Doctoral Dissertation

Essays on Dynamic Panel Data Analysis

March 2017

Graduate School of Social Sciences
Hiroshima University
Meng Qi

Doctoral Dissertation

Essays on Dynamic Panel Data Analysis

Supervisor

Associate Professor
Kazuhiko Hayakawa

March 2017
Graduate School of Social Sciences
Hiroshima University
Meng Qi

Abstract

This dissertation is concerned with the following estimator and some bias corrected estimators of dynamic panel data model. Chapter 2 consider several instrumental variables (IV) and generalized method of moments (GMM) estimators for panel data models with weakly exogenous variables, when the T and N, the sample size of time series and cross-section are large and fixed T and large N seperately. In chapter 3 and 4 , we consider the situation for large N and small T. Fixed effect estimator and conventional is biased and GMM estimator suffer from the weak instrumental problems. In chapter 3, we investigate the weak instruments problem of GMM estimator for dynamic panel data models. Chapter 4 consider the second-order bias corrected estimator of multivariate dynamic panel data models for short-run coefficients and long-run coefficients.

Acknowledgement

This doctoral essay could not have been complicated without the supports of many people. Fisrt of all, I am hearty gratefully to my supervisor, Prof. Hayakawa Kazuhiko. I am also thankful to my supervisor during master course period, Prof. Ginama Isamu. I am also thankful to my condesciples and friends, Gong Zhe, Li Yue, Hou Jie, Ge Bei,Tajima Yuki and Chao Bo. Finaly, this dissertation is delicated to my families, Yu Hua, Huhe Teengesi, Xilin Gaowa and Chairiga.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Corresponding with the vast amount of available data sets, panel data model regression analysis became more significant in Econometrics. Since panel data model includes more data in time series and cross-section dimension, it solves more problems than models in single dimension. Dynamic panel data is important part in panel data model cause the dependent variable could be a significant regressor in panel data model analysis. Therefore, several complicated problems needs proper solutions. Such as correlation problem between error term and regressor, inconsistent estimation or biased asymptotic distribution in short panels. And the weak instrument pro 2 blem caused by persistency of series and larger variance of time-invariant error than idiosyncratic errors in IV and GMM estimations.

Unobservable variable, which treated as one part of error term is a significant reason for poor OLS estimation, cause it has high tendency to be correlated with explaining variables, which leads to inconsistent estimation. As a solution, the unobservable variables can be treated as time-invariant variable for each identical as fixed effect, then here comes the fixed effect model. Another situation, in which the the unobserved variable, which is not correlated with the explaining variable, can be treated as time-invariant error term constructed the random effect model. In cross-section data model, the OLS estimator suffered from inconsistent problem caused by the correlation between the error term including fixed effect and the explaining variable. Consequently, pooled OLS estimation for panel data model is inconsistent.

A common methods for fixed effect panel model is to transform the model to eliminate the fixed effect from model. Such as taking first difference from time series , Fixed Effect (FE) estimation which take Within Group mean or taking Forward Orthogonal Deviation to eliminate the fixed effect from the panel data model. FE estimator hinge on the strict exogenous condition.

Multivariate First differenced model, to guarantee the the dependent regressor and exogenous regressor be uncorrelated with the fixed effect, some stationary restriction should be satisfied. Blundell, Bond, and Windmeijer [2000].

To solve the correlation between fixed effect and regressor, one consideration is to use Instrumental Variable (IV) estimation, Anderson and Hsiao [1981], Anderson and Hsiao [1982] utilize the IV, which correlated with the regressor but uncorrelated with error terms performs well. The most commonly used IV estimations transform the model by first-difference by time series Arellano(1989), or FOD transformation **?** and Hayakawa [2009a]. In both regression estimation the IV is level regressors.

When both the T and N, which denotes the the sample size of time period and cross section, are both large, IV estimator solves the inconsistent problem of fixed effect model, and GMM estimation deals with inefficiency of IV estimation, Holtz-Eakin, Newey, and Rosen [1988], Arellano and Bond [1991]. However when the data set renders a large number of individuals but short time periods, the GMM estimation turned out to be biased. Windmeijer [2005] proposed a finite sample correction for the two step DIF-GMM estimation. In the finite sample case, using too much instruments cause the trade off problem between bias and efficiency, (Bekker [1994]. Many literatures suggests to use a part of instruments than full available instruments to solve the trade off problem, Windmeijer [2005], Bun and Kiviet [2006],Hayakawa [2009a] and Roodman [2009].

On the other side, weak instrumental problem occurs for IV and GMM estimation when the series is persistent or variance of individual effects is larger than idiosyncratic errors. Arellano and Bover [1995] proposed LEV-GMM estimator conquers the weak instrumental problem caused by persistency. System (SYS)-GMM estimation byArellano and Bover [1995], Blundell and Bond [1998] is more efficient than DIF-GMM and LEV-GMM estimator.

Furthermore the multivariate dynamic panel data model includes a heterogeneous variable. The estimation process is familiar with former models, however whether the heterogeneous variable is strictly exogenous, predetermined or endogenous variable make differences on estimation.

1.2 Purpose

Here are many transformed dynamic panel data models and estimation methods for multivariate dynamic panel data model mentioned in former part. The GMM estimation is well utilized. In this doctoral dissertation, we used several ways for better estimation.

In chapter 2,concern about the IV and GMM estimator well used for dynamic panel data models. However here exists a trade off problem between bias and efficiency cased by using many instruments. Double filter instrumental variables estimation (DFIV) is a method for solve this trade off problem. We propose instrumental variables (IV) and generalized method of moments (GMM) estimators for panel data models with weakly exogenous variables. The model is allowed to include heterogeneous time trends in addition to the standard fixed effects. The proposed IV and GMM estimators are obtained by applying a forward filter to the model and a backward filter to the instruments in order to remove fixed effects, thereby called the double filter IV and GMM estimators. We derive the asymptotic properties of the proposed estimators under fixed T and large N, and large T and large N . It is proved that when both N and T are large, the proposed IV estimator has the same asymptotic distribution as the bias corrected fixed effects estimator. Our Monte Carlo simulation results reveal that the proposed estimator performs well in finite samples and outperforms the conventional IV/GMM estimators using instruments in levels in many cases.

Weak instrumental problems happens in dynamic panel data models by many reason, such as the persistency of series in time period dimension, or the variance ratio of fixed effect and error term is large. In chapter 3, we investigate the weak instruments problem of GMM estimator for system GMM estimator for dynamic panel data models cause by the large ratio. Bun and Windmeijer [2010] demonstrates that the system GMM estimator combining LEV and DIF models suffers from the weak instruments problem when the variance ratio is large, which is mainly due to the model in levels. In this chapter, we alternatively consider first-difference and level models transformed by forward GLS transformation, and demonstrate that this transformation yields a higher concentration parameter compared with original models. This indicates that this transformation yields stronger instruments. Our simulation results reveal that the proposed system GMM estimator for the transformed model, called the forward system GMM estimator, performs better than the conventional system GMM estimator for models in DIF and LEV, and the performance of the new system GMM estimator is reasonably well even when the variance ratio is large.

In chapter 4, we propose a bias-corrected mean-group estimator for dynamic heterogeneous panel data models when T is fixed. It is demonstrated that the GMM estimator is inconsistent when coefficient is heterogeneous by Pesaran and Smith [1995] . They proposed to mean group estimator averaged over cross-sections. This estimator is consistent for large T, but inconsistent for small T. The solution proposed by Pesaran and Zhao [1999] is to used the bias correction by Kiviet and Phillips [1993]. For more precise estimation, we propose to use the second-order bias-corrected estimator by Kiviet and Phillips [2012]when constructing the mean-group estimator. Monte Carlo simulation is conducted and it is confirmed that the proposed estimator has smaller bias than the conventional and first-order bias-corrected mean-group estimators when the length of panel data is not so large.

Chapter 2

Double Filter Instrumental Variable Estimation of Panel Data Models with Weakly Exogenous Variables

[1]

In this chapter, we propose instrumental variables (IV) and generalized method of moments (GMM) estimators for panel data models with weakly exogenous variables. The model is allowed to include heterogeneous time trends besides the standard fixed effects. The proposed IV and GMM estimators are obtained by applying a forward filter to the model and a backward filter to the instruments in order to remove fixed effects, thereby called the double filter IV and GMM estimators. We derive the asymptotic properties of the proposed estimators under fixed T and large N, and large T and large N asymptotics where N and T denote the dimensions of cross section and time series, respectively. It is shown that the proposed IV estimator has the same asymptotic distribution as the bias corrected fixed effects estimator when both N and T are large. Monte Carlo simulation results reveal that the proposed estimator performs well in finite samples and outperforms the conventional IV/GMM estimators using instruments in levels in many cases.

[1]This is a joint work with Jörg Breitung and Kazuhiko Hayakawa.

2.1 Introduction

Using panel data in empirical studies has become much more popular than before since many panel data sets are available in these days. Accordingly, many types of panel data models and estimation procedures have been proposed. Among them, most basic approach is the fixed effects (FE) regression model where unobserved individual specific effects are allowed to be correlated with regressors. However, consistency of fixed effects estimator relies on the strict exogeneity assumption, i.e., the regressors and idiosyncratic errors are uncorrelated for all periods when the length of panel data, denoted as T is small. Unfortunately, there are many cases in which the strict exogeneity assumption is violated. A leading example is a dynamic panel model. Regardless of whether the regressors besides the lagged dependent variables are strictly or weakly exogenous, or endogenous, the lagged dependent variable is correlated with the idiosyncratic errors by construction, and hence the fixed effect estimator is inconsistent when T is small [cf. Nickell, 1981]. To address this problem, estimation procedures using instrumental variables (IV) have been extensively considered since the work of Anderson and Hsiao [1981]. These include, among others, Holtz-Eakin et al. [1988], Arellano and Bond [1991], Arellano and Bover [1995], Ahn and Schmidt [1995] and Blundell and Bond [1998] etc. While most of these studies focus on short panels, there are cases where long panel data are available, typically in macro panels. Inspired by the availability of long panel data, several papers study large N and large T asymptotic properties of aforementioned estimators where N is the number of cross-sectional units. Earlier papers that considered large N and large T dynamic panels are Hahn and Kuersteiner [2002] and Alvarez and Arellano [2003]. Hahn and Kuersteiner [2002] and Alvarez and Arellano [2003] demonstrate that, when T and N are large, the fixed effect estimator is consistent but its asymptotic distribution is not centered around the true value in the context of (vector) autoregressive models. To correct for the bias, Hahn and Kuersteiner [2002] also proposes a bias-corrected fixed effects estimator.

More recently, an alternative instrumental variables estimator has been proposed in the literature where a forward demeaning(detrending) is applied to the model while backward demeaning(detrending) is applied to the instruments. We call that IV estimator the double filter IV (DFIV) estimator since unobserved heterogeneity is filtered out forward and backward. Perhaps, the first paper that considers the DFIV estimator is Moon and Phillips [2000] where a near integrated autoregressive panel data model is studied. However, they did not provide distributional results of the DFIV estimator. Hayakawa [2009a] considers the DFIV estimator in a stationary panel autoregressive model and derives the asymptotic properties when both N and T are large. A novel feature of the DFIV estimator is that it has the same asymptotic distribution as the bias-corrected FE estimator when T and N are large. Since the DFIV estimator simply uses variables deviated from past means as instruments, as opposed to the commonly used level variables, it is quite easy to use in practice[2]. From the theoretical point of view, the DFIV estimator has addressed the trade-off problem of using many instruments. Although many instruments are required to improve efficiency, the DFIV estimator becomes efficient despite the same number of instruments as the parameters is used. Hence, the DFIV estimator becomes efficient with the minimal number of instruments. This property has an advantage that it does not cause a large finite sample bias induced by using many instruments. Thereby, the trade-off problem between the bias and efficiency of the generalized method of moments (GMM) estimator is addressed: both the bias and variance of the DFIV estimator become small simultaneously.

The DFIV estimator is also extended to a panel VAR model[Hayakawa, 2016] and an infinite order panel autoregressive model[Lee, Okui, and Shintani, 2013][3]. However, while there are several nice features as above, unfortunately, the asymptotic equivalence between the DFIV and bias-corrected FE estimators are only proved in the context of (V)AR models, which are somewhat restrictive in practice. One of the purposes of this chapter is to demonstrate that this equivalence result holds for more general case with additional regressors. Specifically, we demonstrate that the asymptotic distributions of the DFIV and bias-corrected fixed effect estimator with large N and T are identical for linear panel data models including dynamic models as well as static panel data models with weakly exogenous regressors. Moreover, we demonstrate that this equivalence result holds even when the errors are heteroskedastic and heterogeneous time trends are included in the model, which are not allowed in Hayakawa [2009a, 2016]. We also investigate the efficiency property of the DFIV and related GMM estimators when T is small and N is large. We conduct Monte Carlo simulation to investigate the finite sample behavior of estimators. Consequently, we find that the DFIV/DFGMM estimators tend to outperform the fixed effects estimator and IV/GMM estimators using instruments in levels.

The rest of this chapter is organized as follows. In Section 2, we introduce the models and estimators. In Section 3, the large N and T asymptotic properties of estimators introduced in Section 2 are derived. In Section 4, we carry out Monte Carlo simulation to investigate the finite sample behavior of estimators, and in Section 5, we conclude.

With regard to the notation, we define $T_j = T - j$. For a matrix $\mathbf{A} = \{a_{ij}\}$, a_{ij} denotes the (i, j) element of $\mathbf{A}$. $\|\mathbf{A}\|^2 = \mathrm{tr}\,(\mathbf{A}'\mathbf{A}) = \sum_{ij} a_{ij}^2$ denotes the Euclidean norm of a matrix $\mathbf{A}$.

[2]In the Stata command `xtabond2` by David Roodman, which is routinely used in empirical studies, we can estimate dynamic panel data models by that instruments.

[3]The DFIV estimator is also considered in estimation of panel predictive regression model by Westerlund, Karabiyik, and Narayan [2016].

2.2 Model and estimators

In this section, we introduce models and estimators. We first consider a model with fixed effects and then consider a model with heterogeneous time trends.

2.2.1 Fixed effects model

Consider a panel data model with fixed effects, given by

$$y_{it} = \mathbf{w}_{it}'\boldsymbol{\delta} + \eta_i + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{2.1}$$

where $\boldsymbol{\delta}$ and $\mathbf{w}_{it}$ are $k \times 1$ vectors. Errors v_{it} are serially and cross-sectionally uncorrelated. Fixed effects η_i can be correlated with the regressor $\mathbf{w}_{it}$.

This model includes several models as special cases.

Static model: $\quad y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + \eta_i + v_{it}$

where $\boldsymbol{\delta} = \boldsymbol{\beta}$ and $\mathbf{w}_{it} = \mathbf{x}_{it}$ and $\boldsymbol{\beta}$ and $\mathbf{x}_{it}$ are $r \times 1$ vectors with $k = r$.

AR(p) model: $\quad y_{it} = \alpha_1 y_{i,t-1} + \cdots + \alpha_p y_{i,t-p} + \eta_i + v_{it},$

where $\boldsymbol{\delta} = (\alpha_1, ..., \alpha_p)'$, $\mathbf{w}_{it} = (y_{i,t-1}, ..., y_{i,t-p})'$, and $\boldsymbol{\delta}$ and $\mathbf{w}_{it}$ are $p \times 1$ vectors with $k = p$.

ARX(p) model: $\quad y_{it} = \alpha_1 y_{i,t-1} + \cdots + \alpha_p y_{i,t-p} + \mathbf{x}_{it}'\boldsymbol{\beta} + \eta_i + v_{it}$

where $\boldsymbol{\delta} = \left(\alpha_1, ..., \alpha_p, \boldsymbol{\beta}'\right)'$, $\mathbf{w}_{it} = (y_{i,t-1}, ..., y_{i,t-p}, \mathbf{x}_{it}')'$, and $\boldsymbol{\delta}$ and $\mathbf{w}_{it}$ are $(p + r) \times 1$ vectors with $k = p + r$.

In a matrix form, the model (2.1) can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \eta_i\boldsymbol{\iota}_T + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{2.2}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})'$, $\boldsymbol{\iota}_T = (1, ..., 1)'$ and $\mathbf{v}_i = (v_{i1}, ..., v_{iT})'$.

