Asymptotics and Inference for Possibly Non-Stationary Panel in the Presence of High-Dimensional Nuisance Parameters*

Dingxian Cao† Chihwa Kao‡ Jungbin Hwang §

November 1, 2021

Abstract

This paper studies asymptotic theory for a possibly nonstationary panel AR(1) model when cross-sectional dimension (n) and time dimension (T) are large. We considers the nonstationary case ($\theta_0 = 1$) in the presence of both cross-sectional and time fixed effects, which is not investigated in existing literature of dynamic panel (e.g., Hahn and Kuersteiner (2002) and Hahn and Moon (2006). We derive the limiting distribution of the bias-corrected (quasi-)maximum likelihood estimator with Gaussian or non-Gaussian error terms. Because of the discontinuity between stationary and non-stationary limits, practitioners can face difficulties in choosing between stationary and non-stationary bias corrected confidence intervals. We extend Andrews (1993)' method to a panel data framework and construct a valid confidence interval without knowing the (non)-stationarity.

Keywords: Panel AR(1) model, asymptotic bias, time effects, autoregressive parameter, unit roots, incidental parameter problem, confidence interval, median-unbiased estimator

JEL Codes: C12, C13, C32

*Department of Economics, University of Connecticut, 365 Fairfield Way, U-1063 Storrs, CT 06269-1063, United States
†dingxian.cao@uconn.edu
‡Corresponding author, chih-hwa.kao@uconn.edu
§Corresponding author, jungbin.hwang@uconn.edu

1 Introduction

When the cross-section dimension (n) is large but the time dimension (T) is fixed, the maximum likelihood estimation (MLE) may cause the well-known *incidental parameter problem* pointed out by Neyman et al. (1948) and later studied by Nickell (1981) in the dynamic panel model. One solution to this inconsistency problem is through the instrument variable estimation proposed by Anderson and Hsiao (1981, 1982). Inspired by IV method then, the generalized method of moments (GMM) is studied to accommodate the estimation of panel model about which we can learn from Holtz-Eakin et al. (1988), Arellano and Bond (1991),Blundell and Bond (1998). However, the weak instrument problem in the presence of unit-root and suffering from many moment conditions when time dimension T is large are the drawbacks of GMM method.

As mentioned in Phillips and Moon (1999), we are now faced with more number of collected individual and time-series data such as international financial series, global pandemic spread and popular App's statistics. With one more dimension growing to the limits, Phillips and Moon (1999) also discuss on the difference between the joint asymptotics and sequential asymptotics. Although the original incidental parameter problem can be overcame under such sequentially/jointly asymptotics, the panel MLE is still *asymptotically biased*, which means the asymptotic distribution is not centered at the true value. Hahn and Kuersteiner (2002) propose a bias correction method which helps derive an asymptotically unbiased estimator. Later, Bun (2003), Bruno (2005), Hahn and Moon (2006) investigate such bias correction method in different models under stationarity. On the other hand, Phillips and Sul (2003) extend the median unbiased estimation, first proposed by Andrews (1993), from time series to dynamic panel model in order to bring together the stationary and non-stationary cases. More contributions to the persistent panel model study has been made under different topics like time trend, serial/cross sectional dependence (e.g., Kao (1999), Kao and Chiang (2001), Chang (2002, 2004), Moon and Phillips (2004), Pesaran (2006), Pesaran and Tosetti (2011), Moon et al. (2014)). Since then, many efforts have been put into unifying the inference study of the potentially non-stationary process or panel model. After Mikusheva (2007, 2012) and Phillips (2014) show the uniformity of Andrews' (1993) method and Stock (1991)'s method in time series inference, Phillips (2018), Phillips and Chao (2019) revisit Anderson-Hsiao IV procedure and propose a uniform inference framework for the panel data model.

This paper first goes back to Hahn and Kuersteiner (2002)'s auto-regressive panel model which only contains the cross-sectional fixed effects. Having the *time effects* (f_t) also included in the panel model helps accommodate wider range of structure in practice, but it may create an other incidental parameter problem in the time series domain even when T is large. Hahn and Moon (2006) adopt Hahn and Kuersteiner (2002)'s bias correction method and show that given a stationary linear Panel AR(1) model with both cross-sectional and time effects, the bias-corrected MLE converges to the same distribution as in Hahn and Kuersteiner (2002). Nevertheless it is still incomplete to rule out the nonstationarity when examining the outcome of including the time effects. The first contribution of this paper is we develop a bias corrected MLE in the presence of unit-root and the high-dimensional nuisance parameters of time effects. The asymptotics of such estimator is first studied with Gaussian error terms, then we relax the assumption to non-Gaussian together with some higher order moment conditions, which explicitly illustrates that Gaussian errors are not necessary to derive the asymptotic normality of bias corrected quasi-MLE.

Nonetheless, the bias correction formula and the asymptotic distribution of bias-corrected MLE are different between the stationary and nonstationary scenarios, regardless of the existence of time effects. It means if we have no prior knowledge on the stationarity of the model, there is a choice problem about the bias-correction formula and confidence interval construction.

The thinking of solving such question starts in the literature of time series. Andrews (1993) shows that the MLE estimator of autoregressive parameter is invariant to all the nuisance parameters, suggesting that numerically approximate finite sample quantile functions of the estimator can help construct the exact confidence interval under the i.i.d. normality assumption. Andrews (1993) therefore, introduces the concept *median-unbiasedness*. Although the finite sample distribution of the estimator depends on the sample size, Andrews' (1993) method allows for possibly non-stationary AR(1) process. Hansen (1999) suggest finite sample distribution of t-statistic to construct the confidence interval using bootstrap method without normality assumption. Gonçalves and Kaffo (2015) also consider the bootstrap inference for panel model, but they only prove the validity of their method when $|\theta_0| < 1$ and no time effects involved.

The second contribution of this paper is that we adopt Andrews' (1993) method to the dynamic panel model and construct a valid confidence interval without knowing the (non)-stationarity. The difference between our paper and Phillips and Sul (2003) is that we include the time effects not as the structured trend but as additional fixed effect parameters. The key challenge is to prove the invariance property holds for our panel MLE $\hat{\theta}$, where the nuisance parameters consist of all cross-sectional effects α_i and time effects f_t. We find that under a regular condition on initial values, the panel MLE preserves the invariance property. Given such invariance, we can construct the exact confidence interval by the approximate quantile functions of finite sample MLE. The other implication of the property is, for *unbalanced panel* model where we have missing time effects or cross-sectional effects at random, the invariance property of the MLE make it possible to approximate the distribution of the estimator by simulation. Notice that the i.i.d. normality assumption is strongly imposed only to approximate the distribution of the finite sample estimator but not required for invariance property. Since the exact confidence interval construction is based on the approximate quantile functions under i.i.d.

normality assumption and quantile function estimation are known to be robust to some extent, we can expect such exact confidence interval is also robust to non-Gaussian or some heterogeneity and dependence structure. The robustness of our exact confidence interval is studied through some simulation in this paper.

The rest of this paper is organized as follows. In Section 2, we define a linear panel AR(1) model with cross-sectional effects and time effects, and assumptions required to develop our asymptotic theory. Then we relax the distributional assumption on the innovations into higher-order moment conditions and reveal the same asymptotic behaviour. Section 3 firstly gives out a representation of the panel model and the regularization conditions we need to have the invariance property of the MLE. Secondly, the median-unbiased estimator and the exact confidence interval construction for the autoregressive parameter are introduced. In section 4, we first conduct a Monte Carlo study to investigate the efficiency of the induced asymptotic distribution in the finite sample. Then we provide some Monte Carlo study to compare and test the performance and robustness of different confidence interval methods. Section 5 concludes the paper, and summarizes what contributions this paper makes and future research problems to go further with. All proofs are organized in the Appendix.

2 Asymptotics for Nonstationary Panel AR(1) Model

2.1 Model and Main Results

We format a linear Panel AR(1) model with fixed cross-sectional effects and time effects,

$$y_{it} = \theta_0 y_{it-1} + (1 - \theta_0)\,\alpha_i + f_t + \varepsilon_{it}, \tag{1}$$

where y_{it} are univariate observables, ε_{it} are mean zero scalar error terms, θ_0 is autoregressive parameter of interest, α_i is the fixed cross-sectional effect, f_t is the time effect. The subscripts $i = 1, \ldots, n$ and $t = 1, \ldots T$ denote the cross-sectional unit and the time index, respectively. We denote n and T to be the dimensions of cross section and time, respectively. We impose the following conditions that will be used in deriving the main results in the paper.

Assumption 1 *(i) ε_{it} is i.i.d. $N(0, \sigma^2)$ across i and t for each i. , (ii) $\theta_0 = 1$, (iii) $(n, T) \to \infty$, (iv) $y_{i0} = 0$.*

Assumption 1.(i) impose strong distribution and independence structure on the innovations. Gaussian innovations simplify the calculation of the higher order moments of the quadratic forms in these innovations. We will relax such strong restriction by sufficient moment conditions in section 2.2. Assumption 1.(iv) assumes the initial value are the same across sections. For simplicity, we make the initials to be zero.

Denote $\overline{y}_{i,-1} = \sum_{t=1}^{T} y_{it-1}/T$, $\overline{y}_{t-1} = \sum_{i=1}^{n} y_{it-1}/n$, $\overline{y}_{-1} = \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it-1}/nT$, $\overline{y}_i = \sum_{t=1}^{T} y_{it}/T$, $\overline{y}_t = \sum_{i=1}^{n} y_{it}/n$, $\overline{y} = \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}/nT$, $\tilde{y}_{it-1} = y_{it-1} - \overline{y}_{i,-1} - \overline{y}_{t-1} + \overline{y}_{-1}$, $\tilde{y}_{it} = y_{it} - \overline{y}_i - \overline{y}_t + \overline{y}$.

The same notation applies to $\overline{\varepsilon}_i$, $\overline{\varepsilon}_t$ and $\overline{\varepsilon}$ as well, $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \overline{\varepsilon}_i - \overline{\varepsilon}_t + \overline{\varepsilon}$. The MLE estimator for θ_0 in (1) can be defined as

$$\hat{\theta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}\tilde{y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}^2}. \tag{2}$$

Theorem 1 *Under condition $(i),(ii)$ in **Assumption** 1, and let y_{it} be generated by (1), as $n \to \infty$ and T fixed we have*

$$\hat{\theta} - \theta_0 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}\tilde{\varepsilon}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}^2} \xrightarrow{p} -\frac{3}{1+T} = -\frac{3}{T} + O\left(\frac{1}{T^2}\right). \tag{3}$$

Theorem 1 implies the MLE of θ_0 will be consistent as both n and T go to infinity. When T is fixed, the bias is the same bias found by Harris and Tzavalis (1999) of the linear dynamic panel model without time effects. The fixed T and large n asymptotics in Theorem 1 does not capture the asymptotic unbiasedness of the MLE $\hat{\theta}$ in the sense of Hahn and Kuersteiner (2002)'s findings. Thus we apply the bias correction method when $\theta_0 = 1$ in Hahn and Kuersteiner (2002) to the MLE $\hat{\theta}$ and show the asymptotic unbiasedness of the $\hat{\theta}+3/(T+1)$ in **Theorem** 2. The proof of **Theorem** 1 is given in Appendix B.

Theorem 2 *Under Assumption 1 and let y_{it} be generated by (1), we have*

$$\sqrt{n}T\left(\hat{\theta} - \theta_0 + \frac{3}{T+1}\right) \xrightarrow{d} N\left(0, \frac{51}{5}\right). \tag{4}$$

Theorem 2 implies that the MLE $\hat{\theta}$ is consistent but asymptotically biased with large n and large T. Taking the same bias correction method by Hahn and Kuersteiner (2002) , the bias corrected estimator $\hat{\theta}+3/(T+1)$ is a consistent and asymptotically unbiased estimator for θ_0. By studying the correlation between regressors and error terms, we are able to show that the asymptotic bias of $\hat{\theta}$ is caused by cross-sectional effects (individual effects). When deriving $\hat{\theta}$ in (2), we can first within-transform the model (1) into

$$\underbrace{y_{it} - \overline{y}_t - (\overline{y}_i - \overline{y})}_{\tilde{y}_{it}} = \theta_0 \underbrace{\left(y_{it-1} - \overline{y}_{t-1} - (\overline{y}_{i,-1} - \overline{y}_{-1})\right)}_{\tilde{y}_{it-1}} + \underbrace{\varepsilon_{it} - \overline{\varepsilon}_t - (\overline{\varepsilon}_i - \overline{\varepsilon})}_{\tilde{\varepsilon}_{it}}. \tag{5}$$

The transformation can be regarded as two filter processes along cross-sectional dimension and time dimension aimed to cross-off f_t and α_i in the expression of our MLE. The covariance between the filtered regressor $\tilde{y}_{it-1}$ and $\tilde{\varepsilon}_{it}$ can be divided into four parts: (i) Covariance between y_{it-1} and ε_{it}, (ii) Covariance between y_{it-1} and $\overline{\varepsilon}_i$, (iii) Covariance between y_{it-1} and $\overline{\varepsilon}_t$, (iv) Covariance between y_{it-1} and $\overline{\varepsilon} = \sum_{t=1}^{T} \overline{\varepsilon}_t / T$. Because of the sequential exogeneity $E\left[\varepsilon_{it} | y_{t-1}, y_{t-2}, \ldots, y_0\right] = 0$, the covariances in (i) and (iii) are zeros. Also, as n gets large, the covariance in (iv) will converge to zero because of the cross-sectionally averaged error process $\overline{\varepsilon}_t$. As for the covariance in (ii), we find the following result:

$$E(y_{i,t-1}\overline{\varepsilon}_i) = \frac{t-1}{T}\sigma^2. \tag{6}$$

The equation above shows that the converging rate $\sqrt{n}T$ times the covariance between $y_{i,t-1}$ and $\overline{\varepsilon}_i$ doesn't vanish when n and T grows to infinity. As a result, the asymptotic bias of MLE is sourced from the covariance between $y_{i,t-1}$ and $\overline{\varepsilon}_i$. In other words, we find the asymptotic bias of MLE exists and is mainly sourced from the consideration of cross-sectional effect α_i via the filtering of time-averaged values. The proof of **Theorem** 2 is given in Appendix C.

2.2 Non-Gaussian Innovations

To construct the MLE in previous section, we assume Gaussian innovations. As shown in the proof of Theorem 4 in Hahn and Kuersteiner (2002), the Gaussian innovations help impose enough moment condition we need to derive the asymptotic distribution.The distributional assumption is only for simplifying their poof to apply a proper cross-sectional CLT and should not exclude the consideration of a more general case with non-Gaussian errors. In this section, we fill this technical gap in the literature by considering the same estimator in (2), but relax the strong distributional assumption into higher-order moment conditions, which is sufficient to gain the same result in **Theorem** 2. Under such setting, we call the estimator defined in (2) *quasi-maximum likelihood estimator.*

Assumption 2 *(i) ε_{it} is i.i.d. across i, t, $E(\varepsilon_{it}) = 0$, $Cov(\varepsilon_{it}, \varepsilon_{is}) = \sigma^2 1_{[s=t]}$. (ii)$\mu_k = E(\varepsilon_{it})^k < \infty, k \leq 8$. (iii) $n, T \to \infty$. (iv) $\theta_0 = 1$.*

In the Assumption 2, we keep the identical independent structure of innovations but replace the normality with upto eighth finite moments.

Theorem 3 *Under Assumption 2 and let y_{it} be generated by (1), we have*

$$\sqrt{n}T\left(\hat{\theta}-\theta_0+\frac{3}{T+1}\right)\xrightarrow{d}N\left(0,\frac{51}{5}\right). \tag{7}$$

The previous Gaussian error assumption implicitly assume all moments are finite. The **Theorem** 3 shows the restrictive Gaussian error can be relaxed by imposing the finite moments up to eight. The highest order eight is given by the 4th central moment of the error vectors' quadratic form in the Lyapunov condition that we choose for the CLT proof. The details are included in Appendix D.

