064	0.029	0.048	0.036
(true value:		median bias	-0.053	-0.085	-0.079	-0.015	-0.027	-0.012	-0.041	-0.079	-0.064	-0.022	0.000	-0.004
0.400)		iqr	0.040	0.049	0.049	0.040	0.049	0.041	0.046	0.067	0.065	0.052	0.096	0.071
		cp	0.591	0.331	0.363	0.924	0.861	0.918	0.744	0.540	0.607	0.910	0.957	0.914
	100	med abs dev	0.052	0.085	0.082	0.018	0.028	0.018	0.033	0.053	0.041	0.024	0.032	0.024
		median bias	-0.052	-0.085	-0.082	-0.015	-0.026	-0.012	-0.032	-0.053	-0.041	-0.022	0.000	-0.002
		iqr	0.029	0.035	0.035	0.029	0.035	0.033	0.034	0.049	0.048	0.036	0.063	0.048
		cp	0.308	0.082	0.096	0.873	0.800	0.884	0.713	0.623	0.681	0.852	0.955	0.913
	500	med abs dev	0.052	0.084	0.084	0.014	0.025	0.013	0.022	0.016	0.012	0.022	0.015	0.011
		median bias	-0.052	-0.084	-0.084	-0.014	-0.025	-0.013	-0.022	-0.014	-0.010	-0.022	0.001	0.001
		iqr	0.013	0.016	0.016	0.013	0.016	0.016	0.014	0.022	0.021	0.016	0.030	0.022
		cp	0.000	0.000	0.000	0.663	0.385	0.751	0.480	0.860	0.862	0.544	0.951	0.916
	1000	med abs dev	0.052	0.084	0.084	0.015	0.026	0.013	0.020	0.009	0.008	0.022	0.011	0.008
		median bias	-0.052	-0.084	-0.084	-0.015	-0.026	-0.013	-0.020	-0.007	-0.005	-0.022	0.000	-0.001
		iqr	0.009	0.011	0.011	0.009	0.011	0.013	0.010	0.017	0.014	0.012	0.021	0.016
		cp	0.000	0.000	0.000	0.402	0.119	0.568	0.237	0.893	0.888	0.248	0.959	0.932
(ii) SAR	50	med abs dev	0.114	0.450	0.407	0.048	0.080	0.044	0.088	0.350	0.257	0.057	0.163	0.093
(true value:		median bias	-0.114	-0.450	-0.407	-0.048	-0.078	-0.033	-0.088	-0.350	-0.257	-0.055	-0.001	-0.021
0.286)		iqr	0.051	0.138	0.189	0.047	0.094	0.075	0.060	0.215	0.267	0.068	0.328	0.186
		cp	0.102	0.001	0.009	0.743	0.944	0.892	0.445	0.295	0.257	0.790	0.947	0.823
	100	med abs dev	0.114	0.445	0.434	0.047	0.074	0.038	0.071	0.240	0.158	0.057	0.115	0.068
		median bias	-0.114	-0.445	-0.434	-0.047	-0.074	-0.034	-0.071	-0.240	-0.158	-0.057	-0.003	-0.005
		iqr	0.037	0.093	0.116	0.034	0.063	0.059	0.044	0.160	0.201	0.050	0.231	0.135
		cp	0.005	0.000	0.000	0.523	0.871	0.848	0.363	0.434	0.405	0.609	0.947	0.830
	500	med abs dev	0.113	0.447	0.447	0.047	0.075	0.037	0.054	0.067	0.040	0.055	0.053	0.034
		median bias	-0.113	-0.447	-0.447	-0.047	-0.075	-0.037	-0.054	-0.066	-0.037	-0.055	0.000	0.002
		iqr	0.015	0.043	0.043	0.014	0.029	0.030	0.019	0.076	0.064	0.020	0.107	0.068
		cp	0.000	0.000	0.000	0.008	0.190	0.533	0.036	0.819	0.731	0.057	0.954	0.858
	1000	med abs dev	0.113	0.448	0.448	0.047	0.076	0.037	0.051	0.040	0.026	0.055	0.039	0.025
		median bias	-0.113	-0.448	-0.448	-0.047	-0.076	-0.037	-0.051	-0.035	-0.021	-0.055	-0.002	-0.003
		iqr	0.011	0.032	0.032	0.010	0.022	0.024	0.015	0.064	0.044	0.016	0.079	0.051
		cp	0.000	0.000	0.000	0.000	0.013	0.205	0.001	0.873	0.775	0.002	0.954	0.861

Notes: Lag lengths are $p = 2$ and $p = 8$ for fixed rules with $c = 4$ (fix4) and $c = 12$ (fix12), respectively, or from an automatic sequential procedure (auto). Median absolute deviation (med abs dev). Interquartile range (iqr). Coverage probability (cp) of 95 percent asymptotic confidence interval. 1000 iterations.

Table 2: Finite sample performance of estimators when $T = 25$ (DGP2: $\phi = 0.5$)

			FE			bias corrected FE			GMM			Hayakawa IV		
	N		fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto
(i) α_1	50	med abs dev	0.076	0.091	0.086	0.044	0.036	0.027	0.076	0.092	0.078	0.059	0.053	0.039
(true value:		median bias	-0.076	-0.091	-0.086	-0.044	-0.033	-0.018	-0.076	-0.092	-0.078	-0.059	-0.005	-0.002
0.900)		iqr	0.042	0.051	0.052	0.041	0.052	0.048	0.049	0.067	0.066	0.055	0.107	0.080
		cp	0.271	0.269	0.324	0.666	0.815	0.887	0.363	0.431	0.495	0.642	0.943	0.901
	100	med abs dev	0.075	0.092	0.090	0.043	0.034	0.021	0.065	0.065	0.051	0.058	0.034	0.025
		median bias	-0.075	-0.092	-0.090	-0.043	-0.033	-0.017	-0.065	-0.065	-0.051	-0.058	-0.002	-0.001
		iqr	0.029	0.036	0.037	0.029	0.037	0.035	0.032	0.051	0.053	0.035	0.068	0.050
		cp	0.051	0.059	0.070	0.439	0.720	0.856	0.196	0.518	0.594	0.407	0.953	0.936
	500	med abs dev	0.074	0.090	0.090	0.042	0.031	0.019	0.054	0.019	0.016	0.057	0.017	0.012
		median bias	-0.074	-0.090	-0.090	-0.042	-0.031	-0.019	-0.054	-0.018	-0.014	-0.057	0.001	-0.001
		iqr	0.012	0.017	0.017	0.012	0.016	0.017	0.015	0.023	0.021	0.017	0.034	0.025
		cp	0.000	0.000	0.000	0.005	0.196	0.557	0.000	0.823	0.828	0.002	0.938	0.911
	1000	med abs dev	0.074	0.091	0.091	0.043	0.032	0.019	0.053	0.011	0.009	0.057	0.012	0.009
		median bias	-0.074	-0.091	-0.091	-0.043	-0.032	-0.019	-0.053	-0.009	-0.007	-0.057	0.000	0.001
		iqr	0.009	0.011	0.011	0.009	0.011	0.014	0.011	0.018	0.015	0.012	0.023	0.017
		cp	0.000	0.000	0.000	0.000	0.016	0.298	0.000	0.873	0.869	0.000	0.961	0.933
(ii) SAR	50	med abs dev	0.112	0.294	0.267	0.074	0.085	0.053	0.102	0.243	0.183	0.075	0.114	0.069
(true value:		median bias	-0.112	-0.294	-0.267	-0.074	-0.085	-0.053	-0.102	-0.243	-0.183	-0.075	-0.009	0.003
0.643)		iqr	0.037	0.081	0.106	0.034	0.055	0.050	0.049	0.126	0.157	0.057	0.230	0.140
		cp	0.005	0.000	0.002	0.130	0.722	0.669	0.115	0.186	0.202	0.529	0.954	0.869
	100	med abs dev	0.112	0.296	0.288	0.074	0.087	0.055	0.086	0.171	0.119	0.075	0.075	0.049
		median bias	-0.112	-0.296	-0.288	-0.074	-0.087	-0.055	-0.086	-0.171	-0.119	-0.075	-0.012	-0.003
		iqr	0.027	0.057	0.067	0.025	0.039	0.039	0.035	0.107	0.123	0.041	0.147	0.098
		cp	0.000	0.000	0.000	0.007	0.362	0.439	0.060	0.322	0.343	0.251	0.957	0.891
	500	med abs dev	0.111	0.292	0.292	0.073	0.084	0.059	0.070	0.050	0.034	0.075	0.036	0.024
		median bias	-0.111	-0.292	-0.292	-0.073	-0.084	-0.059	-0.070	-0.049	-0.034	-0.075	0.000	-0.004
		iqr	0.012	0.024	0.024	0.011	0.016	0.019	0.018	0.050	0.046	0.019	0.073	0.048
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.759	0.659	0.000	0.944	0.870
	1000	med abs dev	0.111	0.292	0.292	0.074	0.084	0.060	0.068	0.028	0.020	0.074	0.025	0.018
		median bias	-0.111	-0.292	-0.292	-0.074	-0.084	-0.060	-0.068	-0.025	-0.016	-0.074	-0.004	0.000
		iqr	0.008	0.020	0.020	0.007	0.013	0.017	0.012	0.041	0.032	0.013	0.050	0.036
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.840	0.785	0.000	0.943	0.888

Notes: See notes for Table 1.

Table 3: Finite sample performance of estimators when $T = 25$ (DGP3: $\phi = 0.97$)

			FE			bias corrected FE			GMM			Hayakawa IV		
	N		fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto
(i) α_1	50	med abs dev	0.141	0.139	0.120	0.111	0.078	0.065	0.214	0.188	0.172	0.259	0.212	0.152
(true value:		median bias	-0.141	-0.139	-0.120	-0.111	-0.078	-0.065	-0.214	-0.188	-0.172	-0.246	-0.041	-0.068
1.370)		iqr	0.040	0.052	0.048	0.037	0.053	0.050	0.065	0.075	0.070	0.259	0.410	0.331
		cp	0.000	0.027	0.046	0.011	0.409	0.499	0.000	0.036	0.034	0.607	0.988	0.907
	100	med abs dev	0.139	0.135	0.121	0.110	0.075	0.068	0.224	0.183	0.167	0.268	0.165	0.118
		median bias	-0.139	-0.135	-0.121	-0.110	-0.075	-0.068	-0.224	-0.183	-0.167	-0.263	-0.005	-0.039
		iqr	0.028	0.036	0.037	0.027	0.037	0.036	0.066	0.076	0.075	0.195	0.335	0.247
		cp	0.000	0.000	0.000	0.000	0.173	0.225	0.000	0.020	0.024	0.271	0.989	0.912
	500	med abs dev	0.138	0.134	0.130	0.109	0.073	0.073	0.261	0.146	0.132	0.276	0.071	0.065
		median bias	-0.138	-0.134	-0.130	-0.109	-0.073	-0.073	-0.261	-0.146	-0.132	-0.276	-0.007	-0.008
		iqr	0.013	0.017	0.018	0.012	0.017	0.017	0.064	0.062	0.062	0.078	0.143	0.128
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.026	0.000	0.969	0.928
	1000	med abs dev	0.139	0.134	0.133	0.109	0.073	0.073	0.276	0.118	0.106	0.275	0.051	0.044
		median bias	-0.139	-0.134	-0.133	-0.109	-0.073	-0.073	-0.276	-0.118	-0.106	-0.275	-0.002	0.001
		iqr	0.009	0.011	0.012	0.008	0.011	0.011	0.063	0.046	0.047	0.057	0.102	0.085
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.036	0.055	0.000	0.949	0.924
(ii) SAR	50	med abs dev	0.102	0.150	0.123	0.092	0.095	0.086	0.150	0.163	0.144	0.187	0.176	0.123
(true value:		median bias	-0.102	-0.150	-0.123	-0.092	-0.095	-0.086	-0.150	-0.163	-0.144	-0.179	-0.035	-0.055
0.979)		iqr	0.021	0.036	0.042	0.019	0.025	0.022	0.051	0.066	0.060	0.197	0.344	0.259
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.002	0.673	0.982	0.924
	100	med abs dev	0.101	0.147	0.129	0.091	0.093	0.087	0.157	0.154	0.139	0.190	0.136	0.099
		median bias	-0.101	-0.147	-0.129	-0.091	-0.093	-0.087	-0.157	-0.154	-0.139	-0.188	-0.006	-0.035
		iqr	0.015	0.026	0.035	0.013	0.017	0.015	0.053	0.062	0.058	0.147	0.273	0.192
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.006	0.360	0.985	0.915
	500	med abs dev	0.100	0.146	0.142	0.091	0.092	0.092	0.186	0.123	0.111	0.195	0.059	0.053
		median bias	-0.100	-0.146	-0.142	-0.091	-0.092	-0.092	-0.186	-0.123	-0.111	-0.195	-0.007	-0.008
		iqr	0.006	0.011	0.017	0.006	0.007	0.007	0.051	0.051	0.052	0.061	0.117	0.103
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.012	0.000	0.970	0.927
	1000	med abs dev	0.100	0.145	0.145	0.091	0.092	0.092	0.198	0.099	0.089	0.192	0.042	0.035
		median bias	-0.100	-0.145	-0.145	-0.091	-0.092	-0.092	-0.198	-0.099	-0.089	-0.192	-0.002	0.001
		iqr	0.004	0.007	0.009	0.004	0.005	0.005	0.051	0.040	0.039	0.044	0.085	0.070
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.042	0.000	0.956	0.926

Notes: See notes for Table 1.

Table 4: Finite sample performance of estimators when $T = 50$ and 75 ($N = 100$)

			FE			bias corrected FE			GMM			Hayakawa IV		
	T		fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto
DGP1	50	med abs dev	0.023	0.030	0.028	0.010	0.011	0.010	0.015	0.025	0.022	0.011	0.014	0.012
(i) α_1		median bias	-0.023	-0.030	-0.028	-0.001	-0.005	-0.002	-0.013	-0.025	-0.022	0.003	0.001	0.001
(true value:		iqr	0.019	0.021	0.021	0.019	0.021	0.020	0.020	0.024	0.024	0.022	0.027	0.024
0.400)		cp	0.654	0.545	0.568	0.942	0.930	0.942	0.857	0.698	0.730	0.937	0.947	0.940
	75	med abs dev	0.016	0.018	0.017	0.008	0.009	0.008	0.012	0.019	0.017	0.009	0.010	0.009
		median bias	-0.016	-0.018	-0.017	-0.001	-0.002	-0.001	-0.012	-0.019	-0.017	0.001	0.000	0.001
		iqr	0.016	0.017	0.017	0.016	0.017	0.016	0.016	0.019	0.018	0.017	0.020	0.018
		cp	0.738	0.700	0.712	0.951	0.944	0.951	0.826	0.694	0.725	0.955	0.952	0.953
(ii) SAR	50	med abs dev	0.044	0.212	0.195	0.013	0.030	0.021	0.021	0.154	0.121	0.023	0.050	0.036
(true value:		median bias	-0.044	-0.212	-0.195	0.001	-0.027	-0.011	-0.018	-0.154	-0.121	0.018	0.000	0.004
0.286)		iqr	0.027	0.061	0.076	0.025	0.049	0.041	0.033	0.082	0.105	0.035	0.100	0.072
		cp	0.425	0.001	0.005	0.965	0.941	0.920	0.895	0.243	0.274	0.894	0.967	0.887
	75	med abs dev	0.024	0.139	0.125	0.011	0.022	0.016	0.016	0.127	0.108	0.019	0.037	0.026
		median bias	-0.024	-0.139	-0.125	0.005	-0.014	-0.007	-0.015	-0.127	-0.108	0.017	0.001	0.001
		iqr	0.022	0.047	0.057	0.021	0.040	0.033	0.024	0.058	0.073	0.026	0.073	0.051
		cp	0.686	0.005	0.020	0.960	0.954	0.913	0.872	0.137	0.152	0.871	0.934	0.872
DGP2	50	med abs dev	0.028	0.031	0.030	0.010	0.011	0.010	0.018	0.029	0.025	0.011	0.014	0.012
(i) α_1		median bias	-0.028	-0.031	-0.030	-0.005	-0.006	-0.003	-0.018	-0.028	-0.025	0.003	0.001	0.001
(true value:		iqr	0.019	0.021	0.021	0.019	0.022	0.020	0.020	0.024	0.024	0.022	0.028	0.025
0.900)		cp	0.513	0.507	0.532	0.929	0.925	0.933	0.798	0.651	0.701	0.943	0.949	0.945
	75	med abs dev	0.020	0.019	0.018	0.008	0.009	0.008	0.016	0.020	0.018	0.009	0.010	0.009
		median bias	-0.020	-0.019	-0.018	-0.004	-0.003	-0.001	-0.016	-0.020	-0.018	0.000	0.000	0.000
		iqr	0.016	0.017	0.017	0.016	0.017	0.016	0.017	0.018	0.018	0.017	0.021	0.019
		cp	0.602	0.690	0.701	0.936	0.943	0.953	0.718	0.666	0.693	0.959	0.951	0.950
(ii) SAR	50	med abs dev	0.033	0.129	0.121	0.011	0.032	0.023	0.014	0.097	0.078	0.022	0.031	0.023
(true value:		median bias	-0.033	-0.129	-0.121	-0.010	-0.032	-0.023	-0.012	-0.097	-0.078	0.022	-0.002	-0.002
0.643)		iqr	0.017	0.035	0.042	0.016	0.028	0.024	0.021	0.049	0.059	0.024	0.062	0.046
		cp	0.273	0.000	0.000	0.908	0.758	0.703	0.895	0.161	0.193	0.802	0.954	0.915
	75	med abs dev	0.017	0.083	0.076	0.007	0.018	0.015	0.011	0.078	0.067	0.018	0.022	0.017
		median bias	-0.017	-0.083	-0.076	-0.002	-0.018	-0.015	-0.010	-0.078	-0.067	0.018	0.001	0.000
		iqr	0.014	0.027	0.031	0.014	0.023	0.020	0.016	0.033	0.039	0.017	0.043	0.033
		cp	0.599	0.002	0.007	0.965	0.855	0.783	0.872	0.086	0.114	0.740	0.944	0.896

Table 4: (continued)

			FE			bias corrected FE			GMM			Hayakawa IV		
	T		fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto	fix4	fix12	auto
DGP3	50	med abs dev	0.051	0.052	0.049	0.027	0.027	0.023	0.069	0.073	0.070	0.032	0.036	0.031
(i) α_1		median bias	0.051	-0.052	-0.049	-0.027	-0.027	-0.023	-0.069	-0.073	-0.070	0.026	-0.001	-0.002
(true value:		iqr	0.020	0.023	0.022	0.020	0.022	0.021	0.028	0.030	0.030	0.052	0.072	0.062
1.370)		cp	0.074	0.099	0.120	0.560	0.598	0.660	0.034	0.043	0.049	0.911	0.959	0.942
	75	med abs dev	0.034	0.030	0.029	0.018	0.014	0.013	0.046	0.044	0.043	0.014	0.018	0.015
		median bias	-0.034	-0.030	-0.029	-0.018	-0.014	-0.012	-0.046	-0.044	-0.043	0.006	-0.001	-0.002
		iqr	0.017	0.017	0.017	0.017	0.017	0.017	0.018	0.019	0.019	0.028	0.036	0.031
		cp	0.202	0.341	0.373	0.662	0.792	0.831	0.055	0.098	0.105	0.948	0.953	0.944
(ii) SAR	50	med abs dev	0.041	0.061	0.057	0.037	0.044	0.041	0.050	0.070	0.065	0.032	0.031	0.025
(true value:		median bias	-0.041	-0.061	-0.057	-0.037	-0.044	-0.041	-0.050	-0.070	-0.065	0.030	-0.001	-0.001
0.979)		iqr	0.007	0.010	0.013	0.007	0.008	0.008	0.017	0.019	0.020	0.038	0.060	0.051
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.839	0.962	0.938
	75	med abs dev	0.025	0.036	0.034	0.023	0.028	0.026	0.032	0.044	0.042	0.015	0.014	0.012
		median bias	-0.025	-0.036	-0.034	-0.023	-0.028	-0.026	-0.032	-0.044	-0.042	0.015	0.000	-0.001
		iqr	0.005	0.006	0.007	0.005	0.005	0.005	0.009	0.011	0.012	0.019	0.029	0.025
		cp	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.813	0.948	0.933

Notes: Fixed rules with $c = 4$(fix4) and 12(fix12) imply the lag lengths $p = 3$ and 10 for $T = 50$ and $p = 3$ and 11 for $T = 75$, respectively. See also notes for Table 1.