Define the following matrix that can be used to remove fixed effects:

$$
\underset{(T_1\times T)}{\mathbf{F}_T^{\iota}} = \mathrm{diag}\left(c_1^{\iota}, c_2^{\iota}, ..., c_{T_1}^{\iota}\right)\left[\begin{array}{c|ccccc|c}
1 & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} \\
\hline
0 & 1 & \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} \\
0 & 0 & 1 & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & \frac{-1}{T-3} \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\
\vdots & \vdots & & \ddots & 1 & \frac{-1}{2} & \frac{-1}{2} \\
0 & 0 & \cdots & \cdots & 0 & 1 & -1
\end{array}\right] \tag{2.3}
$$

$$
= \begin{bmatrix} F_{11}^{\iota} & \mathbf{F}_{12}^{\iota} & F_{13}^{\iota} \\ \mathbf{0}_{T_2\times 1} & \mathbf{F}_{22}^{\iota} & \mathbf{F}_{23}^{\iota} \end{bmatrix}
$$

$$
= \{f_{st}^{\iota}\}
$$

$$
= \begin{cases} 0 & \text{if } s > t \\ c_t^{\iota} = 1 + O\left(\frac{1}{T-t}\right) & \text{if } s = t \\ \frac{-c_s^{\iota}}{T-s} = O\left(\frac{1}{T-s}\right) & \text{if } s < t \end{cases}
$$

where $c_t^{\iota} = \sqrt{(T-t)/(T-t+1)}$, F_{11}^{ι} and F_{13}^{ι} are scalars, $\mathbf{F}_{12}^{\iota}$ is $1\times T_2$, $\mathbf{F}_{22}^{\iota}$ is $T_2\times T_2$, and $\mathbf{F}_{23}^{\iota}$ is $T_2\times 1$.

Multiplying $\mathbf{F}_T^{\iota}$ to (2.2), the model to be estimated becomes

$$
\dot{\mathbf{y}}_i^{\iota} = \dot{\mathbf{W}}_i^{\iota}\boldsymbol{\delta} + \dot{\mathbf{v}}_i^{\iota}, \qquad (i = 1, ..., N) \tag{2.4}
$$

where $\dot{\mathbf{y}}_i^{\iota} = \mathbf{F}_T^{\iota}\mathbf{y}_i = \left(\dot{y}_{i1}^{\iota}, ..., \dot{y}_{iT_1}^{\iota}\right)'$, $\dot{\mathbf{W}}_i^{\iota} = \mathbf{F}_T^{\iota}\mathbf{W}_i = \left(\dot{\mathbf{w}}_{i1}^{\iota}, ..., \dot{\mathbf{w}}_{iT_1}^{\iota}\right)'$ and $\dot{\mathbf{v}}_i^{\iota} = \mathbf{F}_T^{\iota}\mathbf{v}_i = \left(\dot{v}_{i1}^{\iota}, ..., \dot{v}_{iT_1}^{\iota}\right)'$ with $\dot{y}_{it}^{\iota} = c_t^{\iota}\left[y_{it} - (y_{i,t+1} + \cdots + y_{iT})/(T-t)\right]$,
$\dot{\mathbf{w}}_{it}^{\iota} = c_t^{\iota}\left[\mathbf{w}_{it} - (\mathbf{w}_{i,t+1} + \cdots + \mathbf{w}_{iT})/(T-t)\right]$ and $\dot{v}_{it}^{\iota} = c_t^{\iota}\left[v_{it} - (v_{i,t+1} + \cdots + v_{iT})/(T-t)\right]$ for $t = 1, ..., T_1$. Note that the fixed effects η_i is removed by taking a deviation from future means. The tth row of (2.4) can be written as

$$
\dot{y}_{it}^{\iota} = \dot{\mathbf{w}}_{it}'\boldsymbol{\delta} + \dot{v}_{it}^{\iota}, \qquad (t = 1, ..., T_1; i = 1, ..., N) \tag{2.5}
$$

This is the model in forward orthogonal deviations(FOD).

Next, we introduce an instrumental variable. In empirical studies, (a subset of) lagged level variables $\mathbf{w}_{i1}, ..., \mathbf{w}_{it}$ are commonly used as instruments. Instead of using variables in levels, Hayakawa [2009a, 2016] suggest to use variables deviated from past means. To introduce variables deviated from past means, let us define

$$\underset{(T_1 \times T)}{\mathbf{B}_T^{\iota}} = \mathrm{diag}\left(c_{T_1}^{\iota}, \cdots, c_2^{\iota}, c_1^{\iota}\right) \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ \frac{-1}{T-3} & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & 1 & 0 & 0 \\ \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} & 1 & 0 \\ \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & 1 \end{bmatrix} = \{b_{st}^{\iota}\} \quad (2.6)$$

$\mathbf{B}_T^{\iota}$ is obtained by rotating $\mathbf{F}_T^{\iota}$. The mathematical relationship between $\mathbf{F}_T^{\iota}$ and $\mathbf{B}_T^{\iota}$ is given in (2.67) in the appendix. Using this, we define an instrumental variable $\ddot{\mathbf{W}}_i^{\iota} = \mathbf{B}_T^{\iota}\mathbf{W}_i = (\ddot{\mathbf{w}}_{i2}^{\iota}, ..., \ddot{\mathbf{w}}_{iT}^{\iota})'$ where[4]

$$\ddot{\mathbf{w}}_{it}^{\iota} = c_{T-t+1}^{\iota}\left[\mathbf{w}_{it} - \frac{\mathbf{w}_{i,t-1} + \cdots + \mathbf{w}_{i1}}{t-1}\right], \qquad (i = 1, ..., N; t = 2, ..., T). \qquad (2.7)$$

Note that the first period is lost due to the difference property of the transformation matrix (2.6). The transformation that induces $\ddot{\mathbf{w}}_{it}^{\iota}$ is called the backward orthogonal deviation (BOD) transformation as opposed to FOD transformation.

Since $E(\ddot{\mathbf{w}}_{is}^{\iota} v_{it}^{\iota}) = \mathbf{0}$ for $2 \leq s \leq t \leq T_1$ holds, we can construct moment conditions from this. Specifically, we consider the moment conditions $E\left(\sum_{t=2}^{T_1} \ddot{\mathbf{w}}_{it}^{\iota} v_{it}^{\iota}\right) = \mathbf{0}$. The corresponding instrumental variable estimator is given by

$$\widehat{\boldsymbol{\delta}}_{IV}^{\iota} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T_1} \ddot{\mathbf{w}}_{it}^{\iota}\dot{\mathbf{w}}_{it}^{\iota\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T_1} \ddot{\mathbf{w}}_{it}^{\iota}\dot{y}_{it}^{\iota}\right). \qquad (2.8)$$

How the moment conditions are derived?

In Hayakawa [2009a], it is shown in the context of AR(p) models that $\ddot{\mathbf{w}}_{it}^{\iota}$ has the same structure as the infeasible optimal instruments which leads to efficient estimation. Here, we provide an alternative explanation how the moment conditions $E\left(\ddot{\mathbf{w}}_{it}^{\iota}\dot{v}_{it}^{\iota}\right) = \mathbf{0}$ are derived. For this, let us define two variables r_{it}^{b} and r_{it}^{f} for some r_{it} such that $r_{it}^{b} = r_{it} - (r_{i,t-1} + \cdots + r_{i1})/(t-1)$ and $r_{it}^{f} = r_{it} - (r_{i,t+1} + \cdots + r_{iT})/(T-t)$. Note that r_{it}^{b} is a variable deviated from backward means while r_{it}^{f} is a variable deviated from forward means. Hence, when $r_{it} = v_{it}$, r_{it}^{f} and $\dot{v}_{it}^{\iota}$ are related such that $r_{it}^{f} = \dot{v}_{it}^{\iota}/c_t^{\iota}$. We demonstrate that the moment conditions $E\left(\ddot{\mathbf{w}}_{it}^{\iota}\dot{v}_{it}^{\iota}\right) = \mathbf{0}$ can be obtained from the fixed effects model:

$$(y_{it} - \bar{y}_i) = (\mathbf{w}_{it} - \bar{\mathbf{w}}_i)'\boldsymbol{\delta} + (v_{it} - \bar{v}_i), \qquad (i = 1, ..., N;\; t = 1, ..., T) \qquad (2.9)$$

[4] Compared with the form given in Hayakawa [2009a], the coefficient c_{T-t+1} is slightly different. However, this is inconsequential and does not affect the main result of this chapter.

where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{\mathbf{w}}_i = T^{-1}\sum_{t=1}^{T} \mathbf{w}_{it}$ and $\bar{v}_i = T^{-1}\sum_{t=1}^{T} v_{it}$. Note that, after some algebra, $(\mathbf{w}_{it} - \bar{\mathbf{w}}_i)$ and $(v_{it} - \bar{v}_i)$ can be written as

$$\begin{aligned}
\mathbf{w}_{it} - \bar{\mathbf{w}}_i &= \frac{t-1}{T}\mathbf{w}_{it}^{b} + \frac{T-t}{T}\mathbf{w}_{it}^{f}, \\
v_{it} - \bar{v}_i &= \frac{t-1}{T}v_{it}^{b} + \frac{T-t}{T}v_{it}^{f}.
\end{aligned}$$

Hence, the covariance between the regressors and error term in (2.9) becomes

$$\begin{aligned}
E\left[(\mathbf{w}_{it} - \bar{\mathbf{w}}_i)(v_{it} - \bar{v}_i)\right] &= \frac{(t-1)^2}{T^2}E(\mathbf{w}_{it}^{b}v_{it}^{b}) + \frac{(T-t)(t-1)}{T^2}E(\mathbf{w}_{it}^{f}v_{it}^{b}) \\
&\quad + \frac{(T-t)(t-1)}{T^2}E(\mathbf{w}_{it}^{b}v_{it}^{f}) + \frac{(T-t)^2}{T^2}E(\mathbf{w}_{it}^{f}v_{it}^{f}) \neq \mathbf{0}.
\end{aligned}$$

This non-zero correlation is the reason why the fixed effect estimator is inconsistent when T is small. However, among the four terms, the third term has zero mean $E(\mathbf{w}_{it}^{b}v_{it}^{f}) = \mathbf{0}$, which can be used to consistently estimate $\boldsymbol{\delta}$ even when T is small. Multiplying $c_{T-t+1}^{\iota}c_t^{\iota}$ to this moment condition in order to account for time series heteroskedasticity, we obtain $c_{T-t+1}^{\iota}c_t^{\iota}E(\mathbf{w}_{it}^{b}v_{it}^{f}) = E\left(\ddot{\mathbf{w}}_{it}^{\iota}\dot{v}_{it}^{\iota}\right) = \mathbf{0}$. This indicates that the proposed moment conditions are derived from the valid part (i.e., no correlation) of the moment conditions of the fixed effects estimator.

2.2.2 Trend model

Next, we consider a panel data model with usual fixed effects and heterogeneous time trends, given by

$$y_{it} = \mathbf{w}_{it}'\boldsymbol{\delta} + \eta_i + \lambda_i t + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T).$$

In this model, both η_i and λ_i can be correlated with $\mathbf{w}_{it}$. Panel data models with heterogeneous time trends are studied by, say, Wansbeek and Knaap [1999], and Phillips and Sul [2007] etc.. In a matrix form, this model can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \eta_i\boldsymbol{\iota}_T + \lambda_i\boldsymbol{\tau}_T + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{2.10}$$

where $\boldsymbol{\tau}_T = (1, 2, ..., T)'$. To remove both η_i and λ_i, we need to multiply a matrix that is orthogonal to both $\boldsymbol{\iota}_T$ and $\boldsymbol{\tau}_T$. While there are several matrices that achieves this (e.g.

the second differences), we consider the following matrix:

$$
\begin{aligned}
\underset{(T_2\times T)}{\mathbf{F}_T^{\tau}}
&= \mathbf{F}_T^{\tau 1}\left[\begin{array}{cc|cccc|cc}
1 & \frac{2(-2T_2)}{T_1T_2} & \frac{2(-2T_2+3)}{T_1T_2} & \frac{2(-2T_2+6)}{T_1T_2} & \cdots & \cdots & \frac{2(-2T_2+3T_3)}{T_1T_2} & \frac{2(-2T_2+3T_2)}{T_1T_2} \\
0 & 1 & \frac{2(-2T_3)}{T_2T_3} & \frac{2(-2T_3+3)}{T_2T_3} & \cdots & \cdots & \frac{2(-2T_3+3T_4)}{T_2T_3} & \frac{2(-2T_3+3T_3)}{T_2T_3} \\
\hline
0 & 0 & 1 & \frac{2(-2T_4)}{T_3T_4} & \cdots & \cdots & \frac{2(-2T_4+3T_5)}{T_3T_4} & \frac{2(-2T_4+3T_4)}{T_3T_4} \\
\vdots & \vdots & 0 & \ddots & \ddots & & \vdots & \vdots \\
\vdots & \vdots & & \ddots & 1 & \frac{2(-4)}{3\cdot 2} & \frac{2(-4+3)}{3\cdot 2} & \frac{2(-4+6)}{3\cdot 2} \\
0 & 0 & \cdots & \cdots & 0 & 1 & \frac{2(-2)}{2\cdot 1} & \frac{2(-2+3)}{2\cdot 1}
\end{array}\right] \\
&= \begin{bmatrix} \mathbf{F}_{11}^{\tau} & \mathbf{F}_{12}^{\tau} & \mathbf{F}_{13}^{\tau} \\ \mathbf{0}_{T_4\times 2} & \mathbf{F}_{22}^{\tau} & \mathbf{F}_{23}^{\tau} \end{bmatrix} \\
&= \{f_{st}^{\tau}\} \\
&= \begin{cases} 0 & \text{if } s > t \\ c_t^{\tau} = 1 + O\left(\frac{1}{T-t}\right) & \text{if } s = t \\ \frac{2c_s^{\tau}[-2(T-s-1)+3(t-s-1)]}{(T-s)(T-s-1)} = O\left(\frac{1}{T-s}\right) + O\left(\frac{t-s}{(T-s)^2}\right) & \text{if } s < t \end{cases}
\end{aligned}
\tag{2.11}
$$

where $\mathbf{F}_T^{\tau 1} = \operatorname{diag}\left(c_1^{\tau}, c_2^{\tau}, ..., c_{T_2}^{\tau}\right)$, $c_t^{\tau} = \sqrt{(T-t)(T-t-1)/(T-t+1)(T-t+2)}$, $\mathbf{F}_{11}^{\tau}$ and $\mathbf{F}_{13}^{\tau}$ are 2×2, $\mathbf{F}_{12}^{\tau}$ is $2\times T_4$, $\mathbf{F}_{22}^{\tau}$ is $T_4\times T_4$, and $\mathbf{F}_{23}^{\tau}$ is $T_4\times 2$. The matrix $\mathbf{F}_T^{\tau}$ is obtained as a GLS transformation of second differences. The formal derivation of $\mathbf{F}_T^{\tau}$ is provided in appendix.

Multiplying (2.11) to (2.10), we have the following transformed model

$$\dot{\mathbf{y}}_i^{\tau} = \dot{\mathbf{W}}_i^{\tau}\boldsymbol{\delta} + \dot{\mathbf{v}}_i^{\tau}, \qquad (i = 1, ..., N) \tag{2.12}$$

where $\dot{\mathbf{y}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{y}_i = (\dot{y}_{i1}^{\tau}, ..., \dot{y}_{iT_2}^{\tau})'$, $\dot{\mathbf{W}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{W}_i = (\dot{\mathbf{w}}_{i1}^{\tau}, ..., \dot{\mathbf{w}}_{iT_2}^{\tau})'$, and $\dot{\mathbf{v}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{v}_i = (\dot{v}_{i1}^{\tau}, ..., \dot{v}_{iT_2}^{\tau})'$. The tth row of $\dot{\mathbf{y}}_i^{\tau}$, $\dot{\mathbf{W}}_i^{\tau}$ and $\dot{\mathbf{v}}_i^{\tau}$ are given by

$$
\begin{aligned}
\dot{y}_{it}^{\tau} &= f_{tt}^{\tau}y_{it} + f_{t,t+1}^{\tau}y_{i,t+1} + \cdots + f_{tT}^{\tau}y_{iT}, \qquad \dot{\mathbf{w}}_{it}^{\tau} = f_{tt}^{\tau}\mathbf{w}_{it} + f_{t,t+1}^{\tau}\mathbf{w}_{i,t+1} + \cdots + f_{tT}^{\tau}\mathbf{w}_{iT}, \\
\dot{v}_{it}^{\tau} &= f_{tt}^{\tau}v_{it} + f_{t,t+1}^{\tau}v_{i,t+1} + \cdots + f_{tT}^{\tau}v_{iT},
\end{aligned}
$$

where f_{st}^{τ} is defined in (2.11).