3 Inference for Possibly Nonstationary Panel

In this section, we revisit Andrews' (1993) method and extend it to the dynamic panel model. As we learn from Hahn and Moon (2006) and first part of this paper, the large-(n,T) asymptotic distribution of the bias-corrected MLE is summarized below,

$$\sqrt{nT}\left(\hat{\theta}-\theta_0+\frac{1}{T}(1+\hat{\theta})\right)\xrightarrow{d}N\left(0,1-\theta_0^2\right)\ \text{when}\ |\theta_0|<1;$$
$$\sqrt{nT^2}\left(\hat{\theta}-\theta_0+\frac{3}{T+1}\right)\xrightarrow{d}N\left(0,\frac{51}{5}\right)\ \text{when}\ \theta_0=1.$$

As a result, the confidence interval constructed by the result above requires prior information about the stationarity. To be more specific, such problem can make practitioners fall into a trap in choosing between two confidence intervals,

$$\left[\hat{\theta}+\frac{1}{T}(1+\hat{\theta})\pm\frac{\sqrt{(1-\hat{\theta}^2)}}{\sqrt{NT}}\times z_{1-\alpha/2}\right];$$

$$\left[\hat{\theta}+\frac{3}{T+1}\pm\frac{1}{\sqrt{NT^2}}\sqrt{\frac{51}{5}}\times z_{1-\alpha/2}\right].$$

The Andrews' (1993) method helps internalize the choice between stationary and nonstationary cases through simulation and approximate quantile functions. Meanwhile, the bias correction is also internalized by the median-unbiased estimator.

We reconsider the linear panel AR(1) model for $\{y_{it}\}$ with cross-sectional effects and time effects same as the model (1). It can be decomposed as following, where we have the time effects contained in a latent process $\{y_{it}^*\}$:

$$\begin{aligned}&y_{it}=\alpha_i+y_{it}^*\ \text{for}\ t=0,...T;\\&y_{it}^*=\theta_0 y_{it-1}^*+f_t+\varepsilon_{it}\ \text{for}\ t=1,...T.\end{aligned} \tag{8}$$

We can represent the latent process $\{y_{it}^*\}$ in (8) further by another latent process $\{y_{it}^{**}\}$:

$$\begin{aligned} y_{it}^* &= f_t + y_{it}^{**}; \\ y_{it}^{**} &= \theta_0 y_{it-1}^{**} + \varepsilon_{it}. \end{aligned} \tag{9}$$

Then we have the observed panel $\{y_{it}\}$ decomposed into cross-sectional effects, time effects and the second latent AR(1) process,

$$\begin{aligned} y_{it} &= \alpha_i + f_t + y_{it}^{**} \text{ for } t = 0, ...T; \\ y_{it}^{**} &= \theta_0 y_{it-1}^{**} + \varepsilon_{it} \text{ for } t = 1, ...T. \end{aligned} \tag{10}$$

And (10) can be written in the form as model (1):

$$y_{it} = \theta_0 y_{it-1} + (1 - \theta_0)\alpha_i + \underbrace{(f_t - \theta_0 f_{t-1})}_{F_t} + \varepsilon_{it}, \tag{11}$$

where y_{it} are observable univariate, ε_{it} are mean zero scalar innovation terms, θ_0 is autoregressive parameter of interest, α_i is the fixed cross-sectional effect, f_t is the time effect. The subscripts in (10) or (11) $i = 1, \ldots, n$ and $t = 1, \ldots T$ denote the cross-sectional unit and the time index respectively. Notice that the model specified above in (11) is a variant version of (1) in the sense that their MLE are the same if the sample data and innovation terms are identical. For the rest of this paper we focus on the model specified in (10) and (11).

3.1 Panel exactly median unbiased estimation

We impose the following conditions that will be used to derive the main results.

Assumption 3 *(i) ε_{it} is i.i.d. $N(0, \sigma^2)$ across i and t for each i, (ii) $\theta_0 \in (-1, 1]$, (iii) For all $i = 1, \cdots, n$, When $\theta_0 \in (-1, 1)$, $y_{i0}^{**} \sim N(0, \frac{\sigma^2}{1-\theta_0})$; when $\theta_0 = 1$, $y_{i0}^{**} = O_p(1)$.*

In **Assumption** 3.(i) above, we do not exclude the potential correlation between α_i, f_t and ε_{it}, because as shown in **Theorem** 4 below, the distribution of MLE is invariant to time effects or cross-sectional effects. The **Assumption** 3.(iii) imposes the same initial condition as in Andrews (1993) and Phillips and Sul (2003) with which the latent process $\{y_{it}^{**}\}$ is a strictly stationary, Gaussian AR(1) process if $\theta_0 \in (-1, 1)$, meanwhile $\{y_{it}^{**}\}$ is a Gaussian random walk with an $O_p(1)$ initial if $\theta_0 = 1$. Given such assumption, we allow heterogeneity for the initials $y_{i0} = \alpha_i + f_0 + y_{i0}^{**}$. The model involves $1 + n + T + 1$ parameters: $(\sigma^2, \alpha_i, f_t, \theta_0), i = 1, ..., n, t = 0, ..., T$.

Denote $\overline{y}_{i,-1} = \sum_{t=1}^{T} y_{it-1}/(T-1)$, $\overline{y}_{t-1} = \sum_{i=1}^{n} y_{it-1}/n$, $\overline{y}_{-1} = \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it-1}/(n(T-1))$, $\overline{y}_i = \sum_{t=1}^{T} y_{it}/(T-1)$, $\overline{y}_t = \sum_{i=1}^{n} y_{it}/n$, $\overline{y} = \sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}/(n(T-1))$, $\tilde{y}_{it-1} = y_{it-1} - \overline{y}_{i,-1} - \overline{y}_{t-1} + \overline{y}_{-1}$, $\tilde{y}_{it} = y_{it} - \overline{y}_i - \overline{y}_t + \overline{y}$.

The same notation applies to $\bar{\varepsilon}_i$, $\bar{\varepsilon}_t$ and $\bar{\varepsilon}$, $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i - \bar{\varepsilon}_t + \bar{\varepsilon}$. The MLE estimator for θ_0 in (10) can be defined as

$$\hat{\theta} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}\tilde{y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it-1}^2}, \tag{12}$$

We exclude the initial values y_{i0} from the MLE to avoid its effects on the distribution of the estimator.

Theorem 4 *(Invariance Property) Under Assumption 3, the sample distribution of MLE $\hat{\theta}$ (12) depends only on θ_0 when model is correct. In particular, for $\theta_0 \in (-1, 1)$, the sample distribution of $\hat{\theta}$ is invariant to the parameters (α_i, f_t, σ), but depends on the initials y_{i0}. When $\theta_0 = 1$, the sample distribution of $\hat{\theta}$ is invariant to the parameters (α_i, f_t, σ) and the initial values y_{i0}.*

The proof of **Theorem** 4 is given in Appendix E. We can see from the proof, the identically and independently normally distributed innovation terms ε_{it} in **Assumption** 3.(i) is not necessary for the invariance property to be preserved, as long as the innovation terms follow some scaled family. It implies as shown in Phillips and Sul (2003) for their model that we can allow heterogeneity and dependence cross-sectionally or along the time dimension for the innovation terms and the result in **Theorem** 4 still holds. However, the i.i.d. normality is imposed here only for the simplicity of approximation by simulation which is shown in the following **Algorithm** 1. The other implication is, if we have unbalanced panel model or the cross-sectional and time effects are missing at random, the invariance property of MLE allows us to get over with those missing parameters.

As illustrated in **Theorem** 4, the distribution of the MLE $\hat{\theta}$ depends only on θ_0, the parameter we want to construct confidence interval for. Such fact allows us to use numerical methods to describe the distribution of the estimator for a grid of θ_0 as shown in **Algorithm** 1. Notice that we approximate the distribution under the i.i.d. normality in **Assumption** 3.(i). If the innovations actually have other distributions, the approximation under iid normality assumption shows some robustness which is studied by simulation in Section 3.2. The approximated quantile functions are summarised in Table 1,2 and the algorithm is summarized in the following:

Algorithm 1: Approximate quantile functions

Let a set G be a grid of θ_0 which we specify as a sequence from 0 to 1 by the chosen step, say, 0.05. ;
for (*each* θ_0 *from the grid* G) {
  $y_{i0}^{**} = 0, \alpha_i = 0$ for all $i = 1, \cdots, n$;
  $f_t = 0$, for all $t = 1, \cdots, T$;
  follow Assumption 3 to generate S samples;
  for (*each of the* S *samples*) {
    Calculate the MLE;
  }
  Treat the empirical quantile function of these S number of MLE estimators as the populational quantile function $q_p(\theta_0)$ of the MLE $\hat{\theta}$.
}

Let $q_p(\theta)$ be the pth quantile of MLE $\hat{\theta}$, P_θ be the probability distribution when θ is the true value of θ_0. And by definition, $P_\theta(\hat{\theta} < q_p(\theta)) = p$. Table 1 and 2 provide approximate values of the quantile function $q_{0.05}(\theta), q_{0.5}(\theta), q_{0.95}(\theta)$ of MLE $\hat{\theta}$ for different sample sizes based on $S = 1,000$ simulated samples under **Assumption** 3.

Corollary 1 *(Exact Median Bias Correction)*

If the quantile function $q_p(\theta)$ of MLE $\hat{\theta}$ is increasing in θ, we can construct a median-unbiased estimator $\hat{\theta}_U$ of which the median is equal to θ_0:

$$\hat{\theta}_U = \begin{cases} 1 & \text{if } \hat{\theta} > q_{0.5}(1) \\ q_{0.5}^{-1}(\hat{\theta}) & \text{if } q_{0.5}(-1) < \hat{\theta} \leq q_{0.5}(1) \\ -1 & \text{if } \hat{\theta} \leq q_{0.5}(-1) \end{cases} . \tag{13}$$

The assumption of increasing quantile function is supported by the numerical results in the tables above. Andrews (1993) has shown that, $\hat{\theta}_U$ is median-unbiased given increasing quantile function because $\forall \theta_0 \in (-1, 1]$, $\hat{\theta}_U \geq \theta_0$ iff $q_{0.5}(\hat{\theta}_U) \geq q_{0.5}(\theta_0)$ iff $\hat{\theta} \geq q_{0.5}(\theta_0)$, and same results holds with $\geq$ replaced by $\leq$.

Theorem 5 *(Exact Confidence Interval/Set) A* $100(1-p)\%$ *confidence interval(set) for* θ_0 *is given by,*

$$\{\theta \in (-1, 1] : q_{p_1}(\theta) \leq \hat{\theta} \leq q_{p_2}(\theta)\}, \tag{14}$$

where $p_2 > p_1 > 0$, and $1 - p = p_2 - p_1$.

If the quantile functions $q_{p_1}(\theta), q_{p_2}(\theta)$ are both increasing in θ, then the set in (14) equals the interval $\{\theta : c_L \leq \theta \leq c_U\}$. $c_L = q_{p_2}^{-1}(\hat{\theta})$, $c_U = q_{p_1}^{-1}(\hat{\theta})$ and if $\hat{\theta} > q_{p_2}(1)$, $c_L > 1, c_U = \infty$. If the quantile functions are not increasing in θ, we can still build a confidence set by (14).

This interval (set) has the correct coverage probability because

$$P_{\theta_0}\left(\theta_0 \in \{\theta \in (-1,1] : q_{p_1}(\theta) \leq \hat{\theta} \leq q_{p_2}(\theta)\}\right) = P_{\theta_0}\left(q_{p_1}(\theta_0) \leq \hat{\theta} \leq q_{p_2}(\theta_0)\right) = 1 - p$$

The values in Table 1 and 2 give us the approximate quantile functions of the MLE. It confirms the assumption in **Theorem** 5 that the quantille functions are increasing in θ_0. This leads us to use the approximate quantile functions to construct two-sided 90% confidence interval or one-sided 95% confidence interval for θ_0 according to **Theorem** 5.

3.2 Panel feasible generalized median unbiased estimator

The cross-sectional independence assumption is hardly true in panel data. In Phillips and Sul (2003), the Andrews' (1993) method is shown to accommodate the cross-sectional dependence. However, their GLS method is restrictive because they impose **one single common factor** g_t in modeling a general form of cross-sectional dependence.

$$\varepsilon_{it} = \delta_i g_t + u_{it}, \quad g_t \sim \text{ i.i.d. } N(0,1) \text{ over } t,$$
$$E(\varepsilon_{it}\varepsilon_{jt}) = \delta_i \delta_j.$$

In this subsection, we extend Andrews' (1993) method in our setting after we relax the cross-sectional i.i.d assumption and extend Phillips and Sul (2003)'s model to allow upto a finite number p of common time factors, which can be regarded as a multiple factor model as below, with g_t to be the common time factor and δ_i to be the factor loading for series i. And since we model the time effects into the error terms as a random effect, the original fixed time effects f_t is removed . Now we consider the model,

$$y_{it} = \theta_0 y_{it-1} + (1 - \theta_0)\alpha_i + \varepsilon_{it}, \tag{15}$$

where

$$\varepsilon_{it} = \delta_i' g_t + u_{it}, \quad g_t = (g_1, ..., g_p)_t' \sim \text{ i.i.d. } N(0, I_p) \text{ over } t$$
$$u_{i,t} \sim i.i.d.N\left(0, \sigma_i^2\right) \text{ over } t, \text{ and } u_{i,t} \text{ is independent of } u_{j,s} \text{ and } g_s \text{ for all } i \neq j \text{ and for all } s, t. \tag{16}$$
$$E\left(\varepsilon_{it}\varepsilon_{jt}\right) = \delta_i'\delta_j, \quad \delta_i = (\delta_1, ..., \delta_p)_i'.$$

Let $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)'$, we have the conditional covariance matrix,

$$\mathbf{V}_\varepsilon = E\left(\varepsilon_t\varepsilon_t' \mid \sigma_1^2, \ldots, \sigma_N^2\right) = \mathbf{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}', \quad \mathbf{\Sigma} = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right) \tag{17}$$

Let $\tilde{y}_{it} = y_{it} - \overline{y}_i$, $\tilde{\mathbf{y}}_t = (\tilde{y}_{1t}, \ldots, \tilde{y}_{Nt})'$, the panel generalized least squares estimator (PGLS) can be defined as similar to Phillips and Sul (2003),

$$\hat{\theta}_{\text{pgls}} = \frac{\sum_{t=1}^T \tilde{\mathbf{y}}_{t-1}' \mathbf{V}_\varepsilon^{-1} \tilde{\mathbf{y}}_t}{\sum_{t=1}^T \tilde{\mathbf{y}}_{t-1}' \mathbf{V}_\varepsilon^{-1} \tilde{\mathbf{y}}_{t-1}} \tag{18}$$

Here we construct a weaker condition for our panel dynamic model (10) allowing cross-sectional dependence,

Assumption 4 *(i) ε_{it} is assumed as a process in (16) for each i, (ii) $\theta_0 \in (-1, 1]$, (iii) For all $i = 1, \cdots, n$, When $\theta_0 \in (-1, 1)$, $y_0^{**} \sim N\left(0, \frac{1}{1-\theta_0^2} V_\varepsilon\right)$; when $\theta_0 = 1$, $y_{i0}^{**} = O_p(1)$.*

Theorem 6 (Invariance Property 2) *Under assumption 4, the distribution of the panel GLS estimator $\hat{\theta}_{\text{pgls}}$ depends only on θ_0. When $\theta_0 = 1$, the distribution of $\hat{\theta}_{\text{pgls}}$ does not depend on the value of y_{i0} either.*

Proof in appendix F.

We now adopt the Phillips and Sul (2003)'s procedure to construct a feasible panel GLS estimator $\hat{\theta}_{fpgls}$

through panel median-unbiased estimation.