We first discuss the results reported in Tables 1-3. The results clearly illustrate the bias properties of the estimators and are consistent with the theoretical predictions. When N is small, both the fixed effects estimator and the GMM estimator suffer from the bias problem. Our bias-correction procedure seems to work well for the fixed effects estimator. For a larger N, however, the GMM estimator dominates the bias-corrected fixed effects estimator. Irrespective of the sample size, Hayakawa's IV estimator performs best among all the estimators in terms of bias.

In terms of the median absolute deviation, the bias corrected fixed effects estimator dominates others when N is small. However, as N increases, the performance of the GMM estimator and Hayakawa's IV estimator improves. However, interquartile ranges of Hayakawa's IV estimator are the largest, with those of the GMM estimator being the second.

In terms of the coverage probability, Hayakawa's IV estimator performs very well. In contrast, the fixed effects estimator has almost zero coverage for a large sample size because of the relatively large bias compared to its small interquartile ranges. It is interesting to note that when the automatic lag selection method is used, the median absolute deviation of the IV estimator becomes smaller than in the case based on the rule of thumb, with some cost of coverage probability.

Table 4 illustrates the effect of increasing the length of time series. The performances of the bias corrected fixed effects estimator improves as T increases. In particular, the improvement of its coverage probability is remarkable. The fixed effects estimator suffers from sizable bias even when $T = 75$, which demonstrates the importance of bias correction. The GMM estimator does not show remarkable improvements in terms of bias and coverage probability when T increases, although its interquartile range shrinks. Hayakawa's estimator still produces the most reliable confidence interval in most of the cases in Table 4. For DGP1 and DGP2, the bias corrected fixed effects estimator also produces a reliable confidence interval, while its coverage probability is not satisfactory for DGP3.

Lastly, we conduct an additional analysis regarding the source of the finite-sample bias. Recall that, in the discussion of Theorem 2, the bias of the fixed effects estimators is decomposed into 'truncation bias' and 'fundamental bias.' In simulation, we can directly evaluate the relative contribution of each component since the information of the true process can be utilized. To be specific, the bias in the simulation can be decomposed as

$$\begin{aligned} &\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\alpha}_F^{(r)}(p)-\alpha(p)\right) \\ =\ &\underbrace{\frac{1}{R}\sum_{r=1}^{R}\left((\hat{\Gamma}_p^{F(r)})^{-1}\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t^{(r)}(p)'\tilde{b}_{t,p}^{(r)}\right)}_{\text{truncation bias}}+\underbrace{\frac{1}{R}\sum_{r=1}^{R}\left((\hat{\Gamma}_p^{F(r)})^{-1}\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t^{(r)}(p)'\tilde{\epsilon}_t^{(r)}\right)}_{\text{fundamental bias}} \end{aligned}$$

where the superscript r signifies the r-th simulated observation in R replications. Analogously, the bias components of the GMM estimator and Hayakawa's efficient IV estimator can be evaluated using the fact that their estimation error, $\hat{\alpha}_G(p)-\alpha(p)$ and $\hat{\alpha}_H(p)-\alpha(p)$, can be decomposed as

$$(\hat{\Gamma}_p^G)^{-1}\frac{1}{NT}\sum_{t=p+1}^{T-1}x_t^*(p)'M_tb_{t,p}^*+(\hat{\Gamma}_p^G)^{-1}\frac{1}{NT}\sum_{t=p+1}^{T-1}x_t^*(p)'M_t\epsilon_t^*$$

and

$$(\hat{\Gamma}_p^H)^{-1}\frac{1}{NT}\sum_{t=p+2}^{T-1}h_t(p)'b_{t,p}^*+(\hat{\Gamma}_p^H)^{-1}\frac{1}{NT}\sum_{t=p+2}^{T-1}h_t(p)'\epsilon_t^*,$$

Table 5: Decomposition of the finite sample bias of estimators when $T = 25$

	N		FE			GMM			Hayakawa IV		
			p=2	p=4	p=6	p=2	p=4	p=6	p=2	p=4	p=6
DGP1	50	total	-0.053	-0.056	-0.068	-0.042	-0.042	-0.058	-0.022	-0.004	-0.001
(i) α_1		trun	-0.010	-0.001	0.000	-0.016	-0.002	0.000	-0.022	-0.003	-0.001
(0.400)		fund	-0.042	-0.055	-0.068	-0.025	-0.040	-0.057	0.000	0.000	0.000
	100	total	-0.053	-0.056	-0.069	-0.033	-0.027	-0.037	-0.023	-0.005	-0.002
		trun	-0.010	-0.001	0.000	-0.018	-0.002	0.000	-0.022	-0.003	-0.001
		fund	-0.042	-0.056	-0.069	-0.015	-0.025	-0.037	-0.001	-0.002	-0.001
	500	total	-0.052	-0.055	-0.068	-0.022	-0.007	-0.009	-0.022	-0.003	0.000
		trun	-0.010	-0.001	0.000	-0.019	-0.002	0.000	-0.022	-0.003	-0.001
		fund	-0.042	-0.055	-0.068	-0.003	-0.005	-0.009	0.000	0.000	0.000
	1000	total	-0.052	-0.055	-0.068	-0.021	-0.005	-0.005	-0.022	-0.003	0.000
		trun	-0.010	-0.001	0.000	-0.019	-0.002	0.000	-0.022	-0.003	-0.001
		fund	-0.042	-0.055	-0.068	-0.002	-0.003	-0.004	0.000	0.000	0.000
(ii) SAR	50	total	-0.114	-0.177	-0.289	-0.089	-0.121	-0.212	-0.056	-0.016	-0.004
(0.286)		trun	-0.033	-0.006	-0.001	-0.044	-0.009	-0.002	-0.055	-0.013	-0.003
		fund	-0.081	-0.171	-0.288	-0.046	-0.113	-0.211	-0.001	-0.003	-0.001
	100	total	-0.114	-0.175	-0.287	-0.072	-0.076	-0.132	-0.057	-0.019	-0.004
		trun	-0.033	-0.006	-0.001	-0.046	-0.009	-0.002	-0.055	-0.013	-0.003
		fund	-0.081	-0.170	-0.286	-0.026	-0.066	-0.130	-0.002	-0.005	-0.001
	500	total	-0.113	-0.175	-0.287	-0.054	-0.026	-0.034	-0.055	-0.014	-0.003
		trun	-0.033	-0.006	-0.001	-0.049	-0.010	-0.002	-0.055	-0.013	-0.003
		fund	-0.080	-0.170	-0.286	-0.005	-0.015	-0.032	0.000	-0.001	0.000
	1000	total	-0.114	-0.175	-0.287	-0.052	-0.018	-0.019	-0.055	-0.014	-0.003
		trun	-0.033	-0.006	-0.001	-0.049	-0.010	-0.002	-0.055	-0.013	-0.003
		fund	-0.081	-0.169	-0.286	-0.003	-0.008	-0.016	-0.001	-0.001	0.000
DGP2	50	total	-0.075	-0.061	-0.074	-0.075	-0.052	-0.068	-0.057	-0.008	-0.002
(i) α_1		trun	-0.033	-0.001	0.000	-0.045	-0.003	0.000	-0.057	-0.008	-0.002
(0.900)		fund	-0.042	-0.060	-0.074	-0.030	-0.049	-0.067	0.000	-0.001	0.000
	100	total	-0.075	-0.061	-0.074	-0.066	-0.035	-0.045	-0.059	-0.010	-0.002
		trun	-0.033	-0.001	0.000	-0.048	-0.004	-0.001	-0.057	-0.008	-0.002
		fund	-0.042	-0.060	-0.074	-0.018	-0.031	-0.045	-0.002	-0.002	-0.001
	500	total	-0.074	-0.060	-0.073	-0.055	-0.012	-0.012	-0.057	-0.008	-0.001
		trun	-0.033	-0.001	0.000	-0.051	-0.005	-0.001	-0.057	-0.008	-0.002
		fund	-0.041	-0.059	-0.073	-0.004	-0.007	-0.011	0.000	0.000	0.001
	1000	total	-0.074	-0.060	-0.073	-0.053	-0.008	-0.006	-0.057	-0.007	-0.001
		trun	-0.033	-0.001	0.000	-0.051	-0.005	-0.001	-0.057	-0.008	-0.002
		fund	-0.041	-0.059	-0.073	-0.002	-0.004	-0.005	0.000	0.000	0.000

Table 5: (continued)

	N		FE p=2	FE p=4	FE p=6	GMM p=2	GMM p=4	GMM p=6	Hayakawa IV p=2	Hayakawa IV p=4	Hayakawa IV p=6
(ii) SAR	50	total	-0.112	-0.136	-0.200	-0.102	-0.102	-0.156	-0.075	-0.019	-0.005
(0.643)		trun	-0.036	-0.006	-0.001	-0.055	-0.010	-0.002	-0.074	-0.017	-0.004
		fund	-0.076	-0.131	-0.199	-0.047	-0.092	-0.154	-0.001	-0.002	-0.001
	100	total	-0.111	-0.134	-0.198	-0.087	-0.067	-0.100	-0.076	-0.020	-0.004
		trun	-0.036	-0.006	-0.001	-0.060	-0.011	-0.002	-0.074	-0.017	-0.004
		fund	-0.075	-0.128	-0.197	-0.027	-0.056	-0.098	-0.002	-0.003	0.000
	500	total	-0.111	-0.134	-0.198	-0.070	-0.026	-0.028	-0.075	-0.018	-0.003
		trun	-0.036	-0.006	-0.001	-0.064	-0.013	-0.003	-0.074	-0.017	-0.004
		fund	-0.075	-0.128	-0.197	-0.006	-0.013	-0.025	0.000	-0.001	0.001
	1000	total	-0.111	-0.134	-0.198	-0.068	-0.019	-0.015	-0.075	-0.017	-0.004
		trun	-0.036	-0.006	-0.001	-0.065	-0.013	-0.003	-0.074	-0.017	-0.004
		fund	-0.075	-0.128	-0.197	-0.003	-0.007	-0.012	-0.001	0.000	0.000
DGP3	50	total	-0.141	-0.108	-0.121	-0.217	-0.165	-0.175	-0.324	0.110	-0.282
(i) α_1		trun	-0.056	-0.002	0.000	-0.077	-0.003	0.000	-0.429	-0.084	-0.011
(1.370)		fund	-0.085	-0.106	-0.121	-0.139	-0.161	-0.174	0.106	0.193	-0.272
	100	total	-0.139	-0.107	-0.120	-0.227	-0.158	-0.169	-0.316	-0.106	-0.052
		trun	-0.056	-0.002	0.000	-0.100	-0.005	-0.001	-0.317	-0.067	-0.031
		fund	-0.084	-0.105	-0.120	-0.127	-0.153	-0.168	0.001	-0.039	-0.021
	500	total	-0.138	-0.105	-0.118	-0.264	-0.120	-0.128	-0.283	-0.057	-0.007
		trun	-0.056	-0.002	0.000	-0.191	-0.014	-0.002	-0.282	-0.058	-0.012
		fund	-0.082	-0.103	-0.118	-0.073	-0.105	-0.126	-0.001	0.001	0.005
	1000	total	-0.138	-0.105	-0.118	-0.280	-0.095	-0.098	-0.278	-0.054	-0.008
		trun	-0.056	-0.002	0.000	-0.232	-0.020	-0.003	-0.277	-0.055	-0.011
		fund	-0.083	-0.103	-0.118	-0.048	-0.075	-0.096	-0.001	0.001	0.003
(ii) SAR	50	total	-0.103	-0.110	-0.127	-0.152	-0.138	-0.150	-0.240	0.071	-0.232
(0.979)		trun	-0.013	-0.002	0.000	-0.033	-0.004	-0.001	-0.305	-0.064	-0.009
		fund	-0.090	-0.108	-0.127	-0.119	-0.135	-0.149	0.065	0.136	-0.223
	100	total	-0.101	-0.107	-0.125	-0.158	-0.130	-0.140	-0.226	-0.086	-0.043
		trun	-0.013	-0.002	0.000	-0.052	-0.005	-0.001	-0.225	-0.055	-0.025
		fund	-0.089	-0.105	-0.124	-0.107	-0.125	-0.140	0.000	-0.031	-0.019
	500	total	-0.100	-0.106	-0.124	-0.188	-0.098	-0.106	-0.199	-0.046	-0.007
		trun	-0.013	-0.002	0.000	-0.127	-0.012	-0.002	-0.198	-0.046	-0.010
		fund	-0.088	-0.104	-0.123	-0.061	-0.085	-0.104	-0.001	0.000	0.004
	1000	total	-0.100	-0.106	-0.124	-0.201	-0.078	-0.081	-0.196	-0.044	-0.008
		trun	-0.013	-0.002	0.000	-0.161	-0.017	-0.002	-0.195	-0.044	-0.009
		fund	-0.088	-0.104	-0.123	-0.040	-0.061	-0.079	-0.001	0.000	0.002

Notes: Mean of the components of finite sample bias. The total finite sample bias (total) is decomposed into the truncation bias (trun) and the fundamental bias (fund). 1000 iterations.

respectively, where

$$
\begin{aligned}
b^*_{t,p} &= \sqrt{(T-t)/(T-t+1)}\left(b_{t,p} - \sum_{\tau=t+1}^{T} b_{\tau,p}/(T-t)\right) \text{ and} \\
\epsilon^*_t &= \sqrt{(T-t)/(T-t+1)}\left(\epsilon_t - \sum_{\tau=t+1}^{T} \epsilon_\tau/(T-t)\right).
\end{aligned}
$$

Table 5 provides such a decomposition of the finite sample bias of the fixed effects estimator, GMM estimator and Hayakawa's IV estimator when the data is generated from DGP1, DGP2 and DGP3 with $N = \{50, 100, 500, 1000\}$ and $T = 25$. Since we expect the decreasing contribution of trunction bias in response to increasing lag length, we report the bias decomposition when the model is estimated using $p = \{2, 4, 6\}$. It should be noted that the total bias presented in Table 5 differs from those reported in Tables 1 to 4 because the latter are median bias rather than mean bias.[14] From the table, we observe that these two types of bias appear very different across three estimators. For given N and p, the relative contribution of truncation bias in total bias is the smallest for the fixed effects estimator and is the largest for Hayakawa's IV estimator. This result is not surprising given that the fixed effects estimator is essentially an OLS estimator and OLS produces the best linear projection. The fixed effects estimator, however, suffers from substantial bias due to fundamental bias. The GMM estimator also exhibits a large fundamental bias when N is small. In contrast, the bias of Hayakawa's estimator is solely from the truncation bias. An important observation is that, for the fixed effects estimator and the GMM estimator, there is a trade-off in the value of p such that, as p increases,

[14]In general, moments may not exist for Hayakawa's estimator but the bias decomposition can still be evaluated in simulation. For this reason, we note that the mean biases of Hayakawa's estimator for DGP3 with $N = 50$ and 100 presented in Table 5 may not be reliable while sensible numbers are obtained for other cases.

the truncation bias quickly vanishes but the fundamental bias increases. Unlike other two estimators, the fundamental bias of Hayakawa's estimator remains negligible even when $p = 6$.

On the whole, the simulation results can be summarized as follows. If our concern is coverage probability or bias, then Hayakawa's estimator is recommended. On the other hand, if we care primarily about precision, then we recommend the bias corrected fixed effects estimator if T is not small compared to N and the GMM estimator when N is large and T is small.

5 Empirical Applications

In this section, we apply our procedure to investigate the relative price adjustment of individual goods across cities. To measure the speed of price adjustment towards the long-run law of one price (LOP) across intranational and international cities, autoregressive models have been often estimated using panel data. For example, Parsley and Wei (1996) use panel price data from 48 cities in the United States and estimate the rate of convergence in terms of the sum of AR coefficients. Applications using a dynamic panel GMM procedure to LOP deviations include Crucini and Shintani (2008), who estimate AR(1) models using both intranational and international city pair data, and Crucini, Shintani and Tsuruga (2008), who use the sum of AR coefficients of the higher order AR models as a measure of persistence.

Our price data is from the American Chamber of Commerce Researchers Association (ACCRA) Cost of Living Index produced by the Council of Community and Economic Research. This is the extended version of the data used by Parsley and Wei (1996) for their studies on intranational LOP deviations. The original ACCRA data

Table 6: List of goods

CPI categorization		ACCRA categorization
1	Food at home	T-bone steak, Ground beef, Frying chicken, Chunk light tuna, Whole milk, Eggs, Margarine, Parmesan cheese, Potatoes, Bananas, Lettuce, Bread, Coffee, Sugar, Corn flakes, Sweat peas, Peaches, Shortening, Frozen corn, Soft drink
2	Food away from home	Hamburger sandwich, Pizza, Fried chicken
3	Alcoholic beverages	Beer, Wine
4	Shelter	Apartment, Home purchase price, Mortgage rate, Monthly payment
5	Fuel and other utilities	Total home energy cost, Telephone
6	Household furnishings and operations	Facial tissues, Dishwashing powder, Dry cleaning, Major appliance repair
7	Men's and boy's apparel	Men's dress shirt
8	Private transportation	Auto maintenance, Gasoline
9	Medical care	Doctor office visit, Dentist office visit
10	Personal care	Haircut, Beauty salon, Toothpaste, Shampoo
11	Entertainment	Newspaper subscription, Movie, Bowling, Tennis balls

includes 75 goods and services and 632 cities, but we focus on 11 CPI categorized good price series from 52 US cities. Our data is a monthly series over 18 years from January 1990 to December 2007 ($T = 72$). In measuring the LOP deviations for each categorized good, we follow Parsley and Wei (1996) and use one benchmark city to compute intercity price differentials over time (our benchmark city is Albuquerque). Let P_{it} and P_{0t} be the price of a good for a city i and that for the benchmark city, respectively. Then, the LOP deviations are computed as $y_{it} = \log P_{it} - \log P_{0t}$ for $i = 1, ..., 51$. Since we pool all the goods in the same category, the total number of

Table 7: Sum of AR coefficients estimates

Goods category	N	FE	bias corrected FE	GMM	Hayakawa IV
1	1020	0.687	0.728	0.717	0.777
		(0.005)	(0.006)	(0.007)	(0.021)
2	153	0.636	0.697	0.614	0.823
		(0.014)	(0.014)	(0.017)	(0.029)
3	102	0.731	0.776	0.714	0.930
		(0.015)	(0.015)	(0.019)	(0.038)
4	204	0.811	0.843	0.791	0.891
		(0.009)	(0.009)	(0.010)	(0.019)
5	102	0.743	0.786	0.674	0.754
		(0.016)	(0.016)	(0.019)	(0.090)
6	204	0.670	0.725	0.561	0.238
		(0.013)	(0.013)	(0.018)	(0.076)
7	51	0.657	0.714	1.048	0.831
		(0.030)	(0.030)	(0.040)	(0.076)
8	102	0.504	0.584	0.358	0.583
		(0.026)	(0.027)	(0.030)	(0.075)
9	102	0.695	0.741	0.619	0.666
		(0.016)	(0.016)	(0.020)	(0.065)
10	204	0.753	0.796	0.726	0.784
		(0.011)	(0.012)	(0.015)	(0.030)
11	204	0.759	0.799	0.652	0.676
		(0.012)	(0.012)	(0.016)	(0.059)

Notes: Numbers in parentheses are standard errors. Sample periods are from January 1990 to December 2007 ($T = 72$).

cross-sectional observations (N) will be multiples of 51. All the names of individual goods in our categorization are presented in Table 6.

Table 7 reports the estimated sum of AR coefficients (SAR) for each categorized good using the fixed effects estimator, the bias-corrected fixed effects estimator, the GMM estimator and the Hayakawa's IV estimator. For each estimator, we use lags selected by sequential rule with the maximum lag set at $p = 11$ based on the formula $p_{\max} = [12(T/100)^{1/4}]$. The results show some variation across the goods categories. At the same time, the SAR also varies among the different estimators. On the whole, the bias-corrected fixed effects estimator and Hayakawa's IV estimator provide somewhat close estimates. In contrast, the fixed effects estimator provides a much smaller number than the bias-corrected estimator, implying a non-negligible downward bias. GMM estimates are also smaller than Hayakawa's IV estimates for most of the goods categories.

6 Conclusion

In this paper, we consider the estimation of a dynamic panel autoregressive (AR) process of possibly infinite order in the presence of individual effects. We approximate and estimate the model by letting the order of the AR process of the fitted model increase with the sample size. We study the asymptotic properties of various estimators and also investigate their finite sample properties in simulations. The results indicate that the fixed effects estimator suffers severely from bias, and is not recommended. The bias-corrected estimator is preferred in terms of mean squared errors when T is not small. On the other hand, the GMM estimator is better when N is large and T is small. Hayakawa's IV estimator shows an excellent performance in terms of bias and coverage probability, but its finite sample distribution is dispersed compared with other estimators. The choice of estimator to be used should depend on the relative magnitude of N and T and on whether we care more about the bias or the overall precision.