Next, to introduce an instrumental variables, we define

$$
\begin{aligned}
\underset{(T_2\times T)}{\mathbf{B}_T^{\tau}} &= \mathbf{B}_T^{\tau 1}\begin{bmatrix}
\frac{2(-2+3)}{2\cdot 1} & \frac{2(-2)}{2\cdot 1} & 1 & 0 & 0 & \cdots & \cdots & 0 \\
\frac{2(-4+6)}{3\cdot 2} & \frac{2(-4+3)}{3\cdot 2} & \frac{2(-4)}{3\cdot 2} & 1 & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\
\frac{2(-2T_4+3T_4)}{T_3T_4} & \frac{2(-2T_4+3T_5)}{T_3T_4} & \cdots & \cdots & \frac{2(-2T_4)}{T_3T_4} & 1 & \ddots & \vdots \\
\frac{2(-2T_3+3T_3)}{T_2T_3} & \frac{2(-2T_3+3T_4)}{T_2T_3} & \cdots & \cdots & \frac{2(-2T_3+3)}{T_2T_3} & \frac{2(-2T_3)}{T_2T_3} & 1 & 0 \\
\frac{2(-2T_2+3T_2)}{T_1T_2} & \frac{2(-2T_2+3T_3)}{T_1T_2} & \cdots & \cdots & \frac{2(-2T_2+6)}{T_1T_2} & \frac{2(-2T_2+3)}{T_1T_2} & \frac{2(-2T_2)}{T_1T_2} & 1
\end{bmatrix} \\
&= \{b_{st}^{\tau}\}
\end{aligned}
\tag{2.13}
$$

where $\mathbf{B}_T^{\tau 1} = \mathrm{diag}\left(c_{T_2}^{\tau}, \cdots c_2^{\tau}, c_1^{\tau}\right)$. Note that $\mathbf{B}_T^{\tau}$ can be obtained by rotating $\mathbf{F}_T^{\tau}$ (see (2.67) in the appendix). Using this, we define an instrumental variable $\ddot{\mathbf{W}}_i^{\tau} = \mathbf{B}_T^{\tau}\mathbf{W}_i = (\mathbf{z}_{i3}^{\tau}, ..., \mathbf{z}_{iT}^{\tau})'$ where its tth row is given by

$$
\ddot{\mathbf{w}}_{it}^{\tau} = b_{t-2,t}^{\tau}\mathbf{w}_{it} + b_{t-2,t-1}^{\tau}\mathbf{w}_{i,t-1} + \cdots + b_{t-2,1}^{\tau}\mathbf{w}_{i1}, \qquad (i = 1, ..., N; t = 3, ..., T) \tag{2.14}
$$

with b_{st}^{τ} being defined in (2.13). Note that the first two periods are lost due to the difference property of the transformation matrix $\mathbf{B}_T^{\tau}$.

Since $E\left(\ddot{\mathbf{w}}_{is}^{\tau}\dot{v}_{it}^{\tau}\right) = \mathbf{0}$, $(3 \leq s \leq t \leq T_2)$ holds, we can construct moment conditions from them. Specifically, we consider the moment conditions $E\left(\sum_{t=3}^{T_2}\ddot{\mathbf{w}}_{it}^{\tau}\dot{v}_{it}^{\tau}\right) = \mathbf{0}$. The corresponding instrumental variable estimator is given by

$$
\widehat{\boldsymbol{\delta}}_{IV}^{\tau} = \left(\sum_{i=1}^{N}\sum_{t=3}^{T_2}\ddot{\mathbf{w}}_{it}^{\tau}\dot{\mathbf{w}}_{it}^{\tau\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=3}^{T_2}\ddot{\mathbf{w}}_{it}^{\tau}\dot{y}_{it}^{\tau}\right). \tag{2.15}
$$

2.2.3 Unified model

To derive the asymptotic properties of the proposed IV estimators (2.8) and (2.15) in the next section, we formulate the above two models (2.1) and (2.10) in a unified framework. For this, let us define a variable d such that $d = 1$ corresponds to the FE model while $d = 2$ corresponds to the trend model. Also, let us define $\mathbf{C}_T$ and $\mathbf{F}_T$ such that $\mathbf{C}_T = \boldsymbol{\iota}_T$ and $\mathbf{F}_T = \mathbf{F}_T^{\iota}$ for the FE model, and $\mathbf{C}_T = (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T)$ and $\mathbf{F}_T = \mathbf{F}_T^{\tau}$ for the trend model. Thereby, the case $(d, T_d, \mathbf{F}_T, \mathbf{C}_T) = (1, T_1, \mathbf{F}_T^{\iota}, \boldsymbol{\iota}_T)$ corresponds to the FE model while $(d, T_d, \mathbf{F}_T, \mathbf{C}_T) = (2, T_2, \mathbf{F}_T^{\tau}, (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T))$ corresponds to the trend model. Note that $\mathbf{F}_T$ has the properties such that $\mathbf{F}_T\mathbf{C}_T = \mathbf{0}$, $\mathbf{F}_T\mathbf{F}_T' = \mathbf{I}_{T_d}$ and

$$
\mathbf{F}_T'\mathbf{F}_T = \mathbf{Q}_T = \mathbf{I}_T - \mathbf{C}_T\left(\mathbf{C}_T'\mathbf{C}_T\right)^{-1}\mathbf{C}_T' = \mathbf{I}_T - \mathbf{R}_T, \tag{2.16}
$$

$$
\mathbf{R}_T = \begin{cases} \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' & \text{FE model} \\ \frac{2(2T+1)}{T(T-1)}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \frac{12}{T(T-1)(T+1)}\boldsymbol{\tau}_T\boldsymbol{\tau}_T' - \frac{6}{T(T-1)}\left(\boldsymbol{\iota}_T\boldsymbol{\tau}_T' + \boldsymbol{\tau}_T\boldsymbol{\iota}_T'\right) & \text{trend model} \end{cases}. \tag{2.17}
$$

Using these, the models (2.1) and (2.10) can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \mathbf{C}_T\boldsymbol{\eta}_i + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{2.18}$$

Multiplying $\mathbf{F}_T$ to (2.18), we have the following transformed model

$$\mathbf{y}_i^* = \mathbf{W}_i^*\boldsymbol{\delta} + \mathbf{v}_i^* \tag{2.19}$$

where $\mathbf{y}_i^* = \mathbf{F}_T\mathbf{y}_i = (y_{i1}, ..., y_{iT_d})'$. $\mathbf{W}_i^* = \mathbf{F}_T\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT_d})'$ and $\mathbf{v}_i^* = \mathbf{F}_T\mathbf{v}_i = (v_{i1}, ..., v_{iT_d})'$. The tth row of (2.19) can be written as

$$y_{it}^* = \mathbf{w}_{it}^{*\prime}\boldsymbol{\delta} + v_{it}^*, \qquad (i = 1, ..., N; t = 1, ..., T_d) \tag{2.20}$$

Note that the models (2.4) and (2.12) are the special cases of (2.20).

Similarly, let $\mathbf{w}_{it}^{**}$, $(i = 1, ..., N; t = d+1, ..., T)$ denote $\ddot{\mathbf{w}}_{it}^{\iota}$ for FE model given by (2.7), and $\ddot{\mathbf{w}}_{it}^{\tau}$ for the trend model given by (2.14), respectively. Since $E\left(\sum_{t=d+1}^{T_d}\mathbf{w}_{it}^{**}v_{it}^*\right) = \mathbf{0}$, we have the following instrumental variable estimator

$$\widehat{\boldsymbol{\delta}}_{IV}^B = \left(\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\mathbf{w}_{it}^{**}\mathbf{w}_{it}^{*\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\mathbf{w}_{it}^{**}y_{it}^*\right). \tag{2.21}$$

Note that the previous two estimators (2.8) and (2.15) are the special case of (2.21). We call $\widehat{\boldsymbol{\delta}}_{IV}^B$ the *double filter instrumental variable* (DFIV) estimator since it is based on forward and backward filtering.

Alternatively, we can consider the GMM estimators that are more efficient than IV estimators. Since the first and last d periods are lost due to difference properties of $\mathbf{F}_T$ and $\mathbf{B}_T$, the middle T_{2d} periods are used in estimation. Hence, the model in a matrix form becomes

$$\check{\mathbf{y}}_i^* = \check{\mathbf{W}}_i^{*\prime}\boldsymbol{\delta} + \check{\mathbf{v}}_i^*, \qquad (i = 1, ..., N) \tag{2.22}$$

where $\check{\mathbf{y}}_i^* = \left(y_{i,d+1}^*, ..., y_{iT_d}^*\right)'$, $\check{\mathbf{W}}_i^* = \left(\mathbf{w}_{i,d+1}^*, ..., \mathbf{w}_{iT_d}^*\right)'$ and $\check{\mathbf{v}}_i^* = \left(v_{i,d+1}^*, ..., v_{iT_d}^*\right)'$. For this model, we consider the moment conditions given by $E(\mathbf{Z}_i^{B\prime}\check{\mathbf{v}}_i^*) = \mathbf{0}$ where $\mathbf{Z}_i^B = \text{diag}(\mathbf{z}_{i,d_0+1}^{B\prime}, ..., \mathbf{z}_{iT_d}^{B\prime})$, $\mathbf{z}_{it}^B = (\mathbf{w}_{i,t-\ell+1}^{**\prime}, ..., \mathbf{w}_{it}^{**\prime})'$, $(1 \leq \ell \leq t-d)$ and $\mathbf{v}_i^* = (v_{i,d_0+1}^*, ..., v_{iT_d}^*)'$. The corresponding one-step GMM estimator is given by[5]

[5]The two-step GMM estimator is not considered in this chapter since it requires a computation of large dimensional weighting matrix. Indeed, if the number of moment conditions $m = \sum_{t=d_0+1}^{T_d} m_t$ exceeds the sample size where m_t denotes the number of instruments used at period t, the optimal weighting matrix cannot be computed. However, the one-step GMM estimator (2.23) can be computed as long as $m_t < N$ for all t even when $m > N$.

$$
\begin{aligned}
\widehat{\boldsymbol{\delta}}^B_{GMM} &= \left[\left(\sum_{i=1}^N \check{\mathbf{W}}_i^{*\prime}\mathbf{Z}_i^B\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\mathbf{Z}_i^B\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\check{\mathbf{W}}_i^*\right)\right]^{-1} \\
&\quad\times\left[\left(\sum_{i=1}^N \check{\mathbf{W}}_i^{*\prime}\mathbf{Z}_i^B\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\mathbf{Z}_i^B\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\check{\mathbf{y}}_i^*\right)\right] \\
&= \left[\sum_{t=d_0+1}^{T_d} \underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^B\left(\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{Z}}_t^B\right)^{-1}\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{W}}_t^*\right]^{-1}\left[\sum_{t=d_0+1}^{T_d} \underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^B\left(\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{Z}}_t^B\right)^{-1}\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{y}}_t^*\right]
\end{aligned} \tag{2.23}
$$

where $\underline{\mathbf{W}}_t^* = (\mathbf{w}_{1t}^*, ..., \mathbf{w}_{Nt}^*)'$, $\underline{\mathbf{Z}}_t^B = (\mathbf{z}_{1t}^B, ..., \mathbf{z}_{Nt}^B)'$ and $\underline{\mathbf{y}}_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$. We call $\widehat{\boldsymbol{\delta}}^B_{GMM}$ the *double filter GMM* (DFGMM) estimator by the same reason as DFIV.

In order to compare the IV estimator with the FE estimator in the next section, we further reformulate (2.22) in terms of $\mathbf{y}_i$, $\mathbf{W}_i$ and $\mathbf{v}_i$. For this, let us define $\mathbf{L}_T = (\mathbf{0}_{T_d\times d}, \mathbf{I}_{T_d})$. Then, by noting that $\mathbf{L}_T\mathbf{y}_i = (y_{i,d+1}, ..., y_{iT})'$ and $\mathbf{L}_T\mathbf{W}_i = (\mathbf{w}_{i,d+1}, ..., \mathbf{w}_{iT})'$, the model (2.22) can be written as

$$
\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{y}_i = \mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i\boldsymbol{\delta} + \mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i, \qquad (i = 1, ..., N). \tag{2.24}
$$

Similarly, by using $\mathbf{K}_T = (\mathbf{I}_{T_d}, \mathbf{0}_{T_d\times d})$ and $\mathbf{K}_T\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT_d})'$, we have $\mathbf{W}_i^{**} = (\mathbf{w}_{i,d+1}^{**}, ..., \mathbf{w}_{iT_d}^{**})' = \mathbf{B}_{T_d}\mathbf{K}_T\mathbf{W}_i$ where $\mathbf{B}_T$ denotes $\mathbf{B}_T^{\iota}$ for the FE model and $\mathbf{B}_T^{\tau}$ for trend model. Using these, the moment conditions $E\left(\sum_{t=d+1}^{T_d}\mathbf{w}_{it}^{**}v_{it}^*\right) = \mathbf{0}$ can be written as $E(\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i) = \mathbf{0}$, and the IV estimator (2.21) can be written as

$$
\begin{aligned}
\widehat{\boldsymbol{\delta}}^B_{IV} &= \left(\sum_{i=1}^N \mathbf{W}_i^{**\prime}\mathbf{W}_i^*\right)^{-1}\left(\sum_{i=1}^N \mathbf{W}_i^{**\prime}\mathbf{y}_i^*\right) \\
&= \left[\sum_{i=1}^N \mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i\right]^{-1}\sum_{i=1}^N \mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{y}_i.
\end{aligned} \tag{2.25}
$$

In the next section, we compare the asymptotic properties of this IV estimator with that of the FE estimator given by

$$
\widehat{\boldsymbol{\delta}}_{FE} = \left[\sum_{i=1}^N \mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i\right]^{-1}\sum_{i=1}^N \mathbf{W}_i'\mathbf{Q}_T\mathbf{y}_i \tag{2.26}
$$

where $\mathbf{Q}_T$ is defined in (2.16).

Also, for later use, we define IV and GMM estimators using instrument in levels.

The IV estimator for model (2.20) using instruments $\mathbf{w}_{it}$ is given by

$$\widehat{\boldsymbol{\delta}}_{IV}^{L} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T_d}\mathbf{w}_{it}\mathbf{w}_{it}^{*\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T_d}\mathbf{w}_{it}y_{it}^{*}\right). \tag{2.27}$$

The GMM estimator based on the moment condition $E(\mathbf{Z}_i^{L\prime}\mathbf{v}_i^*) = \mathbf{0}$ where $\mathbf{Z}_i^L = \text{diag}(\mathbf{z}_{i1}^{L\prime}, ..., \mathbf{z}_{iT_d}^{L\prime})$, $\mathbf{z}_{it}^L = (\mathbf{w}_{i,t-\ell+1}^{\prime}, ..., \mathbf{w}_{it}^{\prime})^{\prime}$, $(1 \leq \ell \leq t)$ is given by

$$\widehat{\boldsymbol{\delta}}_{GMM}^{L} = \left[\sum_{t=1}^{T_d}\underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^{L}\left(\underline{\mathbf{Z}}_t^{L\prime}\underline{\mathbf{Z}}_t^{L}\right)^{-1}\underline{\mathbf{Z}}_t^{L\prime}\underline{\mathbf{W}}_t^{*}\right]^{-1}\left[\sum_{t=1}^{T_d}\underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^{L}\left(\underline{\mathbf{Z}}_t^{L\prime}\underline{\mathbf{Z}}_t^{L}\right)^{-1}\underline{\mathbf{Z}}_t^{L\prime}\underline{\mathbf{y}}_t^{*}\right] \tag{2.28}$$

where $\underline{\mathbf{Z}}_t^L = (\mathbf{z}_{1t}^L, ..., \mathbf{z}_{Nt}^L)^{\prime}$.

2.3 Asymptotic properties

In this section, we derive the asymptotic properties of the IV and GMM estimators introduced in the previous section. Specifically, we consider two asymptotic schemes: fixed T and large N asymptotics and large N and large T asymptotics.

We first consider the case with small T and large N, and then consider large N and large T case.

2.3.1 Fixed T and large N case

Fixed T and large N asymptotic properties of IV and GMM estimators are well established in the literature. Under suitable conditions, the IV and GMM estimators are consistent and asymptotically normally distributed. Moreover, since the GMM estimators exploits more moment conditions than the IV estimator, the GMM estimator is more efficient than the IV estimator. To investigate the efficiency property associated with different form of instruments in detail, let us consider the simple AR(1) model with only fixed effects for illustration as follows:

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (t = 1, ..., T; i = 1, ..., N). \tag{2.29}$$

We make the following assumption

Assumption 1. *We assume that (a)$|\alpha|< 1$, (b) idiosyncratic error term v_{it} is iid with $E(v_{it}) = 0$ and $Var(v_{it}) = \sigma_v^2$, (c) unobserved individual effects η_i is iid with $E(\eta_i) = 0$ and $Var(\eta_i) = \sigma_\eta^2$, (d)initial conditions follow the stationary distribution:*

$$y_{i0} = \frac{\eta_i}{1-\alpha} + e_{i0} \tag{2.30}$$

where $e_{i0} \sim iid(0, \sigma_v^2/(1-\alpha)^2)$.

Most of these assumptions are made just to simplify the theoretical consideration. Indeed, for consistency of IV and GMM estimator, the idiosyncratic term can be heteroskedastic and initial conditions do not need to follow the stationary distribution.

In practice, researchers do not use all past variables as instruments since it causes many instruments problem and resulting GMM estimator is biased. Instead, they used only a few lagged variable in each period. Okui [2009] proposed a statistical procedure to select the number of instruments so that mean-squared error of the GMM estimator is minimized under large T and large N framework. Here, we investigate the effect of lag length of instruments used in each period in terms of efficiency. Since one of the main reasons to use many instruments is to improve efficiency, it would be of interest how efficiency changes depending on the lag length of instruments. To the best of authors' knowledge, such an analysis has not been conducted in the literature even in the simple AR(1) model.