Algorithm 2: Panel feasible generalized median unbiased estimation

1 Construct the panel median-unbiased estimator $\hat{\theta}_U$;

2 Use the residuals from this $\hat{\theta}_U$ to construct the error covariance matrix estimate $\hat{\mathbf{V}}_U$;

3 Perform panel generalized least squares as in (18) using $\hat{\mathbf{V}}_U$, and obtain the feasible estimator,

$$\hat{\theta}_{\text{pfgls}} = \frac{\sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} \hat{\mathbf{V}}_{\text{U}}^{-1} \tilde{\mathbf{y}}_t}{\sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} \hat{\mathbf{V}}_{\text{U}}^{-1} \tilde{\mathbf{y}}_{t-1}} \tag{19}$$

;

4 Using the median function of the $\hat{\theta}_{\text{pfgls}}$ to calculate the feasible panel median-unbiased estimator $\hat{\theta}_{FU}$;

5 Use the $\hat{\theta}_{FU}$ in place of $\hat{\theta}_U$ and repeat steps 2-4 until $\hat{\theta}_{FU}$ converges;

The error covariance matrix estimation in step 2 is following the iterated method of moments procedure in Phillips and Sul (2003).

3.3 PFGLS through thresholding approach

In this subsection, we consider Bai et al. (2021)'s thresholding approach to allow more flexible cross-sectional dependence structure (V_ε) in the model (15) than the parametric method in Phillips and Sul (2003). With such non-parametric method, we can also overcome the high-dimensional estimation problems in Phillips and Sul (2003) when both n and T are large. Besides, the automatic tuning parameter (τ_{ij}) selection using cross-validation can avoid the arbitrariness of the procedure.

Since the serial dependence is not supposed to be included in our AR(1) model (15) as mentioned in Bai (2009), we write the $NT \times NT$ matrix,

$$\Omega = (E\varepsilon_t \varepsilon'_s) = I_T \otimes V_\varepsilon \tag{20}$$

We approximate V_ε by sparsity assumption as in Bickel et al. (2008), to assume cross-sectional weakly dependence,

$$V_\varepsilon^{BL}(i,j) = \begin{cases} E\varepsilon_{it}\varepsilon_{jt}, & \text{if } |E\varepsilon_{it}\varepsilon_{jt}| > \tau_{ij} \\ 0, & \text{if } |E\varepsilon_{it}\varepsilon_{jt}| \leq \tau_{ij} \end{cases} \quad ; \quad \tilde{\Omega} = I_T \otimes V_\varepsilon^{BL} \tag{21}$$

The feasible GLS implementation start with a QMLE residual by,

$$\hat{\theta} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}\tilde{y}_{it-1}\tilde{y}_{it}}{\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}\tilde{y}_{it-1}^{2}} \tag{22}$$
$$\widehat{\varepsilon}_{it} = \tilde{y}_{it} - \widehat{\theta}\tilde{y}_{it-1}$$

Then we estimate the V_{ε},

$$\hat{R}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt}, \quad \text{and } \hat{\sigma}_{ij} = \begin{cases} \hat{R}_{ii}, & \text{if } i = j \\ s_{ij}\left(\hat{R}_{ij}\right), & \text{if } i \neq j \end{cases} \tag{23}$$
$$\widehat{V}_{\varepsilon} = \left(\hat{\sigma}_{ij}\right)_{N\times N}, \quad \widehat{\Omega} = I_T \otimes \widehat{V}_{\varepsilon},$$

where $s_{ij}(z) = \operatorname{sgn}(z)\left(|z| - \tau_{ij}\right)_{+}$, $\tau_{ij} = M \cdot \gamma_T\sqrt{\left|\hat{R}_{ii}\right|\left|\hat{R}_{jj}\right|}$. For some pre-determined value $M > 0$, where $\gamma_T = \sqrt{\frac{logN}{T}}$ is such that $\max_{i,j\leq N} \mid \hat{R}_{ij} - E\varepsilon_{it}\varepsilon_{jt} \mid = O_P\left(\gamma_T\right)$ as in Lemma 5. Note that the constant thresholding parameter could be allowed as Bickel et al. (2008). In practice, however, it is more desirable to have entry dependent threshold, τ_{ij}. M can be chosen by multifold cross-validation (Bai et al. 2021, section 2.2.2).

We have the panel feasible GLS to be,

$$\hat{\theta}_{\text{pfgls}} = \frac{\sum_{t=1}^{T}\tilde{\mathbf{y}}_{t-1}'\widehat{\mathbf{V}}_{\varepsilon}^{-1}\tilde{\mathbf{y}}_{t}}{\sum_{t=1}^{T}\tilde{\mathbf{y}}_{t-1}'\widehat{\mathbf{V}}_{\varepsilon}^{-1}\tilde{\mathbf{y}}_{t-1}} \tag{24}$$

Assumption 5 *(i)* $\{\varepsilon_t\}_{t\geq 1}$ *is strictly stationary. In addition, each* ε_t *has zero mean vector. There exists a constant* $C > 0$ *such that for all* $i \leq N$ *and* $t \leq T$, $E(\varepsilon_{it}^4) < C$. *And* $\max_{i,j\leq N}\left|E\varepsilon_{it}\varepsilon_{jt}\right| = o\left(\sqrt{TlogN}\right)$.

(ii) There are constants $c_1, c_2 > 0$ *such that* $\lambda_{\min}\left(V_{\varepsilon}\right) > c_1$ *and* $\left\|V_{\varepsilon}\right\|_1 < c_2$.

(iii) Exponential tail: There exist $r_1, r_2 > 0$ *and* $b_1, b_2 > 0$, *and for any* $s > 0, i \leq N$,

$$P\left(\left|\varepsilon_{it}\right| > s\right) \leq \exp\left(-\left(s/b_1\right)^{r_1}\right), \quad P\left(\left|\tilde{y}_{it-1}\right| > s\right) \leq \exp\left(-\left(s/b_2\right)^{r_2}\right)$$

(iv) Strong mixing: There exist $\kappa \in (0,1)$ *such that* $r_1^{-1} + r_2^{-1} + \kappa^{-1} > 1$, *and* $C > 0$ *such that for all* $T > 0$,

we have the strong mixing coefficient

$$\alpha(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{T}^{\infty}) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{T}^{\infty}} |P(A)P(B) - P(AB)| < \exp\left(-CT^{\kappa}\right)$$

where $\mathcal{F}_{-\infty}^{0}$ *and* $\mathcal{F}_{T}^{\infty}$ *denote the* σ *-algebras generated by* $\{(\tilde{y}_{t-1}, \varepsilon_t) : t \leq 0\}$ *and* $\{(\tilde{y}_{t-1}, \varepsilon_t) : t \geq T\}$ *respectively.*

(v) Let $\gamma^{-1} = r_1^{-1} + r_2^{-1} + \kappa^{-1} > 1$, $(logN)^{2/\gamma - 1} = o(T)$.

Define a measure of sparsity, for some $q \in [0, 1)$,

$$m_N = \max_{i \leq N} \sum_{j=1}^{N} |V_\varepsilon(i, j)|^q \tag{25}$$

Theorem 7 *Consider the model (15) under the Assumptions 5, when* $\left\|V_\varepsilon^{-1}\right\| = O(1)$, *for some* $q \in [0, 1)$ *such that* $\gamma_T = \sqrt{\frac{logN}{T}}$, $m_N \gamma_T^{1-q} = o(1)$,

$$\|\widehat{V}_\varepsilon - V_\varepsilon\| = O_P\left(m_N \gamma_T^{1-q}\right) = \left\|\widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1}\right\|.$$

Proof in Appendix G.

Assumption 6 *(i)* $\max_{i \leq N, t \leq T} \sum_{s=1}^{T} \sum_{j=1}^{N} \left|V_\varepsilon^{-1}(i, j)\right| = O(1)$,

(ii) There is some $q \in [0, 1)$ *such that* $m_N \gamma_T^{1-q} = o(1)$ *holds. In addition,* $\sqrt{T} m_N^2 \gamma_T^{3-2q} = o(1)$,

(iii) $\sqrt{NT} m_T^3 \gamma_T^{3-3q} = o(1)$.

Theorem 8 *Under the Assumption 5 and Assumption 6 holds, denote* $\tilde{Y}_{-1} = (\tilde{y}_0', \cdots, \tilde{y}_{T-1}')'$, $\mathcal{E} = (\varepsilon_1', \cdots, \varepsilon_T')'$, *we have,*

$$\begin{aligned} \sqrt{NT}\left(\widehat{\theta}_{fpgls} - \theta_0\right) &= \left(\tilde{Y}_{-1}'\Omega^{-1}\tilde{Y}_{-1}/NT\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}_{-1}'\Omega^{-1}\mathcal{E}\right) \\ &+ \left(\tilde{Y}_{-1}'\Omega^{-1}\tilde{Y}_{-1}/NT\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}_{-1}'\Omega^{-1}(\widehat{\Omega} - \Omega)\Omega^{-1}\mathcal{E}\right) + o_P(1) \end{aligned} \tag{26}$$

Proof in Appendix H.

Assumption 7 *Let* $A = \{(i, j) : |E\varepsilon_{it}\varepsilon_{jt}| \neq 0\}$, $w_t' = \mathbf{y}_{t-1}' V_\varepsilon^{-1}$, $u_t = V_\varepsilon^{-1}\varepsilon_t$. *Then*

$$\left\|\frac{1}{\sqrt{NT}} \sum_{i,j \in A} \mathbb{G}_{T,ij}^{1} \mathbb{G}_{T,ij}^{2}\right\| = o_P(1)$$

where $\mathbb{G}^1_{T,ij} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\varepsilon_{it}\varepsilon_{jt} - E\varepsilon_{it}\varepsilon_{jt}\right)$ *and* $\mathbb{G}^2_{T,ij} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_{it}u_{jt}$.

Theorem 9 *Under Assumption 5,6,7,*

$$\begin{aligned}\sqrt{NT}\left(\widehat{\theta}_{fpgls} - \theta_0\right) &= \left(\frac{1}{NT}\tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}\mathcal{E}\right) + o_P(1) \\ &= \sqrt{NT}\left(\widehat{\theta}_{pgls} - \theta_0\right) + o_P(1)\end{aligned}$$

Proof in Appendix I.

The threshold parameter τ_{ij} in (21) or the tuning parameter M can be chosen by a multifold cross-validation procedure (Bai et al. 2021, section 2.2.2).

Algorithm 3: Threshold parameter

We randomly split the data P times;

for (*the split* $p \in (1, 2, \cdots, P)$) {

we divide the data into $P = log(T)$ blocks $(J_1, ..., J_P)$ with block length $T/log(T)$ and take J_p as the validation set;

we denote by $\widetilde{V}^p_\varepsilon = |J_p|^{-1} \sum_{t \in J_p} \widehat{\varepsilon}_t \widehat{\varepsilon}'_t$ the sample covariance matrix based on the validation set;

Let $\widetilde{V}^{S,p}_\varepsilon(M)$ be the thresholding estimator with threshold constant M using the training data set $\{\widehat{\varepsilon}_t\}_{t \notin J_p}$;

}

we choose the constant M^* by minimizing the cross-validation objective function

$$M^* = \arg \min_{c<M<\bar{C}} \frac{1}{P} \sum_{p=1}^{P} \left\| \widetilde{V}^{S,p}_\varepsilon(M) - \widetilde{V}^p_\varepsilon \right\|_F^2$$

where where $\bar{C}$ is a large constant such that $\widetilde{V}^{S,p}_\varepsilon(\bar{C})$ is a diagnal matrix and c is a constant that guarantees the positive definiteness of $\hat{V}_\varepsilon(M)$ for $M > c$;

We propose a comprehensive algorithm that incoporate the thresholding method in the median-unbiased

Algorithm (2) to construct the feasible panel GLS under the flexible cross-sectional dependence.

Algorithm 4: Panel feasible generalized median unbiased estimation with thresholding method

1 Calculate the QMLE $\hat{\theta}$ and residuals $\hat{\varepsilon}_{it}$ as in (22) Use the thresholding method (23) and algorithm 3 to construct the error covariance matrix estimate $\hat{V}_{\varepsilon}$;

2 Perform panel generalized least squares as in (18) using $\hat{V}_{\varepsilon}$, and obtain the feasible estimator,

$$\hat{\theta}_{\text{pfgls}} = \frac{\sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} \hat{V}_{\varepsilon}^{-1} \tilde{\mathbf{y}}_{t}}{\sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} \hat{V}_{\varepsilon}^{-1} \tilde{\mathbf{y}}_{t-1}} \tag{27}$$

;

3 Use the median function of the $\hat{\theta}_{\text{pfgls}}$ to calculate the feasible panel median-unbiased estimator $\hat{\theta}_{FU}$;

4 Monte Carlo Study

In the following two Monte Carlo studies, we first investigate the efficiency of the induced asymptotic distribution from the first part of this paper in the finite sample. Then we include the Andrews' (1993) method from the second part of this paper and Hahn and Moon's (2006) stationary inference to compare the performance of different confidence sets.

4.1 Performance of the bias-corrected MLE in the presence of unit roots and time effects

We conduct two simulation studies to compare the performance of our MLE estimator $\hat{\theta}$ and the bias-corrected estimator $\hat{\theta} + 3/(T+1)$ in finite samples. We generate 5,000 samples from the model

$$y_{it} = \theta_0 y_{it-1} + (1-\theta_0)\alpha_i + f_t + \varepsilon_{it},$$

where $y_{i0} = 0$, $\theta_0 = 1$, $\varepsilon_{it} \sim N(0,1)$ independent across i and t, $f_t \sim N(0,1)$ independent across t. The first simulation has the sample sizes over cross section dimension and time dimension are $n \in \{100, 200, 500, 1000\}$ and $T \in \{5, 10, 20, 100\}$. The second simulation has the sample sizes $n \in \{2, 5\}$ and $T \in \{10, 20, 100, 1000\}$, where T is larger than n. For each sample, we compute the MLE $\hat{\theta}$ based on formulation (2).

The result of the first 5,000 samples simulation study are summarized in Table 3. The result of the second 5,000 samples simulation study are summarized in Table 4. The MLE.MeanBias is the average bias of MLE $\hat{\theta}$ over 5,000 samples. The BCMLE.MeanBias is the average bias of the bias-corrected estimator $\hat{\theta} + 3/(T+1)$

over 5,000 samples. The MLE.RMSE is the root mean square error of MLE $\hat{\theta}$ and the BCMLE.RMSE is the root mean square error of bias-corrected estimator $\hat{\theta} + 3/(T+1)$. We can see from the Table 3 and 2 that with finite sample size, bias-corrected estimator performs much better than the MLE estimator in terms of the finite sample bias and root mean square error.

4.2 Comparison with Andrews' (1993) method and asymptotic inference

Firstly, we want to measure the performance of our exact confidence interval and median-unbiased estimator $\hat{\theta}_U$ under **Assumption** 3. Given some θ_0, we simulate 1,000 samples to compare the 90%-C.I. based on bias-corrected asymptotic distribution to the 90%-C.I. based on **Theorem** 5 with different data size. In particular, as summarized below, when $|\theta_0| < 1$ we use the asymptotic results in Hahn and Moon (2006). When $\theta_0 = 1$, we use the asymptotic results in **Theorem** 2.

$$\sqrt{nT}\left(\hat{\theta} - \theta_0 + \frac{1}{T}(1+\hat{\theta})\right) \xrightarrow{d} N(0, 1-\theta_0^2) \text{ when } |\theta_0| < 1; \tag{28}$$

$$\sqrt{nT^2}\left(\hat{\theta} - \theta_0 + \frac{3}{T+1}\right) \xrightarrow{d} N(0, \frac{51}{5}) \text{ when } \theta_0 = 1. \tag{29}$$

For each sample, we generate the innovation terms $\varepsilon_{it} \overset{i.i.d.}{\sim} N(0,1)$, latent process' initials $y_{10}^{**} = y_{20}^{**} = ... = y_{n0}^{**} \sim N(0,1)$. cross-sectional effects α_i and time effects f_t are assigned by $N(0,1)$ and keep them fixed for each θ_0. Table 5, 6 and 7 contain the coverage frequency of two confidence interval methods over 1,000 samples, where the columns 'Asymp. $\theta_0 < 1$' and 'Asymp. $\theta_0 = 1$' apply the result (28) and (29) to construct the confidence interval, the column 'Andrews' C.I.'. applies **Theorem** 5 to construct the confidence interval.