Our results are useful for making statistical inferences regarding quantities that are important in understanding the dynamic nature of an economic variable, such as the long-run effect, without relying on a strong assumption. Although not discussed in this paper, further applications of our results are possible. For example, our estimators would be useful in constructing a model-free impulse response function. See, e.g., Jordà (2005) and Chang and Sakata (2007) for the model-free impulse response function in time series analysis. It would also be interesting to extend the tests of Granger causality by Lütkepohl and Poskitt (1996) that are based on infinite order AR models to panel data setting. Other applications of an AR model of infinite order would be long-run variance estimation, spectral density estimation as well as unit root tests. These applications seem to be a promising future research agenda.

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**(4): 1121–1159.

Anderson, T. W. and Hsiao, C. (1981). Estimation of dynamics models with error components, *Journal of the American Statistical Association* **76**(375): 598–606.

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

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* **68**: 29–51.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**(3): 657–681.

Berk, K. N. (1974). Consistent autoregressive spectral estimates, *Annals of Statistics* **2**(3): 489–502.

Chang, P.-L. and Sakata, S. (2007). Estimation of impulse response functions using long autoregression, *Econometrics Journal* **10**: 453–469.

Crucini, M. J. and Shintani, M. (2008). Persistence in law of one price deviations: Evidence from micro-data, *Journal of Monetary Economics* **55**(3): 629–644.

Crucini, M. J., Shintani, M. and Tsuruga, T. (2010). Accounting for persistence and volatility of good-level real exchange rates: The role of sticky information, *Journal of International Economics* **51**(5): 48–60.

Gonçalves, S. and Kilian, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity, *Econometric Reviews* **26**(6): 609–641.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* **70**(4): 1639–1657.

Han, C., Phillips, P. C. B. and Sul, D. (2011). X-differencing and dynamic panel model estimation, forthcoming in *Econometric Theory*.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*, John Wiley & Sons, Inc.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator for panel AR(p) models when both N and T are large, *Econometric Theory* **25**: 873–890.

Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton.

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988). Estimating vector autoregressions with panel data, *Econometrica* **56**(6): 1371–1395.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections, *American Economic Review* **95**(1): 161–182.

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* **68**: 53–78.

Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association* **75**: 693–700.

Lee, Y. (2006). *General Approaches to Dynamic Panel Modelling and Bias Correction*, Ph.D. thesis, Yale University.

Lee, Y. (2012a). Bias in dynamic panel models under time series misspecification, *Journal of Econometrics* **169**: 54–60.

Lee, Y. (2012b). Model selection in the presence of incidental parameters, mimeo.

Lewis, R. and Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting, *Journal of Multivariate Analysis* **16**: 393–411.

Lütkepohl, H. and Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models, *Econometric Theory* **7**: 487–496.

Lütkepohl, H. and Poskitt, D. S. (1996). Testing for causation using infinite order vector autoregressive processes, *Econometric Theory* **12**: 61–87.

Lütkepohl, H. and Saikkonen, P. (1997). Impulse response analysis in infinite order vector autoregression processes, *Journal of Econometrics* **81**: 127–157.

Morimune, K. (1983). Approximate distribution of the k-class estimators when the degree of overidentifiability is large compared with the sample size, *Econometrica* **51**(3): 821–841.

Newey, W. K. and Smith, R. (2004). Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica* **72**(1): 219–255.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions, *Econometrica* **77**(3): 687–719.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrica* **16**: 1–32.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncated lag, *Journal of the American Statistical Association* **90**(429): 268–281.

Nickell, S. (1981). Biases in dynamic models with fixed effects, *Econometrica* **49**(6): 1417–1426.

Okui, R. (2009). The optimal choice of moments in dynamic panel data models, *Journal of Econometrics* **151**: 1–16.

Okui, R. (2010). Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data, *Econometric Theory* **26**: 1263–1304.

Parsley, D. C. and Wei, S.-J. (1996). Convergence to the law of one price without trade barriers or currency fluctuations, *Quarterly Journal of Economics* **111**(4): 1211–1236.

Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data, *Econometrica* **67**(5): 1057–1111.

Appendix

Throughout the appendix, $C \in (1, \infty)$ denotes a generic bounded constant, which does not depend on any index and whose actual value varies across occasions. Given a matrix A, we let $||A||$ denote the Euclidean matrix norm defined by $||A||^2 = tr(A'A)$. Also let $||A||_1$ denote the Banach norm so that $||A||_1 = \sup_{x \neq 0}\{||Ax||/||x||\}$, using the Euclidean norm for the vector l, $||l|| = (l'l)^{1/2}$. For any symmetric matrix A, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ be the minimum and the maximum eigenvalues of A, respectively. We note that $||A||_1 = \sqrt{\lambda_{\max}(A'A)}$. When A is symmetric and positive definite, $||A||_1 = \lambda_{\max}(A)$. Define $\gamma_k = E(w_{it}w_{i,t-k})$. We let

$$\bar{w}_{i,t,\tau} = \frac{1}{\tau - t + 1}(w_{i,t} + \cdots + w_{i,\tau}).$$

We also define $\bar{w}_{i,t,\tau}(p) = (\bar{w}_{i,t,\tau}, \ldots, \bar{w}_{i,t-p+1,\tau-p+1})'$, $\bar{w}_{t,\tau} = (\bar{w}_{1,t,\tau}, \ldots, \bar{w}_{N,t,\tau})'$ and $\bar{w}_{t,\tau}(p) = (\bar{w}_{t,\tau}, \ldots, \bar{w}_{t-p+1,\tau-p+1})$. Similarly, define $\bar{\epsilon}_{i,t,\tau} = (\epsilon_{i,t} + \cdots + \epsilon_{i,\tau})/(\tau - t + 1)$ and $\bar{\epsilon}_{t,\tau} = (\bar{\epsilon}_{1,t,\tau}, \ldots, \bar{\epsilon}_{N,t,\tau})'$.

The following inequalities will be used below: $||A||_1 = \sqrt{\lambda_{\max}(A'A)} \leq (tr(A'A))^{1/2} = ||A||$. $||AB||^2 \leq ||A||_1^2||B||^2$ and $||AB||^2 \leq ||A||^2||B||_1^2$ (See Lewis and Reinsel (1985) and Wiener and Masani (1958)). For any conformable matrices A and D and any square matrix B, $||A'BD|| \leq \|B\|_1||A|| \cdot ||D||$.

A Lemmas useful for all the estimators

This section presents several lemmas that are commonly employed in the derivation of the asymptotic properties of all the estimators.

Lemma 1. *Suppose that Assumption 1 is satisfied. Then,*

$$||\bar{w}_{t,\tau}(p)||^2 = O_p\left(\frac{Np}{\tau - t + 1}\right) \text{ and } E||\bar{w}_{t,\tau}(p)||^2 = O\left(\frac{Np}{\tau - t + 1}\right).$$

Proof. The first statement follows from the second statement and the Markov inequality. We thus show the second statement. We observe that

$$E\left\|\bar{w}_{t,\tau}(p)\right\|^2 = Ntr(E(\bar{w}_{i,t,\tau}(p)\bar{w}_{i,t,\tau}(p)')) = NpE(\bar{w}_{i,t,\tau}^2) = O\left(\frac{Np}{\tau - t + 1}\right),$$

where the second last equality follows from the stationarity of w_{it} and the last equality comes from the fact that w_{it} is a short memory process.

□

Lemma 2. *Suppose that Assumption 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/(NT) \to 0$ and $p^3/T^2 \to 0$, then*

$$\left\| \frac{1}{NT} \sum_{t=p+1}^{T} w_{t-1}(p)' w_{t-1}(p) - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} \right)$$

and

$$\left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' w_{t-1}(p) - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} \right).$$

Proof. Noting that $E(w_{t-1}(p)' w_{t-1}(p))) = N\Gamma_p$, we have

$$\begin{aligned} & \left\| \frac{1}{NT} \sum_{t=p+1}^{T} w_{t-1}(p)' w_{t-1}(p) - \Gamma_p \right\| \\ \leq & \left\| \frac{1}{NT} \sum_{t=p+1}^{T} (w_{t-1}(p)' w_{t-1}(p) - E(w_{t-1}(p)' w_{t-1}(p))) \right\| + \left\| \frac{p}{T} \Gamma_p \right\|. \qquad (6) \end{aligned}$$

We see that $\|\Gamma_p\|^2 = \sum_{k=1-p}^{p-1} (p-k) \gamma_{|k|}^2 = O(p)$ so that $\|(p/T)\Gamma_p\| = O(p^{3/2}/T)$. The first term of (6) is

$$\begin{aligned} & E \left\| \frac{1}{NT} \sum_{t=p+1}^{T} (w_{t-1}(p)' w_{t-1}(p) - E(w_{t-1}(p)' w_{t-1}(p))) \right\|^2 \\ = & \sum_{k=0}^{p-1} (p-k) var\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+1}^{T} w_{i,t} w_{i,t-k} \right) = \frac{1}{N} \sum_{k=0}^{p-1} (p-k) var\left(\frac{1}{T} \sum_{t=p+1}^{T} w_{i,t} w_{i,t-k} \right) \end{aligned}$$

by the stationarity assumption. The variance is

$$\begin{aligned} & var\left(\frac{1}{T} \sum_{t=p+1}^{T} w_{i,t} w_{i,t-k} \right) = E\left(\left(\frac{1}{T} \sum_{t=p+1}^{T} (w_{i,t} w_{i,t-k} - \gamma_k) \right)^2 \right) \\ = & \frac{1}{T^2} \sum_{t_1=p+1}^{T} \sum_{t_2=p+1}^{T} (\gamma_{|t_1-t_2|}^2 + \gamma_{|t_1-t_2-k|} \gamma_{|t_1-t_2+k|} + \kappa_w(t_1, t_1-k, t_2, t_2-k)) \\ = & O\left(\frac{1}{T} \right) \end{aligned}$$

uniformly in k, by Assumption 1, where $\kappa_w(t_1, t_1-k, t_2, t_2-k)$ is the fourth order cumulant of $(w_{it_1}, w_{i,t_1-k}, w_{it_2}, w_{i,t_2-k})$.[15] It therefore follows that

$$
\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{t=p+1}^{T}(w_{t-1}(p)'w_{t-1}(p)-E(w_{t-1}(p)'w_{t-1}(p)))\right\|^2 \\
= \; & \frac{1}{N}\sum_{k=0}^{p-1}(p-k)O\left(\frac{1}{T}\right)=O\left(\frac{p^2}{NT}\right).
\end{aligned}
$$

The first statement holds by the Chebyshev inequality.

For the second statement, we observe that

$$
\begin{aligned}
& \left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'w_{t-1}(p)-\Gamma_p\right\| \\
\leq \; & \left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}w_{t-1}(p)'w_{t-1}(p)-\Gamma_p\right\|+\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}w_{t-1}(p)'w_{t-1}(p)\right\| \quad (7)
\end{aligned}
$$

Now, we have

$$
\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}w_{t-1}(p)'w_{t-1}(p)\right\|\leq\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}\left\|w_{i,t-1}(p)w_{i,t-1}(p)'\right\|.
$$

Since

$$
\begin{aligned}
E\left\|w_{i,t-1}(p)w_{i,t-1}(p)'\right\| &= E(tr(w_{i,t-1}(p)w_{i,t-1}(p)'w_{i,t-1}(p)w_{i,t-1}(p)'))^{1/2} \\
&= E\left(\sum_{k=1}^{p}w_{i,t-k}^2\right)=p\gamma_0,
\end{aligned}
$$

[15] We use the fact that $\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\sum_{j_3=-\infty}^{\infty}|\kappa_w(0,j_1,j_2,j_3)|<\infty$. Its proof is the following. By the formula in Anderson (1971, p 467), we have

$$
\kappa_w(0,j_1,\ldots,j_{k-1})=\kappa_\epsilon\sum_{t=-\infty}^{\infty}\psi_t\psi_{t+j_1}\psi_{t+j_2}\psi_{t+j_3}.
$$

It therefore holds that

$$
\begin{aligned}
\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\sum_{j_3=-\infty}^{\infty}|\kappa_w(0,j_1,j_2,j_3)| &= |\kappa_\epsilon|\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\sum_{j_3=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}|\psi_t\psi_{t+j_1}\psi_{t+j_2}\psi_{t+j_3}| \\
&= |\kappa_\epsilon|\sum_{t=-\infty}^{\infty}|\psi_t|\sum_{j_1=-\infty}^{\infty}|\psi_{t+j_1}|\sum_{j_2=-\infty}^{\infty}|\psi_{t+j_2}|\sum_{j_3=-\infty}^{\infty}|\psi_{t+j_3}| \\
&= |\kappa_\epsilon|\left(\sum_{t=-\infty}^{\infty}|\psi_t|\right)^4<\infty
\end{aligned}
$$

by the fourth moment condition $E(\epsilon_{it}^4)<\infty$ and the absolute summability condition $\sum_{t=-\infty}^{\infty}|\psi_t|<\infty$.

the Markov inequality implies that

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}w_{t-1}(p)'w_{t-1}(p)\right\| \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}O_p\left(p\right) = O_p\left(\frac{p\log T}{T}\right).$$

Because $p\log T/T = (p^{3/2}/T)^{2/3}(\log T/T^{1/3}) = o(1)$ when $p^{3/2}/T = o(1)$, (7) and the first statement of the lemma imply the second statement. □

Lemma 3. *Suppose that Assumption 2 is satisfied. Let* $\hat{\Gamma}_p$ *be an estimator of* Γ_p *such that* $\left\|\hat{\Gamma}_p - \Gamma_p\right\| = O_p(\rho_{N,T,p})$ *where* $\rho_{N,T,p} = o(1)$ *as* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$. *Then, as* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$, *we have*

$$\begin{aligned} ||\hat{\Gamma}_p - \Gamma_p||_1 &= O_p\left(\rho_{N,T,p}\right), &(8)\\ ||(\hat{\Gamma}_p)^{-1} - \Gamma_p^{-1}||_1 &= O_p\left(\rho_{N,T,p}\right), &(9)\\ \textit{and } ||(\hat{\Gamma}_p)^{-1}||_1 &= O_p\left(1\right). &(10)\end{aligned}$$

Proof. Since $||A||_1 \leq ||A||$ for any matrix A, (8) holds by the condition of the lemma.

For (9), Assumption 2 states that $||\Gamma_p^{-1}||_1 \leq F < \infty$ for some constant F. Similarly to Lewis and Reinsel (1985, Theorem 1) or Berk (1974, Lemma 3), we can write

$$\Gamma_p^{-1} - \hat{\Gamma}_p^{-1} = \hat{\Gamma}_p^{-1}(\hat{\Gamma}_p - \Gamma_p)\Gamma_p^{-1} = \left[\Gamma_p^{-1} + \left(\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}\right)\right](\hat{\Gamma}_p - \Gamma_p)\Gamma_p^{-1},$$

and thus,

$$\begin{aligned} ||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1 &\leq \left(||\Gamma_p^{-1}||_1 + ||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1\right)||\hat{\Gamma}_p - \Gamma_p||_1||\Gamma_p^{-1}||_1\\ &\leq \left(F + ||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1\right)||\hat{\Gamma}_p - \Gamma_p||_1 F.\end{aligned}$$

We note that $F||\hat{\Gamma}_p - \Gamma_p||_1 \to_p 0$ by the condition of the lemma. Thus, with probability approaching one, we have $F||\hat{\Gamma}_p - \Gamma_p||_1 < 1$ and

$$||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1 \leq \frac{F^2||\hat{\Gamma}_p - \Gamma_p||_1}{1 - F||\hat{\Gamma}_p - \Gamma_p||_1}.$$

The above inequality implies (9).

Lastly, (10) follows because

$$||\hat{\Gamma}_p^{-1}||_1 \leq ||\Gamma_p^{-1}||_1 + ||\hat{\Gamma}_p^{-1} - \Gamma_p^{-1}||_1 \leq F + o_p(1).$$

□

Lemma 4. *Suppose that Assumptions 1 and 2 are satisfied. If* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$ *with* $p^2/T \to 0$, *then*

$$\frac{1}{\sqrt{NT}} \ell_p' \Gamma_p^{-1} \sum_{t=p+1}^{T} w_{t-1}(p)' \epsilon_t / v_p \to_d N(0,1),$$

and

$$\frac{1}{\sqrt{NT}} \ell_p' \Gamma_p^{-1} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' \epsilon_t / v_p \to_d N(0,1).$$

Proof. Let

$$S_{Nt} = \frac{1}{\sqrt{NT}} \ell_p' \Gamma_p^{-1} w_{t-1}(p)' \epsilon_t / v_p.$$

Then, it is easy to see that

$$\frac{1}{\sqrt{NT}} \ell_p' \Gamma_p^{-1} \sum_{t=p+1}^{T} w_{t-1}(p)' \epsilon_t / v_p = \sum_{t=1}^{T} S_{Nt} - \sum_{t=1}^{p} S_{Nt}.$$

We use a central limit theorem for martingale difference sequences (e.g, Davidson's (1994) Theorem 24.3). It is easy to see that S_{Nt} is a martingale difference sequence. The theorem shows that $\sum_{t=1}^{T} S_{Nt} \to_d N(0,1)$ under the three conditions that i) $\sum_{t=1}^{T} var(S_{Nt}) = 1$, which can be verified easily in our case, ii) $\sum_{t=1}^{T} S_{Nt}^2 \to_p 1$ and iii) $\max_{1 \le i \le T} |S_{Nt}| \to_p 0$.

We first verify the condition ii). Note that Assumption 2 and the conditions on ℓ_p imply that $v_p < C$ and $v_p^{-1} < C$. By the definition of v_p, we write

$$\sum_{t=1}^{T} S_{Nt}^2 - 1 = \frac{1}{T} \frac{1}{v_p^2} \ell_p' \Gamma_p^{-1} \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \Gamma_p \right) \Gamma_p^{-1} \ell_p.$$

It holds by Assumption 2 that

$$\begin{aligned} \left\| \sum_{t=1}^{T} S_{Nt}^2 - 1 \right\| &\le \frac{1}{T} |v_p^{-2}| \cdot ||\ell_p||_1 ||\Gamma_p^{-1}||_1 \left\| \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \Gamma_p \right) \right\| \cdot ||\Gamma_p^{-1}||_1 ||\ell_p||_1 \\ &\le C \frac{1}{T} \left\| \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \Gamma_p \right) \right\|. \end{aligned}$$

We now see that

$$
\begin{aligned}
& \left\| \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \Gamma_p \right) \right\| \\
\leq & \left\| \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \frac{1}{N} w_{t-1}(p)' w_{t-1}(p) \right) \right\| \quad (11) \\
& + \left\| \sum_{t=1}^{T} \left(\sigma^2 \frac{1}{N} w_{t-1}(p)' w_{t-1}(p) - \sigma^2 \Gamma_p \right) \right\| . \quad (12)
\end{aligned}
$$

For (11), noting that $w_{t-1}(p)'\epsilon_t\epsilon_t' w_{t-1}(p) - \sigma^2 w_{t-1}(p)' w_{t-1}(p)$ is a martingale difference sequence, we have that

$$
\begin{aligned}
& E \left\| \sum_{t=1}^{T} \left(\frac{1}{N} w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^2 \frac{1}{N} w_{t-1}(p)' w_{t-1}(p) \right) \right\|^2 \\
= & \frac{1}{N^2} \sum_{t=1}^{T} tr \left(E \left(w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) - \sigma^4 w_{t-1}(p)' w_{t-1}(p) w_{t-1}(p)' w_{t-1}(p) \right) \right) .
\end{aligned}
$$

Since the observations are i.i.d. across i, we have

$$
\begin{aligned}
& E \left(w_{t-1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) w_{t_1 - 1}(p)' \epsilon_t \epsilon_t' w_{t-1}(p) \right) \\
= & E \left(\left(\sum_{i=1}^{N} w_{i,t-1}(p) \epsilon_{it} \right) \left(\sum_{i=1}^{N} w_{i,t-1}(p) \epsilon_{it} \right)' \left(\sum_{i=1}^{N} w_{i,t-1}(p) \epsilon_{it} \right) \left(\sum_{i=1}^{N} w_{i,t-1}(p) \epsilon_{it} \right)' \right) \\
= & N E(w_{i,t-1}(p) w_{i,t-1}(p)' w_{i,t-1}(p) w_{i,t-1}(p)' \epsilon_{it}^4) \\
& + 3N(N-1) E(w_{i,t-1}(p) w_{i,t-1}(p) \epsilon_{it}^2) E(w_{i,t-1}(p) w_{i,t-1}(p) \epsilon_{it}^2) \\
= & N E(w_{i,t-1}(p) w_{i,t-1}(p)' w_{i,t-1}(p) w_{i,t-1}(p)') E(\epsilon_{it}^4) + 3N(N-1) \sigma^4 \Gamma_p \Gamma_p
\end{aligned}
$$

and

$$
\begin{aligned}
& E \left(w_{t-1}(p)' w_{t-1}(p) w_{t-1}(p)' w_{t-1}(p) \right) \\
= & E \left(\left(\sum_{i=1}^{N} w_{i,t-1}(p) w_{i,t-1}(p)' \right) \left(\sum_{i=1}^{N} w_{i,t-1}(p) w_{i,t-1}(p)' \right) \right) \\
= & N E(w_{i,t-1}(p) w_{i,t-1}(p)' w_{i,t-1}(p) w_{i,t-1}(p)') + N(N-1) \Gamma_p \Gamma_p .
\end{aligned}
$$