The model after FOD transformation is given by

$$y_{it}^{*} = \alpha y_{i,t-1}^{*} + v_{it}^{*}, \qquad (t = 1, ..., T-1; i = 1, ..., N).$$

Let $\mathbf{z}_{it}^{L(\ell)} = (y_{i,t-\ell}, ..., y_{i,t-1})$ and $\mathbf{z}_{it}^{B(\ell)} = (y_{i,t-\ell}^{**}, ..., y_{i,t-1}^{**})$ be an $m_t^L \times 1$ and $m_t^B \times 1$ vectors of instruments, respectively, where ℓ denotes the maximum length of instruments used in each period and $y_{i,t-l}^{**} = c_{T-t+l+1}^{\iota} [y_{i,t-l} - (y_{i,t-l-1} + \cdots + y_{i0})/(t-l)]$. Note that $m_t^L, m_t^B = t$ for $t < \ell$ and $m_t^L, m_t^B = \ell$ for $t \geq \ell$.

Specifically, the GMM estimator with $\mathbf{z}_{it}^{L(\ell)}$ and $\mathbf{z}_{it}^{B(\ell)}$ as instruments are respectively given as

$$\begin{aligned} \widehat{\alpha}_{GMM}^{L(\ell)} &= \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{L(\ell)} \mathbf{y}_{t-1}^{*} \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{L(\ell)} \mathbf{y}_{t}^{*} \right), \\ \widehat{\alpha}_{GMM}^{B(\ell)} &= \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{B(\ell)} \mathbf{y}_{t-1}^{*} \right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{B(\ell)} \mathbf{y}_{t}^{*} \right) \end{aligned}$$

where $\mathbf{y}_{t-1}^{*} = (y_{1,t-1}^{*}, ..., y_{N,t-1}^{*})'$, $\mathbf{y}_{t}^{*} = (y_{1t}^{*}, ..., y_{Nt}^{*})'$, $\mathbf{P}_t^{L(\ell)} = \mathbf{Z}_t^{L(\ell)} \left(\mathbf{Z}_t^{L(\ell)\prime} \mathbf{Z}_t^{L(\ell)} \right)^{-1} \mathbf{Z}_t^{L(\ell)\prime}$, $\mathbf{P}_t^{B(\ell)} = \mathbf{Z}_t^{B(\ell)} \left(\mathbf{Z}_t^{B(\ell)\prime} \mathbf{Z}_t^{B(\ell)} \right)^{-1} \mathbf{Z}_t^{B(\ell)\prime}$, $\mathbf{Z}_t^{L(\ell)} = \left(\mathbf{z}_{1t}^{L(\ell)}, ..., \mathbf{z}_{Nt}^{L(\ell)} \right)'$ and $\mathbf{Z}_t^{B(\ell)} = \left(\mathbf{z}_{1t}^{B(\ell)}, ..., \mathbf{z}_{Nt}^{B(\ell)} \right)'$.

The asymptotic variances of $\widehat{\alpha}_{GMM}^{L(\ell)}$ and $\widehat{\alpha}_{GMM}^{B(\ell)}$ with fixed T and large N are given by

$$Avar\left(\widehat{\alpha}_{GMM}^{L(\ell)}\right) = \sigma_v^2 \left[\sum_{t=1}^{T-1} E\left(y_{i,t-1}^{*} \mathbf{z}_{it}^{L(\ell)\prime} \right) \left[E\left(\mathbf{z}_{it}^{L(\ell)} \mathbf{z}_{it}^{L(\ell)\prime} \right) \right]^{-1} E\left(\mathbf{z}_{it}^{L(\ell)} y_{i,t-1}^{*} \right) \right]^{-1} \quad (2.31)$$

$$Avar\left(\widehat{\alpha}_{GMM}^{B(\ell)}\right) = \sigma_v^2 \left[\sum_{t=1}^{T-1} E\left(y_{i,t-1}^{*} \mathbf{z}_{it}^{B(\ell)\prime} \right) \left[E\left(\mathbf{z}_{it}^{B(\ell)} \mathbf{z}_{it}^{B(\ell)\prime} \right) \right]^{-1} E\left(\mathbf{z}_{it}^{B(\ell)} y_{i,t-1}^{*} \right) \right]^{-1} \quad (2.32)$$

Figure 1 shows the asymptotic variances of $\widehat{\alpha}_{GMM}^{L(\ell)}$ and $\widehat{\alpha}_{GMM}^{B(\ell)}$ based on (2.31) and (2.32) with various lag length of instruments for the cases with $\alpha = 0.3, 0.6, 0.9$, $r = 0.2, 1, 5$ and $T = 10$. For the detail of computation of (2.31) and (2.32), see appendix. From the figure, it is found that when $r = 0.2$, the asymptotic variances of two GMM estimators are very similar regardless of lag length of instruments ℓ. This implies that many lags are not required to improve efficiency. Also it is found that $\widehat{\alpha}_{GMM}^{L}$ tends to be more efficient than $\widehat{\alpha}_{GMM}^{B}$. This is because one estimation period is lost in $\widehat{\alpha}_{GMM}^{B}$ compared with $\widehat{\alpha}_{GMM}^{L}$. However, when $r = 1$ and $r = 5$, the result dramatically changes. From the figure, it is found that $\widehat{\alpha}_{GMM}^{B}$ is little affected by lag length and hence, we do not need to use many lags; in view of the figure, one lagged instrument is enough to obtain nearly efficient estimator. However, this is not the case for $\widehat{\alpha}_{GMM}^{L}$. When $r = 1$ and $r = 5$, the lag length of instruments substantially effects the asymptotic variance. When

r is large, $\widehat{\alpha}^L_{GMM}$ with one or two lagged instruments are far less efficient than $\widehat{\alpha}^B_{GMM}$ with the same number of instruments. But as the lag length increases, efficiency improve. What is striking is that efficiency gain when lag length is increased from one to two or two from three is substantial. If lag length is more than four, the asymptotic variances of $\widehat{\alpha}^L_{GMM}$ and $\widehat{\alpha}^B_{GMM}$ are very similar. This implies that although many instruments leads to efficiency gain, in the current case, three or four lags are sufficient to obtain reasonably efficient $\widehat{\alpha}^L_{GMM}$. In other words, using higher order lags does not contribute to efficiency gain so much. Thus, this result supports the use of a few lagged variable as instruments. Based on this results, in the following, we mainly consider $\widehat{\alpha}^L_{GMM}$ with $\ell = 3$ and $\widehat{\alpha}^B_{GMM}$ with $\ell = 1$.

Next, we investigate the effect of time length T on the efficiency of IV and GMM estimators with level and BOD instruments. In addition to the GMM estimator, we consider IV estimators given by

$$\widehat{\alpha}^L_{IV} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} y_{i,t-1}y^*_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} y_{i,t-1}y^*_{it}\right), \tag{2.33}$$

$$\widehat{\alpha}^B_{IV} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} y^{**}_{i,t-1}y^*_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} y^{**}_{i,t}y^*_{it}\right). \tag{2.34}$$

Theorem 1. *Asymptotic variances of IV and GMM estimators with fixed T and large N asymptotics are given by*

$$Avar\left(\widehat{\alpha}^{L(t)}_{GMM}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1}\psi_t^2\left(1-\frac{r(\frac{1+\alpha}{1-\alpha})}{1+r\{(\frac{1+\alpha}{1-\alpha})+(t-1)\}}\right)\right]^{-1}, \tag{2.35}$$

$$Avar\left(\widehat{\alpha}^{L(3)}_{GMM}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1}\psi_t^2\left(1-\frac{r(\frac{1+\alpha}{1-\alpha})}{1+r\{(\frac{1+\alpha}{1-\alpha})+m^L_t\}}\right)\right]^{-1}, \tag{2.36}$$

$$Avar\left(\widehat{\alpha}^{L(1)}_{GMM}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1}\psi_t^2\left(\frac{1}{1+r(\frac{1+\alpha}{1-\alpha})}\right)\right]^{-1}, \tag{2.37}$$

$$Avar\left(\widehat{\alpha}^{L}_{IV}\right) = (1-\alpha^2)\left[1+r\left(\frac{1+\alpha}{1-\alpha}\right)\right]\left[\sum_{t=1}^{T-1}\psi_t\right]^{-2}, \tag{2.38}$$

$$Avar\left(\widehat{\alpha}^{B(1)}_{GMM}\right) = (1-\alpha^2)\left[\sum_{t=2}^{T-1}\psi_t^2\left(1-\frac{\alpha\phi_{t-1}}{t-1}\right)^2 A_t^{-1}\right]^{-1}, \tag{2.39}$$

$$Avar\left(\widehat{\alpha}^{B}_{IV}\right) = (1-\alpha^2)\left(\sum_{t=2}^{T-1} c^{t2}_{T-t+1}A_t\right)\left[\sum_{t=2}^{T-1}\psi_t c^t_{T-t+1}\left(1-\frac{\alpha\phi_{t-1}}{t-1}\right)^2\right]^{-2} \tag{2.40}$$

where $r = \sigma_\eta^2/\sigma_v^2$ *and*

$$\psi_t = c_t^\iota\left[1 - \frac{\alpha\phi_{T-t}}{T-t}\right], \quad (2.41)$$

$$\phi_j = \frac{1-\alpha^j}{1-\alpha} = 1 + \alpha + \cdots + \alpha^{j-1}, \quad (2.42)$$

$$A_t = \left[1 - \frac{2\alpha\phi_{t-1}}{t-1} + \frac{1}{(t-1)^2}\left\{\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}\right]. \quad (2.43)$$

Figure 2 depicts $Avar\left(\widehat{\alpha}_{GMM}^{L(3)}\right)$, $Avar\left(\widehat{\alpha}_{IV}^{L}\right)$, $Avar\left(\widehat{\alpha}_{GMM}^{B(1)}\right)$ and $Avar\left(\widehat{\alpha}_{IV}^{B}\right)$ for $\alpha = 0.3, 0.6, 0.9$, $r = 0.2, 1, 5$ and $T = 5, 6, ..., 20$ based on Theorem 1[6]. From the figure, it is found that the efficiency of $\widehat{\alpha}_{IV}^{L}$ is substantially affected by the variance ratio r. When r is large, $\widehat{\alpha}_{IV}^{L}$ is much less efficient than other estimators. Also, it is found that the variances of $\widehat{\alpha}_{GMM}^{B(1)}$ and $\widehat{\alpha}_{IV}^{B}$ are almost identical in all cases. With regard to the effect of T, we find that the difference in efficiency between GMM estimators using instruments in levels and new instruments are not small when T is less than 10 and $r = 0.2$. However, that difference becomes smaller as r increases. Indeed, when r is larger than 1 and T is larger than 10, $\widehat{\alpha}_{GMM}^{L(3)}$, $\widehat{\alpha}_{GMM}^{B(1)}$ and $\widehat{\alpha}_{IV}^{B}$ have a very similar efficiency property. However, it should be note that $\widehat{\alpha}_{GMM}^{B(1)}$ and $\widehat{\alpha}_{IV}^{B}$ use less instruments than $\widehat{\alpha}_{GMM}^{L(3)}$.

Also, from Theorem 1, we heuristically find that $T^{-1}Avar\left(\widehat{\alpha}_{GMM}^{L(t)}\right)$, $T^{-1}Avar\left(\widehat{\alpha}_{GMM}^{B(1)}\right)$ and $T^{-1}Avar\left(\widehat{\alpha}_{IV}^{B}\right)$ tend to $(1-\alpha^2)$, which coincides with the asymptotic variance under large T and large N, whereas it is not the case for $Avar\left(\widehat{\alpha}_{GMM}^{L(3)}\right)$, $Avar\left(\widehat{\alpha}_{GMM}^{L(1)}\right)$ and $Avar\left(\widehat{\alpha}_{IV}^{L}\right)$. A formal discussion under large N and large T asymptotics is given next.

2.3.2 Large T and large N case

when N and T are large. We first make the following assumptions.

Assumption 2. *The error term* v_{it} *are serially and cross-sectionally uncorrelated and satisfy*

$$E(v_{it}|\mathbf{w}_{it}, ..., \mathbf{w}_{i1}, \eta_i) = 0. \quad (2.44)$$

[6]Since $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell=3)}\right)$ and $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell=t)}\right)$, and $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell=1)}\right)$ and $Avar\left(\widehat{\alpha}_{IV}^{L}\right)$ are very similar, $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell=t)}\right)$ and $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell=1)}\right)$ are excluded in the figure.

Assumption 3. *The regressor* $\mathbf{w}_{it}$ *follows the process:*

$$\mathbf{w}_{it} = \begin{cases} \boldsymbol{\mu}_i + \boldsymbol{\xi}_{it} & \text{FE model} \\ \boldsymbol{\mu}_i + \boldsymbol{\kappa}_i t + \boldsymbol{\xi}_{it} & \text{trend model} \end{cases} \tag{2.45}$$

where $E(\boldsymbol{\xi}_{it}) = \mathbf{0}$, $E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}'_{i,t+s}\right) = \boldsymbol{\Gamma}_{i,s}$ *and* $\sum_{\ell=-\infty}^{\infty}\|\boldsymbol{\Gamma}_{i,\ell}\| < \infty$ *for all* i. *Also, for all* i, *assume that* $E(\boldsymbol{\xi}_{it}v_{is}) = \mathbf{0}$ *for* $t \le s$ *and* $E(\boldsymbol{\xi}_{it}v_{is}) = \boldsymbol{\phi}_{i,t-s} \neq \mathbf{0}$ *for* $t > s$ *where* $\sum_{\ell=1}^{\infty}\|\boldsymbol{\phi}_{i,\ell}\| < \infty$. $\boldsymbol{\mu}_i$ *and* $\boldsymbol{\kappa}_i$ *are uncorrelated with* v_{it} *for all* i *and* t, *but can be correlated with* η_i *and* λ_i *in an unrestricted manner.*

Assumption 4. *As* $N, T \to \infty$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}\boldsymbol{\xi}'_{it} \to^p \boldsymbol{\Gamma}_0, \tag{2.46}$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} \to^d \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \tag{2.47}$$

where $\boldsymbol{\Gamma}_0 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\boldsymbol{\Gamma}_{i0}$, $\boldsymbol{\Omega} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(v_{it}^2\boldsymbol{\xi}_{it}\boldsymbol{\xi}'_{it})$ *and both* $\boldsymbol{\Gamma}_0$ *and* $\boldsymbol{\Omega}$ *are positive definite.*

Assumption 2 indicates that the regressor $\mathbf{w}_{it}$ is weakly exogenous. The correlation structure between regressors and errors are specified in Assumption 3. Assumption 4 is a high-level assumption that can be used to derive the large N and T asymptotic properties. More primitive assumptions can be found in Phillips and Moon [1999].

The following Lemma 1 is useful to understand the relationship between $\widehat{\boldsymbol{\delta}}_{FE}$ and $\widehat{\boldsymbol{\delta}}_{IV}^{B}$.

Lemma 1. *Let Assumptions 2 and 3 hold. Then,*

$$(a)\quad \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}'_i\mathbf{K}'_T\mathbf{B}'_{T_d}\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}'_i\mathbf{Q}_T\mathbf{W}_i + O_p\left(\frac{\log T}{T}\right), \tag{2.48}$$

$$(b)\quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}'_i\mathbf{Q}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} - O_p\left(\sqrt{\frac{N}{T}}\right), \tag{2.49}$$

$$(c)\quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}'_i\mathbf{K}'_T\mathbf{B}'_{T_d}\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} + o_p(1). \tag{2.50}$$

Lemma 1(a) indicates that the denominators of $\widehat{\boldsymbol{\delta}}_{FE}$ and $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ are asymptotically equivalent when T is large. Also, comparing (b) and (c), we find that the second term of the right-hand side makes a significant difference in FE and DFIV estimators. When

N/T converges to a non-zero constant, the second term becomes $O_p(1)$, and because of this, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{FE}$ is not centered around the true value as shown in Theorem 2 below. This bias is due the incidental parameter problem. Contrary to the FE estimator, the second term of (c) vanishes asymptotically. Hence, as shown in Theorem 4 below, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ is centered around the true value.

Specifically, the asymptotic distributions of $\widehat{\boldsymbol{\delta}}_{FE}$ and $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ are given in the following theorems.

Theorem 2. *Let Assumptions 2, 3 and 4 hold. Also assume that $N/T \to \kappa, (0 < \kappa < \infty)$. Then the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{FE}$ as $N, T \to \infty$ is given by*

$$\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{FE} - \boldsymbol{\delta}) \xrightarrow{d} \mathcal{N}(\mathbf{b}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^{-1})$$

where $\mathbf{b} = \sqrt{\kappa}\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{h}}$ *and* $\bar{\mathbf{h}} = \mathrm{plim}_{N,T\to\infty} N^{-1}\sum_{i=1}^{N} E(\mathbf{W}_i'\mathbf{Q}_T\mathbf{v}_i)$.