We also calculate the corresponding bias-corrected estimator according to asymptotic results in (28),(29) and median-unbiased estimator $\hat{\theta}_U$ by **Corollary** 1. Table 8 and 9 show the bias of two types of bias-corrected estimators rounded upto four digits, where the columns 'Asymptotic $\theta_0 < 1$' and 'Asymptotic $\theta_0 = 1$' apply the result (28) and (29) to construct the bias-corrected estimator, the column 'Andrews' C.I.'. applies **Corollary** 1 to construct the median-unbiased estimator.

Secondly, we want to see the robustness of the exact confidence interval assuming i.i.d. normal innovations when the innovations ε_{it} actually follow some other non-Gaussian distributions or have heterogeneity structure. In Table 11, we have the coverage frequency of our exact confidence interval for different data generation process with data size to be $n = 10, T = 50$. t_{10}, t_3, t_1 is when the innovations follow some central t distribution with degree of freedom to be 10,3 or 1. $ARCH(0.3), ARCH(0.85)$ is when the innovations are uncorrelated but conditionally heteroskedastic modelled by ARCH with order 1 and parameter b to be 0.3 or 0.85. That is

$\varepsilon_{it} \sim N(0, \sigma_t^2)$ where $\sigma_t^2 = a + b * \varepsilon_{it-1}^2$. We make $a = 1$ for the fact that the MLE is invariant to a.

4.3 Performance of thresholding PFGLS approach with cross-sectionally dependent error

Following the Algorithm 4, we construct the cross-sectional dependent simulation sample by assuming a clustered block diagonal covariance matrix V_ε. Let the number of clusters to be fixed as $G = 25$ and each diagonal block is one $N/G \times N/G$ matrix with the off-diagonal entries (i, j) in the same cluster, $V_{i,j}$ for $i \neq j$, which are generated from i.i.d. $Uniform(0, \gamma)$. In this study, we set the level of cross-sectional correlation in each cluster as $\gamma = 0.7$. As for the diagonal entries in each cluster, we fix that to be $\sigma^2 = 2$.

In this simulation study, we choose the sample sizes $(N, T) = (50, 50), (100, 50), (50, 100)$ and replicated for $S = 1000$ times. For each sample, we choose the data-driven tuning parameter M and estimate the median functions of $\hat{\theta}_{pfgls}$ by Andrews' method. The simulation process for the median functions is to fix the $G = 25, \gamma = 0.6, \sigma^2 = 1$. To access the performance of this PFGLS, we build the 90% confidence set of θ_0, following the Theorem 5 and investigate the coverage rate. Results are organized in Table 12.

5 Conclusion

The first part of this paper investigates the bias correction method of Hahn and Kuersteiner (2002) when $\theta_0 = 1$ and time effects are linearly included into the panel AR(1) model. We show that the asymptotic bias of MLE can be eliminated if the cross-sectional dimension n and time dimension T are both large. Also, We fill the gap in the literature by relaxing the Gaussian distribution of the errors to general distributions. We explicitly show that Gaussian errors are not required to show the asymptotic normality of bias corrected QMLE.

As for the inference of auto-regressive parameter, one long-standing question is how to obtain a valid confidence interval without knowing the (non)-stationarity. The second part of this paper offers one solution to this question based on the invariance property of the MLE with respect to all nuisance parameters under the regularization condition. This finding allows us to extend Andrews' (1993) method in panel data framework and construct exact confidence interval of panel auto-regressive parameter. When the innovations have some structure other than the i.i.d. normality, the simulation results show that the exact confidence interval is robust under some circumstances. Nevertheless, for theoretical guarantee on the robustness, further investigation of Bootstrap procedure in Hansen (1999), when the innovation distribution exhibits non-Gaussian or i.i.d. structure is suggested. The other routine to accommodate the choice question, for example local-to-unity method and uniformity, is considered in a separate paper we are working on.

References

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

T. W. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, 18(1):47–82, 1982.

D. W. Andrews. Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica: Journal of the Econometric Society*, pages 139–165, 1993. doi: 10.2307/2951781.

M. Arellano and S. Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297, 1991.

J. Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

J. Bai, S. H. Choi, and Y. Liao. Feasible generalized least squares for panel data with cross-sectional and serial correlations. *Empirical Economics*, 60(1):309–326, 2021.

P. J. Bickel, E. Levina, et al. Covariance regularization by thresholding. *The Annals of Statistics*, 36(6): 2577–2604, 2008.

R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143, 1998.

G. S. Bruno. Approximating the bias of the lsdv estimator for dynamic unbalanced panel data models. *Economics letters*, 87(3):361–366, 2005.

M. J. Bun. Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews*, 22(1):29–58, 2003.

Y. Chang. Nonlinear iv unit root tests in panels with cross-sectional dependency. *Journal of econometrics*, 110 (2):261–292, 2002.

Y. Chang. Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of econometrics*, 120 (2):263–293, 2004.

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320, 2011.

J. Fan, Y. Liao, and M. Mincheva. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4), 2013.

S. Gonçalves and M. Kaffo. Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of econometrics*, 186(2):407–426, 2015.

J. Hahn and G. Kuersteiner. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica*, 70(4):1639–1657, 2002. doi: 10.1111/1468-0262.00344.

J. Hahn and H. R. Moon. Reducing bias of mle in a dynamic panel model. *Econometric Theory*, 22(3):499–512, 2006. doi: 10.1017/S0266466606060245.

B. E. Hansen. The grid bootstrap and the autoregressive model. *Review of Economics and Statistics*, 81(4): 594–607, 1999.

R. D. Harris and E. Tzavalis. Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of econometrics*, 91(2):201–226, 1999. doi: 10.1016/S0304-4076(98)00076-1.

D. Holtz-Eakin, W. Newey, and H. S. Rosen. Estimating vector autoregressions with panel data. *Econometrica: Journal of the econometric society*, pages 1371–1395, 1988.

C. Kao. Spurious regression and residual-based tests for cointegration in panel data. *Journal of econometrics*, 90(1):1–44, 1999.

C. Kao and M.-H. Chiang. On the estimation and inference of a cointegrated regression in panel data. In *Nonstationary panels, panel cointegration, and dynamic panels*. Emerald Group Publishing Limited, 2001.

F. Merlevède, M. Peligrad, and E. Rio. A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474, 2011.

A. Mikusheva. Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452, 2007.

A. Mikusheva. One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica*, 80(1):173–212, 2012.

H. R. Moon and P. C. Phillips. Gmm estimation of autoregressive roots near unity with panel data. *Econometrica*, 72(2):467–522, 2004.

H. R. Moon, B. Perron, and P. C. Phillips. Point-optimal panel unit root tests with serially correlated errors. *The Econometrics Journal*, 17(3):338–372, 2014.

J. Neyman, E. L. Scott, et al. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32, 1948. doi: 10.2307/1914288.

S. Nickell. Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426, 1981. doi: 10.2307/1911408.

M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

M. H. Pesaran and E. Tosetti. Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202, 2011.

P. C. Phillips. On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3): 1177–1195, 2014.

P. C. Phillips. Dynamic panel anderson-hsaio estimation with roots near unity. *Econometric Theory*, 34(2): 253–276, 2018.

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

P. C. Phillips and D. Sul. Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal*, 6(1):217–259, 2003.

P. C. B. Phillips and J. C. Chao. Uniform inference in panel autoregression. *Econometrics*, 7(4), 2019.

J. H. Stock. Confidence intervals for the largest autoregressive root in us macroeconomic time series. *Journal of monetary economics*, 28(3):435–459, 1991.

A Tables

	n=2, T=50			n=5, T=50			n=10, T=50			n=100, T=50		
θ_0	0.05	0.5	0.95	0.05	0.5	0.95	0.05	0.5	0.95	0.05	0.5	0.95
0	-0.2655	-0.0173	0.2087	-0.1328	-0.0189	0.1016	-0.0968	-0.0204	0.0601	-0.0429	-0.0214	0.0016
0.05	-0.2208	0.0303	0.2536	-0.0851	0.0299	0.1491	-0.0481	0.028	0.1099	0.0064	0.0277	0.0507
0.1	-0.1791	0.0795	0.2993	-0.0394	0.079	0.1958	-1e-04	0.0769	0.1581	0.0558	0.0769	0.0992
0.15	-0.1393	0.1278	0.3439	0.0081	0.128	0.2415	0.0481	0.1256	0.2059	0.1048	0.126	0.1478
0.2	-0.091	0.1769	0.389	0.058	0.1775	0.2911	0.0982	0.1754	0.2544	0.1532	0.1751	0.1971
0.25	-0.0442	0.2242	0.4331	0.109	0.2248	0.3353	0.1476	0.2244	0.3022	0.2021	0.224	0.2455
0.3	0.0036	0.2727	0.4777	0.1579	0.2733	0.3846	0.1967	0.2727	0.3498	0.2512	0.2728	0.2939
0.35	0.0533	0.3218	0.5191	0.2064	0.3213	0.4294	0.246	0.3214	0.3982	0.3	0.3215	0.3423
0.4	0.1022	0.3698	0.56	0.257	0.3689	0.4752	0.2958	0.3706	0.447	0.349	0.3704	0.3905
0.45	0.1476	0.4168	0.6028	0.3063	0.4167	0.5195	0.3443	0.4198	0.4914	0.3981	0.4192	0.4391
0.5	0.1988	0.4631	0.6434	0.3554	0.4655	0.5632	0.3925	0.4689	0.5395	0.4475	0.4677	0.4874
0.55	0.2514	0.5074	0.6814	0.4055	0.5138	0.608	0.442	0.5166	0.5857	0.4969	0.5165	0.5361
0.6	0.2962	0.5512	0.7191	0.4546	0.5632	0.6546	0.4926	0.5648	0.6315	0.5462	0.5651	0.5846
0.65	0.347	0.5992	0.7582	0.506	0.6137	0.6997	0.5423	0.6135	0.6792	0.5951	0.6138	0.6321
0.7	0.4021	0.6463	0.7938	0.5548	0.6626	0.744	0.593	0.6623	0.7253	0.6445	0.6624	0.6796
0.75	0.4498	0.6963	0.8337	0.6076	0.7094	0.7863	0.6434	0.7107	0.7694	0.693	0.7103	0.7273
0.8	0.5093	0.7426	0.8697	0.66	0.7569	0.8277	0.6961	0.7585	0.813	0.7417	0.7583	0.7742
0.85	0.5632	0.7891	0.9041	0.71	0.8031	0.868	0.7464	0.8048	0.8538	0.7899	0.8053	0.8201
0.9	0.6088	0.8291	0.9356	0.7621	0.8484	0.9056	0.7952	0.8498	0.894	0.8362	0.8506	0.864
0.95	0.679	0.8673	0.9649	0.8127	0.8892	0.9405	0.8415	0.8907	0.9288	0.8798	0.8921	0.9045
0.99	0.7245	0.9045	0.9929	0.8513	0.924	0.9668	0.8828	0.9243	0.9562	0.9156	0.9267	0.9378
1	0.7355	0.9153	1.0005	0.8674	0.9349	0.9781	0.8978	0.9374	0.9672	0.9289	0.9397	0.9498

Table 1: Quantiles of MLE in Small Sample

n=1000, T=100			
θ_0	0.05	0.5	0.95
0	-0.01477	-0.01013	-0.00493
0.05	0.03474	0.03937	0.04462
0.1	0.08406	0.08888	0.09404
0.15	0.13375	0.1384	0.14348
0.2	0.1832	0.18786	0.1929
0.25	0.23269	0.2373	0.24233
0.3	0.28228	0.28672	0.29165
0.35	0.33172	0.33619	0.34104
0.4	0.38131	0.38571	0.3904
0.45	0.43084	0.43524	0.43968
0.5	0.4804	0.48471	0.48908
0.55	0.52999	0.53409	0.53846
0.6	0.5796	0.58355	0.58793
0.65	0.62919	0.63295	0.6372
0.7	0.67871	0.68223	0.68634
0.75	0.72821	0.73151	0.73537
0.8	0.77763	0.78069	0.78431
0.85	0.82683	0.8297	0.83294
0.9	0.87545	0.87813	0.881
0.95	0.9226	0.92482	0.92728
0.99	0.95646	0.95845	0.9603
1	0.96841	0.97005	0.97168

Table 2: Quantiles of MLE in Large Sample

$y_{it} = \theta_0 y_{it-1} + (1-\theta_0)\alpha_i + f_t + \varepsilon_{it}$, $\theta_0 = 1$					
n	T	MLE.MeanBias	BCMLE.MeanBias	MLE.RMSE	BCMLE.RMSE
100	5	-0.502348	-0.002348	0.504769	0.049434
200	5	-0.501182	-0.001182	0.502387	0.034793
500	5	-0.500247	-0.000247	0.500707	0.021472
1,000	5	-0.500078	-0.000078	0.500314	0.015340
100	10	-0.274443	-0.001716	0.275872	0.028096
200	10	-0.273428	-0.000701	0.274123	0.019517
500	10	-0.273033	-0.000306	0.273300	0.012084
1,000	10	-0.272928	-0.000201	0.273071	0.008818
100	20	-0.143758	-0.000900	0.144515	0.014802
200	20	-0.143065	-0.000208	0.143435	0.010295
500	20	-0.143084	-0.000227	0.143238	0.006645
1,000	20	-0.143045	-0.000188	0.143119	0.004629
100	100	-0.029890	-0.000187	0.030054	0.003141
200	100	-0.029771	-0.000068	0.029853	0.002216
500	100	-0.029747	-0.000044	0.029779	0.001389
1,000	100	-0.029730	-0.000027	0.029747	0.000995

Table 3: Performance of bias-corrected estimator

$y_{it} = \theta_0 y_{it-1} + (1-\theta_0)\alpha_i + f_t + \varepsilon_{it}$, $\theta_0 = 1$					
n	T	MLE.MeanBias	BCMLE.MeanBias	MLE.RMSE	BCMLE.RMSE
2	10	-0.412048	-0.139320	0.519567	0.345799
2	20	-0.231160	-0.088303	0.292678	0.200056
2	100	-0.052853	-0.023150	0.069023	0.050067
2	1,000	-0.005489	-0.002492	0.007131	0.005189
5	10	-0.313370	-0.040643	0.347254	0.155036
5	20	-0.166498	-0.023641	0.185510	0.085155
5	100	-0.035557	-0.005854	0.039784	0.018783
5	1,000	-0.003558	-0.000561	0.003977	0.001865

Table 4: Performance of bias-corrected estimator

n=2, T=50				n=5, T=50		
θ_0	Asymp $\theta_0 < 1$	Asymp $\theta_0 = 1$	Andrews' C.I.	Asymp $\theta_0 < 1$	Asymp $\theta_0 = 1$	Andrews' C.I.
0.5	0.665	0.431	0.919	0.756	0.493	0.895
0.75	0.603	0.516	0.909	0.697	0.589	0.904
0.9	0.551	0.633	0.906	0.601	0.713	0.899
0.95	0.498	0.678	0.896	0.527	0.768	0.884
0.99	0.432	0.699	0.913	0.435	0.806	0.895
1	0.416	0.721	0.914	0.432	0.824	0.909