It therefore holds that

$$\begin{aligned}
& E\left\|\sum_{t=1}^{T}\left(\frac{1}{N}w_{t-1}(p)'\epsilon_t\epsilon_t'w_{t-1}(p)-\sigma^2\frac{1}{N}w_{t-1}(p)'w_{t-1}(p)\right)\right\|^2 \\
= & \frac{T}{N}tr(E(w_{i,t-1}(p)w_{i,t-1}(p)'w_{i,t-1}(p)w_{i,t-1}(p)'))E(\epsilon_{it}^4)+\frac{3T(N-1)}{N}\sigma^4 tr(\Gamma_p\Gamma_p) \\
& -\frac{T}{N}\sigma^4 tr(E(w_{i,t-1}(p)w_{i,t-1}(p)'w_{i,t-1}(p)w_{i,t-1}(p)'))-\frac{T(N-1)}{N}\sigma^4 tr(\Gamma_p\Gamma_p) \\
= & \frac{T}{N}tr(E(w_{i,t-1}(p)w_{i,t-1}(p)'w_{i,t-1}(p)w_{i,t-1}(p)'))\left(E(\epsilon_{it}^4)-\sigma^4\right)+\frac{2T(N-1)}{N}\sigma^4 tr(\Gamma_p\Gamma_p).
\end{aligned}$$

Since

$$tr(E(w_{i,t-1}(p)w_{i,t-1}(p)'w_{i,t-1}(p)w_{i,t-1}(p)'))=E\left(\left(\sum_{k=1}^{p}w_{i,t-k}^2\right)^2\right)=O(p^2)$$

and $tr(\Gamma_p\Gamma_p)=\|\Gamma_p\|^2=O(p^2)$, we have

$$E\left\|\sum_{t=1}^{T}\left(\frac{1}{N}w_{t-1}(p)'\epsilon_t\epsilon_t'w_{t-1}(p)-\sigma^2\frac{1}{N}w_{t-1}(p)'w_{t-1}(p)\right)\right\|^2=O(p^2T).$$

Therefore, the Chebyshev inequality implies that (11) is $O(p\sqrt{T})$. Next, the proof of Lemma 2 implies that (12) is $O_p(p\sqrt{T}/\sqrt{N})$. We thus have that

$$\frac{1}{T}\left\|\sum_{t=1}^{T}\left(\frac{1}{N}w_{t-1}(p)'\epsilon_t\epsilon_t'w_{t-1}(p)-\sigma^2\Gamma_p\right)\right\|=O_p\left(\frac{p}{\sqrt{T}}\right), \tag{13}$$

which is $o_p(1)$ if $p^2/T\to 0$ so that condition ii) is satisfied.

Next, we verify condition iii). Note that for any $\delta>0$, the Chebyshev inequality gives

$$P\left(\max_{1\le t\le T}|S_{Nt}|>\delta\right)\le\sum_{t=1}^{T}P\left(|S_{Nt}|>\delta\right)\le\sum_{t=1}^{T}\frac{E\left|S_{Nt}\right|^4}{\delta^4}. \tag{14}$$

We have

$$\begin{aligned}
E|S_{Nt}|^4 &= \frac{1}{N^2T^2v_p^4}E\left(\sum_{i=1}^{N}\ell_p'\Gamma_p^{-1}w_{i,t-1}(p)\epsilon_{it}\right)^4 \\
&= \frac{1}{N^2T^2v_p^4}\sum_{i=1}^{N}E(\ell_p'\Gamma_p^{-1}w_{i,t-1}(p)\epsilon_{it})^4 \\
&\quad +\frac{6}{N^2T^2v_p^4}\sum_{i_1\neq i_2}E((\ell_p'\Gamma_p^{-1}w_{i_1,t-1}(p)\epsilon_{i_1,t})^2(\ell_p'\Gamma_p^{-1}w_{i_2,t-1}(p)\epsilon_{i_2,t})^2),
\end{aligned}$$

where the other terms are zeros because of the i.i.d. assumption. Since $v_p^2 = E((\ell_p'\Gamma_p^{-1}w_{i,t-1}(p)\epsilon_{i,t})^2)$, the i.i.d. assumption further implies that

$$\frac{6}{N^2T^2v_p^4}\sum_{i_1\neq i_2}E((\ell_p'\Gamma_p^{-1}w_{i_1,t-1}(p)\epsilon_{i_1,t})^2(\ell_p'\Gamma_p^{-1}w_{i_2,t-1}(p)\epsilon_{i_2,t})^2) = \frac{6(N-1)}{T^2N} = O\left(\frac{1}{T^2}\right).$$

For the term involving the fourth moments, it follows that

$$\frac{1}{N^2T^2v_p^4}\sum_{i=1}^{N}E(\ell_p'\Gamma_p^{-1}w_{i,t-1}(p)\epsilon_{it})^4 = \frac{E(\epsilon_{it}^4)}{N^2T^2v_p^4}\sum_{i=1}^{N}E(\ell_p'\Gamma_p^{-1}w_{i,t-1}(p))^4.$$

We now see that

$$E(\ell_p'\Gamma_p^{-1}w_{i,t-1}(p))^4 \leq ||\ell_p||_1^4||\Gamma_p^{-1}||_1^4E||w_{i,t-1}(p)||^4.$$

It holds that $||\ell_p||_1^4 = O(1)$ and $||\Gamma_p^{-1}||_1^4 = O(1)$ by Assumption 2. Moreover,

$$E||w_{i,t-1}(p)||^4 = E\left(\left(\sum_{\tau=t-p}^{t-1}w_{i,\tau}^2\right)^2\right) = E\left(\sum_{\tau_1=t-p}^{t-1}\sum_{\tau_2=t-p}^{t-1}w_{i,\tau_1}^2w_{i,\tau_2}^2\right) = O(p^2).$$

Thus, we have

$$\frac{1}{N^2T^2v_p^4}\sum_{i=1}^{N}E(\ell_p'\Gamma_p^{-1}w_{i,t-1}(p)\epsilon_{it})^4 = O\left(\frac{1}{N^2T^2}\sum_{i=1}^{N}p^2\right) = O\left(\frac{p^2}{NT^2}\right).$$

Thus, if $p^2/(NT^2) \to 0$,

$$E|S_{Nt}|^4 = O\left(\frac{p^2}{NT^2}\right) + O\left(\frac{1}{T^2}\right) = o\left(\frac{1}{T}\right)$$

so that condition iii) holds by (14). Lastly, note that $\sum_{t=1}^{p}S_{Nt} \to_p 0$ because $var(\sum_{t=1}^{p}S_{Nt}) = \sum_{t=1}^{p}var(S_{Nt}) = p/T$ and $p/T \to 0$. We thus have shown the first statement of the lemma.

Next, we consider the second statement of the lemma. We observe that

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\ell_p'\Gamma_p^{-1}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\epsilon_t/v_p \\
=\ &\frac{1}{\sqrt{NT}}\ell_p'\Gamma_p^{-1}\sum_{t=p+1}^{T-1}w_{t-1}(p)'\epsilon_t/v_p + \ell_p'\Gamma_p^{-1}\frac{1}{\sqrt{NT}}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}w_{t-1}(p)'\epsilon_t/v_p.
\end{aligned}$$

The first term of the right hand side converges to $N(0,1)$ by the first statement of the lemma. The expectation of the second term is zero and the variance is

$$
\begin{aligned}
& E\left(\ell_p'\Gamma_p^{-1}\frac{1}{\sqrt{NT}}\sum_{t=p+1}^{T-1}\frac{1}{T-t+1}w_{t-1}(p)'\epsilon_t/v_p\right)^2 \\
= & \frac{1}{NT}\frac{1}{v_p^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}E\left(\ell_p'\Gamma_p^{-1}w_{t-1}(p)'\epsilon_t\epsilon_t'w_{t-1}(p)'\Gamma_p^{-1}\ell_p\right) \\
= & \frac{1}{NT}\frac{1}{v_p^2}N\sigma^2\ell_p'\Gamma_p^{-1}\Gamma_p\Gamma_p^{-1}\ell_p\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}=O\left(\frac{1}{T}\right)=o(1).
\end{aligned}
$$

Therefore, the second term is $o_p(1)$ and the second statement of the lemma is proven. □

B The fixed effects estimator

This section presents several lemmas and the proofs of Theorems 1, 2, 3 and 4. Let $\tilde{b}_{t,p}=b_{t,p}-\sum_{t'=p+1}^{T}b_{t',p}/(T-p)$ and $\tilde{\epsilon}_t=\epsilon_t-\bar{\epsilon}_{p+1,T}$.

The estimation error of the fixed effects estimator can be decomposed as

$$\hat{\alpha}_F(p)-\alpha(p)=(\hat{\Gamma}_p^F)^{-1}F_1+(\hat{\Gamma}_p^F)^{-1}F_2$$

where

$$\hat{\Gamma}_p^F=\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{x}_t(p),\quad F_1=\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{b}_{t,p}\text{ and }F_2=\frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{\epsilon}_t.$$

Note that we can write

$$\tilde{x}_t(p)=w_{t-1}(p)-\bar{w}_{p,T-1}(p).$$

Lemma 5. *Suppose that Assumptions 1 and 2 are satisfied. If $N\to\infty$, $T\to\infty$ and $p\to\infty$ with $p^2/(NT)\to 0$ and $p^3/T^2\to 0$, then*

$$||\hat{\Gamma}_p^F-\Gamma_p||=O_p\left(\frac{p}{\sqrt{NT}}+\frac{p^{3/2}}{T}\right).$$

Proof. We observe that

$$
\begin{aligned}
\hat{\Gamma}_p^F-\Gamma_p &= \frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{x}_t(p)-\Gamma_p \\
&= \left(\frac{1}{NT}\sum_{t=p+1}^{T}w_{t-1}(p)'w_{t-1}(p)-\Gamma_p\right)-\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{w}_{p,T-1}(p).
\end{aligned}
$$

Lemma 2 states that the first term is of order $O_p(p/\sqrt{NT} + p^{3/2}/T)$. For the second term, we observe that

$$\left\| \frac{T-p}{NT} \bar{w}_{p,T-1}(p)' \bar{w}_{p,T-1}(p) \right\| \leq \frac{T-p}{NT} \left\| \bar{w}_{p,T-1}(p) \right\|^2 = O_p\left(\frac{T-p}{NT} \frac{Np}{T-p} \right) = O_p\left(\frac{p}{T} \right),$$

where the second equality follows from Lemma 1. To sum up, it holds that

$$\left\| \hat{\Gamma}_p^F - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} + \frac{p}{T} \right) = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} \right).$$

□

Lemma 6. *Suppose that Assumption 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$, then*

$$||F_1|| = O_p\left(\sqrt{p} \sum_{k=p+1}^{\infty} |\alpha_k| \right) = o_p(1).$$

Proof. We have that

$$\begin{aligned} ||F_1|| &= \left\| \frac{1}{NT} \sum_{t=p+1}^{T} \tilde{x}_t(p)' \tilde{b}_{t,p} \right\| \leq \frac{1}{NT} \sum_{t=p+1}^{T} \left\| \tilde{x}_t(p)' \tilde{b}_{t,p} \right\| \\ &\leq \frac{1}{NT} \sum_{t=p+1}^{T} \left\| \tilde{x}_t(p) \right\| \cdot \left\| \tilde{b}_{t,p} \right\|. \end{aligned}$$

We observe that

$$E \left\| \tilde{x}_t(p) \right\|^2 = NE(tr(\tilde{x}_{it}(p)\tilde{x}_{it}(p)')) = N \sum_{k=1}^{p} E\left((\tilde{x}_{it,k})^2 \right),$$

where $\tilde{x}_{it,k}$ is the kth element of $\tilde{x}_{it}(p)$. Since $E((\tilde{x}_{it,k})^2) < C$ for any k, it holds that

$$\left\| \tilde{x}_t(p) \right\| = O_p(\sqrt{Np})$$

uniformly in t. Let $w_{i,t-k}^{\dagger} = w_{i,t-k} - \bar{w}_{i,p+1-k,T-k}$. Then,

$$\begin{aligned} E \left\| \tilde{b}_{t,p} \right\|^2 &= NE((\tilde{b}_{it,p})^2) = NE\left(\left(\sum_{k=p+1}^{\infty} \alpha_k w_{i,t-k}^{\dagger} \right)^2 \right) \\ &\leq N \sum_{k=p+1}^{\infty} \sum_{k'=p+1}^{\infty} |\alpha_k| \cdot |\alpha_{k'}| \cdot |E(w_{i,t-k}^{\dagger} w_{i,t-k'}^{\dagger})| \\ &\leq CN \left(\sum_{k=p+1}^{\infty} |\alpha_k| \right)^2, \end{aligned}$$

by observing that $|E(w^{\dagger}_{i,t-k}w^{\dagger}_{i,t-k'})|$ is uniformly bounded. Therefore, we have

$$\left\|\tilde{b}_{t,p}\right\| = O_p\left(\sqrt{N}\sum_{k=p+1}^{\infty}|\alpha_k|\right).$$

To sum up, we have that

$$||F_1|| = \frac{1}{NT}\sum_{t=p+1}^{T} O_p\left(\sqrt{Np}\right) O_p\left(\sqrt{N}\sum_{k=p+1}^{\infty}|\alpha_k|\right) = O_p\left(\sqrt{p}\sum_{k=p+1}^{\infty}|\alpha_k|\right).$$

□

Lemma 7. *Suppose that Assumption 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p/T \to 0$, we have*

$$\|B\| = O\left(\frac{\sqrt{p}}{T}\right),$$

$$\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}\right\| = O_p\left(\frac{\sqrt{p}}{T}\right)$$

and

$$\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right\| = O_p\left(\frac{\sqrt{p}}{\sqrt{N}T}\right).$$

Proof. We note that

$$\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} = \frac{1}{NT(T-p)}\sum_{t=p+1}^{T}\sum_{m=p+1}^{T} w_{t-1}(p)'\epsilon_m.$$

We observe that $E(w_{t-1}(p)'\epsilon_m) = 0$ if $t-1 < m$. Let $\psi_k(p-) = (\psi_k, \ldots, \psi_{k+p-1})'$. Since $w_{t-1} = \sum_{k=0}^{\infty}\psi_k\epsilon_{t-1-k}$, we have $E(w_{t-1}(p)'\epsilon_m) = N\sigma^2\psi^{-}_{t-1-m}(p)$ if $t-1 \geq m$. Thus, we have that

$$B = E\left(\frac{1}{NT}\sum_{t=p+1}^{T} w_{t-1}(p)'\bar{\epsilon}_{p+1,T}\right) = \frac{1}{T(T-p)}\sum_{t=p+1}^{T}\sum_{m=p+1}^{t-1}\sigma^2\psi^{-}_{t-1-m}(p).$$

We observe that

$$\begin{aligned}
\|B\|^2 = tr(BB') &= \sigma^4\frac{1}{T^2(T-p)^2}\sum_{k=0}^{p-1}\left(\sum_{t=p+1}^{T}\sum_{m=p+1}^{t-1}\sigma^2\psi_{t-1-m+k}\right)^2 \\
&\leq \sigma^4\frac{1}{T^2(T-p)^2}\sum_{k=0}^{p-1}\left(\sum_{t=p+1}^{T}\sum_{m=0}^{\infty}\sigma^2|\psi_m|\right)^2 \\
&= O\left(\frac{p}{T^2}\right). \qquad (15)
\end{aligned}$$

Therefore, we have $\|B\| = O(\sqrt{p}/T)$.

Next, we examine

$$
\begin{aligned}
E\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right\|^2 &= tr\left(var\left(\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}\right)\right) \\
&= \frac{1}{N}\left(\frac{T-p}{T}\right)^2 tr\left(var(\bar{w}_{i,p,T-1}(p)\bar{\epsilon}_{i,p+1,T})\right) \\
&= \frac{1}{N}\left(\frac{T-p}{T}\right)^2 \sum_{k=0}^{p-1} var(\bar{w}_{i,p-k,T-1-k}\bar{\epsilon}_{i,p+1,T}).
\end{aligned}
$$

We see that $((T-p)/T)^2 = O(1)$. We also see that

$$
\begin{aligned}
var(\bar{w}_{i,p-k,T-1-k}\bar{\epsilon}_{i,p+1,T}) &\leq E(\bar{w}^2_{i,p-k,T-1-k}\bar{\epsilon}^2_{i,p+1,T}) \\
&\leq \sqrt{E(\bar{w}^4_{i,p-k,T-1-k})}\sqrt{E(\bar{\epsilon}^4_{i,p+1,T})}.
\end{aligned}
$$

It holds that

$$
\begin{aligned}
E(\bar{w}^4_{i,p-k,T-1-k}) &= \frac{1}{(T-p)^4}\sum_{t_1=p-k}^{T-1-k}\sum_{t_2=p-k}^{T-1-k}\sum_{t_3=p-k}^{T-1-k}\sum_{t_4=p-k}^{T-1-k} E(w_{i,t_1}w_{i,t_2}w_{i,t_3}w_{i,t_4}) \\
&= \frac{3}{(T-p)^4}\left(\sum_{t_1=p-k}^{T-1-k}\sum_{t_2=p-k}^{T-1-k} E(w_{i,t_1}w_{i,t_2})\right)^2 \\
&\quad + \frac{1}{(T-p)^4}\sum_{t_1=p-k}^{T-1-k}\sum_{t_2=p-k}^{T-1-k}\sum_{t_3=p-k}^{T-1-k}\sum_{t_4=p-k}^{T-1-k} \kappa_w(t_1,t_2,t_3,t_4) \\
&= O\left(\frac{1}{T^2}\right)
\end{aligned}
$$

by Assumption 1 and the argument given in footnote 15. Moreover,

$$
E(\bar{\epsilon}^4_{i,p+1,T}) = \frac{1}{(T-p)^4}((T-p)E(\epsilon^4_{it}) + 3(T-p)(T-p-1)\sigma^4) = O\left(\frac{1}{T^2}\right).
$$

It therefore follows that

$$
E\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right\|^2 = O\left(\frac{1}{N}\sum_{k=0}^{p-1}\sqrt{\frac{1}{T^2}\frac{1}{T^2}}\right) = O\left(\frac{p}{NT^2}\right). \qquad (16)
$$

Therefore, the Chebyshev inequality shows that

$$
\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right\| = O_p\left(\frac{\sqrt{p}}{\sqrt{N}T}\right).
$$

Lastly, by (16) and (15), we have

$$E\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}\right\|^2 = tr\left(var\left(\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}\right)\right)+tr\left(BB'\right)$$
$$= O\left(\frac{p}{NT^2}\right)+O\left(\frac{p}{T^2}\right)=O\left(\frac{p}{T^2}\right).$$

The Chebyshev inequality gives

$$\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}\right\| = O_p\left(\frac{\sqrt{p}}{T}\right).$$

□

Lemma 8. *Suppose that Assumptions 1 and 2 are satisfied. If $N\to\infty$, $T\to\infty$ and $p\to\infty$ with $p/T\to 0$, then*

$$||F_2|| = O_p\left(\sqrt{\frac{p}{NT}}+\frac{\sqrt{p}}{T}\right) = o_p(1)$$

and

$$||F_2+B|| = O_p\left(\sqrt{\frac{p}{NT}}\right) = o_p(1).$$

Proof. We observe that

$$F_2 = \frac{1}{NT}\sum_{t=p+1}^{T}\tilde{x}_t(p)'\tilde{\epsilon}_t = \frac{1}{NT}\sum_{t=p+1}^{T}w_{t-1}(p)'\epsilon_t - \frac{1}{NT}\sum_{t=p+1}^{T}w_{t-1}(p)'\bar{\epsilon}_{p+1,T}$$
$$-\frac{1}{NT}\sum_{t=p+1}^{T}\bar{w}_{p,T-1}(p)'\epsilon_t + \frac{1}{NT}\sum_{t=p+1}^{T}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}$$
$$= \frac{1}{NT}\sum_{t=p+1}^{T}w_{t-1}(p)'\epsilon_t - \frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T}.$$