This result implies that the asymptotic distribution of the FE estimator is not centered around the true value due to the bias caused by the incidental parameter problem. To correct for this bias, we consider a bias-corrected FE estimator:

$$\widehat{\boldsymbol{\delta}}_{BCFE} = \widehat{\boldsymbol{\delta}}_{FE} - \frac{1}{T}\widehat{\boldsymbol{\Gamma}}_0^{-1}\widehat{\bar{\mathbf{h}}} \tag{2.51}$$

where $\widehat{\boldsymbol{\Gamma}}_0 = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i$ and $\widehat{\bar{\mathbf{h}}}$ is a consistent estimator of $\bar{\mathbf{h}}$. This bias correction is not always possible in practice and feasibility depends on the model specification. For instance, if the model is assumed to be AR(1), then, it is possible to correct the bias as proposed in Hahn and Kuersteiner [2002]. However, for other cases, say, for a model with weakly exogenous regressors, bias-correction is infeasible unless a specific form is assumed for the regressors, which is undesirable in practice, since the form of bias depends on the correlation structure between the regressors and errors. Apart from the feasibility, the asymptotic distribution of bias-corrected FE estimator is given in the following theorem.

Theorem 3. *Let Assumptions 2, 3 and 4 hold. Then the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{BCFE}$ as $N, T \to \infty$ is given by*

$$\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{BCFE} - \boldsymbol{\delta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^{-1}).$$

Finally, the asymptotic distribution of the DFIV estimator is given in the following theorem.

Theorem 4. *Let Assumptions 2, 3 and 4 hold. Then, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ as $N, T \to \infty$ is given by*

$$\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{IV}^{B} - \boldsymbol{\delta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^{-1}).$$

This theorem implies that $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ has the same asymptotic distribution $\widehat{\boldsymbol{\delta}}_{BCFE}$ when both N and T are large.

For the GMM estimator, since it is quite involved to derive the large T and large N asymptotic properties for the general model (2.18), we instead consider a simple AR(1) model given by (2.29). The asymptotic distributions of IV/GMM estimators for the AR(1) model (2.29) under large N and large T asymptotics is given in the following theorem.

Theorem 5. *The asymptotic distributions of FE, IV and GMM estimators under large N and large T are given as follows:*

$$
\begin{aligned}
&(a) \quad \sqrt{NT}\left[\widehat{\alpha}_{FE} - \left(\alpha - \frac{1}{T}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \qquad \textit{if } T^3/N \to 0,\\
&(b) \quad \sqrt{NT}\left[\widehat{\alpha}_{GMM}^{L(t)} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2),\\
&\qquad\qquad\qquad\qquad \textit{if } (\log T)^2/N \to \infty \textit{ and } T/N \to c, (0 \le c \le 1),\\
&(c) \quad \sqrt{NT}\left(\widehat{\alpha}_{GMM}^{L(1)} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\left(1 + \frac{r(1+\alpha)}{1-\alpha}\right)\right), \qquad \textit{if } T/N \to c, (0 \le c \le 1),\\
&(d) \quad \sqrt{NT}\left(\widehat{\alpha}_{IV}^{L} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\left(1 + \frac{r(1+\alpha)}{1-\alpha}\right)\right),\\
&(e) \quad \sqrt{NT}\left(\widehat{\alpha}_{GMM}^{B(1)} - \alpha\right) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \qquad \textit{if } T/N \to c, (0 \le c \le 1),\\
&(f) \quad \sqrt{NT}\left(\widehat{\alpha}_{IV}^{B} - \alpha\right) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2).
\end{aligned}
$$

These results are already derived in the literature. (a) is due to Hahn and Kuersteiner [2002] and Alvarez and Arellano [2003], and (b) is due to Alvarez and Arellano [2003], (c) is derived by Hayakawa [2006] and Hsiao and Zhou [2015], (d) is derived by Hsiao and Zhou [2015], (e) is derived by Hayakawa [2006], and (f) is derived by Hayakawa [2009a].

Note that while GMM estimators require conditions on the relative speed between N and T, such a condition is not required for IV estimators $\widehat{\alpha}_{IV}^{L}$ and $\widehat{\alpha}_{IV}^{B}$. From Theorem 5, we find that the GMM estimator using instruments in levels becomes efficient[Hahn and Kuersteiner, 2002] if all past variables are used as instruments in each period. However, if only one lagged variable is used as an instrument, the GMM estimator is not efficient and also it has the same asymptotic distribution as IV estimator $\widehat{\alpha}_{IV}^{L}$. However, for IV and GMM estimators using BOD filtered instruments, both $\widehat{\alpha}_{GMM}^{B(1)}$ and $\widehat{\alpha}_{IV}^{B}$ become efficient, which implies that we do not need to use many instruments to enhance efficiency.

2.4 Monte Carlo simulation

In this section, we investigate the finite sample properties of the proposed estimators in the context of dynamic panel data models with/without time trends.

2.4.1 Design

The data are generated as

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + \varphi \lambda_i t + v_{it}, \\ x_{it} &= \rho x_{i,t-1} + \tau_\eta \eta_i + \varphi \tau_\lambda \lambda_i t + \theta v_{i,t-1} + e_{it}. \end{aligned}$$

Note that the case with $\varphi = 0$ corresponds to the FE model while that with $\varphi = 1$ corresponds to the trend model. In a matrix form, this can be written as

$$\begin{aligned} \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} &= \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix} \begin{pmatrix} y_{i,t-1} \\ x_{i,t-1} \end{pmatrix} + \begin{pmatrix} (1+\beta\tau_\eta) \\ \tau_\eta \end{pmatrix} \eta_i \\ &\quad + \varphi \begin{pmatrix} (1+\beta\tau_\lambda) \\ \tau_\lambda \end{pmatrix} \lambda_i t + \begin{pmatrix} v_{it} + \beta\theta v_{i,t-1} + \beta e_{it} \\ \theta v_{i,t-1} + e_{it} \end{pmatrix} \end{aligned}$$

or

$$\mathbf{p}_{it} = \mathbf{\Phi} \mathbf{p}_{i,t-1} + \mathbf{c}_\eta \eta_i + \varphi \mathbf{c}_\lambda \lambda_i t + \boldsymbol{\varepsilon}_{it} \tag{2.52}$$

where $\mathbf{p}_{it} = (y_{it}, x_{it})'$, $\mathbf{c}_\eta = (1+\beta\tau_\eta, \tau_\eta)'$, $\mathbf{c}_\lambda = (1+\beta\tau_\lambda, \tau_\lambda)'$, $\boldsymbol{\varepsilon}_{it} = (v_{it} + \beta\theta v_{i,t-1} + \beta e_{it}, \theta v_{i,t-1} + e_{it})'$ and

$$\mathbf{\Phi} = \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix}. \tag{2.53}$$

Alternatively, $\mathbf{p}_{it}$ can be written as in a component form:

$$\begin{aligned} \mathbf{p}_{it} &= \mathbf{a}_i + \varphi \mathbf{b}_i t + \boldsymbol{\zeta}_{it}, \\ \boldsymbol{\zeta}_{it} &= \mathbf{\Phi} \boldsymbol{\zeta}_{i,t-1} + \boldsymbol{\varepsilon}_{it} \end{aligned} \tag{2.54}$$

where

$$\begin{aligned} \mathbf{a}_i &= (\mathbf{I} - \mathbf{\Phi})^{-1} \mathbf{c}_\eta \eta_i - (\mathbf{I} - \mathbf{\Phi})^{-1} \mathbf{\Phi} (\mathbf{I} - \mathbf{\Phi})^{-1} \mathbf{c}_\lambda \lambda_i, \\ \mathbf{b}_i &= (\mathbf{I} - \mathbf{\Phi})^{-1} \mathbf{c}_\lambda \lambda_i, \\ Var(\boldsymbol{\varepsilon}_{it}) &= \begin{bmatrix} (1+\beta^2\theta^2)\sigma_v^2 + \beta^2\sigma_e^2 & \beta\theta^2\sigma_v^2 + \beta\sigma_e^2 \\ \beta\theta^2\sigma_v^2 + \beta\sigma_e^2 & \theta^2\sigma_v^2 + \sigma_e^2 \end{bmatrix}. \end{aligned}$$

Data for y_{it} and x_{it} are generated from (2.54). For the sample size, we consider $T = 10, 25, 50, 100$ and $N = 50, 100, 250$. For parameter values, we consider $\alpha = 0.4, 0.8$,

$\beta = 1$, $\rho = 0.5$, $\theta = -0.2$, $\tau_\eta = 0.5$, $\tau_\lambda = 0.5$. v_{it}, e_{it}, η_i and λ_i are independently generated as $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$, $e_{it} \sim \mathcal{N}(0, \sigma_e^2)$, $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$ and $\lambda_i \sim \mathcal{N}(0, \sigma_\lambda^2)$ with $\sigma_v^2 = 1$ $\sigma_e^2 = 0.16$, $\sigma_\eta^2 = 1, 5$ and $\sigma_\lambda^2 = 1$[7]. We report the median bias, interquartile range (IQR), median absolute error (MAE) and empirical size with 5% significance level based on 2,000 replications.

2.4.2 Estimators to be compared

We consider seven estimators. The first is the FE estimator $\widehat{\boldsymbol{\delta}}_{FE}$ given in (2.26)[8]. The second is the IV estimator $\widehat{\boldsymbol{\delta}}_{IV}^{L}$ given in (2.27) where instruments in levels are used. The third and fourth are the GMM estimator $\widehat{\boldsymbol{\delta}}_{GMM}^{L}$ given in (2.28) where instruments in levels are used. For the choice of lag length of instruments, we consider $\ell = 1$ and 3. The corresponding GMM estimators are denoted as "LEV1" and "LEV3", respectively. The fifth estimator is the DFIV estimator $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ defined in (2.21). The last two estimators are the GMM estimator $\widehat{\boldsymbol{\delta}}_{GMM}^{B}$ defined in (2.23) where backward filtered instruments are used. For the choice of lag length of instruments, we consider $\ell = 1$ and 3. The corresponding GMM estimators are denoted as "BOD1" and "BOD3", respectively. For the computation of standard errors, we use those obtained under large N and fixed T since they are more accurate than those obtained under large N and large T[see Hayakawa, 2015].

2.4.3 Results

Simulation results are provided in Tables 1-4. We first consider the model with fixed effects only. From Tables 1 and 2, we find that the FE estimator for α is severely biased when $T = 10$. However, as T gets larger, the bias becomes small as expected since the FE estimator is consistent when T is large. However, in terms of accuracy of inference, the sizes are severely distorted even when T is large, say, $T = 100$. This is because the asymptotic distribution of the FE estimator is not centered around the true value due to the incidental parameter problem. Also, note that increase in N does not reduce the bias since the bias of FE estimator does not depend on N. With regard to the FE estimator of β, the performance is better than those of α. However, it still shows some bias and size distortions. This result implies that the FE estimator does not work even when T is large. Also, note that a widely acceptable bias-correction method is not available since the regressor is weakly exogenous[9]. With regard to the IV and GMM

[7] Although we tried the cases with $(\sigma_\eta^2, \sigma_\lambda^2) = (1, 5), (5, 5)$, the results are very similar to those with $(\sigma_\eta^2, \sigma_\lambda^2) = (5, 1)$. Hence the results of these cases are not reported to save space.

[8] A bias corrected FE estimator is not compared since it is not available in the current case where the regressors is weakly exogenous.

[9] If the regressors are strictly exogenous, bias-corrected FE estimators such as Bun and Carree [2005] or Breitung and Hayakawa [2015] can be used.

estimators, in terms of MAE, the IV estimators using instruments in levels perform worst among the four estimators mainly due to the large dispersions. With regard to the remaining three estimators, they perform very similarly in terms of MAE when $T = 10$. However, as T gets larger, IV and GMM estimators using new instruments outperform the GMM estimator using instruments in levels. With regard to the choice fo IV or GMM estimators using new instruments, it is observed that GMM estimator tends to slightly smaller MAEs than IV estimator. In terms of accuracy of inference, IV and GMM estimators using new instruments have almost correct empirical sizes in all cases while the GMM estimator using instruments in levels have large size distortions especially when $T = 10$ and $\alpha = 0.8$. With regard to the effects of lags of instruments ℓ, we find that the efficiency of GMM estimator using instruments in levels substantially depends on ℓ. Comparing the IQRs with $\ell = 1$ and 3, the reduction of dispersion with $\ell = 3$ is substantial though it induces many instruments. Contrary to IV/GMM with instruments in levels, the effects of ℓ in GMM with new instruments are minor and the IQRs are relatively smaller than those of GMM with instruments in levels. This result is consistent with the theoretical implication that using new instruments leads to efficient estimation. Considering overall performance, we may conclude that the IV estimator or GMM estimator using new instruments with $\ell = 1$ tend to perform best in many cases.

Next, we consider the models with both fixed effects and heterogeneous time trends. The results are provided in Tables 3 and 4. Compared with the models with fixed effects only, the FE estimator is severely biased when T is small in this model too, and the magnitude of bias is larger. This also can be seen in the substantial size distortions even for a large $T = 100$. This implies that the FE estimator deteriorates further if time trends are included in the model. With regard to the IV and GMM estimators, IV estimator using instruments in levels perform poorly compared with other estimators. However, contrary to the previous model, other three IV and GMM estimators perform poorly when $T = 10$. Compared with the previous model with fixed effects, the dispersion is much larger when $T = 10$. However, the performances of these estimators improve as T gets larger. When $T = 25$ or larger, three estimators perform reasonably well when $\alpha = 0.4$ while more than $T = 50$ is required when $\alpha = 0.8$. For the relative performance among the three estimators, we find that the GMM estimator using instruments in levels perform best when $T = 10$. However, for all other cases, the GMM estimator using new instruments perform best.

2.5 Conclusion

In this chapter, we have proposed a new instrumental variable estimator for panel data models including static and dynamic models with weakly exogenous variables and with fixed effects and/or heterogeneous time trends. We showed that the new IV estimator called the DFIV estimator is consistent and has the same asymptotic distribution as the bias-corrected fixed effects estimator, which is sometimes infeasible, when both N and T are large. This implies that the DFIV estimator is as efficient as the fixed effects estimator. Monte Carlo simulation results revealed that the DFIV and DFGMM estimators tend to perform better than the conventional IV/GMM estimators using instruments in levels in almost all cases.

Appendix

2.A Derivation of $\mathbf{F}_T$

We derive the form of $\mathbf{F}_T^\iota$ and $\mathbf{F}_T^\tau$. Although a brief derivation of $\mathbf{F}_T^\iota$ is given in Arellano [2003], a complete derivation is not provided. Hence, we fill that gap. Let us define the following $T_1 \times T$ matrix that takes the first difference:

$$\mathbf{D}_T = \begin{bmatrix} -1 & 1 & 0 & & & 0 \\ 0 & -1 & 1 & & & \\ & & & \ddots & \ddots & \\ 0 & & & & -1 & 1 \end{bmatrix}.$$

Multiplying $\mathbf{D}_T$ by (2.1) and noting that $\mathbf{D}_T \boldsymbol{\iota}_T = \mathbf{0}$, we have

$$\mathbf{D}_T \mathbf{y}_i = \mathbf{D}_T \mathbf{W}_i + \mathbf{D}_T \mathbf{v}_i.$$

where it is simply assumed that $Var(\mathbf{v}_i) = \sigma_v^2 \mathbf{I}_T$. Since $Var(\mathbf{D}_T \mathbf{v}_i) = \sigma_v^2 \mathbf{D}_T \mathbf{D}_T'$, the transformed error is serially correlated. To correct for the serial correlation, we use the following transformation matrix, which is a GLS transformation:

$$\mathbf{F}_T^\iota = \left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2} \mathbf{D}_T,$$

where $(\mathbf{D}_T \mathbf{D}_T')^{-1/2}$ is the *upper* triangular Cholesky factorization of $(\mathbf{D}_T \mathbf{D}_T')^{-1}$ with[10][11]

$$\mathbf{D}_T \mathbf{D}_T' = \begin{bmatrix} 2 & -1 & 0 & & & 0 \\ -1 & 2 & -1 & & & \\ 0 & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & 0 \\ & & & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 2 \end{bmatrix}.$$

To compute $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2}$, we need to derive the inverse matrix $\boldsymbol{\Phi}^\iota = \left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1} = \{\phi_{st}^\iota\}$. Using the results by El-Mikkawy and Karawia [2006], we have

$$\phi_{st}^\iota = \begin{cases} \frac{n}{n+1} & \text{if } s = t = 1 \text{ or } s = t = n \\ \frac{s(n-s+1)}{n+1} & \text{if } s = t < n \\ \frac{s(n-t+1)}{n+1} & \text{if } s < t \\ \frac{t(n-s+1)}{n+1} & \text{if } s > t \end{cases}$$

[10] A matrix with this structure is called *tridiagonal* matrix.