Table 5: Coverage Frequency

n=10, T=50				n=100, T=50		
θ_0	Asymp $\theta_0 < 1$	Asymp $\theta_0 = 1$	Andrews' C.I.	Asymp $\theta_0 < 1$	Asymp $\theta_0 = 1$	Andrews' C.I.
0.5	0.783	0.464	0.913	0.8	0.081	0.871
0.75	0.722	0.568	0.914	0.702	0.191	0.881
0.9	0.623	0.729	0.907	0.387	0.509	0.891
0.95	0.51	0.807	0.906	0.054	0.815	0.887
0.99	0.362	0.849	0.886	0	0.809	0.892
1	0.377	0.88	0.893	0.003	0.892	0.894

Table 6: Coverage Frequency

n=1000, T=100			
θ_0	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.
0.5	0.799	0	0.884
0.75	0.713	0	0.903
0.9	0.222	0	0.906
0.95	0	0.016	0.904
0.99	0	0.404	0.915
1	0	0.887	0.888

Table 7: Coverage Frequency

n=2, T=50				n=5, T=50		
θ_0	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.
0.5	-0.0219	0.0079	-0.0126	-0.0051	0.0244	0
0.75	-0.036	-0.0108	-0.0145	-0.0116	0.0131	-0.0052
0.9	-0.0515	-0.029	-0.0197	-0.0219	1e-04	-0.0062
0.95	-0.061	-0.0392	-0.0306	-0.0302	-0.009	-0.0099
0.99	-0.068	-0.0468	-0.0439	-0.0353	-0.0148	-0.0169
1	-0.0671	-0.0461	-0.0474	-0.0324	-0.0122	-0.0182

Table 8: Bias

n=10, T=50				n=100, T=50		
θ_0	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.
0.5	-0.0037	0.0258	-0.0015	-0.0018	0.0276	0.0011
0.75	-0.0077	0.0169	-0.0024	-0.0047	0.0199	8e-04
0.9	-0.0164	0.0055	-0.0023	-0.0122	0.0096	6e-04
0.95	-0.0244	-0.0033	-0.0037	-0.0199	0.001	0
0.99	-0.0293	-0.009	-0.0081	-0.0249	-0.0046	-5e-04
1	-0.0262	-0.006	-0.0095	-0.0217	-0.0017	-0.002

Table 9: Bias

n=1000, T=100			
θ_0	Asymptotic $\theta_0 < 1$	Asymptotic $\theta_0 = 1$	Andrews' C.I.
0.5	-5e-04	0.0143	-1e-04
0.75	-0.0011	0.0112	0
0.9	-0.003	0.0079	1e-04
0.95	-0.0059	0.0046	0
0.99	-0.0121	-0.0019	-1e-04
1	-0.0103	-3e-04	-4e-04

Table 10: Bias

n=10, T=50					
θ_0	t_{10}	t_3	t_1	$ARCH(0.3)$	$ARCH(0.85)$
0.5	0.9	0.926	0.96	0.918	0.926
0.75	0.911	0.926	0.964	0.937	0.942
0.9	0.9	0.928	0.947	0.918	0.914
0.95	0.9	0.912	0.887	0.914	0.901
0.99	0.8821	0.873	0.648	0.884	0.862
1	0.878	0.868	0.655	0.875	0.857

Table 11: Coverage Frequency

PFGLS			
θ_0	N=50, T=50	N=100, T=50	N=50, T=100
0.5	0.895	0.927	0.905
0.75	0.898	0.879	0.897
0.9	0.866	0.852	0.895
0.95	0.877	0.903	0.878
0.99	0.880	0.887	0.893
1	0.889	0.908	0.895

Table 12: Coverage Frequency

B Proof of Theorem 1

Notations are consistent with the notations defined below Assumption 1. To help simplify the proof, we transform the model $y_{it} = \theta_0 y_{it-1} + (1-\theta_0)\,\alpha_i + f_t + \varepsilon_{it}$ equivalently as following.

$$y_{it} = \alpha_i + y_{it}^*$$
$$y_{it}^* = \theta_0 y_{i,t-1}^* + f_t + \varepsilon_{it}$$

and it is obvious that $\tilde{y}_{it} = y_{it} - \overline{y}_i - \overline{y}_t + \overline{y} = y_{it}^* - \overline{y^*}_i - \overline{y^*}_t + \overline{y^*} = \tilde{y^*}_{it}$.

The MLE is expressed as

$$\hat{\theta}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}^{2}}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}\tilde{y^*}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\theta_0\tilde{y^*}_{it-1}^{2}+\tilde{y^*}_{it-1}\tilde{\varepsilon}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}}=\theta_0+\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}\tilde{\varepsilon}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}}$$

$$\hat{\theta}-\theta_0=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}^*_{it-1}\tilde{\varepsilon}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}}=\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}^*_{it-1}\varepsilon_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}} \tag{30}$$

Proof 1 *When* $\theta_0 = 1$, $y^*_{it} = \sum_{j=1}^{t}(\varepsilon_{ij}+f_j)$, $y^*_{it-1}-\overline{y^*}_{t-1}=\sum_{j=1}^{t-1}\varepsilon_{ij}-\sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij}/n$, $\sum_{t=1}^{T}\sum_{j=1}^{t-1}\varepsilon_{ij} = \sum_{t=1}^{T}(T-t)\varepsilon_{it}$. *We have the following simplification.*

For the denominator in equation Eq.(30):

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^{2}=\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\overline{y^*}_{t-1}-\overline{y^*}_{i-1}+\overline{y^*}_{-1}\right)^2\\
&=\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}(y^*_{it-1}-\overline{y^*}_{t-1})^2-\frac{1}{T}[\sum_{t=1}^{T}y^*_{it-1}-\overline{y^*}_{t-1}]^2\right\}\\
&=\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1}\varepsilon_{ij}-\frac{\sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij}}{n}\right)^2-\frac{1}{T}\left[\sum_{t=1}^{T}(T-t)\varepsilon_{it}-\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it}\right]^2\right\}
\end{aligned}$$

$$\begin{aligned}
&=\sum_{i=1}^{n}\left\{\underbrace{\sum_{t=1}^{T}(\sum_{j=1}^{t-1}\varepsilon_{ij})^2}_{a_i}+\underbrace{\sum_{t=1}^{T}(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij})^2}_{a'}-\underbrace{2\sum_{t=1}^{T}\sum_{j=1}^{t-1}\varepsilon_{ij}\frac{\sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij}}{n}}_{a_i''}\right\}\\
&-\sum_{i=1}^{n}\left\{\frac{1}{T}\left[\underbrace{(\sum_{t=1}^{T}(T-t)\varepsilon_{it})^2}_{b_i}+\underbrace{(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it})^2}_{b'}-\underbrace{2\sum_{t=1}^{T}(T-t)\varepsilon_{it}\frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it}}{n}}_{b_i''}\right]\right\}
\end{aligned} \tag{31}$$

Eq.(31) can be partitioned into 3 parts:

$$\sum_{i=1}^{n}(a_i-\frac{1}{T}b_i)+\sum_{i=1}^{n}(a^{'}-\frac{1}{T}b^{'})-\sum_{i=1}^{n}(a_i^{''}-\frac{1}{T}b_i^{''}),$$

$$
\begin{aligned}
&a_i - \frac{1}{T}b_i = \sum_{t=1}^{T}\frac{1}{T}t(T-t)\varepsilon_{it}^2 + \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)\varepsilon_{it}\varepsilon_{is},\\
&a^{'} - \frac{1}{T}b^{'} = \sum_{t=1}^{T}\frac{1}{T}t(T-t)\overline{\varepsilon}_t^2 + \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)\overline{\varepsilon}_t\overline{\varepsilon}_s,\\
&\sum_{i=1}^{n}a_i^{''} = \frac{2}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sum_{j=1}^{t-1}\varepsilon_{ij}\sum_{j=1}^{t-1}\sum_{i=1}^{n}\varepsilon_{ij}\right] = \frac{2}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[\sum_{j=1}^{t-1}\varepsilon_{ij}\sum_{j=1}^{t-1}\sum_{i=1}^{n}\varepsilon_{ij}\right] = \frac{2}{n}\sum_{t=1}^{T}\left[\sum_{j=1}^{t-1}\sum_{i=1}^{n}\varepsilon_{ij}\right]^2\\
&= 2n\sum_{t=1}^{T}(\sum_{j=1}^{t-1}\overline{u}_j)^2 = 2na^{'},\\
&\sum_{i=1}^{n}b_i^{''} = 2nb^{'}.
\end{aligned}
$$

So that the denominator in Eq.(30) = Eq.(31):

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^2 = \sum_{i=1}^{n}(a_i - \frac{1}{T}b_i) + \sum_{i=1}^{n}(a^{'} - \frac{1}{T}b^{'}) - \sum_{i=1}^{n}(a_i^{''} - \frac{1}{T}b_i^{''})\\
&= \sum_{i=1}^{n}(a_i - \frac{1}{T}b_i) + n(a^{'} - \frac{1}{T}b^{'}) - 2n(a^{'} - \frac{1}{T}b^{'})\\
&= \sum_{i=1}^{n}(a_i - \frac{1}{T}b_i) - n(a^{'} - \frac{1}{T}b^{'})
\end{aligned}
$$

As T fixed and $n \to \infty$, we have:

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - \frac{1}{T}b_i) - (a^{'} - \frac{1}{T}b^{'})\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\frac{1}{T}t(T-t)\varepsilon_{it}^2 + \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)\varepsilon_{it}\varepsilon_{is}\right]\\
&- \sum_{t=1}^{T}\frac{1}{T}t(T-t)\overline{\varepsilon}_t^2 - \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)\overline{\varepsilon}_t\overline{\varepsilon}_s\\
&\xrightarrow{n\to\infty,p}\left[\sum_{t=1}^{T}\frac{1}{T}t(T-t)E(\varepsilon_{it}^2) + \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)E(\varepsilon_{it}\varepsilon_{is})\right]\\
&- \sum_{t=1}^{T}\frac{1}{T}t(T-t)E^2(\varepsilon_{it}) - \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{2}{T}t(T-s)E(\varepsilon_{it})E(\varepsilon_{is})\\
&= \sigma^2\sum_{t=1}^{T}\frac{t(T-t)}{T} = \sigma^2\frac{T^2-1}{6}
\end{aligned} \tag{32}
$$

For the numerator in Eq.(30) :

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}\varepsilon_{it}\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}y^*_{it-1}\varepsilon_{it}-\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\overline{y^*}_{i-1}\varepsilon_{it}-\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\overline{y^*}_{t-1}\varepsilon_{it}+\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\overline{y^*}_{-1}\varepsilon_{it}\\
&=\sum_{t=1}^{T}\frac{\sum\limits_{i=1}^{n}y^*_{it-1}\varepsilon_{it}}{n}-\sum_{t=1}^{T}\frac{\sum\limits_{i=1}^{n}\overline{y^*}_{i-1}\varepsilon_{it}}{n}-\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\sum\limits_{j=1}^{n}y^*_{jt-1}\varepsilon_{it}}{n}\\
&+\frac{1}{n^2T}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it}+\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)f_t\right)\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}
\end{aligned}
$$

As T fixed and $n\to\infty$, by Week Law of Large Number, we have the following:

$$
\begin{aligned}
&\sum_{t=1}^{T}\frac{\sum\limits_{i=1}^{n}y^*_{it-1}\varepsilon_{it}}{n}\xrightarrow{P}\sum_{t=1}^{T}E(y^*_{i,t-1}\varepsilon_{it})=0,\\
&\sum_{t=1}^{T}\frac{\sum\limits_{i=1}^{n}\overline{y^*}_{i-1}\varepsilon_{it}}{n}\xrightarrow{P}E(\overline{y^*}_{i-1}\sum_{t=1}^{T}\varepsilon_{it})=E(\frac{\sum\limits_{t=1}^{T}y^*_{it-1}}{T}\sum_{t=1}^{T}\varepsilon_{it})=\frac{1}{T}E(\sum_{t=1}^{T}(T-t)\varepsilon_{it}\sum_{t=1}^{T}\varepsilon_{it})+0=\frac{T-1}{2}\sigma^2,\\
&\frac{1}{n}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\sum\limits_{j=1}^{n}y^*_{jt-1}}{n}\varepsilon_{it}=\sum_{t=1}^{T}\frac{\sum\limits_{j=1}^{n}y^*_{jt-1}}{n}\frac{\sum\limits_{i=1}^{n}\varepsilon_{it}}{n}\xrightarrow{P}\sum_{t=1}^{T}E(y^*_{it-1})E(\varepsilon_{it})=0,\\
&\frac{1}{n^2T}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)f_t\right]\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}=\left[\sum_{t=1}^{T}(T-t)f_t\right]\frac{\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}\varepsilon_{it}}{nT}\xrightarrow{p}0,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\sum\limits_{t=1}^{T}\sum\limits_{i=1}^{n}(T-t)\varepsilon_{it}\sum\limits_{t=1}^{T}\sum\limits_{i=1}^{n}\varepsilon_{it}}{Tn^2}&=\frac{\sum\limits_{t=1}^{T}(T-t)\bar{\varepsilon}_t\sum\limits_{t=1}^{T}\bar{\varepsilon}_t}{T}=\frac{\sum\limits_{t=1}^{T}(T-t)\bar{\varepsilon}_t^2+\sum\limits_{t=1}^{T-1}\sum\limits_{s=t+1}^{T}(2T-s-t)\bar{\varepsilon}_t\bar{\varepsilon}_s}{T},\\
&\xrightarrow{P}\frac{\sum\limits_{t=1}^{T}(T-t)E(\varepsilon_{it})^2+\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(2T-s-t)E(\varepsilon_{it})E(\varepsilon_{is})}{T}=0.
\end{aligned}
$$

So that

$$
\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^*}_{it-1}\varepsilon_{it}\xrightarrow{P}-\frac{T-1}{2}\sigma^2. \tag{33}
$$

Given Eq.(32) and Eq.(33), as T fixed and $n \to \infty$, by Slutsky's theorem, we have

$$\hat{\theta} - \theta_0 = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}^{*}_{it-1}\tilde{\varepsilon}_{it}}{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y^{*}}^{2}_{it-1}} \xrightarrow{P} \frac{-\frac{T-1}{2}\sigma^2}{\sigma^2\frac{T^2-1}{6}} = -\frac{3}{T+1}. \tag{34}$$

C Proof of Theorem 2

To be consistent with the notations defined below **Assumption** 1, let $F_{t-1} = \sum_{j=1}^{t-1} f_j$, $\overline{F}_{-1} = \sum_{t=1}^{T} F_{t-1}/T$, $u_{it-1} = \sum_{j=1}^{t-1}\varepsilon_{ij}$, $\overline{u}_{t-1} = \sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij}/n$, $\overline{u}_{i,-1} = \sum_{t=1}^{T} u_{it-1}/T$, $\overline{u}_{-1} = \sum_{i=1}^{n}\sum_{t=1}^{T} u_{it-1}/nT$. When $\theta_0 = 1$, we have following:

$y_{it} = y_{it-1} + f_t + \varepsilon_{it}$

And,

$$y_{it-1} = \sum_{j=1}^{t-1} f_j + \sum_{j=1}^{t-1}\varepsilon_{ij} = F_{t-1} + u_{it-1}$$

$$\overline{y}_{t-1} = \sum_{j=1}^{t-1} f_j + \sum_{i=1}^{n}\sum_{j=1}^{t-1}\varepsilon_{ij}/n = F_{t-1} + \overline{u}_{t-1}$$

$$\overline{y}_{i,-1} = \sum_{t=1}^{T}\sum_{j=1}^{t-1} f_j/T + \sum_{t=1}^{T}\sum_{j=1}^{t-1}\varepsilon_{ij}/T = \sum_{t=1}^{T} F_{t-1}/T + \sum_{t=1}^{T} u_{it-1}/T = \overline{F}_{-1} + \overline{u}_{i,-1}$$

$$\overline{y}_{-1} = \overline{F}_{-1} + \sum_{i=1}^{n}\sum_{t=1}^{T} u_{it-1}/nT = \overline{F}_{-1} + \overline{u}_{-1}$$

$$\tilde{y}_{it-1} = F_{t-1} + u_{it-1} - (F_{t-1} + \overline{u}_{t-1}) - (\overline{F}_{-1} + \overline{u}_{i,-1}) + \overline{F}_{-1} + \overline{u}_{-1} = u_{it-1} - \overline{u}_{i,-1} - \overline{u}_{t-1} + \overline{u}_{-1}.$$

Proof 2 *We thus find that in case of unit root, the values of $\tilde{y}_{it-1}$ are invariant to both time and cross-sectional effects α_i 's and time effects f_t 's for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$. And this immediately implies that the probability distribution of MLE does not change with respect to α_i 's and f_t 's and enables us to assume they are all zeros in Eq.(1) without loss of generality, i.e.*

$$y_{it} = \theta_0 y_{it-1} + (1 - \theta_0)\alpha_i + \varepsilon_{it} = y_{it-1} + \varepsilon_{it}. \tag{35}$$

Simplify $\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{\varepsilon}_{it}$ into two parts denoted as a and A:

$$\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{\varepsilon}_{it} = \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1}-(\overline{y}_{t-1}-\overline{y}_{-1}))(\varepsilon_{it}-\overline{\varepsilon}_{i}-(\overline{\varepsilon}_{t}-\overline{\varepsilon}))$$

$$= \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\varepsilon_{it}-\overline{\varepsilon}_{i}) - \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\overline{\varepsilon}_{t}-\overline{\varepsilon})$$

$$-\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\varepsilon_{it}-\overline{\varepsilon}_{i}) + \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{\varepsilon}_{t}-\overline{\varepsilon})$$

$$= \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\varepsilon_{it}-\overline{\varepsilon}_{i}) - \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\overline{\varepsilon}_{t}-\overline{\varepsilon})$$

$$= \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\varepsilon_{it}-\overline{\varepsilon}_{i}) - \frac{\sqrt{n}}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{\varepsilon}_{t}-\overline{\varepsilon}) = a - A,$$

where $a = \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\varepsilon_{it}-\overline{\varepsilon}_{i})$, $A = \frac{\sqrt{n}}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{\varepsilon}_{t}-\overline{\varepsilon})$.