Noting that $w_{t-1}(p)'\epsilon_t$ is a martingale difference sequence, we have

$$E\left\|\frac{1}{NT}\sum_{t=p+1}^{T}w_{t-1}(p)'\epsilon_t\right\|^2 = \frac{1}{N^2T^2}\sum_{t=p+1}^{T}\sum_{t'=p+1}^{T}tr(E(w_{t-1}(p)'\epsilon_t\epsilon_{t'}'w_{t'-1}(p)))$$
$$= \frac{1}{N^2T^2}\sum_{t=p+1}^{T}\sigma^2 tr(E(w_{t-1}(p)'w_{t-1}(p))).$$

Since $tr(E(w_{t-1}(p)'w_{t-1}(p))) = E||w_{t-1}(p)||^2 = O(Np)$ by Lemma 1, we have

$$E\left\|\frac{1}{NT}\sum_{t=p+1}^{T} w_{t-1}(p)'\epsilon_t\right\|^2 = O\left(\frac{1}{N^2T^2}\sum_{t=p+1}^{T} Np\right) = O\left(\frac{p}{NT}\right).$$

Therefore, the Markov inequality gives

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T} w_{t-1}(p)'\epsilon_t\right\| = O_p\left(\sqrt{\frac{p}{NT}}\right).$$

Hence, Lemma 7 implies that

$$||F_2|| = O_p\left(\sqrt{\frac{p}{NT}} + \frac{\sqrt{p}}{T}\right)$$

and

$$||F_2 + B|| = O_p\left(\sqrt{\frac{p}{NT}}\right) + O_p\left(\frac{\sqrt{p}}{\sqrt{N}T}\right) = O_p\left(\sqrt{\frac{p}{NT}}\right).$$

□

Lemma 9. *Suppose that Assumptions 1 and 2 are satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/T \to 0$, then*

$$\sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B)/v_p \to_d N(0,1).$$

Proof. As in the proof of Lemma 8, we observe that

$$\begin{aligned}\sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B) &= \frac{1}{\sqrt{NT}}\ell_p'\Gamma_p^{-1}\sum_{t=p+1}^{T} w_{t-1}(p)'\epsilon_t \\ &\quad -\sqrt{NT}\ell_p'\Gamma_p^{-1}\left(\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right).\end{aligned}$$

Lemma 4 gives

$$\frac{1}{\sqrt{NT}}\ell_p'\Gamma_p^{-1}\sum_{t=p+1}^{T} w_{t-1}(p)'\epsilon_t/v_p \to_d N(0,1).$$

Lemma 7 and Assumption 2 imply that

$$\begin{aligned}&\left\|\sqrt{NT}\ell_p'\Gamma_p^{-1}\left(\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right)\right\| \\ \le\ & \sqrt{NT}\,\|\ell_p\|_1\,\left\|\Gamma_p^{-1}\right\|\left\|\frac{T-p}{NT}\bar{w}_{p,T-1}(p)'\bar{\epsilon}_{p+1,T} - B\right\| \\ =\ & O_p\left(\sqrt{NT}\frac{\sqrt{p}}{\sqrt{N}T}\right) = O_p\left(\sqrt{\frac{p}{T}}\right) = o_p(1).\end{aligned}$$

Since $v_p^{-1} = O(1)$ by Assumption 2, the result follows. □

B.1 Proof of Theorem 1

Proof. We have

$$||\hat{\alpha}_F(p) - \alpha(p)|| = ||(\hat{\Gamma}_p^F)^{-1}(F_1 + F_2)|| \leq ||(\hat{\Gamma}_p^F)^{-1}||_1||F_1|| + ||(\hat{\Gamma}_p^F)^{-1}||_1||F_2||.$$

Lemmas 3 and 5 give that $||(\hat{\Gamma}_p^F)^{-1}||_1 = O_p(1)$. Lemma 6 gives that $||F_1|| = o_p(1)$. Lastly, $||F_2|| = o_p(1)$ follows by Lemma 8. □

B.2 Proof of Theorem 2

Proof. We note that

$$\begin{aligned} & \sqrt{NT}(\ell_p'\hat{\alpha}_F(p) - \ell_p'\alpha(p) + \ell_p'\Gamma_p^{-1}B) \\ = \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1 + \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_2 + \sqrt{NT}\ell_p'\Gamma_p^{-1}B \\ = \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1 + \sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})F_2 + \sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B). \end{aligned}$$

Lemma 9 gives

$$\sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B)/v_p \to_d N(0,1).$$

Next, we consider

$$||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1|| \leq ||\ell_p||_1||\sqrt{NT}(\hat{\Gamma}_p^F)^{-1}F_1|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^F)^{-1}||_1||\sqrt{NT}F_1||.$$

We have $||\ell_p||_1 = O(1)$ by the assumption. $||(\hat{\Gamma}_p^F)^{-1}||_1 = O_p(1)$ by Lemmas 3 and 5. $||\sqrt{NT}F_1|| = o_p(1)$ by Lemma 6 because $\sqrt{NTp}\sum_{k=p+1}^{\infty}|\alpha_k| \to 0$. Therefore, we have $||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1|| = o_p(1)$.

Lastly, we see that

$$||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})F_2|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1}||_1||\sqrt{NT}F_2||.$$

We have $||\ell_p||_1 = O(1)$, $||(\hat{\Gamma}_p^F)^{-1} - \hat{\Gamma}_p^{-1}||_1 = O_p(p/\sqrt{NT} + p^{3/2}/T)$ by Lemmas 3 and 5 and $||\sqrt{NT}F_2|| = O_p(\sqrt{p} + \sqrt{p/(NT)})$ by Lemma 8. Therefore, we have $||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})F_2|| = O_p(p^{3/2}/\sqrt{NT} + p^{3/2}/T + p^2/T + p^2\sqrt{N}/T^{3/2}) = O_p(p^{3/2}/\sqrt{NT} + p^2/T + p^2\sqrt{N}/T^{3/2})$, which is $o_p(1)$ under the assumption of the theorem.

□

B.3 Proof of Theorem 3

Proof. We have

$$\begin{aligned} ||\hat{\alpha}_{BF}(p) - \alpha(p)|| &= ||(\hat{\Gamma}_p^F)^{-1}(F_1 + F_2 + \hat{B})|| \\ &\leq ||(\hat{\Gamma}_p^F)^{-1}||_1(||F_1|| + ||F_2 + B|| + ||\hat{B} - B||). \end{aligned}$$

Lemmas 3 and 5 give that $||(\hat{\Gamma}_p^F)^{-1}||_1 = O_p(1)$. Lemma 6 gives that $||F_1|| = o_p(1)$. By Lemma 8, we have $||F_2 + B|| = O_p(\sqrt{p/(NT)})) = o_p(1)$. The assumption of the theorem gives $||\hat{B} - B|| = o_p(1)$. □

B.4 Proof of Theorem 4

Proof. We note that

$$\begin{aligned} \sqrt{NT}(\ell_p'\hat{\alpha}_{BF}(p) - \ell_p'\alpha(p)) &= \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}(F_1 + F_2 + \hat{B}) \\ &= \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1 + \sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})(F_2 + B) \\ &\quad + \sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B) + \sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}(\hat{B} - B). \end{aligned}$$

Similarly to the proof of Theorem 2, we have $\sqrt{NT}\ell_p'\Gamma_p^{-1}(F_2 + B)/v_p \to_d N(0,1)$ by Lemma 9 and $||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}F_1|| = o_p(1)$ by Lemmas 3, 5 and 6. We also have

$$||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})(F_2 + B)|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1}||_1||\sqrt{NT}(F_2 + B)||.$$

We have $||\ell_p||_1 = O(1)$, $||(\hat{\Gamma}_p^F)^{-1} - \hat{\Gamma}_p^{-1}||_1 = O_p(p/\sqrt{NT} + p^{3/2}/T)$ by Lemmas 3 and 5 and $||\sqrt{NT}(F_2 + B)|| = O_p(\sqrt{p})$ by Lemma 8. Therefore, we have $||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^F)^{-1} - \Gamma_p^{-1})(F_2 + B)|| = O_p(p^{3/2}/\sqrt{NT} + p^2/T)$, which is $o_p(1)$ under the assumption of the theorem. Lastly, we have

$$||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^F)^{-1}(\hat{B} - B)|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^F)^{-1}||_1||\sqrt{NT}(\hat{B} - B)|| = O_p(\sqrt{NT}\nu_{NTp}) = o_p(1).$$

□

C The GMM estimator

This section presents several lemmas and the proofs of Theorems 5 and 6. Note that variables with superscript "*" are transformed by the forward filter so that $b_{t,p}^* = \sqrt{(T-t)/(T-t+1)}(b_{t,p} - \sum_{\tau=t+1}^T b_{\tau,p}/(T-t))$ and $\epsilon_t^* = \sqrt{(T-t)/(T-t+1)}(\epsilon_t - \sum_{\tau=t+1}^T \epsilon_t/(T-t))$. The estimation error of the GMM estimator can be decomposed as

$$\hat{\alpha}(p) - \alpha(p) = (\hat{\Gamma}_p^G)^{-1}G_1 + (\hat{\Gamma}_p^G)^{-1}G_2$$

where

$$\hat{\Gamma}_p^G = \frac{1}{NT}\sum_{t=p+1}^{T-1} x_t^*(p)' M_t x_t^*(p), \quad G_1 = \frac{1}{NT}\sum_{t=p+1}^{T-1} x_t^*(p)' M_t b_{t,p}^* \text{ and } G_2 = \frac{1}{NT}\sum_{t=p+1}^{T-1} x_t^*(p)' M_t \epsilon_t^*.$$

Note that we can write

$$x_t^*(p) = \sqrt{\frac{T-t}{T-t+1}}\left(w_{t-1}(p) - \bar{w}_{t,T-1}(p)\right).$$

Lemma 10. *Suppose that Assumptions 1, 2 and 3 are satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/T \to 0$, then*

$$||\hat{\Gamma}_p^G - \Gamma_p|| = O_p\left(\frac{p}{\sqrt{T}}\right)$$

Proof. We observe that

$$\begin{aligned}
\hat{\Gamma}_p &= \frac{1}{NT}\sum_{t=p+1}^{T-1} x_t^*(p)' M_t x_t^*(p) \\
&= \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' M_t w_{t-1}(p) - \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' M_t \bar{w}_{t,T-1}(p) \\
&\quad - \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{t,T-1}(p)' M_t w_{t-1}(p) + \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{t,T-1}(p)' M_t \bar{w}_{t,T-1}(p).
\end{aligned}$$

The first term in the decomposition is that

$$\begin{aligned}
&\frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' M_t w_{t-1}(p) \\
= &\frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' w_{t-1}(p) + \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)'(I - M_t) w_{t-1}(p).
\end{aligned}$$

Lemma 2 gives

$$\left\| \frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' w_{t-1}(p) - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T}\right).$$

We consider the term

$$\frac{1}{NT}\sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)'(I - M_t) w_{t-1}(p).$$

Let η_t^* is the $N \times 1$ vector of the errors of the population linear projection of η on Z_t so that $\eta_t^* = \eta - Z_t'\delta_t$, where

$$\delta_t = E((z_{it}z_{it}'))^{-1}E(z_{it}\eta_i) = (\sigma_\eta^2 \iota_t \iota_t + \Gamma_t)^2 \sigma_\eta^2 \iota_t = \frac{\sigma_\eta^2}{1 + \sigma_\eta^2 \iota_t' \Gamma_t^{-1} \iota_t} \Gamma_t^{-1} \iota.$$

We have that

$$E(\eta_t^{*\prime}\eta_t) = N(\sigma_\eta^2 - \delta_t' E(z_{it}\eta_i)) = N\left(\frac{\sigma_\eta^2}{1 + \sigma_\eta^2 \iota_t' \Gamma_t^{-1} \iota_t}\right) = \left(\frac{N}{t}\right),$$

where $1/(1 + \iota_t'\Gamma_t^{-1}\iota_t) = O(1/t)$ by Assumption 2. Since $w_{t-1}(p) = x_{t-1}(p) - \eta\iota_p'$ and $x_{t-1}(p)(I - M_t) = 0$, we have that

$$\begin{aligned}
& \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)'(I - M_t) w_{t-1}(p) \right\| \\
= & \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \iota_p \eta_t^{*\prime} (I - M_t) \eta_t^* \iota_p \right\| \\
\leq & \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \left\| \iota_p \eta_t^{*\prime} (I - M_t) \eta_t^* \iota_p \right\| \\
\leq & \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \lambda_{\max}(I - M_t) \left\| \eta_t^* \iota_p' \right\|^2.
\end{aligned}$$

We observe that $\lambda_{\max}(I - M_t) = 1$ because $I - M_t$ is an idempotent matrix and that $\left\|\eta_t^*\iota_p'\right\|^2 \leq \|\eta_t^*\|^2 \cdot \|\iota_p\|^2 = O_p(Np/t)$. Thus, it holds that

$$\begin{aligned}
\left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)'(I - M_t) w_{t-1}(p) \right\| &= O_p\left(\frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \frac{Np}{t} \right) \\
&= O_p\left(\frac{p \log T}{T} \right).
\end{aligned}$$

Next, we have

$$\begin{aligned}
& \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' M_t \bar{w}_{t,T-1}(p) \right\| \\
\leq & \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \left\| w_{t-1}(p)' M_t \bar{w}_{t,T-1}(p) \right\| \\
\leq & \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \lambda_{\max}(M_t) \left\| w_{t-1}(p) \right\| \cdot \left\| \bar{w}_{t,T-1}(p) \right\|.
\end{aligned}$$

Since M_t is an idempotent matrix, we have $\lambda_{\max}(M_t) = 1$. Lemma 1 gives that $\|w_{t-1}(p)\| = O_p((Np)^{1/2})$ and $\|\bar{w}_{t,T-1}(p)\| = O_p(\sqrt{Np/(T-t)})$. Thus, we have

$$\begin{aligned} & \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' M_t \bar{w}_{t,T-1}(p) \right\| \\ = \; & O_p\left(\frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \sqrt{Np} \sqrt{Np \frac{1}{T-t}} \right) = O_p\left(\frac{p}{\sqrt{T}} \right). \end{aligned}$$

Similarly, we have

$$\left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{t,T-1}(p)' M_t w_{t-1}(p) \right\| = O_p\left(\frac{p}{\sqrt{T}} \right).$$

Lastly, we have

$$\begin{aligned} \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{t,T-1}(p)' M_t \bar{w}_{t,T-1}(p) \right\| \; & \leq \; \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \left\| \bar{w}_{t,T-1}(p)' M_t \bar{w}_{t,T-1}(p) \right\| \\ & \leq \; \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \lambda_{\max}(M_t) \left\| \bar{w}_{t,T-1}(p) \right\|^2 . \end{aligned}$$

Since M_t is an idempotent matrix, we have $\lambda_{\max}(M_t) = 1$. Lemma 1 implies that $\|\bar{w}_{t,T-1}(p)\|^2 = O_p(Np/(T-t))$. Thus, we have

$$\begin{aligned} & \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{t-1,T-1}(p)' M_t \bar{w}_{t-1,T-1}(p) \right\| \\ = \; & O_p\left(\frac{1}{NT} \sum_{t=p+1}^{T-1} \frac{T-t}{T-t+1} Np \frac{1}{T-t} \right) = O_p\left(\frac{p \log T}{T} \right). \end{aligned}$$

To sum up, noting that $p \log /T = o(p/\sqrt{T})$, it holds that

$$\left\| \hat{\Gamma}_p^G - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} + \frac{p \log T}{T} + \frac{p}{\sqrt{T}} + \frac{p \log T}{T} \right) = O_p\left(\frac{p}{\sqrt{T}} \right).$$

□

Lemma 11. *Suppose that Assumption 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$, then*

$$||G_1|| = O_p\left(\sqrt{p} \sum_{k=p+1}^{\infty} |\alpha_k| \right) = o_p(1).$$

Proof. We have that

$$
\begin{aligned}
||G_1|| &= \left\| \frac{1}{NT} \sum_{t=p+1}^{T-1} x_t^*(p)' M_t b_{t,p}^* \right\| \leq \frac{1}{NT} \sum_{t=p+1}^{T-1} \left\| x_t^*(p)' M_t b_{t,p}^* \right\| \\
&\leq \frac{1}{NT} \sum_{t=p+1}^{T-1} \lambda_{\max}(M_t) \left\| x_t^*(p) \right\| \cdot \left\| b_{t,p}^* \right\|.
\end{aligned}
$$

Since M_t is an idempotent matrix, $\lambda_{\max}(M_t) = 1$. We also let

$$
w_{i,t-k,t}^{\dagger} = \sqrt{\frac{T-t}{T-t+1}} \left(w_{i,t-k} - \bar{w}_{i,t-k+1,T-k} \right).
$$

Then, we write $x_{it}^*(p) = (w_{i,t-1,1}^{\dagger}, \ldots, w_{i,t-p,1}^{\dagger})'$ and $b_{i,t,p}^* = \sum_{k=p+1}^{\infty} \alpha_k w_{i,t-k,t}^{\dagger}$. We have that

$$
E \left\| x_t^*(p) \right\|^2 = NE(tr(x_{it}^*(p) x_{it}^*(p)')) = N \sum_{k=1}^{p} E\left(\left(w_{i,t-k,t}^{\dagger} \right)^2 \right) = NpE((w_{i,t-k,t}^{\dagger})^2)
$$

by the stationarity of $y_{i,t-k,t}^{\dagger}$. Since $E((y_{i,t-k,t}^{\dagger})^2)$ is uniformly bounded, we have that

$$
\left\| x_t^*(p) \right\| = O_p(\sqrt{Np})
$$

uniformly in t. We also have

$$
\begin{aligned}
E \left\| b_{t,p}^* \right\|^2 &= NE((b_{it,p}^*)^2) = NE\left(\left(\sum_{k=p+1}^{\infty} \alpha_k w_{i,t-k,t}^{\dagger} \right)^2 \right) \\
&\leq N \sum_{k=p+1}^{\infty} \sum_{k'=p+1}^{\infty} |\alpha_k| \cdot |\alpha_{k'}| \cdot |E(w_{i,t-k,t}^{\dagger} w_{i,t-k',t}^{\dagger})| \\
&\leq CN \left(\sum_{k=p+1}^{\infty} |\alpha_k| \right)^2,
\end{aligned}
$$

by observing that $|E(w_{i,t-k,t}^{\dagger} w_{i,t-k',t}^{\dagger})|$ is uniformly bounded. Therefore, we have

$$
\left\| b_{t,p}^* \right\| = O_p\left(\sqrt{N} \sum_{k=p+1}^{\infty} |\alpha_k| \right). \tag{17}
$$

To sum up, we have that

$$
||G_1|| = \frac{1}{NT} \sum_{t=p+1}^{T-1} O_p\left(\sqrt{Np}\right) O_p\left(\sqrt{N} \sum_{k=p+1}^{\infty} |\alpha_k| \right) = O_p\left(\sqrt{p} \sum_{k=p+1}^{\infty} |\alpha_k| \right).
$$

□

Lemma 12. *Suppose that Assumption 1 is satisfied. Let d_t be the $N \times 1$ vector containing the diagonal elements of M_t. Define κ_3 and κ_4 be the third and fourth cumulants of ϵ_{it}, respectively. Suppose that $r \geq t$, $q \geq s$ and $t \geq s$. Then, $cov(\epsilon_l' M_t \epsilon_r, \epsilon_p' M_s \epsilon_q)$ is equal to*

$$\begin{aligned} 2\sigma^4 s + \kappa_4 E(d_t' d_s) \leq (2\sigma^4 + \kappa_4)s \quad & \text{if} \quad l = r = p = q, \\ \kappa_3 E(d_t' M_s \epsilon_q) \quad & \text{if} \quad l = r = p \neq q < t, \\ \kappa_3 E(d_s' M_t \epsilon_l) \quad & \text{if} \quad r = p = q \neq l < t, \\ \kappa_3 E(d_t' M_s \epsilon_p) \quad & \text{if} \quad l = r = q \neq p < t, \\ \sigma^4 s \quad & \text{if} \quad s \leq l = p \neq r = q \text{ or } l = q \neq r = p, \\ \sigma^2 E(\epsilon_l' M_s \epsilon_p) \quad & \text{if} \quad r = q, l < s \text{ and } p < s, \end{aligned}$$

where $|E(d_t' M_s \epsilon_q)| < \sqrt{st}\sigma$ if $q < t$, $|E(d_s' M_t \epsilon_l)| \leq (sN)^{1/2}\sigma$, $|E(d_t' M_s \epsilon_p)| < \sqrt{st}\sigma$ if $s \leq p < t$ and $|E(d_t' M_s \epsilon_p)| < \sqrt{sN}\sigma$ if $p < s$ and $|E(\epsilon_l' M_s \epsilon_p)| \leq N\sigma$ if $l < s$ and $p < s$.