[11] Arellano [2003] does not provide the details how the upper triangular Cholesky factorization can be computed.

where $n = T_1$. Next, we need to compute the Cholesky factorization to $\mathbf{\Phi}^{\iota}$. For a $K \times K$ matrix $\mathbf{A} = \{a_{ij}\}$, its Cholesky factorization is given by

$$\mathbf{A} = \mathbf{L}\mathbf{L}'$$

where $\mathbf{L} = (\ell_{ij})$ is the lower triangular matrix. Then using ℓ_{ij}, we can write the elements of $\mathbf{A}$ as follows:

$$\begin{array}{rcl} a_{11} & = & \ell_{11}^2, \\ a_{21} & = & \ell_{21}\ell_{11}, \qquad a_{22} = \ell_{21}^2 + \ell_{22}^2, \\ a_{31} & = & \ell_{31}\ell_{11}, \qquad a_{32} = \ell_{31}\ell_{21} + \ell_{32}\ell_{22}, \qquad \ell_{33} = \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2, \\ & \vdots & \qquad\qquad \vdots \qquad\qquad\qquad\qquad \ddots \\ a_{K1} & = & \ell_{K1}\ell_{11}, \qquad a_{K2} = \ell_{K1}\ell_{21} + \ell_{K2}\ell_{22}, \qquad \cdots \qquad , a_{KK} = \ell_{K1}^2 + \cdots + \ell_{KK}^2. \end{array}$$

ℓ_{ij} can be solved sequentially as follows:

$$\begin{array}{rcl} \ell_{11} & = & \sqrt{a_{11}}, \\ \ell_{21} & = & a_{21}/\ell_{11}, \qquad \ell_{22} = \sqrt{a_{22} - \ell_{21}^2}, \\ \ell_{31} & = & a_{31}/\ell_{11}, \qquad \ell_{32} = (a_{32} - \ell_{31}\ell_{21})/\ell_{22}, \qquad \ell_{33} = \sqrt{a_{33} - \ell_{31}^2 - \ell_{32}^2}, \\ & \vdots & \qquad\qquad \vdots \qquad\qquad\qquad\qquad \ddots \\ \ell_{K1} & = & a_{K1}/\ell_{11}, \qquad \ell_{K2} = (a_{K2} - \ell_{K1}\ell_{21})/a_{K2}, \ \cdots \ \ell_{KK} = \sqrt{a_{KK} - \ell_{K1}^2 - \cdots - \ell_{K,K-1}^2}. \end{array}$$

The explicit form of $\mathbf{F}_T^{\iota}$ is obtained by letting $\mathbf{A} = \mathbf{\Phi}^{\iota}$.

Next, we consider a model with individual effects and heterogeneous time trends given by (2.10). To remove both η_i and λ_i from the model, we need to take second differences. In terms of a model in matrix, this corresponds to multiplying by $\mathbf{D}_{T-1}\mathbf{D}_T$, we have

$$\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{y}_i = \mathbf{D}_{T-1}\mathbf{D}_T\mathbf{W}_i\boldsymbol{\delta} + \mathbf{D}_{T-1}\mathbf{D}_T\mathbf{v}_i.$$

Since the transformed error is serially correlated, we consider the following GLS-type transformation matrix:

$$\mathbf{F}_T^{\tau} = \left(\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T-1}'\right)^{-1/2}\mathbf{D}_{T-1}\mathbf{D}_T,$$

where $\left(\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T-1}'\right)^{-1/2}$ is the *upper* triangular Cholesky factorization of $\left(\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T-1}'\right)^{-1}$. To compute $\mathbf{F}_T^{\tau}$, we need to derive the inverse matrix $\mathbf{\Phi}^{\tau} =$

$\left(\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T-1}'\right)^{-1} = \{\phi_{st}^{\tau}\}$ with[12]

$$\mathbf{D}_{T-1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T-1}' = \begin{bmatrix} 6 & -4 & 1 & & & 0 \\ -4 & 6 & -4 & & & \\ 1 & -4 & 6 & -4 & & \\ & & \ddots & \ddots & \ddots & 1 \\ & & & -4 & 6 & -4 \\ 0 & & & 1 & -4 & 6 \end{bmatrix}.$$

Using the results by Dow [2003], we have[13]

$$\phi_{st}^{\tau} = \begin{cases} a_{t0}s^3 + a_{t1}s^2 + a_{t2}s, & s \leq t+1 \\ b_{t0}s^3 + b_{t1}s^2 + b_{t2}s + b_{t3}, & s \geq t+1 \end{cases}$$

where

$$\begin{aligned} a_{t0} &= -(3+2t+n)d_t/c, \qquad a_{t1} = 3t(1+n)d_t/c, \qquad a_{t2} = (3+5t+n+3tn)d_t/c, \\ b_{t0} &= (5-2t+3n)e_t/c, \qquad b_{t1} = -3(1+n)(4-t+2n)e_t/c, \\ b_{t2} &= (1+5t+12n+3tn+12n^2+3n^3)e_t/c, \qquad b_{t3} = (1-t)e_t/6, \\ d_t &= (n-t+1)(n-t+2), \qquad e_t = t(t+1), \qquad c = 6(n+1)(n+2)(n+3), \quad n = T_2. \end{aligned}$$

Using these and applying the algorithm of Cholesky factorization introduced above where $\mathbf{A} = \mathbf{\Phi}^{\tau}$, after a lengthy calculation, we obtain the explicit expression of $\mathbf{F}_T^{\tau}$ as in (2.11).

2.B Derivation of asymptotic variances (2.31) and (2.32)

Let us define $\mathbf{y}_{i,-1} = (y_{i0}, ..., y_{i,T-1})$ and $Var(\mathbf{y}_{i,-1}) = \mathbf{V}_T = \sigma_\mu^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma_v^2 \mathbf{\Phi}_T$ where $\sigma_\mu^2 = \sigma_\eta^2/(1-\alpha)^2$ and $\mathbf{\Phi}_T = \{\phi_{st}\} = \alpha^{|s-t|}/(1-\alpha^2)$. Also, let $\mathbf{f}_t'$ be the tth row of $\mathbf{F}_T^{\iota}$, $\mathbf{L}_{(s:t)}$ be the sth to tth rows of $\mathbf{I}_T$ and $\mathbf{B}_{(s:t)}$ be the sth to tth rows of $\mathbf{B}_T^{\iota}$. Then, we obtain $y_{i,t-1}^{*} = \mathbf{f}_t'\mathbf{y}_{i,-1}$ and

$$\mathbf{z}_{it}^{L} = (y_{i,t-\ell}, ..., y_{i,t-1})' = \mathbf{L}_{(t-\ell+1:t)}\mathbf{y}_{i,-1}, \qquad \mathbf{z}_{it}^{B} = \left(y_{i,t-\ell}^{**}, ..., y_{i,t-1}^{**}\right)' = \mathbf{B}_{(t-\ell+1:t)}\mathbf{y}_{i,-1}.$$

Using this, we have

$$\begin{aligned} E\left(\mathbf{z}_{it}^{L(\ell)} y_{i,t-1}^{*}\right) &= \mathbf{L}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{f}_t, \qquad E\left(\mathbf{z}_{it}^{B} y_{i,t-1}^{*}\right) = \mathbf{B}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{f}_t, \\ E\left(\mathbf{z}_{it}^{L(\ell)} \mathbf{z}_{it}^{L\prime}\right) &= \mathbf{L}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{L}_{(t-\ell+1:t)}', \qquad E\left(\mathbf{z}_{it}^{B(\ell)} \mathbf{z}_{it}^{B\prime}\right) = \mathbf{B}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{B}_{(t-\ell+1:t)}'. \end{aligned}$$

[12] A matrix with this structure is called *pentadiagonal* matrix.

[13] See also Chen [2013] for an alternative expression.

Hence, (2.31) and (2.32) can be written as

$$Avar\left(\widehat{\alpha}_{GMM}^{L}\right)=\sigma_v^2\left[\sum_{t=1}^{T-1}\mathbf{f}_t'\mathbf{V}_T\mathbf{L}_{(t-\ell+1:t)}'\left[\mathbf{L}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{L}_{(t-\ell+1:t)}'\right]^{-1}\mathbf{L}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{f}_t\right]^{-1}, \tag{2.55}$$

$$Avar\left(\widehat{\alpha}_{GMM}^{B}\right)=\sigma_v^2\left[\sum_{t=2}^{T-1}\mathbf{f}_t'\mathbf{V}_T\mathbf{B}_{(t-\ell+1:t)}'\left[\mathbf{B}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{B}_{(t-\ell+1:t)}'\right]^{-1}\mathbf{B}_{(t-\ell+1:t)}\mathbf{V}_T\mathbf{f}_t\right]^{-1} \tag{2.56}$$

Figure 1 is described based on (2.55) and (2.56).

2.C Proof of Theorem 1

We derive the explicit formula for asymptotic variances of IV and GMM estimators. First, condier $Avar\left(\widehat{\alpha}_{GMM}^{L(t)}\right)$. Note that under Assumption 1, $y_{i,t-1}$ can be written as

$$y_{i,t-1}=\frac{\eta_i}{1-\alpha}+\xi_{i,t-1} \tag{2.57}$$

where $\xi_{i,t-1}=\sum_{j=0}^{\infty}\alpha^j v_{i,t-1-j}$. Also, from (A43) of Alvarez and Arellano [2003], we have

$$y_{i,t-1}^{*}=\psi_t\xi_{i,t-1}-c_t\left(\frac{\phi_{T-t}v_{it}+\cdots+\phi_1 v_{i,T-1}}{T-t}\right). \tag{2.58}$$

Using (2.57) and (2.58), and under Assumption 1, we have

$$\begin{aligned} E(\mathbf{z}_{it}^{L}y_{i,t-1}^{*}) &= \psi_t\left(\frac{\sigma_v^2}{1-\alpha^2}\right)(\alpha^{\ell-1},\cdots,1)', && (2.59)\\ \left[E\left(\mathbf{z}_{it}^{L}\mathbf{z}_{it}^{L\prime}\right)\right]^{-1} &= \mathbf{V}_\ell^{-1}=\frac{1}{\sigma_v^2}\left[(\sqrt{\lambda}\boldsymbol{\iota}_\ell)(\sqrt{\lambda}\boldsymbol{\iota}_\ell)'+\boldsymbol{\Phi}_\ell\right]^{-1} && (2.60)\end{aligned}$$

where $\lambda=\sigma_\mu^2/\sigma_v^2$, $\boldsymbol{\iota}_\ell$ is an ℓ dimensional column vector of ones, and $\mathbf{V}_\ell$ is the upper-left $\ell\times\ell$ matrix of $\mathbf{V}_T$. The explicit expression of (2.60) is obtained as follows. By using the Sherman-Morrison-Woodbury inversion formula

$$[\mathbf{A}+\mathbf{b}\mathbf{b}']^{-1}=\mathbf{A}^{-1}-\left[\frac{1}{1+\mathbf{b}'\mathbf{A}^{-1}\mathbf{b}}\right]\mathbf{A}^{-1}\mathbf{b}\mathbf{b}'\mathbf{A}^{-1}$$

and the decomposition of $\mathbf{V}_\ell^{-1}$ [14]

$$\mathbf{V}_\ell^{-1}=\mathbf{C}'\mathbf{C}$$

[14] See Amemiya (1985, p.164), Hamilton (1994, p.120) and Greene (2001, p.822).

where

$$
\mathbf{C} = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 \end{bmatrix}
$$

we obtain

$$
\left[E(\mathbf{z}_{it}^{L}\mathbf{z}_{it}^{L\prime})\right]^{-1} = \sigma_v^{-2}\left[\mathbf{C}'\mathbf{C} - \frac{\lambda}{1+\lambda\boldsymbol{\iota}_\ell'\mathbf{C}'\mathbf{C}\boldsymbol{\iota}_\ell}\mathbf{C}'\mathbf{C}\boldsymbol{\iota}_\ell\boldsymbol{\iota}_\ell'\mathbf{C}'\mathbf{C}\right]. \tag{2.61}
$$

By substituting (2.59) and (2.61) into (2.31), we obtain (2.35). (2.36) and (2.37) are obtained from (2.35). Next, we consider $Var(\widehat{\alpha}_{GMM}^{B(1)})$. First, note that $y_{i,t-1}^{**}$ can be written as

$$
y_{i,t-1}^{**} = c_{T-t+1}^{\iota}\left[\xi_{i,t-1} - \frac{\xi_{i,t-2}+\cdots+\xi_{i0}}{t-1}\right]. \tag{2.62}
$$

Then, using (2.58) and (2.62), we obtain

$$
E\left(y_{i,t-1}^{**}y_{i,t-1}^{*}\right) = \left(\frac{\sigma_v^2}{1-\alpha^2}\right)\psi_t c_{T-t+1}^{\iota}\left(1-\frac{\phi_{t-1}}{t-1}\right). \tag{2.63}
$$

Also, from (2.57), we obtain

$$
\begin{aligned}
E[(y_{i,t-1}^{**})^2] &= c_{T-t+1}^{\iota 2}E\left[\xi_{i,t-1} - \frac{1}{t-1}(\xi_{i,0}+\cdots+w_{i,t-2})\right]^2 \\
&= c_{T-t+1}^{\iota 2}\left[\frac{\sigma_v^2}{1-\alpha^2}\left(1-\frac{2\alpha\phi_{t-1}}{t-1}\right) + \frac{1}{(t-1)^2}E(\xi_{i0}+\cdots+\xi_{i,t-1})^2\right] \qquad (2.64)
\end{aligned}
$$

Using the result of (A8) in Alvarez and Arellano [2003], we have

$$
E(\xi_{i0}+\cdots+\xi_{i,t-1})^2 = \frac{\sigma_v^2}{1-\alpha^2}\left[\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right]. \tag{2.65}
$$

By substituting this into (2.64), we get

$$
E[(y_{i,t-1}^{**})^2] = \left(\frac{\sigma_v^2}{1-\alpha^2}\right)c_{T-t+1}^{\iota 2}A_t \tag{2.66}
$$

where A_t is defined in (2.43). (2.39) is obtained by substituting (2.63) and (2.66) into (2.31).

Next, we derive the asymptotic variances of $\widehat{\alpha}_{IV}^{L}$ and $\widehat{\alpha}_{IV}^{B}$. Using (2.57) and (2.58) and the fact that v_{it}^{*} is serially uncorrelated, we obtain

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T-1}E\left(y_{i,t-1}y_{i,t-1}^{*}\right) &= \left(\frac{\sigma_v^2}{1-\alpha^2}\right)\sum_{t=1}^{T-1}\psi_t \\
Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}y_{i,t-1}v_{it}^{*}\right) &= Var\left(\sum_{t=1}^{T-1}y_{i,t-1}v_{it}^{*}\right)=\sigma_v^2\sum_{t=1}^{T-1}E\left(y_{i,t-1}^2\right)=\sigma_v^2\sigma_\mu^2+\frac{\sigma_v^4}{1-\alpha^2}
\end{aligned}
$$

Using these, we obtain the asymptotic variance of $\widehat{\alpha}_{IV}^{L}$ as in (2.38). The asymptotic variance of $\widehat{\alpha}_{IV}^{B}$ can be derived similarly. Using (2.63) and (2.66), and under Assumption 1, we obtain

$$
\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T-1}E\left(y_{i,t-1}^{**}y_{i,t-1}^{*}\right)=\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\sum_{t=2}^{T-1}\psi_t c_{T-t+1}^{\iota}\left(1-\frac{\phi_{t-1}}{t-1}\right),
$$

$$
\begin{aligned}
&Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=2}^{T-1}y_{i,t-1}^{**}v_{it}^{*}\right)\\
&=Var\left(\sum_{t=2}^{T-1}y_{i,t-1}^{**}v_{it}^{*}\right)\\
&=\sigma_v^2\sum_{t=2}^{T-1}E\left[\left(y_{i,t-1}^{**}\right)^2\right]\\
&=\left(\frac{\sigma_v^4}{1-\alpha^2}\right)\sum_{t=2}^{T-1}c_{T-t+1}^{\iota 2}A_t.
\end{aligned}
$$

From these, the asymptotic variance of $\widehat{\alpha}_{IV}^{L}$ is obtained as (2.40).

2.D Proof of Lemma 1

First, we decompose $T_d\times T$ matrix $\mathbf{F}_T$ as

$$
\mathbf{F}_T=\begin{bmatrix}\mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{F}_{13}\\ \mathbf{0}_{T_{2d}\times d} & \mathbf{F}_{22} & \mathbf{F}_{23}\end{bmatrix}=\{f_{st}\},(s=1,...,T_d;t=1,...,T)
$$

where $\mathbf{F}_{11}$ is $d\times d$, $\mathbf{F}_{12}$ is $d\times T_{2d}$, $\mathbf{F}_{13}$ is $d\times d$, $\mathbf{F}_{22}$ is $T_{2d}\times T_{2d}$, and $\mathbf{F}_{23}$ is $T_{2d}\times d$. Note that $\mathbf{B}_T$ and $\mathbf{F}_T$ have the following relationship

$$
\mathbf{B}_T=\mathcal{I}_{T_d}\mathbf{F}_T\mathcal{I}_T \qquad (2.67)
$$

where

$$\mathcal{I}_T = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}$$

and $\mathcal{I}_T^2 = \mathcal{I}_T'\mathcal{I}_T = \mathbf{I}_T$. Furthermore, using (2.45), $\mathbf{W}_i$ can be written as

$$\mathbf{W}_i = \boldsymbol{\iota}_T\boldsymbol{\mu}_i' + \boldsymbol{\tau}_T\boldsymbol{\kappa}_i' + \boldsymbol{\Xi}_i = \mathbf{C}_T\boldsymbol{\Psi}_i + \boldsymbol{\Xi}_i$$

where $\boldsymbol{\Xi}_i = (\boldsymbol{\xi}_{i1}', ..., \boldsymbol{\xi}_{iT}')'$ and $\boldsymbol{\Psi}_i = (\boldsymbol{\mu}_i, \boldsymbol{\kappa}_i)'$.