Simplify $\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}^2$ *into two parts denoted as* b *and* B*:*

$$\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it-1}^2 = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1}-(\overline{y}_{t-1}-\overline{y}_{-1}))^2$$

$$= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})^2 - \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\overline{y}_{t-1}-\overline{y}_{-1})$$

$$-\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(y_{it-1}-\overline{y}_{i,-1}) + \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{y}_{t-1}-\overline{y}_{-1})$$

$$= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})^2 - \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(y_{it-1}-\overline{y}_{i,-1})$$

$$= \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})^2 - \frac{1}{T^2}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{y}_{t-1}-\overline{y}_{-1}) = b - B,$$

where $b = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})^2$, $B = \frac{1}{T^2}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{y}_{t-1}-\overline{y}_{-1})$.

Given the panel model of $\{y_{it}\}$ *in (35) and by Theorem 4 in Hahn and Kuersteiner (2002) :*

$$\frac{\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})(\varepsilon_{it}-\overline{\varepsilon}_{i})}{\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}-\overline{y}_{i,-1})^2} + \sqrt{n}T\frac{3}{T+1} \xrightarrow{d} N\left(0, \frac{51}{5}\right)$$

Or with notations defined above,

$$\frac{a}{b} + \sqrt{n}T\frac{3}{T+1} \xrightarrow{d} N\left(0, \frac{51}{5}\right) \tag{36}$$

We can conclude that if $A = o_p(1)$*,* $B = o_p(1)$*, then* $\sqrt{n}T(\hat{\theta} - \theta_0 + \frac{3}{T+1}) = \frac{a-A}{b-B} + \sqrt{n}T\frac{3}{T+1} \xrightarrow{d} N\left(0, \frac{51}{5}\right)$*, since* $\sqrt{n}T(\hat{\theta} - \theta_0 + \frac{3}{T+1}) = \frac{a-A}{b-B} + \sqrt{n}T\frac{3}{T+1}$.

Consider A*,*

$$A = \frac{\sqrt{n}}{T}\sum_{i=1}^{n}\sum_{t=1}^{T}(\overline{y}_{t-1} - \overline{y}_{-1})(\overline{\varepsilon}_t - \overline{\varepsilon}) = \frac{1}{n\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1} - \overline{y}_{i,-1})(\varepsilon_{it} - \overline{\varepsilon}_i)$$
$$+ \frac{1}{n\sqrt{n}T}\sum_{i\neq j}^{n}\sum_{t=1}^{T}(y_{it-1} - \overline{y}_{i,-1})(\varepsilon_{jt} - \overline{\varepsilon}_j).$$

Because we know from Lemma 12 in Hahn and Kuersteiner (2002) $\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1} - \overline{y}_{i,-1})(\varepsilon_{it} - \overline{\varepsilon}_i) \xrightarrow{d} \mathcal{N}\left(0, \frac{17\sigma^4}{60}\right)$*, so that the first term in* A *is of order* $O_p\left(n^{-1}\right) = o_p(1)$*. And* $y_{it-1} - \overline{y}_{i,-1} = u_{it-1} - \overline{u}_{i,-1}$*, if* $\frac{1}{n\sqrt{n}T}\sum_{i\neq j}^{n}\sum_{t=1}^{T}\left(u_{it-1} - \overline{u}_{i,-1}\right)\varepsilon_{jt} = o_p(1)$ *then the second term in* A *is of order* $o_p(1)$*.*

$$\frac{1}{n\sqrt{n}T}\sum_{i\neq j}^{n}\sum_{t=1}^{T}\left(u_{it-1} - \overline{u}_{i,-1}\right)\varepsilon_{jt} = \frac{1}{n\sqrt{n}T}\sum_{i\neq j}^{n}\sum_{t=1}^{T}u_{it-1}\varepsilon_{jt}$$
$$- \frac{1}{n^2}\sum_{i\neq j}^{n}\overline{u}_{i,-1}\varepsilon_{jt}. \tag{37}$$

The first term on the right of Eq.(37) has mean zero and variance equal to:

$$
\begin{aligned}
E\left[\left(\frac{1}{n\sqrt{n}T}\sum_{i\neq j}^{n}\sum_{t=1}^{T}u_{it-1}\varepsilon_{jt}\right)^2\right] &= \frac{1}{n^3T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i\neq j}^{n}E[u_{it-1}\varepsilon_{jt}u_{is-1}\varepsilon_{js}] \\
&= \frac{n(n-1)}{n^3}\frac{1}{T^2}\sum_{t=1}^{T}E[u_{it-1}^2\sigma^2] \\
&= \frac{n(n-1)}{n^3}\frac{\sigma^2}{T^2}\sum_{t=1}^{T}E[(\sum_{j=1}^{t-1}\varepsilon_{ij})^2] \\
&= \frac{n(n-1)}{n^3}\frac{\sigma^2}{T^2}\sum_{t=1}^{T}E(\sum_{j=1}^{t-1}\varepsilon_{ij}^2) \\
&= \frac{n(n-1)}{n^3}\frac{\sigma^2}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\sigma^2 \\
&= \frac{n(n-1)}{n^3}\frac{\sigma^4}{T^2}\frac{T(T-1)}{2} = O\left(\frac{1}{n}\right) = o(1).
\end{aligned}
$$

The first term on the right of Eq.(37) is of order $o_p(1)$ by Markov Inequality.

The second term on the right of Eq.(37) has mean zero and by Week Law of Large Number, we have:

$$
\frac{1}{n^2}\sum_{i\neq j}^{n}\overline{u}_{i,-1}\varepsilon_{jt} = \frac{1}{n^2T}\sum_{i\neq j}^{n}\sum_{s=1}^{T}\varepsilon_{i,s-1}\varepsilon_{j,t} = o_p(1).
$$

As a result, A is of order $o_p(1)$.

Consider B,

$$
\begin{aligned}
B = \frac{1}{T^2}\sum_{t=1}^{T}(\overline{y}_{t-1}-\overline{y}_{-1})(\overline{y}_{t-1}-\overline{y}_{-1}) &= \frac{1}{T^2}\sum_{t=1}^{T}(\overline{u}_{t-1}-\overline{u}_{-1})(\overline{u}_{t-1}-\overline{u}_{-1}) \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\overline{u}_{t-1}^2 - \frac{1}{T}\overline{u}_{-1}^2.
\end{aligned} \tag{38}
$$

and note that we have:

$$
E\left[\left|\frac{1}{T^2}\sum_{t=1}^{T}\overline{u}_{t-1}^2\right|\right] \leq \frac{1}{nT^2}\sum_{t=1}^{T}E\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_{it-1}\right|^2\right] = \frac{1}{nT^2}\frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{t-1}E(\varepsilon_{ij}^2)}{n} = O\left(\frac{1}{n}\right).
$$

By Markov Inequality, we know the first part on the right side of Eq.(38) is of order $o_p(1)$.

$\frac{1}{\sqrt{T}}\overline{u}_{-1} = \sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it}/nT^{3/2}$ has mean zero and variance equal to:

$$\frac{1}{n^2T^3}E(\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)\varepsilon_{it})^2 = \frac{1}{n^2T^3}\sum_{i=1}^{n}\sum_{t=1}^{T}(T-t)^2E(\varepsilon_{it}^2) = \frac{\sigma^2}{n^2T^3}O(nT^3) = O\left(\frac{1}{n}\right).$$

Therefore, by Markov inequality, we know

$$\frac{1}{\sqrt{T}}\overline{u}_{-1} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

The second part on the right side of Eq.(38) is of order $o_p(1)$. Then B is of order $o_p(1)$.

Given the proof above,

$$\sqrt{n}T(\hat{\theta} - \theta_0 + \frac{3}{T+1}) = \frac{a-A}{b-B} + \sqrt{n}T\frac{3}{T+1} \xrightarrow{d} N\left(0, \frac{51}{5}\right).$$

D Proof of Theorem 3

Lemma 1,2,3 show the fact that the result of Lemma 11 and Lemma 12 in Hahn and Kuersteiner (2002) hold under our Assumption 2. Then, we can follow the steps since (36) to finish the proof of **Theorem** 3.

Let's consider the model: $y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}$ and assumptions:

Assumption 8 *(1) ε_{it} is i.i.d. across i, t, $E(\varepsilon_{it}) = 0$, $Cov(\varepsilon_{it}, \varepsilon_{is}) = \sigma^2 1_{[s=t]}$. (2) $\alpha_i = 0$ (3)$\theta_0 = 1$ (4) $n, T \to \infty$ (5) $\mu_k := E(\varepsilon_{it})^k < \infty, k = 1, 2, 3, 4, 5, 6, 7, 8$.*

Ignore the subscription i whenever obvious, we have: $y_t = \alpha + y_{t-1} + \varepsilon_t$

Use the same notation in Hahn and Kuersteiner(2002), let $y = (y_1, \ldots, y_T)', y_- = (y_0, \ldots, y_{T-1})', \varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, we have the matrix version:

$$y_- = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} y_0 + \begin{pmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ T-1 \end{pmatrix} \alpha + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \varepsilon \tag{39}$$

$$= \xi_{1T} y_0 + \xi_{2T}\alpha + A_T\varepsilon$$

Let the within transform operator over time be $H_T = I_T - \frac{1}{T}\ell_T\ell_T'$, $D_T \equiv H_T A_T$. We have when $\alpha = 0$,

$$\varepsilon' H_T y_- = \varepsilon' D_T \varepsilon, \quad y_-' H_T y_- = \varepsilon' D_T' D_T \varepsilon \tag{40}$$

$$D_T = \begin{bmatrix} \frac{1}{T}-1 & \frac{2}{T}-1 & \frac{3}{T}-1 & \cdots & 0 \\ \frac{1}{T} & \frac{2}{T}-1 & \frac{3}{T}-1 & \cdots & 0 \\ \frac{1}{T} & \frac{2}{T} & \frac{3}{T}-1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \cdots \\ \frac{1}{T} & \frac{2}{T} & \frac{3}{T} & ... & 0 \end{bmatrix} \tag{41}$$

Lemma 1 *(Quadratic Moments of Non-Gaussian) Let $M = \frac{D_T + D_T'}{2}$ s.t. Under Assumption 8, $\varepsilon' D_T \varepsilon = \varepsilon' M \varepsilon$.*

$$E(\varepsilon' M \varepsilon)^i = O(T^i), i = 1, 2, 3, 4.$$

Proof 3

$$E(\varepsilon' D_T \varepsilon) = trace(D_T E(\varepsilon\varepsilon')) = \sigma^2 trace(D_T) = \sigma^2 \frac{1-T}{2}$$
$$Var(\varepsilon' M \varepsilon) = (\mu_4 - 3\sigma^4)\sum_{t=1}^{T} m_{tt}^2 + 2\sigma^4 trace(M^2) = (\mu_4 - 3\sigma^4)\frac{T(T-1)(2T-1)}{6T^2} + 2\sigma^4(\frac{T^2}{24} + \frac{T}{4} - \frac{7}{24})$$
$$= O(T^2)$$

Given the result from Lemma 10 in Hahn and Kuersteiner (2002), $trace(M)^3 = O(T^3)$, $trace(M)trace(M^2) = O(T^3)$, $trace(M^3) = O(T^2)$,

$$
\begin{aligned}
E(\varepsilon' M\varepsilon)^3 &= E\sum_i\sum_j\sum_k\sum_l\sum_a\sum_b m_{ij}m_{kl}m_{ab}\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l\varepsilon_a\varepsilon_b \\
&= \mu_6\sum_t m_{tt}^3 + \mu_2\mu_4\sum_i\sum_{j\neq i}\left[3m_{ii}m_{jj}^2 + 12m_{ij}^2m_{jj}\right] + \mu_3^2\sum_i\sum_{j\neq i}\left[6m_{ii}m_{jj}m_{ij} + 4m_{ij}^3\right] + \\
&\mu_2^3\sum_i\sum_{j\neq i}\sum_{k\neq j,i}\left[m_{ii}m_{jj}m_{kk} + 6m_{ij}^2m_{kk} + 4m_{ij}m_{ik}m_{jk}\right] \\
&= C\sum_t m_{tt}^3 + \mu_2\mu_4\sum_{i,j}\left[3m_{ii}m_{jj}^2 + 12m_{ij}^2m_{jj}\right] + \mu_3^2\sum_{i,j}\left[6m_{ii}m_{jj}m_{ij} + 4m_{ij}^3\right] + \\
&\mu_2^3\sum_{i,j,k}\left[m_{ii}m_{jj}m_{kk} + 6m_{ij}^2m_{kk} + 4m_{ij}m_{ik}m_{jk}\right] \\
&= C\sum_t m_{tt}^3 + \mu_2\mu_4\left[3\sum_{i,j}m_{ii}m_{jj}^2 + 12\sum_{i,j}m_{ij}^2m_{jj}\right] + \mu_3^2\left[6\sum_{i,j}m_{ii}m_{jj}m_{ij} + 4\sum_{i,j}m_{ij}^3\right] + \\
&\mu_2^3\left[trace(M)^3 + 6trace(M)trace(M^2) + 4trace(M^3)\right]
\end{aligned}
$$

Easy to show: $\left|\frac{\sum_{ij}m_{ij}^2m_{jj}}{\sum_{ijk}m_{ij}^2m_{kk}}\right| < 1$, $\left|\frac{\sum_{ij}m_{ii}m_{jj}m_{ij}}{\sum_{ijk}m_{ii}m_{jj}m_{kk}}\right| < 1$, $\left|\frac{\sum_{ij}m_{ij}^3}{\sum_{ijk}m_{ij}^2m_{kk}}\right| < 1$,