Proof. Alvarez and Arellano (2003, Lemma C1) show that for $l \geq r$ and $p \geq q$,

$$cov(\epsilon_l' M_t \epsilon_r, \epsilon_p' M_s \epsilon_q) = \begin{cases} 2\sigma^4 s + \kappa_4 E(d_t' d_s) \leq (2\sigma^4 + \kappa_4)s & \text{if } l = r = p = q, \\ \kappa_3 E(d_t' M_s \epsilon_q) & \text{if } l = r = p \neq q < t, \\ \sigma^4 s & \text{if } l = p \neq r = q, \\ 0 & \text{otherwise,} \end{cases}$$

where $|E(d_t' M_s \epsilon_q)| \leq \sqrt{st}\sigma$.

Therefore, we consider cases with $l < r$ or $p < q$. We note that in these cases $cov(\epsilon_l' M_t \epsilon_r, \epsilon_p' M_s \epsilon_q) = E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q)$ because $E(\epsilon_l' M_t \epsilon_r) = 0$if $l < r$, and $E(\epsilon_p' M_t \epsilon_q) = 0$ if $p < q$.

Let E_t be the conditional expectation operator given $(\epsilon_{t-1}, \dots, \epsilon_0, \dots)$.

First, we consider the case in which $l < r = p = q$. Noting that $r \geq t$, it follows that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = E(\epsilon_l' M_t \epsilon_r \epsilon_r' M_s \epsilon_r) = E\left(tr\left(\epsilon_l M_t E_t(\epsilon_r \epsilon_r' M_s \epsilon_r)\right)\right) = \kappa_3 E(\epsilon_l' M_t d_s).$$

We have that $E(\epsilon_l' M_t d_s) = 0$ if $l \geq t$. If $l < t$, then, we have that

$$(E(\epsilon_l' M_t d_s))^2 \leq E((\epsilon_l' M_t d_s)^2) \leq E(d_s' d_s \epsilon_l' M_t \epsilon_l) = sE(\epsilon_l' M_t \epsilon_l) \leq \sigma^2 sN.$$

Next, we suppose that $p < l = r = q$. Noting that $r \geq t$, it follows that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = E(\epsilon_r' M_t \epsilon_r \epsilon_p' M_s \epsilon_r) = E\left(tr\left(E_t(\epsilon_r M_t \epsilon_r \epsilon_r') M_s \epsilon_p\right)\right) = \kappa_3 E(d_t M_s \epsilon_p).$$

We have that $E(d_t' M_s \epsilon_p) = 0$ if $p \geq t$. If $p < t$, then, we have that

$$(E(d_t' M_s \epsilon_p))^2 \leq E((d_t' M_s \epsilon_p)^2) \leq E(d_t' d_t \epsilon_p' M_s \epsilon_p) = tE(\epsilon_p' M_s \epsilon_p) \leq \begin{cases} \sigma^2 st & \text{if } p \geq s, \\ \sigma^2 sN & \text{if } p < s. \end{cases}$$

For the case with $l = p = q < r$ and the case with $l = r = p < q$, it is obvious that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0.$$

We now suppose that $l = p < r = q$. We observe that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = E(\epsilon_l' M_t \epsilon_r \epsilon_l' M_s \epsilon_r) = E\left(tr\left(M_t E_r(\epsilon_r \epsilon_r') M_s \epsilon_l \epsilon_l'\right)\right) = \sigma^2 E\left(tr(M_s \epsilon_l \epsilon_l')\right)$$

If $l \geq s$, then $\sigma^2 E\left(tr(M_s \epsilon_l \epsilon_l')\right) = \sigma^4 s$. If $l < s$, then $\sigma^2 E\left(tr(M_s \epsilon_l \epsilon_l')\right) = \sigma^2 E(\epsilon_l' M_s \epsilon_l) \leq \sigma^2 E(\epsilon_l' \epsilon_l) \leq \sigma^4 N$.

We examine the case in which $l = q < r = p$. Noting that $l = q \geq s$, we have

$$E(\epsilon_l' M_t \epsilon_r \epsilon_r' M_s \epsilon_l) = E\left(tr\left(M_t E_r(\epsilon_r \epsilon_r') M_s \epsilon_l \epsilon_l'\right)\right) = \sigma^2 E\left(tr(M_s \epsilon_l \epsilon_l')\right) = \sigma^4 s.$$

The case in which $l = q > r = p$ can be considered similarly. Noting that $l > t$, we have

$$E(\epsilon_l' M_t \epsilon_r \epsilon_r' M_s \epsilon_l) = E\left(tr\left(M_t \epsilon_r \epsilon_r' M_s E_l(\epsilon_l \epsilon_l')\right)\right) = \sigma^2 E\left(tr(M_s \epsilon_r \epsilon_r')\right) = \sigma^4 s.$$

Consider the case in which $l = r$, $l \neq p$, $l \neq q$ and $p < q$. If $q > l$, it is easy to see that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0.$$

On the other hand, if $l > q$, then

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = trE(E_t(\epsilon_l' M_t \epsilon_l) \epsilon_p' M_s \epsilon_q) = \sigma^2 t E(\epsilon_p' M_s \epsilon_q) = 0,$$

where the third equality follows because $l = r \geq t$.

Similarly, if $p = q$, $l < r$, $l \neq p$ and $r \neq p$, then $E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0$.

For the case with $l = p$, $l \neq r$, $l \neq q$ and $r \neq q$, we have

$$\begin{aligned} E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = E(\epsilon_l' M_t \epsilon_r \epsilon_l' M_s \epsilon_q) &= \begin{cases} 0 & \text{if } r > l \text{ or } q > l, \\ tr(E(M_t \epsilon_r \epsilon_q' M_s \epsilon_l \epsilon_l')) & \text{if } l > r \text{ and } l > q, \end{cases} \\ &= \begin{cases} 0 & \text{if } r > l \text{ or } q > l, \\ \sigma^2 tr(E(M_s \epsilon_r \epsilon_q')) & \text{if } l > r \text{ and } l > q, \end{cases} \\ &= 0, \end{aligned}$$

where the third equality follows by noting $l > t$ if $l > r$.

Similarly, we have $E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0$ for the case with $l = q$, $l \neq r$, $l \neq p$ and $r \neq p$, and for the case $r = p$, $l \neq p$, $l \neq q$ and $r \neq q$, $E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0$.

For the case that $r = q$, $l \neq r$, $l \neq p$ and $r \neq p$, we have that

$$\begin{aligned} E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_r) &= \begin{cases} 0 & \text{if } l > r \text{ or } p > r, \\ tr(E(M_t \epsilon_l \epsilon_p' M_s \epsilon_r \epsilon_r')) & \text{if } r > l \text{ and } r > p, \end{cases} \\ &= \begin{cases} 0 & \text{if } l > r \text{ or } p > r, \\ \sigma^2 tr(E(M_s \epsilon_l \epsilon_p')) & \text{if } r > l \text{ and } r > p. \end{cases} \end{aligned}$$

If $l \geq s$ or $p \geq s$, then $tr(E(M_s \epsilon_l \epsilon_p')) = E(\epsilon_l' M_s \epsilon_p) = 0$. However, if $l < s$ and $p < s$, then $tr(E(M_s \epsilon_l \epsilon_p')) = E(\epsilon_l' M_s \epsilon_p)$ may not be zero but $|E(\epsilon_l' M_s \epsilon_p)| \leq \sigma^2 N$.

Lastly, when l, q, r, p are all different, it holds that

$$E(\epsilon_l' M_t \epsilon_r \epsilon_p' M_s \epsilon_q) = 0.$$

□

Lemma 13. *Suppose that Assumption 1 is satisfied. Suppose that $r \geq t$, $q \geq s$ and $t \geq s$. Then, it holds that, if $r = q$,*

$$\begin{aligned} & cov(w_l' M_t \epsilon_r, w_p' M_s \epsilon_q) \\ = \; & O\left(\sqrt{st} + \sqrt{sN}\right) + \sigma^2 \sum_{k=l-s+1}^{\infty} \sum_{m=p-s+1}^{\infty} \psi_k \psi_m E(\epsilon_{l-k}' M_s \epsilon_{p-m}), \end{aligned}$$

where $E(\epsilon_{l-k}' M_s \epsilon_{p-m}) < \sigma N$, if $r \neq q$ and $q \geq t$,

$$cov(w_l' M_t \epsilon_r, w_p' M_s \epsilon_q) = \psi_{l-q} \psi_{p-r} (2\sigma^4 s + \kappa_4 E(d_t' d_s)) = O(s),$$

and, if $q < t$,

$$cov(w_l' M_t \epsilon_r, w_p' M_s \epsilon_q) = \psi_{l-r} \psi_{p-r} \kappa_3 E(d_t' M_s \epsilon_q) + s\sigma^4 \psi_{l-q} \psi_{p-r} = O(\sqrt{st}).$$

Proof. We first consider cases with $r = q$. Lemma 12 implies that

$$\begin{aligned} & cov(w_l' M_t \epsilon_r, w_p' M_s \epsilon_r) \\ = \; & \psi_{l-r} \psi_{p-r} cov(\epsilon_r' M_t \epsilon_r, \epsilon_r' M_s \epsilon_r) + \sum_{k=0}^{\infty} \psi_k \psi_{p-r} cov(\epsilon_{l-k}' M_t \epsilon_r, \epsilon_r' M_s \epsilon_r) \\ & + \psi_{l-r} \sum_{m=0}^{\infty} \psi_m cov(\epsilon_r' M_t \epsilon_r, \epsilon_{p-m}' M_s \epsilon_r) + \sum_{k=0, k \neq l-r}^{\infty} \psi_k \psi_{p-l+k} cov(\epsilon_{l-k}' M_t \epsilon_r, \epsilon_{l-k}' M_s \epsilon_r) \\ & + \sum_{k=l-s+1}^{\infty} \sum_{m=p-s+1}^{\infty} \psi_k \psi_m cov(\epsilon_{l-k}' M_t \epsilon_r, \epsilon_{p-m}' M_s \epsilon_r). \end{aligned}$$

We have that

$$
\begin{aligned}
& \psi_{l-r}\psi_{p-r}cov(\epsilon_r' M_t\epsilon_r, \epsilon_r' M_s\epsilon_r) + \sum_{k=0}^{\infty}\psi_k\psi_{p-r}cov(\epsilon_{l-k}' M_t\epsilon_r, \epsilon_r' M_s\epsilon_r) \\
& + \psi_{l-r}\sum_{m=0}^{\infty}\psi_m cov(\epsilon_r' M_t\epsilon_r, \epsilon_{p-m}' M_s\epsilon_r) + \sum_{k=0,k\neq l-r}^{\infty}\psi_k\psi_{p-l+k}cov(\epsilon_{l-k}' M_t\epsilon_r, \epsilon_{l-k}' M_s\epsilon_r) \\
= & \ \psi_{l-r}\psi_{p-r}(2\sigma^4 + \kappa_4 E(d_t' d_s)) + \sum_{k=0}^{\infty}\psi_k\psi_{p-r}\kappa_3 E(d_s' M_t\epsilon_{l-k}) \\
& + \psi_{l-r}\sum_{m=0}^{\infty}\psi_{mk}E(d_t' M_s\epsilon_{p-m}) + s\sigma^4\sum_{k=0,k\neq l-r}^{\infty}\psi_k\psi_{p-l+k} \\
= & \ O\left(s + (sN)^{1/2} + \sqrt{st} + (sN)^{1/2} + s\right) = O\left(\sqrt{st} + (sN)^{1/2}\right),
\end{aligned}
$$

noting that $t \geq s$. We also have

$$
\sum_{k=l-s+1}^{\infty}\sum_{m=p-s+1}^{\infty}\psi_k\psi_m cov(\epsilon_{l-k}' M_t\epsilon_r, \epsilon_{p-m}' M_s\epsilon_r) = \sigma^2\sum_{k=l-s+1}^{\infty}\sum_{m=p-s+1}^{\infty}\psi_k\psi_m E(\epsilon_{l-k}' M_s\epsilon_{p-m}).
$$

Next, we suppose that $r \neq q$ and $q \geq t$. Lemma 12 gives that

$$
\begin{aligned}
cov(w_l' M_t\epsilon_r, w_p' M_s\epsilon_r) &= \psi_{l-r}\psi_{p-r}cov(\epsilon_r' M_t\epsilon_r, \epsilon_r' M_s\epsilon_r) \\
&= \psi_{l-q}\psi_{p-r}(2\sigma^4 s + \kappa_4 E(d_t' d_s)) = O(s).
\end{aligned}
$$

For cases with $q < t$ (note that in this case $r \neq q$), Lemma 12 provides that

$$
\begin{aligned}
cov(w_l' M_t\epsilon_r, w_p' M_s\epsilon_r) &= \psi_{l-r}\psi_{p-r}\kappa_3 E(d_t' M_s\epsilon_q) + s\sigma^4\psi_{l-q}\psi_{p-r} \\
&= O\left(\sqrt{st} + s\right) = O(\sqrt{st}).
\end{aligned}
$$

□

Lemma 14. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. If* $N \to \infty$, $T \to \infty$ *and* $p \to \infty$ *with* $p \log T/T \to 0$ *and* $T/N \to 0$, *then*

$$
||G_2|| = O_p\left(\sqrt{\frac{p}{NT}} + \frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}} + \frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right) = o_p(1).
$$

Proof. We observe that

$$
\begin{aligned}
G_2 &= \frac{1}{NT}\sum_{t=p+1}^{T-1} x_t^*(p)' M_t\epsilon_t^* \\
&= \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)' M_t\epsilon_t - \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)' M_t\bar{\epsilon}_{t,T} \\
&\quad - \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t-1,T-1}(p)' M_t\epsilon_t + \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t-1,T-1}(p)' M_t\bar{\epsilon}_{t,T}.
\end{aligned}
$$

Noting that $w_{t-1}(p)'M_t\epsilon_t$ is a martingale difference sequence, we have

$$\begin{aligned} & E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\epsilon_t\right\|^2 \\ = & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2 tr(E(w_{t-1}(p)'M_t\epsilon_t\epsilon_t'M_tw_{t-1}(p))) \\ = & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2 \sigma^2 tr(E(w_{t-1}(p)'M_tw_{t-1}(p))). \end{aligned}$$

Since

$$tr(E(w_{t-1}(p)'M_tw_{t-1}(p))) \leq tr(E(w_{t-1}(p)'w_{t-1}(p))) = E||w_{t-1}(p)||^2 = O(Np),$$

where the last equality follows from Lemma 1, and $(T-t)^2/(T-t+1)^2 < 1$, we have

$$E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\epsilon_t\right\|^2 = O\left(\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}Np\right) = O\left(\frac{p}{NT}\right).$$

Therefore, the Markov inequality gives

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\epsilon_t\right\| = O_p\left(\sqrt{\frac{p}{NT}}\right).$$

We consider the term

$$\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\bar{\epsilon}_{t+1,T} = \frac{1}{NT}\sum_{t=p+1}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}w_{t-1}(p)'M_t\epsilon_m.$$

Now, we have that

$$\begin{aligned} & E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}w_{t-1}(p)'M_t\epsilon_m\right\|^2 \\ = & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t+1}^{T}\sum_{m'=t'+1}^{T} tr\left(E\left(w_{t-1}(p)'M_t\epsilon_m\epsilon_{m'}'M_{t'}w_{t'-1}(p)\right)\right). \end{aligned}$$

Now, we have that $E\left(w_{t-1}(p)'M_t\epsilon_m\epsilon_{m'}'M_{t'}w_{t'-1}(p)\right) = 0$ if $m \neq m'$ and $E\left(w_{t-1}(p)'M_t\epsilon_m\epsilon_m'M_tw_{t'-1}(p)\right) = \sigma^2 E\left(w_{t-1}(p)'M_{\min(t,t')}w_{t'-1}(p)\right)$. Therefore, we have

$$\begin{aligned} & E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}w_{t-1}(p)'M_t\epsilon_m\right\|^2 \\ = & \frac{\sigma^2}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)} tr\left(E\left(w_{t-1}(p)'M_{\min(t,t')}w_{t'-1}(p)\right)\right). \end{aligned}$$

We observe that

$$\begin{aligned}
& \frac{\sigma^2}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)}tr\left(E\left(w_{t-1}(p)'M_{\min(t,t')}w_{t'-1}(p)\right)\right) \\
\leq\ & \frac{2\sigma^2}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=t}^{T-1}\frac{T-t'}{(T-t+1)(T-t'+1)}tr\left(E\left(w_{t-1}(p)'M_t w_{t'-1}(p)\right)\right) \\
=\ & \frac{2\sigma^2}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}tr\left(E\left(w_{t-1}(p)'M_t\tilde{w}_{t-1,T-2}(p)\right)\right),
\end{aligned}$$

where

$$\tilde{w}_{t-1,T-2}(p)=\frac{1}{T-t}\sum_{t'=t}^{T-1}\frac{T-t'}{T-t'+1}w_{t'-1}(p).$$

It holds that

$$\begin{aligned}
tr(E\left(w_{t-1}(p)'M_t\tilde{w}_{t-1,T-2}(p)\right)) &\leq tr(E\left(w_{t-1}(p)'\tilde{w}_{t-1,T-2}(p)\right)) \\
&\leq N\sqrt{E(||w_{i,t-1}(p)||^2)}\sqrt{E(||\tilde{w}_{i,t-1,T-2}(p)||^2)} \\
&= O\left(\frac{Np}{\sqrt{T-t}}\right),
\end{aligned}$$

where $E(||\tilde{w}_{i,t-1,T-2}(p)||^2=O\left(Np/(T-t)\right)$ by the short memory assumption in Assumption 1. Thus, it holds that

$$\begin{aligned}
E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}w_{t-1}(p)'M_t\epsilon_m\right\|^2 &= \frac{2\sigma^2}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}O\left(\frac{Np}{\sqrt{T-t}}\right) \\
&= O\left(\frac{p}{NT^{3/2}}\right).
\end{aligned}$$

The Chebyshev inequality gives that

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\bar{\epsilon}_{t+1,T}\right\|=O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}}\right).$$

Next, we consider the term

$$\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\epsilon_t=\frac{1}{NT}\sum_{t=p+1}^{T-1}\sum_{m=t}^{T-1}\frac{1}{T-t+1}w_m(p)'M_t\epsilon_t.$$

We have

$$
\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\epsilon_t\right\|^2 \\
= & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1} tr\left(E\left(w_m(p)'M_t\epsilon_t\epsilon_{t'}'M_{t'}w_{m'}(p)\right)\right).
\end{aligned}
$$

We have that

$$
\begin{aligned}
& E\left(w_m(p)'M_t\epsilon_t\epsilon_{t'}'M_{t'}w_{m'}(p)\right) \\
= & \; cov\left(w_m(p)'M_t\epsilon_t, w_{m'}(p)'M_{t'}\epsilon_{t'}\right)+E\left(w_m(p)'M_t\epsilon_t\right)E\left(\epsilon_{t'}'M_{t'}w_{m'}(p)\right).
\end{aligned}
$$

Let $\psi_t(p)=(\psi_t,\ldots,\psi_{t-p})'$. Since $E\left(\epsilon_{t'}'M_t\epsilon_t\right)=0$ for $t'\neq t$, we have that

$$
E\left(w_m'M_t\epsilon_t\right)=E\left(\sum_{k=0}^{\infty}\psi_k\epsilon_{m-k}'M_t\epsilon_t\right)=\psi_{m-t}E(\epsilon_t'M_t\epsilon_t)=\psi_{m-t}\sigma^2(t-1).
$$

It therefore holds that

$$
E\left(w_m(p)'M_t\epsilon_t\right)=\sigma^2(t-1)\psi_{m-t}(p).
$$

Thus, we can write

$$
\begin{aligned}
& \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1} tr\left(E\left(w_m(p)'M_t\epsilon_t\right)E\left(\epsilon_{t'}'M_{t'}w_{m'}(p)\right)\right) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{t-1}{T-t+1}\frac{t'-1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1} tr\left(\psi_{m-t}(p)\psi_{m'-t'}(p)'\right) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{t-1}{T-t+1}\frac{t'-1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1}\psi_{m-t}(p)'\psi_{m'-t'}(p) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{t-1}{T-t+1}\frac{t'-1}{T-t'+1}\left(\sum_{m=t}^{T-1}\psi_{m-t}(p)\right)'\left(\sum_{m'=t'}^{T-1}\psi_{m'-t'}(p)\right).
\end{aligned}
$$

Since $\left\|\sum_{m=t}^{T-1}\psi_{m-t}(p)\right\|=O(\sqrt{p})$ uniformly in m,t,T and

$$
\sum_{t=p+1}^{T-1}\frac{t-1}{T-t+1}=\sum_{s=2}^{T-p}\frac{T-s+1}{s}=O\left(T\log T\right),
$$

we have

$$\begin{aligned}
&\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{t-1}{T-t+1}\frac{t'-1}{T-t'+1}\left(\sum_{m=t}^{T-1}\psi_{m-t}(p)\right)'\left(\sum_{m'=t'}^{T-1}\psi_{m'-t'}(p)\right)\\
=\ & O\left(\frac{pT^2(\log T)^2}{N^2T^2}\right)=O\left(\frac{p(\log T)^2}{N^2}\right).
\end{aligned}$$