(a): Note the following decomposition:

$$\begin{aligned} \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{Q}_T\right)\mathbf{W}_i. \end{aligned} \tag{2.68}$$

Using $\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{C}_T = \mathbf{B}_{T_d}\mathbf{K}_T\mathbf{C}_T = \mathbf{Q}_T\mathbf{C}_T = \mathbf{0}$ and (2.16), the second term of (2.68) can be further decomposed as

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{Q}_T\right)\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{I}_T\right)\boldsymbol{\Xi}_i + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{R}_T\boldsymbol{\Xi}_i. \tag{2.69}$$

To consider the first term of right-hand side of (2.69), we derive the explicit form of $\mathbf{A}_T = \mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{I}_T$. Using (2.67), $\mathbf{F}_{T_d} = \mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'$ and

$$\begin{aligned} \mathbf{K}_T'\mathcal{I}_{T_d}\mathbf{L}_T &= \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & \mathcal{I}_d \\ \mathbf{0}_{T_{2d}\times d} & \mathcal{I}_{T_{2d}} & \mathbf{0}_{T_{2d}\times d} \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & \mathbf{0}_{d\times d} \end{bmatrix}, \quad \mathbf{L}_{T_d}'\mathcal{I}_{T_{2d}}\mathbf{L}_{T_d} = \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} \\ \mathbf{0}_{T_{2d}\times d} & \mathcal{I}_{T_{2d}} \end{bmatrix}, \\ \mathbf{L}_T'\mathbf{L}_T &= \begin{bmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & \mathbf{0}_{d\times d} \\ \mathbf{0}_{T_{2d}\times d} & \mathbf{I}_{T_{2d}} & \mathbf{0}_{T_{2d}\times d} \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & \mathbf{I}_d \end{bmatrix}, \end{aligned}$$

we have

$$
\begin{aligned}
\mathbf{A}_T &= \mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'\mathbf{L}_T - \mathbf{I}_T = \mathbf{K}_T'\mathcal{I}_{T_d}\mathbf{F}_{T_d}'\mathcal{I}_{T_{2d}}\mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'\mathbf{L}_T - \mathbf{I}_T \\
&= \left(\mathbf{K}_T'\mathcal{I}_{T_d}\mathbf{L}_T\right)\mathbf{F}_T'\left(\mathbf{L}_{T_d}'\mathcal{I}_{T_{2d}}\mathbf{L}_{T_d}\right)\mathbf{F}_T\left(\mathbf{L}_T'\mathbf{L}_T\right) - \mathbf{I}_T \\
&= \begin{bmatrix} -\mathbf{I}_d & \mathcal{I}_d\mathbf{F}_{23}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} & \mathcal{I}_d\mathbf{F}_{23}'\mathcal{I}_{T_{2d}}\mathbf{F}_{23} \\ \mathbf{0}_{T_{2d}\times d} & \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} - \mathbf{I}_{T_{2d}} & \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{23} \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & -\mathbf{I}_d \end{bmatrix} \\
&= \{\mathbf{A}_{ij}\}, (i,j = 1,2,3).
\end{aligned} \tag{2.70}
$$

Next, we derive the form of each $\mathbf{A}_{ij}$. Using

$$
\mathcal{I}_{T_{2d}}\mathbf{F}_{22} = \begin{bmatrix} 0 & \cdots & 0 & f_{T_dT_d} \\ \vdots & \iddots & \iddots & \vdots \\ 0 & f_{d+2,d+2} & \cdots & f_{d+2,T_d} \\ f_{d+1,d+1} & f_{d+1,d+2} & \cdots & f_{d+1,T_d} \end{bmatrix}, \quad \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'
$$

$$
= \begin{bmatrix} f_{d+1,T_d} & f_{d+2,T_d} & \cdots & f_{T_dT_d} \\ \vdots & \vdots & \iddots & 0 \\ f_{d+1,d+2} & f_{d+2,d+2} & & \vdots \\ f_{d+1,d+1} & 0 & \cdots & 0 \end{bmatrix}
$$

$$
\mathcal{I}_{T_{2d}}\mathbf{F}_{23} = \begin{bmatrix} f_{T_d,T_d+1} & f_{T_d,T} \\ \vdots & \vdots \\ f_{d+2,T_d+1} & f_{d+2,T} \\ f_{d+1,T_d+1} & f_{d+1,T} \end{bmatrix}, \quad \mathcal{I}_d\mathbf{F}_{23}' = \begin{bmatrix} f_{d+1,T} & f_{d+2,T} & \cdots & f_{T_d,T} \\ f_{d+1,T_d+1} & f_{d+2,T_d+1} & \cdots & f_{T_d,T_d+1} \end{bmatrix}
$$

we have

$$\underset{(d\times T_{2d})}{\mathbf{A}_{12}} = \mathcal{I}_d\mathbf{F}'_{23}\mathcal{I}_{T_{2d}}\mathbf{F}_{22}$$

$$= \begin{bmatrix} f_{T_d,T}f_{d+1,d+1} & f_{T_d-1,T}f_{d+2,d+2}+f_{T_d,T}f_{d+1,d+2} & \cdots & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T}f_{T_d-\ell+1,T_d} \\ f_{T_d,T_d+1}f_{d+1,d+1} & f_{T_d-1,T_d+1}f_{d+2,d+2}+f_{T_d,T_d+1}f_{d+1,d+2} & \cdots & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d+1}f_{T_d-\ell+1,T_d} \end{bmatrix}$$

$$= \left\{a_{12}^{jk}\right\}$$

$$= \sum_{\ell=1}^{k} f_{T_d-\ell+1,T-j+1}f_{d+\ell,d+k}$$

$$= \sum_{\ell=1}^{k-1} f_{T_d-\ell+1,T-j+1}f_{d+\ell,d+k} + f_{T_d-k+1,T-j+1}f_{d+k,d+k}$$

$$= \begin{cases} \sum_{\ell=1}^{k-1}\frac{c_{T-\ell}^{\iota}c_{\ell+1}^{\iota}}{\ell(T-\ell-1)} - \frac{c_{T-k}^{\iota}c_{k+1}^{\iota}}{k} < \sum_{\ell=1}^{T_3}\frac{c_{T-\ell}^{\iota}c_{\ell+1}^{\iota}}{\ell(T-\ell-1)} & \text{for FE model} \\ \qquad (j=1;k=1,...,T_{2d}) & \\ \sum_{\ell=1}^{k-1}\frac{4c_{\ell+2}^{\tau}c_{T-\ell-1}^{\tau}(3j-\ell-3)(2T-3k+\ell-3)}{\ell(\ell+1)(T-\ell-2)(T-\ell-3)} - \frac{2c_{k+2}^{\tau}c_{T-k-1}^{\tau}(3j-k-3)}{k(k+1)} & \text{for trend model} \\ \qquad (j=1,2;k=1,...,T_{2d}) & \end{cases}$$

$$= O\left(\frac{\log T}{T}\right),$$

$$\begin{aligned} \underset{(d\times d)}{\mathbf{A}_{13}} &= \mathcal{I}_d\mathbf{F}'_{23}\mathcal{I}_{T_{2d}}\mathbf{F}_{23} = \begin{bmatrix} \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T}f_{T_d-\ell+1,T_d+1} & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T}f_{T_d-\ell+1,T} \\ \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d+1}f_{T_d-\ell+1,T_d+1} & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d+1}f_{T_d-\ell+1,T} \end{bmatrix} \\ &= \left\{a_{13}^{jk}\right\} = \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T-j+1}f_{T_d-\ell+1,T_d+k} \\ &= \begin{cases} \sum_{\ell=1}^{T-2}\frac{c_{\ell+1}^{\iota}c_{T-\ell}^{\iota}}{(T-\ell-1)\ell}, & (j=1,k=1) \quad \text{for FE model} \\ \sum_{\ell=1}^{T_{2d}}\frac{4c_{\ell+2}^{\tau}c_{T-\ell-1}^{\tau}(3k+\ell-6)(T-3j-\ell)}{\ell(\ell+1)(T-\ell-2)(T-\ell-3)}, & (j=1,2;k=1,2) \quad \text{for trend model} \end{cases} \\ &= O\left(\frac{\log T}{T}\right), \end{aligned}$$

$$\underset{(T_{2d}\times T_{2d})}{\mathbf{A}_{22}} = \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} - \mathbf{I}_{T_{2d}}$$

$$= \begin{bmatrix} a_{22}^{11} & a_{22}^{12} & \cdots & a_{22}^{1,T_{2d}} \\ 0 & a_{22}^{22} & \cdots & a_{22}^{2,T_{2d}} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{22}^{T_{2d},T_{2d}} \end{bmatrix}$$

$$= \left\{ a_{22}^{jk} \right\}$$

$$= \begin{cases} 0 & \text{if } j > k \\ f_{T_d-j+1,T_d-j+1} f_{d+j,d+j} - 1 & \text{if } j = k \\ \sum_{\ell=1}^{k} f_{T_d-\ell+1,T_d-j+1} f_{d+\ell,d+k} & \text{if } j < k \end{cases}$$

$$= \begin{cases} \begin{cases} 0 & \text{if } j > k \\ c_{T-j}^{\iota} c_{j+1}^{\iota} - 1 = O\left(\frac{1}{j}\right) + O\left(\frac{1}{(T-j)}\right) + O\left(\frac{1}{j^2} + \frac{1}{(T-j)^2}\right) & \text{if } j = k \\ \sum_{\ell=1}^{k-j+1} \frac{c_{\ell+1}^{\iota} c_{T-\ell}^{\iota}}{(T-\ell-1)\ell} = O\left(\frac{\log T}{T}\right) & \text{if } j < k \\ (j,k = 1,...,T_{2d}) & \end{cases} & \text{for FE model} \\ \begin{cases} 0 & \text{if } j > k \\ c_{T-j-1}^{\tau} c_{j+2}^{\tau} - 1 = O\left(\frac{1}{j}\right) + O\left(\frac{1}{(T-j)}\right) & \text{if } j = k \\ \sum_{\ell=1,\ell>j,\ell<k}^{k} \frac{4c_{\ell+2}^{\tau} c_{T-\ell-1}^{\tau}(3j-\ell+3)(2T-3k+\ell-3)}{\ell(\ell+1)(T-\ell-2)(T-\ell-3)} & \\ -\frac{2c_{j+2}^{\tau} c_{T-j-1}^{\tau}(2T+j-3k-3)}{(T-j-2)(T-j-3)} - \frac{2c_{k+2}^{\tau} c_{T-k-1}^{\tau}(3j-k+3)}{k(k+1)} = O\left(\frac{\log T}{T}\right) & \text{if } j < k \\ (j,k = 1,...,T_{2d}) & \end{cases} & \text{for trend model} \end{cases},$$

$$
\begin{aligned}
&\underset{(T_{2d}\times d)}{\mathbf{A}_{23}}\\
&= \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{23}\\
&= \begin{bmatrix} \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d} f_{T_d-\ell+1,T_d+1} & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d} f_{T_d-\ell+1,T} \\ \vdots & \vdots \\ \sum_{\ell=1}^{2} f_{d+\ell,d+2} f_{T_d-\ell+1,T_d+1} & \sum_{\ell=1}^{2} f_{d+\ell,d+2} f_{T_d-\ell+1,T} \\ \sum_{\ell=1}^{1} f_{d+\ell,d+1} f_{T_d-\ell+1,T_d+1} & \sum_{\ell=1}^{1} f_{d+\ell,d+1} f_{T_d-\ell+1,T} \end{bmatrix}\\
&= \left\{ a_{23}^{jk} \right\}\\
&= \sum_{\ell=1}^{T_{2d}-j+1} f_{d+\ell,T_d-j+1} f_{T_d-\ell+1,T_d+k}\\
&= \sum_{\ell=1}^{T_{2d}-j} f_{d+\ell,T_d-j+1} f_{T_d-\ell+1,T_d+k} + f_{T_d-j+1,T_d-j+1} f_{d+j,T_d+k}
\end{aligned}
$$

$$
= \begin{cases} \sum_{\ell=1}^{T_{2d}-j} \frac{c_{\ell+1}^{\iota} c_{T-\ell}^{\iota}}{(T-\ell-1)\ell} - \frac{c_{T_d-j+1}^{\iota} c_{d+j}^{\iota}}{(T_d-j)\ell}, \qquad (j=1,...,T_{2d}; k=1) & \text{for FE model} \\ \sum_{\ell=1}^{T_{2d}-j} \frac{4c_{\ell+2}^{\tau} c_{T-\ell-1}^{\tau}(3k+\ell-6)(T-3j-\ell-6)}{\ell(\ell+1)(T-\ell-2)(T-\ell-3)} + \frac{2c_{j+2}^{\tau} c_{T-j-1}^{\tau}(T-j+3k-9)}{(T-j-2)(T-j-3)} & \text{for trend model} \\ \qquad (j=1,...,T_{2d}; k=1,2) & \end{cases}
$$

$$
= O\left(\frac{\log T}{T}\right).
$$

We now assess the first term of (2.69). Using $\mathbf{\Xi}_i = (\mathbf{\Xi}_{1i}, \mathbf{\Xi}_{2i}, \mathbf{\Xi}_{3i})'$ where $\mathbf{\Xi}_{1i}$ is $d \times k$, $\mathbf{\Xi}_{2i}$ is $T_{2d} \times k$, and $\mathbf{\Xi}_{3i}$ is $d \times k$, we have

$$
\mathbf{S}_i = \mathbf{\Xi}_i' \mathbf{A}_T \mathbf{\Xi}_i = \mathbf{S}_{1i} + \mathbf{S}_{2i} + \mathbf{S}_{3i} + \mathbf{S}_{4i} + \mathbf{S}_{5i} + \mathbf{S}_{6i}
$$

where

$$
\begin{aligned}
\mathbf{S}_{1i} &= -\mathbf{\Xi}_{1i}'\mathbf{\Xi}_{1i}, \quad \mathbf{S}_{2i} = \mathbf{\Xi}_{1i}'\mathbf{A}_{12}\mathbf{\Xi}_{2i}, \quad \mathbf{S}_{3i} = \mathbf{\Xi}_{1i}'\mathbf{A}_{13}\mathbf{\Xi}_{3i}, \quad \mathbf{S}_{4i} = \mathbf{\Xi}_{2i}'\mathbf{A}_{22}\mathbf{\Xi}_{2i},\\
\mathbf{S}_{5i} &= \mathbf{\Xi}_{2i}'\mathbf{A}_{23}\mathbf{\Xi}_{3i}, \quad \mathbf{S}_{6i} = -\mathbf{\Xi}_{3i}'\mathbf{\Xi}_{3i}.
\end{aligned}
$$

We now evaluate each term. We consider the FE model and trend model separately below.