So that,

$$E(\varepsilon' M\varepsilon)^3 = O(T^3) \tag{42}$$

$$E\varepsilon' M\varepsilon^4 = E\sum_i\sum_j\sum_k\sum_l\sum_a\sum_b\sum_c\sum_d m_{ij}m_{kl}m_{ab}m_{cd}\varepsilon_i\varepsilon_j\varepsilon_k\varepsilon_l\varepsilon_a\varepsilon_b\varepsilon_c\varepsilon_d$$

$$= \mu_8\sum_t m_{tt}^4 + \mu_2\mu_6\left[4\sum_i\sum_{j\neq i} m_{ii}m_{jj}^3 + 12\sum_i\sum_{j\neq i} m_{ij}^2 m_{jj}^2\right] +$$

$$\mu_3\mu_5\left[8\sum_i\sum_{j\neq i} m_{ii}m_{ij}m_{jj}^2 + 32\sum_i\sum_{j\neq i} m_{ij}^3 m_{jj}\right] +$$

$$\mu_4^2\sum_i\sum_{j\neq i}\left[4m_{ii}^2m_{jj}^2 + 12m_{ii}m_{ij}^2m_{jj}\right] +$$

$$\mu_2\mu_3^2\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\left[8m_{ii}m_{jj}m_{jk}m_{kk} + 16m_{ii}m_{jk}^3 + 48m_{ij}^2m_{jk}m_{kk} + 192m_{ij}m_{ik}m_{jk}^2\right]$$

$$\mu_2^2\mu_4\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\left[6m_{ii}m_{jj}m_{kk}^2 + 12m_{ij}^2m_{kk}^2 + 24m_{ii}m_{jk}^2m_{kk} + 48m_{ij}m_{ik}m_{jk}m_{kk} + 48m_{ik}^2m_{jk}^2\right] +$$

$$\mu_2^4\sum_i\sum_{j\neq i}\sum_{k\neq i,j}\sum_{l\neq i,j,k}[m_{ii}m_{jj}m_{kk}m_{ll} + 12m_{ii}m_{jj}m_{kl}^2 + 32m_{ii}m_{jk}m_{jl}m_{kl} + 8m_{ij}^2m_{kl}^2 +$$

$$24m_{ij}m_{ik}m_{jl}m_{kl}]$$

$$= C\sum_t m_{tt}^4 + \mu_2\mu_6\sum_{ij}\left[4m_{ii}m_{jj}^3 + 12m_{ij}^2m_{jj}^2\right] +$$

$$\mu_3\mu_5\sum_{ij}\left[8m_{ii}m_{ij}m_{jj}^2 + 32m_{ij}^3m_{jj}\right] +$$

$$\mu_4^2\sum_{ij}\left[4m_{ii}^2m_{jj}^2 + 12m_{ii}m_{ij}^2m_{jj}\right] +$$

$$\mu_2\mu_3^2\sum_{ijk}\left[8m_{ii}m_{jj}m_{jk}m_{kk} + 16m_{ii}m_{jk}^3 + 48m_{ij}^2m_{jk}m_{kk} + 192m_{ij}m_{ik}m_{jk}^2\right] +$$

$$\mu_2^2\mu_4\sum_{ijk}\left[6m_{ii}m_{jj}m_{kk}^2 + 12m_{ij}^2m_{kk}^2 + 24m_{ii}m_{jk}^2m_{kk} + 48m_{ij}m_{ik}m_{jk}m_{kk} + 48m_{ik}^2m_{jk}^2\right] +$$

$$\mu_2^4[trace(M)^4 + 12trace(M)^2trace(M^2) + 32trace(M)trace(M^3) + 8trace(M^2)^2 +$$

$$24trace(M^4)]$$

Given the result from Lemma 10 in Hahn and Kuersteiner (2002),

$trace(M)^4 = \sum_{ijkl} m_{ii}m_{jj}m_{kk}m_{ll} = O(T)$, $trace(M)^2trace(M^2) = \sum_{ijkl} m_{ii}m_{jj}m_{kl}^2 = O(T^3)$, $trace(M)trace(M^3) = \sum_{ijkl} m_{ii}m_{jk}m_{jl}m_{kl} = O(T^3)$, $trace(M^2)^2 = \sum_{ijkl} m_{ij}^2m_{kl}^2 = O(T^4)$, $trace(M^4) = \sum_{ijkl} m_{ij}m_{ik}m_{jl}m_{kl} = O(T^4)$

We can conclude the order of the fourth moment is dominated by the last term (μ_2^4) which means that $E\varepsilon' M\varepsilon^4 = O(T^4)$, *thus* $E(\varepsilon' M\varepsilon - E(\varepsilon' M\varepsilon))^4 = O(T^4)$.

Lemma 2 *Under Assumption 8,*

$$\frac{1}{nT^2}\sum_i \varepsilon D_T' D_T \varepsilon \xrightarrow{p} \frac{\sigma^2}{6}$$

Proof 4 $E(\varepsilon D_T' D_T \varepsilon) = \sigma^2 trace(D_T' D_T) = \sigma^2 \frac{T^2-1}{6}$,

Denote the t-th diagnal entry of $D_T' D_T := d'd_{tt} = (\frac{t}{T}-1)^2 t + (\frac{t}{T})^2(T-t)$, *we have* $\sum_t d'd_{tt}^2 = O(T^2)$ *and,*

$$\begin{aligned} Var(\varepsilon D_T' D_T \varepsilon) &= (\mu_4 - 3\sigma^4)\sum_{t=1}^{T} d'd_{tt}^2 + 2\sigma^4 trace((D_T' D_T)^2) \\ &= (\mu_4 - 3\sigma^4)O(T^2) + 2\sigma^4 O(T^4) = O(T^4). \end{aligned}$$

So that $Var(\frac{1}{nT^2}\sum_i \varepsilon D_T' D_T \varepsilon) = \frac{nO(T^4)}{n^2T^4} \to 0$,

By Chebyshev inequality,

$$\frac{1}{nT^2}\sum_i \varepsilon D_T' D_T \varepsilon \xrightarrow{p} plim E(\frac{1}{nT^2}\sum_i \varepsilon D' D \varepsilon) = \frac{\sigma^2}{6} \tag{43}$$

Lemma 3 *Under Assumption 8,*

$$\frac{1}{\sqrt{n}T}\sum_i \varepsilon^{'}(D_T + \frac{3}{T+1}D_T' D_T)\varepsilon \xrightarrow{d} N(0, \frac{17\sigma^4}{60})$$

Proof 5 *Let* $G_T := D_T + \frac{3}{T+1}D_T' D_T$, *we have* $E(\varepsilon^{'} G_T \varepsilon) = 0$. *And by ineuqlity* $(a+b)^4 \leq 8(a^4+b^4)$, *we have*

$$E(\varepsilon^{'} G_T \varepsilon - E(\varepsilon^{'} G_T \varepsilon))^4 = E(\varepsilon^{'} G_T \varepsilon)^4 \leq 8E(\varepsilon' D_T \varepsilon)^4 + 8E(\frac{3}{T+1}\varepsilon' D_T' D_T \varepsilon)^4$$

Learn from the proof of Lemma 1, $E[\varepsilon^{'} G_T \varepsilon]^4 = O(T^4)$.

And, let the t-th diagnal entry of $G_T := g_{tt}$, *we have* $\sum_t g_{tt} = O(T)$:

$$\begin{aligned} Var(\varepsilon' G_T \varepsilon) &= \frac{1}{4}Var(\varepsilon'(G_T + G_T')\varepsilon) \\ &= (\mu_4 - 3\sigma^4)\sum_{t=1}^{T} g_{tt}^2 + \frac{1}{2}\sigma^4 trace((G_T + G_T')^2) \\ &= (\mu_4 - 3\sigma^4)O(T) + \frac{1}{2}\sigma^4\frac{17}{30}O(T^2) = \sigma^4\frac{17}{60}O(T^2) \end{aligned}$$

By Lyapunov CLT with condition $\frac{\sum_i E[\varepsilon^{'} G_T \varepsilon]^{2+2}}{(\sum_i Var(\varepsilon^{'} G_T \varepsilon))^2} = \frac{nO(T^4)}{n^2 O(T^4)} \to 0$,

$$\frac{\sum_i \varepsilon^{'} G_T \varepsilon - E(\varepsilon^{'} G_T \varepsilon)}{\sqrt{nVar(\varepsilon^{'} G_T \varepsilon)}} \xrightarrow{d} N(0,1) \tag{44}$$

$$\frac{\sum_i \varepsilon^{'} G_T \varepsilon}{\sqrt{nT^2}} \xrightarrow{d} N(0, \frac{17\sigma^4}{60}) \tag{45}$$

By Slusky theory and the result of Lemma 2 and 3 above, we have

$$\frac{\frac{1}{\sqrt{nT^2}} \sum_i \varepsilon^{'} G_T \varepsilon - E(\varepsilon^{'} G_T \varepsilon)}{\frac{1}{nT^2} \sum_i \varepsilon' D_T' D_T \varepsilon} \xrightarrow{d} \frac{N(0, \frac{17\sigma^4}{60})}{\sigma^2/6} = N(0, \frac{51}{5}) \tag{46}$$

E Proof of Theorem 4

Proof 6 *The notations are consistent with the list defined below **Assumption** 3.*

By the definition of MLE (12),

$$\hat{\theta} = \frac{\sum_{i=1}^n \sum_{t=1}^T \tilde{y}_{it-1} \tilde{y}_{it}}{\sum_{i=1}^n \sum_{t=1}^T \tilde{y}_{it-1}^2} = \frac{\sum_{i=1}^n \sum_{t=1}^T \tilde{y}_{it-1}^{**} \tilde{y}_{it}^{**}}{\sum_{i=1}^n \sum_{t=1}^T \tilde{y}_{it-1}^{**2}}. \tag{47}$$

Let $u_{i,t-1} = \varepsilon_{it-1} + \theta_0 \varepsilon_{it-2} + ... + \theta_0^{t-2} \varepsilon_{i1}$*. According to the model (10), we have the following,*

$$\begin{aligned}
y_{it-1}^{**} &= u_{it-1} + \theta_0^{t-1} y_{i0}^{**} \\
\overline{y}_{\cdot,t-1}^{**} &= \overline{u}_{\cdot,t-1} + \theta_0^{t-1} \overline{y}_{\cdot,0}^{**} \\
\overline{y}_{i,-1}^{**} &= \overline{u}_{i,-1} + \left(\frac{1}{T} \sum_{t=1}^T \theta_0^{t-1} \right) y_{i0}^{**} \\
\overline{y}_{-1}^{**} &= \overline{u}_{-1} + \left(\frac{1}{T} \sum_{t=1}^T \theta_0^{t-1} \right) \overline{y}_{\cdot,0}^{**}
\end{aligned}$$

Let $\tilde{u}_{it-1} = u_{it-1} - \overline{u}_{\cdot,t-1} - \overline{u}_{i,-1} + \overline{u}_{-1}$*, and* $y_{i0}^{**} = \overline{y}_{,0}^{**}$*,*

$$\tilde{y}_{it-1}^{**} = \tilde{u}_{it-1} + \left(\theta_0^{t-1} - \left(\frac{1}{T} \sum_{t=1}^T \theta_0^{t-1} \right) \right) (y_{i0}^{**} - \overline{y}_{,0}^{**}) = \tilde{u}_{it-1}, \tag{48}$$

From(47) and (48), the distribution of the MLE is invariant to $(\alpha_i, f_t, y_{i0}, y_{i0}^{**})$ *for* θ_0 *in* $(-1, 1]$ *under **Assumption** 3. In particular, if* $\theta_0 = 1$*, we can impose no assumption on initial values* y_{i0} *to preserve the invariance.*

As for σ^2*, we can multiply the model by a non-zero constant c, so that the innovation term becomes* $c\varepsilon \sim N(0, c^2\sigma^2)$*. Given the definition of MLE (12), we find the constant c is crossed off. As a result, the distribution* *of MLE is invariant to the innovation term's variance* σ^2.

F Proof of Theorem 6

Proof 7 *Given* $\tilde{y}_{it} = \tilde{y}^{**}_{it}$, $\mathbf{y}^{**}_t = (y^{**}_{1t}, \ldots, y^{**}_{Nt})'$, $y^{**} = vec\,(y^{**}_1, \ldots, y^{**}_T)$, $y^{**}_{-1} = vec\left(y^{**}_0, \ldots, y^{**}_{T-1}\right)$, *we can rewrite the PGLS (18) as:*

$$\hat{\theta}_{\text{pgls}} = \frac{\sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1}\mathbf{V}_\varepsilon^{-1}\tilde{\mathbf{y}}_t}{\sum_{t=1}^{T} \hat{\mathbf{y}}'_{t-1}\mathbf{V}_\varepsilon^{-1}\tilde{\mathbf{y}}_{t-1}} = \frac{\sum_{t=1}^{T} \tilde{\mathbf{y}}^{**\prime}_{t-1}\mathbf{V}_\varepsilon^{-1}\tilde{\mathbf{y}}^{**}_t}{\sum_{t=1}^{T} \tilde{\mathbf{y}}^{**\prime}_{t-1}\mathbf{V}_\varepsilon^{-1}\tilde{\mathbf{y}}^{**}_{t-1}} = \frac{\tilde{\boldsymbol{y}}^{**\prime}_{-1}\left(\mathbf{I}_T \otimes \mathbf{V}_\varepsilon^{-1}\right)\tilde{\boldsymbol{y}}^{**}}{\tilde{\boldsymbol{y}}^{**\prime}_{-1}\left(\mathbf{I}_T \otimes \mathbf{V}_\varepsilon^{-1}\right)\tilde{\boldsymbol{y}}^{**}_{-1}}$$

The invariance to the (α_i, f_t) *is same as the proof for Theorem 4. To show the invariance to dependence, we first check the stationary case. Denote* $T \times T$ *matrix* Ω *of which the* $ijth$ *element to be* $\theta_0^{|i-j|}/(1-\theta_0^2)$. *Then obviously,* $y^{**}_t \sim N\left(\mathbf{0}, \frac{1}{1-\theta_0^2}\mathbf{V}_\varepsilon\right)$ *and* $y^{**} \sim N\left(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{V}_\varepsilon\right)$. *As a result,*

$$\begin{aligned} \mathbf{V}_\varepsilon^{-\frac{1}{2}} y^{**}_t &\sim N\left(\mathbf{0}, \frac{1}{1-\theta_0^2} I_N\right) \\ \left(\mathbf{I}_N \otimes \mathbf{V}_u^{-\frac{1}{2}}\right) y^{**} &\sim N\left(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{I}_N\right) \end{aligned} \tag{49}$$

Let Q *be the orthogonal projection matrix that do the within-transformation for that* $Qy^{**} = \tilde{y}^{**}$.

$$\left(\mathbf{I}_N \otimes \mathbf{V}_\varepsilon^{-\frac{1}{2}}\right)\tilde{y}^{**}, \left(\mathbf{I}_N \otimes \mathbf{V}_\varepsilon^{-\frac{1}{2}}\right)\tilde{y}^{**}_{-1} \sim N\left(\mathbf{0}, \mathbf{Q}(\mathbf{\Omega} \otimes \mathbf{I}_N)\mathbf{Q}\right) \tag{50}$$

As shown above, the distribution of the GLS estimator only depends on θ_0 *contained in the* Ω.