Next, we consider the covariances. Lemma 13 gives

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1}tr\left(cov\left(w_m(p)'M_t\epsilon_t,w_{m'}(p)'M_{t'}\epsilon_{t'}\right)\right)\\
=\ & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}cov\left(w'_{m-j}M_t\epsilon_t,w'_{m'-j}M_{t'}\epsilon_{t'}\right)\\
=\ & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}O\left(t+\sqrt{tN}\right)\\
&+\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}\sum_{k=m-j-t+1}^{\infty}\sum_{k=m'-j-t+1}^{\infty}\psi_k\psi'_kE(\epsilon'_{m-j-k}M_t\epsilon_{m'-j-k'})\\
&+\frac{2}{N^2T^2}\sum_{t'=p+1}^{T-1}\sum_{t=t'+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\\
&\times\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}(\psi_{m-j-t'}\psi_{m'-j-t'}\kappa_3E(d'_tM_{t'}\epsilon_{t'})+\sigma^4t'\psi_{m-j-t'}\psi_{m'-j-t}).
\end{aligned}$$

We observe that

$$\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}O\left(t+\sqrt{tN}\right)=O\left(\frac{p}{N^2}+\frac{p}{N^{3/2}\sqrt{T}}\right).$$

For the second term, we have

$$\begin{aligned}
&\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}\sum_{k=m-j-t+1}^{\infty}\sum_{k=m'-j-t+1}^{\infty}\psi_k\psi'_kE(\epsilon'_{m-j-k}M_t\epsilon_{m'-j-k'})\\
=\ & O\left(\frac{1}{NT^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}\sum_{k=m-j-t+1}^{\infty}\sum_{k=m'-j-t+1}^{\infty}\psi_k\psi'_k\right)\\
=\ & O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\left(\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}\psi_k\right)^2\right).
\end{aligned}$$

Since $\left|\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}\psi_k\right| \le \sum_{k=1}^{\infty} k|\psi_{-j+k}|$, which exists for any j by Assumption 4.

$$O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\left(\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}\psi_k\right)^2\right) = O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2\right).$$

Now, we observe that

$$\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2 \le p\left(\sum_{k=1}^{\infty}k|\psi_{-p+1+k}|\right)^2 = p\left(\sum_{k=0}^{\infty}(k+p-1)|\psi_k|\right)^2 = O(p^3),$$

which implies that

$$O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2\right) = O\left(\frac{p^3}{NT^2}\right).$$

Therefore we have that

$$\begin{aligned}&\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\frac{1}{(T-t+1)^2}\sum_{m=t}^{T-1}\sum_{m'=t}^{T-1}\sum_{j=0}^{p-1}\sum_{k=m-j-t+1}^{\infty}\sum_{k=m'-j-t+1}^{\infty}\psi_k\psi_k' E(\epsilon_{m-j-k}' M_t\epsilon_{m'-j-k'})\\ =\;& O\left(\frac{p^3}{NT^2}\right).\end{aligned}$$

Since $\sum_{m=t}^{T-1}\psi_{m-j-t'} = O(1)$ by Assumption 1, and $|E(d_tM_{t'}\epsilon_{t'-1})| = O(\sqrt{tt'})$, we have

$$\begin{aligned}&\frac{2}{N^2T^2}\sum_{t'=p+1}^{T-1}\sum_{t=t'+1}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\\ &\times\sum_{m=t}^{T-1}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}(\psi_{m-j-t'}\psi_{m'-j-t'}\kappa_3E(d_t'M_{t'}\epsilon_{t'}) + \sigma^4 t'\psi_{m-j-t'}\psi_{m'-j-t})\\ =\;& O\left(\frac{p}{N^2T^2}\sum_{t'=p+1}^{T-1}\sum_{t=t'+1}^{T-1}\frac{\sqrt{t}}{T-t+1}\frac{\sqrt{t'}}{T-t'+1}\right) = O\left(\frac{p}{N^2T}\right).\end{aligned}$$

Therefore, it holds that

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\epsilon_t\right\| &= O_p\left(\frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p}}{N} + \frac{\sqrt{p}}{N^{3/4}T^{1/4}} + \frac{p^{3/2}}{\sqrt{NT}} + \frac{\sqrt{p}}{N\sqrt{T}}\right)\\ &= O_p\left(\frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p}}{N^{3/4}T^{1/4}} + \frac{p^{3/2}}{\sqrt{NT}}\right),\end{aligned}$$

because $T/N \to 0$.

Lastly, we consider the term

$$\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\bar{\epsilon}_{t+1,T} = \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\sum_{m=t}^{T-1}\sum_{l=t+1}^{T} w_m(p)'M_t\epsilon_l.$$

By an argument similar to what shown above, we have

$$E(w_m'M_t\epsilon_l) = E\left(\sum_{k=0}^{\infty}\psi_k\epsilon_{m-k}'M_t\epsilon_l\right) = \psi_{m-l}E(\epsilon_l'M_t\epsilon_l) = \psi_{m-l}\sigma^2(t-1).$$

so that

$$E\left(w_m(p)'M_t\epsilon_l\right) = \sigma^2(t-1)\psi_{m-l}(p).$$

Therefore we have that

$$\begin{aligned}
&\left\|E\left(\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\sum_{m=t}^{T-1}\sum_{l=t+1}^{T} w_m(p)'M_t\epsilon_l\right)\right\|^2 \\
= \;& \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{t-1}{(T-t)(T-t+1)}\frac{t-1}{(T-t')(T-t'+1)} \\
& \times\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1}^{T} tr\left(\psi_{m-l}(p)\psi_{m'-l'}(p)'\right).
\end{aligned}$$

Now, we observe that

$$\begin{aligned}
&\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1}^{T} tr\left(\psi_{m-l}(p)\psi_{m'-l'}(p)'\right) \\
= \;& \left(\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\psi_{m-l}(p)\right)'\left(\sum_{m'=t'}^{T-1}\sum_{l'=t'+1}^{T}\psi_{m'-l'}(p)\right) = O\left(p\sqrt{(T-t)(T-t')}\right),
\end{aligned}$$

uniformly in m, m', l, l'. Since

$$\sum_{t=p+1}^{T-1}\frac{t-1}{\sqrt{T-t}(T-t+1)} = \sum_{s=1}^{T-p-1}\frac{T-s}{\sqrt{s}(s+1)} = O\left(T\right),$$

we have that

$$\left\|E\left(\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\sum_{m=t}^{T-1}\sum_{l=t+1}^{T} w_m(p)'M_t\epsilon_l\right)\right\|^2 = O\left(\frac{p}{N^2}\right).$$

We then consider the variance. Let $g_T(t) = (T-t)^{-1}(T-t+1)^{-1}$. We observe that

$$
\begin{aligned}
& tr\left(var\left(\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\sum_{m=t}^{T-1}\sum_{l=t+1}^{T} w_m(p)'M_t\epsilon_l\right)\right) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1}^{T} tr\left(cov(w_m(p)'M_t\epsilon_l, w_{m'}(p)'M_{t'}\epsilon_{l'})\right) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1}^{T}\sum_{j=0}^{p-1} cov(w'_{m-j}M_t\epsilon_l, w'_{m'-j}M_{t'}\epsilon_{l'}).
\end{aligned}
$$

Noting that $O(\min(t,t')) = O(\sqrt{tt'})$, Lemma 13 shows that

$$
\begin{aligned}
& tr\left(var\left(\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\sum_{m=t}^{T-1}\sum_{l=t+1}^{T} w_m(p)'M_t\epsilon_l\right)\right) \\
= & \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1} O(\sqrt{tt'}+\sqrt{\min(t,t')N}) \\
& +\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1} \\
& \times\sum_{j=0}^{p-1}\sum_{k=m-j-\min(t,t')+1}^{\infty}\sum_{k=m'-j-\min(t,t')+1}^{\infty} \psi_k\psi_{k'}E(\epsilon'_{m-j-k}M_{\min(t,t')}\epsilon_{m'-j-k'}) \\
& +\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1,l'\neq l}^{T}\sum_{j=0}^{p-1} O((\sqrt{tt'}).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& \frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1} O(\sqrt{tt'}+\sqrt{\min(t,t')N}) \\
= & O\left(\frac{p}{N^2}\right) + O\left(\frac{p}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)}(\sqrt{\min(t,t')N}\right) \\
= & O\left(\frac{p}{N^2}+\frac{p\log T}{N^{3/2}T}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)}\sqrt{\min(t,t')N}\\
=\;&\sum_{t=p+1}^{T-1}\frac{T-t}{(T-t+1)^2}\sqrt{tN}+2\sum_{t=p+1}^{T-1}\sum_{t'=t+1}^{T-1}\frac{T-t'}{(T-t+1)(T-t'+1)}\sqrt{tN}\\
=\;&O\left(\sqrt{N}T\log T\right)+O\left(\sqrt{N}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\sqrt{t}\right)\\
=\;&O\left(\sqrt{N}T\log T+\sqrt{NT}\right)=O\left(\sqrt{N}T\log T\right).
\end{aligned}
$$

For the second term, it holds that

$$
\begin{aligned}
&\Big|\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}\\
&\times\sum_{k=m-j-\min(t,t')+1}^{\infty}\sum_{k=m'-j-\min(t,t')+1}^{\infty}\psi_k\psi_{k'}E(\epsilon'_{m-j-k}M_{\min(t,t')}\epsilon_{m'-j-k'})\Big|\\
\le\;&\frac{\sigma^6}{NT^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1}g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}\sum_{k=m-j-\min(t,t')+1}^{\infty}\sum_{k'=m'-j-\min(t,t')+1}^{\infty}|\psi_k|\cdot|\psi_{k'}|\\
=\;&\frac{\sigma^6}{NT^2}\sum_{j=0}^{p-1}\sum_{t=p+1}^{T-1}g_T(t)\frac{1}{(T-t+1)}\left(\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}|\psi_k|\right)^2\\
&+2\frac{\sigma^6}{NT^2}\sum_{j=0}^{p-1}\sum_{t=p+1}^{T-1}\sum_{t'=t+1}^{T-1}g_T(t)\frac{1}{(T-t'+1)}\left(\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}|\psi_k|\right)\left(\sum_{m'=t'}^{T-1}\sum_{k'=m'-j-t+1}^{\infty}|\psi_{k'}|\right).
\end{aligned}
$$

Observing that

$$
\begin{aligned}
\sum_{m=t}^{T-1}\sum_{k=m-j-t+1}^{\infty}|\psi_k| &\le \sum_{k=1}^{\infty}k|\psi_{-j+k}|,\\
\sum_{m'=t'}^{T-1}\sum_{k'=m'-j-t+1}^{\infty}|\psi_k| &\le \sum_{k=1}^{\infty}k|\psi_{t'-t-j+k}|,
\end{aligned}
$$

it follows that

$$\Big|\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}$$

$$\times\sum_{k=m-j-\min(t,t')+1}^{\infty}\sum_{k=m'-j-\min(t,t')+1}^{\infty}\psi_k\psi_{k'}E(\epsilon'_{m-j-k}M_{\min(t,t')}\epsilon_{m'-j-k'})\Big|$$

$$= O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2\right)$$

$$+O\left(\frac{1}{NT^2}\sum_{j=0}^{p-1}\sum_{t=p+1}^{T-1}\sum_{t'=t+1}^{T-1}\frac{1}{(T-t)(T-t+1)}\frac{1}{(T-t'+1)}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)\left(\sum_{k=1}^{\infty}k|\psi_{t'-t-j+k}|\right)\right).$$

Now, we observe that

$$\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2 \le p\left(\sum_{k=1}^{\infty}k|\psi_{-p+1+k}|\right)^2 = p\left(\sum_{k=0}^{\infty}(k+p-1)|\psi_k|\right)^2 = O(p^3),$$

and

$$\sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)\left(\sum_{k=1}^{\infty}k|\psi_{t'-t-j+k}|\right) \le \sum_{j=0}^{p-1}\left(\sum_{k=1}^{\infty}k|\psi_{-j+k}|\right)^2 = O(p^3).$$

It therefore follows that

$$\Big|\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=\max(t,t')+1}^{T}\sum_{m'=t'}^{T-1}\sum_{j=0}^{p-1}$$

$$\times\sum_{k=m-j-\min(t,t')+1}^{\infty}\sum_{k=m'-j-\min(t,t')+1}^{\infty}\psi_k\psi_{k'}E(\epsilon'_{m-j-k}M_{\min(t,t')}\epsilon_{m'-j-k'})\Big|$$

$$= O\left(\frac{p^3}{NT^2}\right)+O\left(\frac{p^3\log T}{NT^2}\right)=O\left(\frac{p^3\log T}{NT^2}\right).$$

For the third term, we have

$$\frac{\sigma^4}{N^2T^2}\sum_{t=p+1}^{T-1}\sum_{t'=p+1}^{T-1} g_T(t)g_T(t')\sum_{m=t}^{T-1}\sum_{l=t+1}^{T}\sum_{m'=t'}^{T-1}\sum_{l'=t'+1,l'\neq l}^{T}\sum_{j=0}^{p-1}O((tt')^{1/2}) = O\left(\frac{p}{N^2}\right).$$

Therefore we have that

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\bar{\epsilon}_{t+1,T}\right\| = O_p\left(\frac{\sqrt{p}}{N}+\frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}}+\frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right).$$

To sum up, we have

$$\begin{aligned} ||G_2|| &= O_p\left(\sqrt{\frac{p}{NT}}\right) + O_p\left(\frac{\sqrt{p}}{N^{1/2}T^{3/4}}\right) + O_p\left(\frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p}}{N^{3/4}T^{1/4}} + \frac{p^{3/2}}{\sqrt{N}T}\right) \\ &\quad + O_p\left(\frac{\sqrt{p}}{N} + \frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}} + \frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right) \\ &= O_p\left(\sqrt{\frac{p}{NT}} + \frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}} + \frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right) \end{aligned}$$

by $T/N \to 0$.

□

Lemma 15. *Suppose that Assumptions 1, 2, 3 and 4 are satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2T/N \to 0$ and $p^3 \log T/T \to 0$, then*

$$\sqrt{NT}\ell_p'\Gamma_p^{-1}G_2/v_p \to_d N(0,1).$$

Proof. We observe that

$$G_2 = \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\epsilon_t + \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t + G_{22},$$

where

$$\begin{aligned} G_{22} &= -\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'M_t\bar{\epsilon}_{t+1,T} - \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\epsilon_t \\ &\quad + \frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{t,T-1}(p)'M_t\bar{\epsilon}_{t+1,T}. \end{aligned}$$

The proof of Lemma 14 shows that

$$\begin{aligned} \|G_{22}\| &= O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}}\right) + O_p\left(\frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p}}{N^{3/4}T^{1/4}} + \frac{p^{3/2}}{\sqrt{N}T}\right) \\ &\quad + O_p\left(\frac{\sqrt{p}}{N} + \frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}} + \frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right) \\ &= O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}} + \frac{\sqrt{p}\log T}{N} + \frac{\sqrt{p}}{N^{3/4}T^{1/4}} + \frac{\sqrt{p\log T}}{N^{3/4}\sqrt{T}} + \frac{p^{3/2}\sqrt{\log T}}{\sqrt{N}T}\right) \end{aligned}$$

so that $\sqrt{NT}||G_{22}|| = o_p(1)$ if $p^2/T \to 0$, $pT(\log T)^2/N \to 0$, $p^2T/N \to 0$, $p^2 \log T/N \to 0$ and $p^3 \log T/T \to 0$, which are satisfied by the conditions of the lemma. It therefore follows that

$$\left\|\sqrt{NT}\ell_p'\Gamma_p^{-1}G_{22}\right\| \le \|\ell_p\|_1 \cdot \left\|\Gamma_p^{-1}\right\|_1 \cdot \left\|\sqrt{NT}G_{22}\right\| = o_p(1),$$

by the assumption that $\|\ell_p\|_1 = O(1)$ and Assumption 2.

We consider the first term in the decomposition of G_2. Lemma 4 gives

$$\sqrt{NT}\ell_p'\Gamma_p^{-1}\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\epsilon_t/v_p \to_d N(0,1).$$

Next, we consider the second term in the decomposition of G_2. Since $w_{t-1}(p)'(I-M_t)\epsilon_t$ is a martingale difference sequence, we have that

$$\begin{aligned} & E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t\right\|^2 \\ = & \frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2\sigma^2 tr\left(E(w_{t-1}(p)'(I-M_t)w_{t-1}(p))\right). \end{aligned}$$

Hence, as in the proof of Lemma 10, we have that

$$tr\left(E(w_{t-1}(p)'(I-M_t)w_{t-1}(p))\right) = O\left(\frac{Np^2}{t}\right).$$

It therefore follows that

$$\begin{aligned} E\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t\right\|^2 &= O\left(\frac{1}{N^2T^2}\sum_{t=p+1}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2\frac{Np^2}{t}\right) \\ &= O\left(\frac{p^2\log T}{NT^2}\right). \end{aligned}$$

By the Chebyshev inequality, it holds that

$$\left\|\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t\right\| = O_p\left(\frac{p(\log T)^{1/2}}{N^{1/2}T}\right).$$

This result implies that

$$\begin{aligned} & \left\|\sqrt{NT}\ell_p'\Gamma_p^{-1}\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t\right\| \\ \leq & \|\ell_p\|_1\cdot\left\|\Gamma_p^{-1}\right\|_1\cdot\left\|\sqrt{NT}\frac{1}{NT}\sum_{t=p+1}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'(I-M_t)\epsilon_t\right\| \\ = & O_p\left(\frac{p(\log T)^{1/2}}{T^{1/2}}\right) = o_p(1), \end{aligned}$$

by by the assumption that $\|\ell_p\|_1 = O(1)$, Assumption 2 and the assumption that $p^2\log T/T \to 0$.

□

C.1 Proof of Theorem 5

Proof. We have

$$||\hat{\alpha}_G(p) - \alpha(p)|| \leq ||(\hat{\Gamma}_p^G)^{-1}||_1 ||G_1|| + ||(\hat{\Gamma}_p^G)^{-1}||_1 ||G_2||.$$

Lemmas 3 and 10 give that $||(\hat{\Gamma}_p^G)^{-1}||_1 = O_p(1)$. Lemma 11 gives that $||G_1|| = o_p(1)$ when $p^{1/2} \sum_{k=p+1}^{\infty} \alpha_k \to 0$ and $||G_2|| = o_p(1)$ follows by Lemma 14. □

C.2 Proof of Theorem 6

Proof. We note that

$$\begin{aligned}
& \sqrt{NT}(\ell_p' \hat{\alpha}_G(p) - \ell_p' \alpha(p)) \\
= \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^G)^{-1} G_1 + \sqrt{NT}\ell_p'(\hat{\Gamma}_p^G)^{-1} G_2 \\
= \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^G)^{-1} G_1 + \sqrt{NT}\ell_p'((\hat{\Gamma}_p^G)^{-1} - \Gamma_p^{-1}) G_2 + \sqrt{NT}\ell_p' \Gamma_p^{-1} G_2.
\end{aligned}$$

Lemma 15 gives

$$\sqrt{NT}\ell_p' \Gamma_p^{-1} G_2 / v_p \to_d N(0,1).$$

Next, we show that

$$||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^G)^{-1} G_1|| \leq ||\ell_p||_1 ||(\hat{\Gamma}_p^G)^{-1}||_1 ||\sqrt{NT} G_1|| = o_p(1),$$

where we have $||\ell_p||_1 = O(1)$ by the assumption, $||(\hat{\Gamma}_p^G)^{-1}||_1 = O_p(1)$ by Lemmas 3 and 10 and $||\sqrt{NT} G_1|| = o_p(1)$ by Lemma 11 because $\sqrt{NTp} \sum_{k=T-1}^{\infty} |\alpha_k| \to 0$.

Lastly, we see that

$$||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^G)^{-1} - \Gamma_p^{-1}) G_2|| \leq ||\ell_p||_1 ||(\hat{\Gamma}_p^G)^{-1} - \Gamma_p^{-1}||_1 ||\sqrt{NT} G_2||.$$

We have $||\ell_p||_1 = \ell_p' \ell_p \leq C$. $||(\hat{\Gamma}_p^G)^{-1} - (\hat{\Gamma}_p^G)^{-1}||_1 = O_p(p/T^{1/2})$ by Lemmas 3 and 10. $||\sqrt{NT} G_2|| = O_p(p^{1/2} + p^{1/2} T^{1/2} \log T / N^{1/2} + p^{1/2} (\log T)^{1/2} / N^{1/4} + p^{3/2} (\log T)^{1/2} / T^{1/2}) = O_p(p^{1/2})$ by Lemma 14 and the condition that $p^2 T/N \to 0$ and $p^3 \log T / T \to 0$. Therefore we have $||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^G)^{-1} - \Gamma_p^{-1}) G_2|| = O_p(p^{3/2}/T^{1/2})$ which is of order $o_p(1)$ by the condition $p^3 \log T / T \to 0$. □

D Hayakawa's efficient IV estimator

This section presents several lemmas and the proofs of Theorem 7 and 8.

The estimation error of the Hayakawa's IV estimator can be decomposed as

$$\hat{\alpha}(p) - \alpha(p) = (\hat{\Gamma}_p^H)^{-1} H_1 + (\hat{\Gamma}_p^H)^{-1} H_2$$

where

$$\hat{\Gamma}_p^H = \frac{1}{NT}\sum_{t=p+2}^{T-1} h_t(p)'x_t^*(p), \quad H_1 = \frac{1}{NT}\sum_{t=p+2}^{T-1} h_t(p)'b_{t,p}^* \text{ and } H_2 = \frac{1}{NT}\sum_{t=p+2}^{T-1} h_t(p)'\epsilon_t^*.$$

Note that we can write

$$x_t^*(p) = \sqrt{\frac{T-t}{T-t+1}}\left(w_{t-1}(p) - \bar{w}_{t,T-1}(p)\right) \text{ and } h_t(p) = \sqrt{\frac{T-t}{T-t+1}}\left(w_{t-1}(p) - \bar{w}_{p,t-2}(p)\right).$$

Lemma 16. *Suppose that Assumptions 1 and 2 are satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/T \to 0$, then*

$$||\hat{\Gamma}_p^H - \Gamma_p|| = O_p\left(\frac{p}{\sqrt{T}}\right)$$

Proof. We observe that

$$\begin{aligned}
\hat{\Gamma}_p^H &= \frac{1}{NT}\sum_{t=p+2}^{T-1} h_t(p)'x_t^*(p) \\
&= \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'w_{t-1}(p) - \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\bar{w}_{t,T-1}(p) \\
&\quad -\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'w_{t-1}(p) + \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'\bar{w}_{t,T-1}(p).
\end{aligned}$$

For the first term in the decomposition, Lemma 2 gives

$$\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'w_{t-1}(p) - \Gamma_p\right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T}\right).$$

The second term is

$$\begin{aligned}
&\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\bar{w}_{t,T-1}(p)\right\| \le \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\|w_{t-1}(p)\|\cdot\|\bar{w}_{t,T-1}(p)\| \\
&= O_p\left(\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\sqrt{Np}\sqrt{Np\frac{1}{T-t}}\right) = O_p\left(\frac{p}{\sqrt{T}}\right),
\end{aligned}$$

where the first equality follows from Lemma 1.

For the third term, we observe that

$$\begin{aligned}
&\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'w_{t-1}(p)\right\| \le \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\|w_{t-1}(p)\|\cdot\|\bar{w}_{p,t-2}(p)\| \\
&= O_p\left(\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\sqrt{Np}\sqrt{Np\frac{1}{t-p}}\right) = O_p\left(\frac{p}{\sqrt{T}}\right),
\end{aligned}$$

where the first equality is given by Lemma 1.

Lastly, again applying Lemma 1, we obtain

$$\begin{aligned}
& \left\| \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{p,t-2}(p)' \bar{w}_{t,T-1}(p) \right\| \\
\leq & \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} \|\bar{w}_{p,t-2}(p)\| \cdot \|\bar{w}_{t,T-1}(p)\| \\
= & O_p\left(\frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} \sqrt{Np\frac{1}{T-t}} \sqrt{Np\frac{1}{t-p}} \right) = O_p\left(\frac{p}{\sqrt{T}}\right).
\end{aligned}$$

To sum up, it holds that

$$\left\| \hat{\Gamma}_p^H - \Gamma_p \right\| = O_p\left(\frac{p}{\sqrt{NT}} + \frac{p^{3/2}}{T} + \frac{p}{\sqrt{T}} \right) = O_p\left(\frac{p}{\sqrt{T}}\right).$$

□

Lemma 17. *Suppose that Assumption 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$, then*

$$||H_1|| = O_p\left(\sqrt{p} \sum_{k=p+1}^{\infty} |\alpha_k| \right) = o_p(1).$$

Proof. Similarly to the proof of Lemma 11, we have We have that

$$E\|h_t(p)\|^2 = NE(tr(h_{it}(p)h_{it}(p)')) = N\sum_{k=1}^{p} E\left((h_{it,k})^2\right) = NpE((h_{it,1})^2) = O_p(Np),$$

so that $\|h_t(p)\| = O_p(\sqrt{Np})$ and

$$\left\|b_{t,p}^*\right\| = O_p\left(\sqrt{N} \sum_{k=p+1}^{\infty} |\alpha_k|\right)$$

by (17). Thus, it holds that

$$\begin{aligned}
||H_1|| &= \left\| \frac{1}{NT} \sum_{t=p+2}^{T-1} h_t(p)' b_{t,p}^* \right\| \leq \frac{1}{NT} \sum_{t=p+2}^{T-1} \left\| h_t(p)' b_{t,p}^* \right\| \leq \frac{1}{NT} \sum_{t=p+2}^{T-1} \|h_t(p)\| \cdot \left\|b_{t,p}^*\right\| \\
&= \frac{1}{NT} \sum_{t=p+2}^{T-1} O_p\left(\sqrt{Np}\right) O_p\left(\sqrt{N} \sum_{k=p+1}^{\infty} |\alpha_k|\right) \\
&= O_p\left(\sqrt{p} \sum_{k=p+1}^{\infty} |\alpha_k|\right).
\end{aligned}$$

□

Lemma 18. *Suppose that Assumptions 1 is satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p/T \to 0$, then*

$$||H_2|| = O_p\left(\sqrt{\frac{p}{NT}}\right) = o_p(1).$$

Proof. We observe that

$$\begin{aligned}
H_2 &= \frac{1}{NT}\sum_{t=p+2}^{T-1} h_t(p)'\epsilon_t^* \\
&= \frac{1}{NT}\sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)'\epsilon_t - \frac{1}{NT}\sum_{t=p+2}^{T-1}\sqrt{\frac{T-t}{T-t+1}} h_t(p)'\bar{\epsilon}_{t+1,T} \\
&\quad - \frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'\epsilon_t.
\end{aligned}$$

As in the proof of Lemma 14, it holds that

$$\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}w_{t-1}(p)'\epsilon_t\right\| = O_p\left(\sqrt{\frac{p}{NT}}\right).$$

Let $\dot{h}_t(p) = \sqrt{(T-t+1)/(T-t)}h_t(p) = w_{t-1}(p) - \bar{w}_{p,t-2}$. Then, the second term in the decomposition is

$$\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\dot{h}_t(p)'\bar{\epsilon}_{t+1,T} = \frac{1}{NT}\sum_{t=p+2}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}\dot{h}_t(p)'\epsilon_m.$$

Now, we have that

$$\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}\dot{h}_t(p)'\epsilon_m\right\|^2 \\
=\ & \frac{1}{N^2T^2}\sum_{t=p+2}^{T-1}\sum_{t'=p+2}^{T-1}\frac{1}{T-t+1}\frac{1}{T-t'+1}\sum_{m=t+1}^{T}\sum_{m'=t'+1}^{T} tr\left(E\left(\dot{h}_t(p)'\epsilon_m\epsilon_{m'}'\dot{h}_{t'}(p)\right)\right) \\
=\ & \frac{\sigma^2}{N^2T^2}\sum_{t=p+2}^{T-1}\sum_{t'=p+2}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)} tr\left(E\left(\dot{h}_t(p)'\dot{h}_{t'}(p)\right)\right)
\end{aligned}$$

since $E\left(\dot{h}_t(p)'\epsilon_m\epsilon_{m'}'\dot{h}_{t'}(p)\right) = 0$ if $m \neq m'$ and $E\left(\dot{h}_t(p)'\epsilon_m\epsilon_m'\dot{h}_{t'}(p)\right) = \sigma^2 E\left(\dot{h}_t(p)'\dot{h}t'(p)\right)$.

We observe that

$$\begin{aligned}
& \frac{\sigma^2}{N^2T^2}\sum_{t=p+2}^{T-1}\sum_{t'=p+2}^{T-1}\frac{T-\max(t,t')}{(T-t+1)(T-t'+1)}tr\left(E\left(\dot{h}_t(p)'\dot{h}_{t'}(p)\right)\right) \\
\leq \; & \frac{2\sigma^2}{N^2T^2}\sum_{t=p+2}^{T-1}\sum_{t'=t}^{T-1}\frac{T-t'}{(T-t+1)(T-t'+1)}tr\left(E\left(\dot{h}_t(p)'\dot{h}_{t'}(p)\right)\right) \\
= \; & \frac{2\sigma^2}{N^2T^2}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}tr\left(E\left(\dot{h}_t(p)'\tilde{\dot{h}}_{t,T-1}(p)\right)\right),
\end{aligned}$$

where

$$\tilde{\dot{h}}_{t,T-1}(p)=\frac{1}{T-t}\sum_{t'=t}^{T-1}\frac{T-t'}{T-t'+1}\dot{h}_{t'}(p).$$

It holds that

$$tr(E\left(\dot{h}_t(p)'\tilde{\dot{h}}_{t,T-1}(p)\right))\leq N(E(||\dot{h}_{i,t}(p)||^2))^{1/2}(E(||\tilde{\dot{h}}_{i,t,T-1}(p)||^2))^{1/2}=O\left(\frac{Np}{\sqrt{T-t}}\right),$$

where $E||\tilde{\dot{h}}_{i,t,T-1}(p)||^2 = O\left(Np/(T-t)\right)$ by the short memory assumption in Assumption 1. Thus, it holds that

$$\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\sum_{m=t+1}^{T}\frac{1}{T-t+1}\dot{h}_t(p)'\epsilon_m\right\|^2 \\
= \; & O\left(\frac{1}{N^2T^2}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\frac{Np}{\sqrt{T-t}}\right)=O\left(\frac{p}{NT^{3/2}}\right).
\end{aligned}$$

The Chebyshev inequality gives that

$$\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\dot{h}_t(p)'\bar{\epsilon}_{t+1,T}\right\|=O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}}\right).$$

For the third term, since $\bar{w}_{p,t-2}(p)'\epsilon_t$ is a martingale difference sequence, we observe that

$$E\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'\epsilon_t\right\|^2=\frac{1}{N^2T^2}\sum_{t=p+2}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2\sigma^2tr(E(\bar{w}_{p,t-2}(p)'\bar{w}_{p,t-2}(p))).$$

Since $tr(E(\bar{w}_{p,t-2}(p)'\bar{w}_{p,t-2}(p)))=O(Np/(t-p))$ by Lemma 1 and $(T-t)^2/(T-t+1)^2<1$, we have

$$E\left\|\frac{1}{NT}\sum_{t=p+2}^{T-1}\frac{T-t}{T-t+1}\bar{w}_{p,t-2}(p)'\epsilon_t\right\|^2=O\left(\frac{1}{N^2T^2}\sum_{t=p+2}^{T-1}\left(\frac{T-t}{T-t+1}\right)^2\frac{Np}{t-p}\right)=O\left(\frac{p\log T}{NT^2}\right).$$

Therefore, the Chebyshev inequality gives

$$\left\| \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{p,t-2}(p)' \epsilon_t \right\| = O_p\left(\frac{\sqrt{p \log T}}{\sqrt{N}T} \right).$$

To sum up, we have

$$||H_2|| = O_p\left(\sqrt{\frac{p}{NT}} \right) + O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}} \right) + O\left(\frac{\sqrt{p \log T}}{\sqrt{N}T} \right) = O_p\left(\sqrt{\frac{p}{NT}} \right).$$

□

Lemma 19. *Suppose that Assumptions 1 and 2 are satisfied. If $N \to \infty$, $T \to \infty$ and $p \to \infty$ with $p^2/T \to 0$, then*

$$\sqrt{NT} \ell_p' \Gamma_p^{-1} H_2 / v_p \to_d N(0,1).$$

Proof. We observe that

$$H_2 = \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' \epsilon_t + H_{22},$$

where

$$H_{22} = -\frac{1}{NT} \sum_{t=p+2}^{T-1} \sqrt{\frac{T-t}{T-t+1}} h_t(p)' \bar{\epsilon}_{t+1,T} - \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} \bar{w}_{p,t-2}(p)' \epsilon_t.$$

First, Lemma 4 gives

$$\sqrt{NT} \ell_p' \Gamma_p^{-1} \frac{1}{NT} \sum_{t=p+2}^{T-1} \frac{T-t}{T-t+1} w_{t-1}(p)' \epsilon_t / v_p \to_d N(0,1).$$

The proof of Lemma 18 shows that

$$\|H_{22}\| = O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}} \right) + O\left(\frac{\sqrt{p \log T}}{\sqrt{N}T} \right) = O_p\left(\frac{\sqrt{p}}{\sqrt{N}T^{3/4}} \right)$$

so that $\sqrt{NT}||H_{22}|| = o_p(1)$ if $p^2/T \to 0$. It therefore follows that

$$\left\| \sqrt{NT} \ell_p' \Gamma_p^{-1} H_{22} \right\| \leq \|\ell_p\|_1 \cdot \left\|\Gamma_p^{-1}\right\|_1 \cdot \left\| \sqrt{NT} H_{22} \right\| = o_p(1),$$

by the assumption that $\|\ell_p\|_1 = O(1)$ and Assumption 2.

□

D.1 Proof of Theorem 7

Proof. We have

$$||\hat{\alpha}_H(p) - \alpha(p)|| = ||(\hat{\Gamma}_p^H)^{-1}(H_1 + H_2)|| \leq ||(\hat{\Gamma}_p^H)^{-1}||_1||H_1|| + ||(\hat{\Gamma}_p^H)^{-1}||_1||H_2||.$$

Lemmas 3 and 16 give that $||(\hat{\Gamma}_p^H)^{-1}||_1 = O_p(1)$. Lemma 17 gives that $||H_1|| = o_p(1)$ and $||H_2|| = o_p(1)$ follows by Lemma 18. □

D.2 Proof of Theorem 8

Proof. We note that

$$\begin{aligned}
& \sqrt{NT}(\ell_p'\hat{\alpha}_H(p) - \ell_p'\alpha(p)) \\
= \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^H)^{-1}H_1 + \sqrt{NT}\ell_p'(\hat{\Gamma}_p^H)^{-1}H_2 \\
= \; & \sqrt{NT}\ell_p'(\hat{\Gamma}_p^H)^{-1}H_1 + \sqrt{NT}\ell_p'((\hat{\Gamma}_p^H)^{-1} - \Gamma_p^{-1})H_2 + \sqrt{NT}\ell_p'\Gamma_p^{-1}H_2
\end{aligned}$$

Lemma 19 gives

$$\sqrt{NT}\ell_p'\Gamma_p^{-1}H_2/v_p \to_d N(0,1).$$

Next, we consider

$$||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^H)^{-1}H_1|| \leq ||\ell_p||_1||\sqrt{NT}(\hat{\Gamma}_p^H)^{-1}H_1|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^H)^{-1}||_1||\sqrt{NT}H_1||.$$

We have $||\ell_p||_1 = O(1)$ by the assumption. $||(\hat{\Gamma}_p^H)^{-1}||_1 = O_p(1)$ by Lemmas 3 and 16. $||\sqrt{NT}H_1|| = o_p(1)$ by Lemma 17 because $\sqrt{NT}p\sum_{k=p+1}^{\infty}|\alpha_k| \to 0$. Therefore, we have $||\sqrt{NT}\ell_p'(\hat{\Gamma}_p^H)^{-1}H_1|| = o_p(1)$.

Lastly, we see that

$$||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^H)^{-1} - \Gamma_p^{-1})H_2|| \leq ||\ell_p||_1||(\hat{\Gamma}_p^H)^{-1} - \Gamma_p^{-1}||_1||\sqrt{NT}H_2||.$$

We have $||(\hat{\Gamma}_p^H)^{-1} - \Gamma_p^{-1}||_1 = O_p(p/\sqrt{T})$ by Lemmas 3 and 16 and $||\sqrt{NT}H_2|| = O_p(\sqrt{p})$ by Lemma 18. These results imply that $||\sqrt{NT}\ell_p'((\hat{\Gamma}_p^H)^{-1} - \Gamma_p^{-1})H_2|| = O_p(p^{3/2}/\sqrt{T})$ which is of order $o_p(1)$ if $p^3/T \to 0$. □

E Proof of Theorem 9

Proof. We present only the proof of the consistency of $\hat{v}_{p,F}$. The consistency of the other estimators can be established analogously.

It is sufficient to show that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}(\tilde{y}_{it} - \tilde{x}_{it}(p)'\hat{\alpha}_F(p))^2 \to_p \sigma^2 \tag{18}$$

and

$$\left\| \left(\frac{1}{NT} \sum_{t=p+1}^{T} \tilde{x}_t(p)'\tilde{x}_t(p) \right)^{-1} - \Gamma_p^{-1} \right\|_1 \to_p 0. \tag{19}$$

The convergence (19) holds by Lemma 5. For (18), observe that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}(\tilde{y}_{it}-\tilde{x}_{it}(p)'\hat{\alpha}_F(p))^2 \\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}(\tilde{x}_{it}(p)'(\alpha(p)-\hat{\alpha}_F(p))+\tilde{b}_{it,p}+\tilde{\epsilon}_{it})^2 \\
=\ &(\alpha(p)-\hat{\alpha}_F(p))'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{x}_{it}(p)\tilde{x}_{it}(p)'(\alpha(p)-\hat{\alpha}_F(p))'+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{b}_{it,p}^2 \\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{\epsilon}_{it}^2+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{b}_{it,p}\tilde{x}_{it}(p)'(\alpha(p)-\hat{\alpha}_F(p)) \\
&+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{\epsilon}_{it}\tilde{x}_{it}(p)'(\alpha(p)-\hat{\alpha}_F(p))+2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{b}_{it,p}\tilde{\epsilon}_{it}.
\end{aligned}$$

The first term converges to zero in probability by Theorem 1, Lemma 5 and Assumption 2. For the second term, as in the proof of Lemma 6, $\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{b}_{it,p}^2/(NT) = O_p((\sum_{k=p+1}^{\infty}|\alpha_k|)^2/T) = o_p(1)$. It is easy to see that $\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{\epsilon}_{it}^2/(NT) \to_p \sigma^2$. Theorem 1 and Lemma 6 imply that the fourth term converges to zero in probability. Similarly, Theorem 1 and Lemma 8 imply that the fifth term converges to zero in probability. For the sixth term, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\tilde{b}_{it,p}\tilde{\epsilon}_{it} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}b_{it,p}\epsilon_{it} - \frac{T-p}{NT}\bar{b}_{p+1,T}'\bar{\epsilon}_{p+1,T}.$$

Since $E(b_{it,p}\epsilon_{it}) = 0$ and $var(\sum_{i=1}^{N}\sum_{t=p+1}^{T}b_{it,p}\epsilon_{it}/(NT)) = O(1/(NT))$, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}b_{it,p}\epsilon_{it} \to_p 0.$$

Moreover, we have

$$\left\|\frac{T-p}{NT}\bar{b}_{p+1,T}'\bar{\epsilon}_{p+1,T}\right\| \le \frac{T-p}{NT}\left\|\bar{b}_{p+1,T}\right\|\cdot\left\|\bar{\epsilon}_{p+1,T}\right\| = O_p\left(\frac{\sqrt{p}}{T}\sum_{k=p+1}^{\infty}|\alpha_k|\right) = o_p(1).$$

Thus, the sixth term also converges to zero in probability. Therefore, the convergence (18) holds.

□

F Proof of Theorem 10

Proof. We first show that $t_p(\hat{\alpha}(p)) \to_d N(0,1)$ as $p \to \infty$. Since

$$t_p(\hat{\alpha}) = \frac{\sqrt{NT}(e_p'\hat{\alpha}(p) - e_p'\alpha(p))}{\hat{v}_p} + \frac{\sqrt{NT}e_p'\alpha(p)}{\hat{v}_p},$$

it holds by noting that $\hat{\alpha}(p)$ is asymptotically normal and that $\sqrt{NT}e_p'\alpha(p) = \sqrt{NT}\alpha_p \to 0$ by the assumption. The rest of the proof is exactly the same as the proof of Lemma 5.2 of Ng and Perron (1995) and thus is omitted. □

Additional References

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley and Sons, Inc., New York.

Davidson, J. (1994). *Stochastic Limit Theory*, Oxford University Press, Oxford.

Wiener, N. and Masani, P. (1958). The prediction theory of multivariate stochastic processes, ii: The linear predictor, *Acta Mathematica* **99**: 93–137.