Again, proofs in the non-stationary case $\theta_0 = 1$ *carry over in a similar fashion as we prove in the Theorem 4, and we can see in (48) the initials are removed.*

G Proof of theorem 7

Lemma 4 *For model (15), under Assumption 5,*

$$\max_{i,j \le N}\left|\frac{1}{T}\sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{jt}\right| = O_P\left(\sqrt{\frac{logN}{T}}\right)$$

Proof 8 (Lemma 4) *By (Fan et al. 2011, Lemma A.2), the product of two RV with exponential tails satisfies* *the exponential tail condition. For some b_3 and $r_3 \in (0, r_1 r_2/(r_1+r_2))$, we have for any $s \geq 0$.*

$$P\left(|\tilde{y}_{it-1}\varepsilon_{jt}| > s\right) \leq \exp\left(-\left(s/b_3\right)^{r_3}\right).$$

Let centered $\gamma_{ij,t} = \tilde{y}_{it-1}\varepsilon_{jt} - E\tilde{y}_{it-1}\varepsilon_{jt}$, by Bonferroni method we have,

$$P\left(\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\gamma_{ij,t}\right| > x\right) \leq N^2 \max_{ij} P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\gamma_{ij,t}\right| > x\right) \leq N^2 \max_{ij} P\left(\sup_{\tau}\left|\frac{1}{T}\sum_{t=1}^{\tau}\gamma_{ij,t}\right| > x\right),$$

Then using Bernstein inequality for weakly dependant data in (Merlevède et al. 2011, Theorem 1), and for some $c > 0$, let $x = c\sqrt{\frac{logN}{T}}$, $\gamma^{-1} = r_3^{-1} + \kappa^{-1} > 1$, $(logN)^{(2/\gamma)-1} = o(T)$, we have (Fan et al. 2011, LEMMA A.3),

$$\begin{aligned} P\left(\sup_{\tau \leq T}\left|\sum_{t=1}^{\tau}\gamma_{ij,t}\right| > Tx\right) &\leq T\exp\left(-\frac{(Tx)^{\gamma}}{C_1}\right) + \exp\left(-\frac{(Tx)^2}{C_2(1+TC_5)}\right) + \exp\left(-\frac{(Tx)^2}{C_3 T}\exp\left(\frac{(Tx)^{\gamma(1-\gamma)}}{C_4(\log Tx)^{\gamma}}\right)\right) \\ &= O\left(\frac{1}{N^4}\right) \end{aligned}$$

So that, $P\left(\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\gamma_{ij,t}\right| > c\sqrt{\frac{logN}{T}}\right) \to 0$, and $\max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\gamma_{ij,t}\right| = O_P\left(\sqrt{\frac{logN}{T}}\right)$.

$$\begin{aligned} \max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt}\right| &\leq \max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt} - E(\tilde{y}_{it-1}\varepsilon_{jt})\right| + \max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}E(\tilde{y}_{it-1}\varepsilon_{jt})\right| \\ &= \max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt} - E(\tilde{y}_{it-1}\varepsilon_{jt})\right| + \max_{i,j\leq N}\left|\frac{E\varepsilon_{it}\varepsilon_{jt}}{T^2}\left(\frac{T}{1-\theta_0} - \frac{(1-\theta_0^T)}{(1-\theta_0)^2}\right)\right| \\ &= O_P\left(\sqrt{\frac{logN}{T}}\right) + \max_{i,j\leq N}|E\varepsilon_{it}\varepsilon_{jt}|O\left(\frac{1}{T}\right). \end{aligned}$$

Lemma 5 *(i)* $\max_{i,j\leq N}\left|\widehat{R}_{ij} - V_\varepsilon(i,j)\right| = O_P\left(\sqrt{\frac{logN}{T}}\right)$,

(ii) $\left\|\widehat{V}_\varepsilon - V_\varepsilon\right\| \leq \left\|\widehat{V}_\varepsilon - V_\varepsilon\right\|_1 = O_P\left(m_N\sqrt{\frac{logN}{T}}^{1-q}\right)$.

Proof 9 (Lemma 5.(i))

$$\begin{aligned} \max_{i,j\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt} - E\varepsilon_{it}\varepsilon_{jt}\right| &\leq \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}(\widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt} - \varepsilon_{it}\varepsilon_{jt})\right| + \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt} - E\varepsilon_{it}\varepsilon_{jt}\right| \\ &:= a_1 + a_2 \end{aligned}$$

For a_1*,* $\hat{\varepsilon}_{it} = \tilde{y}_{it} - \hat{\theta}\tilde{y}_{it-1}$*,*

$$\begin{aligned}
a_1 &= \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{it}-\varepsilon_{it}\right)\left(\widehat{\varepsilon}_{jt}-\varepsilon_{jt}\right)+\left(\widehat{\varepsilon}_{it}-\varepsilon_{it}\right)\varepsilon_{jt}+\varepsilon_{it}\left(\widehat{\varepsilon}_{jt}-\varepsilon_{jt}\right)\right| \\
&= \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}[\tilde{y}_{it-1}(\theta-\hat{\theta})-\bar{\varepsilon}_i][\tilde{y}_{jt-1}(\theta-\hat{\theta})-\bar{\varepsilon}_j]+[\tilde{y}_{it-1}(\theta-\hat{\theta})-\bar{\varepsilon}_i]\varepsilon_{jt}+\varepsilon_{it}[\tilde{y}_{jt-1}(\theta-\hat{\theta})-\bar{\varepsilon}_j]\right| \\
&= \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{y}_{jt-1}(\theta-\hat{\theta})^2+\bar{\varepsilon}_i\bar{\varepsilon}_j+2\tilde{y}_{it-1}\varepsilon_{jt}(\theta-\hat{\theta})-2\bar{\varepsilon}_i\varepsilon_{jt}\right| \\
&= \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{y}_{jt-1}(\theta-\hat{\theta})^2+2\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt}(\theta-\hat{\theta})-\bar{\varepsilon}_i\bar{\varepsilon}_j\right| \\
&\leq \max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\tilde{y}_{jt-1}(\theta-\hat{\theta})^2\right|+2\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt}(\theta-\hat{\theta})\right|+\max_{ij}|\bar{\varepsilon}_i\bar{\varepsilon}_j| \\
&:= a_{11}+a_{12}+a_{13}
\end{aligned} \tag{51}$$

$$\begin{aligned}
a_{11} &\leq \|\widehat{\theta}-\theta\|^2\max_{ij}\frac{1}{T}\sum_{t=1}^{T}\|\tilde{y}_{it-1}\|\,\|\tilde{y}_{jt-1}\| \\
&\leq O_P\left(\frac{1}{NT}\right)\max_i\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}^2
\end{aligned} \tag{52}$$

Note that $\max_i \frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}^2$ *is bounded by the exponential tail conditions according to (Fan et al. 2011, Lemma 3.1), so that* $a_{11} = O_P\left(\frac{1}{NT}\right)$*.*

And by previous lemma 4,

$$\begin{aligned}
a_{12} &\leq 2|\theta-\hat{\theta}|\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt}\right| \\
&\leq O_P(\frac{1}{\sqrt{NT}})\max_{ij}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{jt}\right| \\
&= O_P\left(\frac{1}{T}\sqrt{\frac{logN}{N}}\right),
\end{aligned}$$

$$a_{13} = \max_i\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right|^2 = O_p\left(\frac{logN}{T}\right).$$

So that

$$a_1 = O_p\left(\frac{logN}{T}\right)$$

For a_2*, similar to the Lemma 4, we have,*

$$a_2 = \max_{ij} \left| \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{jt} - E\varepsilon_{it}\varepsilon_{jt} \right| = O_P\left(\sqrt{\frac{logN}{T}}\right)$$

So that given $\frac{logN}{T} = o(1)$, $\max_{i,j\leq N} \left| \frac{1}{T} \sum_{t=1}^{T} \widehat{\varepsilon}_{it}\widehat{\varepsilon}_{jt} - E\varepsilon_{it}\varepsilon_{jt} \right| = O_P\left(\sqrt{\frac{logN}{T}}\right)$.

Proof 10 (Lemma 5.(ii)) *Following (Fan et al. 2013, Theorem 5) and Hölder's inequality, we can show that,*

$$\left\| \widehat{V}_\varepsilon - V_\varepsilon \right\| \leq \left\| \widehat{V}_\varepsilon - V_\varepsilon \right\|_1 = O_P\left(m_N \sqrt{\frac{logN}{T}}^{1-q} \right)$$

Proof 11 (Theorem 7) *By the triangular inequality, submultiplicativity and Lemma 5.(ii), we have*

$$\begin{aligned}
\left\| \widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1} \right\| &\leq \left\| \left(\widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1} \right) (\widehat{V}_\varepsilon - V_\varepsilon) V_\varepsilon^{-1} \right\| + \left\| V_\varepsilon^{-1} (\widehat{V}_\varepsilon - V_\varepsilon) V_\varepsilon^{-1} \right\| \\
&\leq \left\| \widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1} \right\| \left\| V_\varepsilon^{-1} \right\| \left\| \widehat{V}_\varepsilon - V_\varepsilon \right\| + \left\| V_\varepsilon^{-1} \right\|^2 \left\| \widehat{V}_\varepsilon - V_\varepsilon \right\| \\
&= \left\| \widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1} \right\| O(1) O_P\left(m_N \sqrt{\frac{logN}{T}}^{1-q} \right) + O(1) O_P\left(m_N \sqrt{\frac{logN}{T}}^{1-q} \right)
\end{aligned}$$

$$\left\| \widehat{V}_\varepsilon^{-1} - V_\varepsilon^{-1} \right\| (1 + o_P(1)) = O_P\left(m_N \sqrt{\frac{logN}{T}}^{1-q} \right)$$

H Proof of Theorem 8

Proof 12

$$\sqrt{NT}(\hat{\theta}_{fpgls} - \theta_0) = \frac{\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} \widehat{V}_\varepsilon^{-1} \varepsilon_t}{\frac{1}{NT} \sum_{t=1}^{T} \hat{\mathbf{y}}'_{t-1} \widehat{V}_\varepsilon^{-1} \tilde{\mathbf{y}}_{t-1}} = \left(\frac{1}{NT} \tilde{Y}'_{-1} \widehat{\Omega}^{-1} \tilde{Y}_{-1} \right)^{-1} \left(\frac{1}{\sqrt{NT}} \tilde{Y}'_{-1} \widehat{\Omega}^{-1} \mathcal{E} \right)$$

Consider,

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\tilde{\mathbf{y}}'_{t-1}(\hat{V}_\varepsilon^{-1}-V_\varepsilon^{-1})\varepsilon_t &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\tilde{\mathbf{y}}'_{t-1}V_\varepsilon^{-1}(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t\\
&+\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\tilde{\mathbf{y}}'_{t-1}V_\varepsilon^{-1}(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t\\
&+\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\tilde{\mathbf{y}}'_{t-1}(\hat{V}_\varepsilon^{-1}-V_\varepsilon^{-1})(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t\\
&\equiv a+b+c.\end{aligned}$$

Let $w'_t=\tilde{\mathbf{y}}'_{t-1}V_\varepsilon^{-1}$*,* $u_t=V_\varepsilon^{-1}\varepsilon_t$*, and by Assumption 6.(ii),*

$$\begin{aligned}|b| &= \frac{1}{\sqrt{NT}}\left|\sum_{t=1}^{T}w'_t(\hat{V}_\varepsilon-V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon-V_\varepsilon)u_t\right|\\
&=\frac{1}{\sqrt{NT}}\left|\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{p=1}^{N}\sum_{q=1}^{N}(\hat{V}_\varepsilon-V_\varepsilon)_{ij}(\hat{V}_\varepsilon-V_\varepsilon)_{pq}\sum_{t=1}^{T}w_{it}u_{qt}\left(V_\varepsilon^{-1}\right)_{jp}\right|\\
&\le\frac{1}{\sqrt{NT}}\left(\max_j\sum_{i=1}^{N}|(\hat{V}_\varepsilon-V_\varepsilon)_{ij}|\right)\left(\max_q\sum_{p=1}^{N}|(\hat{V}_\varepsilon-V_\varepsilon)_{pq}|\right)\left(\max_i\max_p\left|\sum_{j=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}w_{it}u_{qt}\left(V_\varepsilon^{-1}\right)_{jp}\right|\right)\\
&\le\frac{1}{\sqrt{NT}}\|\hat{V}_\varepsilon-V_\varepsilon\|_1^2\max_{i,p}\left|\sum_{q=1}^{N}\sum_{t=1}^{T}\left(u_{qt}w_{it}\sum_{j=1}^{N}\left(V_\varepsilon^{-1}\right)_{jp}\right)\right|\\
&\le O_P(\sqrt{NT}m_N^2\gamma_T^{2-2q})\max_{i,p}\left|\frac{1}{NT}\sum_{q=1}^{N}\sum_{t=1}^{T}\left(u_{qt}w_{it}\sum_{j=1}^{N}\left(V_\varepsilon^{-1}\right)_{jp}\right)\right|\\
&=O_P(\sqrt{NT}m_N^2\gamma_T^{2-2q})O_P\left(\sqrt{\frac{logN}{NT}}\right)\\
&=O_P(\sqrt{T}m_N^2\gamma_T^{3-2q})=o_p(1)\end{aligned}$$

By Theorem 7, submultiplicity and assumption 6.(iii), when $\|V^{-1}\|_1 = O(1)$

$$
\begin{aligned}
|c| &= \left| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1}(\hat{V}_\varepsilon^{-1} - V_\varepsilon^{-1})(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t \right| \\
&\leq \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \left| \tilde{\mathbf{y}}'_{t-1}(\hat{V}_\varepsilon^{-1} - V_\varepsilon^{-1})(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t \right| \\
&\leq \frac{T}{\sqrt{NT}} \left\| (\hat{V}_\varepsilon^{-1} - V_\varepsilon^{-1})(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1} \right\| O_P(N) \\
&\leq \left\| \widehat{V}_\varepsilon^{-1} - V^{-1} \right\| \|\widehat{V}_\varepsilon - V_\varepsilon\|^2 O_P(\sqrt{NT}) = O_P\left(\sqrt{NT} m_N^3 \gamma_T^{3-3q}\right) = o_P(1)
\end{aligned}
$$

Therefore, we have

$$
\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1}(\hat{V}_\varepsilon^{-1} - V_\varepsilon^{-1})\varepsilon_t = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} V_\varepsilon^{-1}(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t + o_p(1)
$$

And by Theorem 7, we can show that, $\tilde{Y}'_{-1}\widehat{\Omega}^{-1}\tilde{Y}_{-1}/NT = \tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}/NT + o_P(1)$. *Then,*

$$
\begin{aligned}
\sqrt{NT}\left(\widehat{\theta}_{fpgls} - \theta_0\right) &= \left(\frac{1}{NT}\tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\widehat{\Omega}^{-1}\mathcal{E}\right) + o_P(1) \\
&= \left(\frac{1}{NT}\tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}\mathcal{E} + \frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}(\widehat{\Omega}^{-1} - \Omega^{-1})\mathcal{E}\right) + o_P(1) \\
&= \left(\frac{1}{NT}\tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}\mathcal{E}\right) \\
&+ \left(\frac{1}{NT}\tilde{Y}'_{-1}\Omega^{-1}\tilde{Y}_{-1}\right)^{-1}\left(\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}(\widehat{\Omega} - \Omega)\Omega^{-1}\mathcal{E}\right) + o_P(1)
\end{aligned}
$$

I Proof of Theorem 9

It suffices to prove $\left\|\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}(\widehat{\Omega} - \Omega)\Omega^{-1}\mathcal{E}\right\| = o_p(1)$

Proof 13

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{NT}}\tilde{Y}'_{-1}\Omega^{-1}(\widehat{\Omega} - \Omega)\Omega^{-1}\mathcal{E}\right\| &= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \tilde{\mathbf{y}}'_{t-1} V_\varepsilon^{-1}(\hat{V}_\varepsilon - V_\varepsilon)V_\varepsilon^{-1}\varepsilon_t \right\| \\
&= \left\| \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} w'_t(\hat{V}_\varepsilon - V_\varepsilon)u_t \right\| \\
&\leq \left\| \frac{C}{\sqrt{NT}} \sum_{i,j\in A} \sum_{t=1}^{T} w_{it}u_{jt} \frac{1}{T} \sum_{s=1}^{T} (\varepsilon_{is}\varepsilon_{js} - E\varepsilon_{it}\varepsilon_{jt}) \right\| + o_P(1) \\
&= o_P(1)
\end{aligned}
$$