2.D.1 FE model

From the definition of $\mathbf{\Xi}_{1i}$ and Assumption 3, we have

$$
E(\mathbf{S}_{1i}) = -E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{i1}'\right) = -\mathbf{\Gamma}_{i0} = O(1).
$$

Using $a_{12}^{1,t-1} = O\left(\frac{\log T}{T}\right)$ for all t, $a_{13}^{11} = O\left(\frac{\log T}{T}\right)$ and Assumption 3, we have

$$E(\mathbf{S}_{2i}) = \sum_{t=2}^{T_d} a_{12}^{1,t-1} E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it}'\right) = O\left(\frac{\log T}{T}\right) \sum_{t=2}^{T_1} \boldsymbol{\Gamma}_{i,t-1} = O\left(\frac{\log T}{T}\right),$$

$$E(\mathbf{S}_{3i}) = a_{13}^{1,1} E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{iT}'\right) = O\left(\frac{\log T}{T}\right) \boldsymbol{\Gamma}_{i,T-1} = O\left(\frac{\log T}{T}\right).$$

Similarly, using $a_{22}^{t-1,t-1} = O(1/(t+1)) + O(1/(T-t))$ and $a_{22}^{s-1,t-1} = O(\log T/T)$ for all $s \neq t$, we have

$$\begin{aligned}
E(\mathbf{S}_{4i}) &= \sum_{s=2}^{T_d}\sum_{t=2}^{T_d} a_{22}^{s-1,t-1} E\left(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it}'\right) = \sum_{t=2}^{T_d} a_{22}^{t-1,t-1} E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'\right) + \sum_{s=2}^{T_d-1}\sum_{t=s+1}^{T_d} a_{22}^{s-1,t-1} E\left(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it}'\right) \\
&= \sum_{t=2}^{T_d}\left[O\left(\frac{1}{t+1}\right) + O\left(\frac{1}{T-t}\right)\right] E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'\right) + O\left(\frac{\log T}{T}\right) \sum_{s=2}^{T_d-1}\sum_{t=s+1}^{T_d} \boldsymbol{\Gamma}_{t-s,i} \\
&= O\left(\log T\right).
\end{aligned}$$

Finally, using $a_{23}^{t-1,1} = O(\log T/T)$ for all t, and the definition of $\boldsymbol{\Xi}_{3i}$, we have

$$E(\mathbf{S}_{5i}) = \sum_{t=2}^{T_d} a_{23}^{t-1,1} E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{iT}'\right) = \sum_{t=2}^{T_d} O\left(\frac{\log T}{T}\right) \boldsymbol{\Gamma}_{i,T-t} = O\left(\frac{\log T}{T}\right),$$

$$E(\mathbf{S}_{6i}) = -E\left(\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\right) = -\boldsymbol{\Gamma}_{i0} = O\left(1\right).$$

Thus, for the FE model, we have $\mathbf{S}_i = \sum_{l=1}^{6} \mathbf{S}_{li} = O(\log T)$ for all i and obtain

$$\frac{1}{NT}\sum_{i=1}^{N} \boldsymbol{\Xi}_i' \mathbf{A}_T \boldsymbol{\Xi}_i = O_p\left(\frac{\log T}{T}\right). \tag{2.71}$$

2.D.2 Trend model

From the definition of $\boldsymbol{\Xi}_{1i}$, we have

$$E(\mathbf{S}_{1i}) = -E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{i1}'\right) - E\left(\boldsymbol{\xi}_{i2}\boldsymbol{\xi}_{i2}'\right) = -2\boldsymbol{\Gamma}_{i0} = O\left(1\right).$$

Since $a_{12}^{1,t-1}$, $a_{12}^{2,t-1}$, a_{13}^{1k} and a_{13}^{2k} are $O\left(\frac{\log T}{T}\right)$ for all t and k, using Assumption 3, we have

$$\begin{aligned}
E(\mathbf{S}_{2i}) &= \sum_{t=2}^{T_d}\left[a_{12}^{1,t-1}E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it}'\right)+a_{12}^{2,t-1}E\left(\boldsymbol{\xi}_{i2}\boldsymbol{\xi}_{it}'\right)\right] \\
&= O\left(\frac{\log T}{T}\right)\sum_{t=2}^{T_1}\boldsymbol{\Gamma}_{i,t-1}+O\left(\frac{\log T}{T}\right)\sum_{t=2}^{T_1}\boldsymbol{\Gamma}_{i,t-2}=O\left(\frac{\log T}{T}\right),
\end{aligned}$$

$$\begin{aligned}
E(\mathbf{S}_{3i}) &= \sum_{t=T-1}^{T}\left[\left(a_{13}^{1,1}+a_{13}^{1,2}\right)E\left(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it}'\right)+\left(a_{13}^{2,1}+a_{13}^{2,2}\right)E\left(\boldsymbol{\xi}_{i2}\boldsymbol{\xi}_{it}'\right)\right] \\
&= O\left(\frac{\log T}{T}\right)\left(\boldsymbol{\Gamma}_{i,T-1}+2\boldsymbol{\Gamma}_{i,T-2}+\boldsymbol{\Gamma}_{i,T-3}\right)=O\left(\frac{\log T}{T}\right).
\end{aligned}$$

Similarly, using $a_{22}^{t-1,t-1}=O(1/(t+1))+O(1/(T-t))$ and $a_{22}^{s-1,t-1}=O(\log T/T)$ for all $s\neq t$, we have

$$\begin{aligned}
E(\mathbf{S}_{4i}) &= \sum_{s=3}^{T_d}\sum_{t=3}^{T_d}a_{22}^{s-1,t-1}E\left(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it}'\right) \\
&= \sum_{t=3}^{T_d}a_{22}^{t-1,t-1}E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'\right)+\sum_{s=3}^{T_d-1}\sum_{t=s+1}^{T_d}a_{22}^{s-1,t-1}E\left(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it}'\right) \\
&= \sum_{t=3}^{T_d}\left[O\left(\frac{1}{t+1}\right)+O\left(\frac{1}{T-t}\right)\right]E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}'\right)+O\left(\frac{\log T}{T}\right)\sum_{s=3}^{T_d-1}\sum_{t=s+1}^{T_d}\boldsymbol{\Gamma}_{t-s,i} \\
&= O\left(\log T\right)
\end{aligned}$$

Finally, using $a_{23}^{t-1,1}=O(\log T/T)$ and $a_{23}^{t-1,2}=O(\log T/T)$ for all t, and the definition of $\boldsymbol{\Xi}_{3i}$, we have

$$\begin{aligned}
E(\mathbf{S}_{5i}) &= \sum_{t=2}^{T_d}a_{23}^{t-1,1}E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{iT_1}'\right)+\sum_{t=2}^{T_d}a_{23}^{t-1,2}E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{iT}'\right) \\
&= \sum_{t=2}^{T_d}O\left(\frac{\log T}{T}\right)\left(\boldsymbol{\Gamma}_{i,T-t-1}+\boldsymbol{\Gamma}_{i,T-t}\right)=O\left(\frac{\log T}{T}\right),
\end{aligned}$$

$$E(\mathbf{S}_{6i}) = -E\left(\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\right)-E\left(\boldsymbol{\xi}_{iT_1}\boldsymbol{\xi}_{iT_1}'\right)=-2\boldsymbol{\Gamma}_{i0}=O\left(1\right).$$

Thus, for the trend model, we have $\mathbf{S}_i=\sum_{l=1}^{6}\mathbf{S}_{li}=O(\log T)$ for all i and obtain

$$\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{A}_T\boldsymbol{\Xi}_i=O_p\left(\frac{\log T}{T}\right). \tag{2.72}$$

Next, we consider the second term of (2.69). Let us define $\mathbf{H}_i = \mathbf{\Xi}_i'\mathbf{R}_T\mathbf{\Xi}_i$. Then, for the FE model, using (2.17), we have

$$\begin{aligned} E(\mathbf{H}_i) &= \frac{1}{T}E(\mathbf{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{\Xi}_i) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}'\right) \\ &= \mathbf{\Gamma}_{i0} + \frac{1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}(\mathbf{\Gamma}_{i,t-s} + \mathbf{\Gamma}_{i,t-s}') = O(1) \end{aligned} \tag{2.73}$$

where we used $\|\sum_{t=s+1}^{T}\mathbf{\Gamma}_{i,t-s}\| = \|\sum_{l=1}^{T-s}\mathbf{\Gamma}_{i,l}\| \leq \sum_{l=1}^{T-s}\|\mathbf{\Gamma}_{i,l}\| < \sum_{l=1}^{\infty}\|\mathbf{\Gamma}_{i,l}\| < \infty$. For the trend model, using (2.17), we have

$$\begin{aligned} \mathbf{H}_i &= \frac{2(2T+1)\mathbf{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{\Xi}_i}{T(T-1)} + \frac{12\mathbf{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\tau}_T'\mathbf{\Xi}_i}{T(T-1)(T+1)} - \frac{6(\mathbf{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\tau}_T'\mathbf{\Xi}_i + \mathbf{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\iota}_T'\mathbf{\Xi}_i)}{T(T-1)} \\ &= \mathbf{H}_{1i} + \mathbf{H}_{2i} + \mathbf{H}_{3i}. \end{aligned}$$

Using Assumption 3 and (2.73), we have

$$\begin{aligned} E(\mathbf{H}_{1i}) &= \frac{2(2T+1)}{T(T-1)}E\left[\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}\boldsymbol{\xi}_{is}'\right)\right] = O(1), \\ E(\mathbf{H}_{2i}) &= \frac{12}{T(T-1)(T+1)}E\left(\left(\sum_{t=1}^{T}t\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}s\boldsymbol{\xi}_{is}'\right)\right) \\ &= \frac{12T}{(T-1)(T+1)}\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{t}{T}\right)\left(\frac{s}{T}\right)E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}') = O(1), \\ E(\mathbf{H}_{3i}) &= \frac{6}{T(T-1)}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}sE(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}') + \sum_{t=1}^{T}\sum_{s=1}^{T}tE(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}')\right) \\ &= \frac{6}{(T-1)}\left(\sum_{s=1}^{T}\sum_{t=1}^{T}\left(\frac{s}{T}\right)E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}') + \sum_{s=1}^{T}\sum_{s=1}^{T}\left(\frac{t}{T}\right)E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}')\right) = O(1) \end{aligned}$$

where we used $0 < t/T \leq 1$ and $0 < s/T \leq 1$ for all s and t. Hence, for each i, we have $E(\mathbf{H}_i) = O(1)$ for both FE and trend models, and we obtain

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Xi}_i'\mathbf{R}_T\mathbf{\Xi}_i = O_p\left(\frac{1}{T}\right). \tag{2.74}$$

By combining (2.71), (2.72), and (2.74), we obtain

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i + O_p\left(\frac{\log T}{T}\right).$$

(b): Using $\mathbf{Q}_T\left(\boldsymbol{\iota}_T, \boldsymbol{\tau}_T\right)=\mathbf{0}$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{Q}_T\mathbf{v}_i=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{v}_i-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{R}_T\mathbf{v}_i.$$

The first term converges in distribution to $\mathcal{N}(\mathbf{0},\boldsymbol{\Omega})$ by Assumption 4. To assess the second term, let us define $\mathbf{h}_i=-\boldsymbol{\Xi}_i'\mathbf{R}_T\mathbf{v}_i$. Then, for the case of FE model, using Assumption 3, we have

$$\begin{aligned}
E(\mathbf{h}_i) &= \frac{-1}{T}E\left[\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}v_{is}\right)\right]=\frac{-1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}E\left(\boldsymbol{\xi}_{it}v_{is}\right)+\frac{-1}{T}\sum_{t=1}^{T}\sum_{s=t}^{T}E\left(\boldsymbol{\xi}_{it}v_{is}\right)\\
&= \frac{-1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\boldsymbol{\phi}_{i,t-s}=O\left(1\right).
\end{aligned} \tag{2.75}$$

where we used $\|\sum_{t=s+1}^{T}\boldsymbol{\phi}_{i,t-s}\|=\|\sum_{l=1}^{T-s}\boldsymbol{\phi}_{i,l}\|\leq\sum_{l=1}^{T-s}\|\boldsymbol{\phi}_{i,l}\|<\sum_{l=1}^{\infty}\|\boldsymbol{\phi}_{i,l}\|<\infty$. For trend model, we have

$$\begin{aligned}
\mathbf{h}_i &= -\frac{2\left(2T+1\right)\boldsymbol{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{v}_i}{T\left(T-1\right)}-\frac{12\boldsymbol{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\tau}_T'\mathbf{v}_i}{T\left(T-1\right)\left(T+1\right)}+\frac{6\left(\boldsymbol{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\tau}_T'\mathbf{v}_i+\boldsymbol{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\iota}_T'\mathbf{v}_i\right)}{T\left(T-1\right)}\\
&= -\mathbf{h}_{1i}-\mathbf{h}_{2i}+\mathbf{h}_{3i}.
\end{aligned}$$

Using Assumption 3 and (2.75), we have

$$\begin{aligned}
E(\mathbf{h}_{1i}) &= \frac{2\left(2T+1\right)}{T\left(T-1\right)}E\left[\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}v_{is}\right)\right]=\frac{2\left(2T+1\right)}{T\left(T-1\right)}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\boldsymbol{\phi}_{i,t-s}=O\left(1\right),\\
E(\mathbf{h}_{2i}) &= \frac{12}{T\left(T-1\right)\left(T+1\right)}E\left(\left(\sum_{t=1}^{T}t\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}sv_{is}\right)\right)\\
&= \frac{12}{\left(T+1\right)}\left(\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\left(\frac{t}{T}\right)\left(\frac{s}{T-1}\right)\boldsymbol{\phi}_{i,t-s}\right)=O\left(1\right),\\
E(\mathbf{h}_{3i}) &= \frac{6}{T\left(T-1\right)}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}sE(\boldsymbol{\xi}_{it}v_{is})+\sum_{t=1}^{T}\sum_{s=1}^{T}tE(\boldsymbol{\xi}_{it}v_{is})\right)\\
&= \frac{6}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\left(\frac{s}{T-1}\right)\boldsymbol{\phi}_{i,t-s}+\frac{6}{\left(T-1\right)}\sum_{s=1}^{T-1}\sum_{s=t+1}^{T}\left(\frac{t}{T}\right)\boldsymbol{\phi}_{i,t-s}=O\left(1\right)
\end{aligned}$$

where we used $0<t/T\leq 1$ and $0<s/(T-1)\leq 1$ for all s and t. Thus, both for FE and trend models, $E(\mathbf{h}_i)=O(1)$ and obtain

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{R}_T\mathbf{v}_i=\sqrt{\frac{N}{T}}\bar{\mathbf{h}}_N=O_p\left(\sqrt{\frac{N}{T}}\right)$$

where $\bar{\mathbf{h}}_N = N^{-1}\sum_{i=1}^N \mathbf{h}_i$.

(c) Noting $\mathbf{B}_{T_d}\mathbf{K}_T\mathbf{C}_T = \mathbf{0}$, we have the following decomposition:

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Xi}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Xi}_i'\mathbf{v}_i + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Xi}_i'\mathbf{A}_T\mathbf{v}_i.
\end{aligned}$$

The first term converges in distribution to $\mathcal{N}(\mathbf{0},\boldsymbol{\Omega})$ by Assumption 4. To derive the order of the second term, let us define $\mathbf{s}_i = \boldsymbol{\Xi}_i'\mathbf{A}_T\mathbf{v}_i$. Then, using (3.40), $\mathbf{s}_i$ can be decomposed as

$$\mathbf{s}_i = \mathbf{s}_{1i} + \mathbf{s}_{2i} + \mathbf{s}_{3i} + \mathbf{s}_{4i} + \mathbf{s}_{5i} + \mathbf{s}_{6i}$$

where

$$\begin{aligned}
\mathbf{s}_{1i} &= -\boldsymbol{\Xi}_{1i}'\mathbf{v}_{1i}, \quad \mathbf{s}_{2i} = \boldsymbol{\Xi}_{1i}'\mathbf{A}_{12}\mathbf{v}_{2i}, \quad \mathbf{s}_{3i} = \boldsymbol{\Xi}_{1i}'\mathbf{A}_{13}\mathbf{v}_{3i}, \quad \mathbf{s}_{4i} = \boldsymbol{\Xi}_{2i}\mathbf{A}_{22}\mathbf{v}_{2i}, \\
\mathbf{s}_{5i} &= \boldsymbol{\Xi}_{2i}'\mathbf{A}_{23}\mathbf{v}_{3i}, \quad \mathbf{s}_{6i} = -\boldsymbol{\Xi}_{3i}'\mathbf{v}_{3i}.
\end{aligned}$$

To derive the variance of $\mathbf{s}_i$, we need to calculate $Var(\mathbf{s}_{ki})$ and $Cov(\mathbf{s}_{ki},\mathbf{s}_{li}),(k \neq l)$ for $k,l = 1,...,6$. We consider the FE and trend models separately.

2.D.3 FE model

We have

$$Var(\mathbf{s}_{1i}) = Var(\boldsymbol{\xi}_{i1}v_{i1}) = O(1).$$

Using $a_{12}^{1,t-1} = O\left(\frac{\log T}{T}\right)$ for all t, $a_{13}^{11} = O\left(\frac{\log T}{T}\right)$, we have

$$Var(\mathbf{s}_{2i}) = Var\left(\sum_{t=2}^{T_d} a_{12}^{1,t-1}\boldsymbol{\xi}_{i1}v_{it}\right) = \sum_{t=2}^{T_d}\left(a_{12}^{1,t-1}\right)^2 Var\left(\boldsymbol{\xi}_{i1}v_{it}\right) = O\left(\frac{(\log T)^2}{T}\right),$$

$$Var(\mathbf{s}_{3i}) = Var\left(a_{13}^{11}\boldsymbol{\xi}_{i1}v_{iT}\right) = (a_{13}^{11})^2 Var\left(\boldsymbol{\xi}_{i1}v_{iT}\right) = O\left(\frac{(\log T)^2}{T^2}\right).$$

Similarly, using $a_{22}^{t-1,t-1} = O(1/(t+1)) + O(1/(T-t))$ and $a_{22}^{s-1,t-1} = O(\log T/T)$ for all $s \neq t$, we have

$$\begin{aligned}
Var(\mathbf{s}_{4i}) &= Var\left[\sum_{s=2}^{T_d-1}\sum_{t=s}^{T_d} a_{22}^{s-1,t-1}\boldsymbol{\xi}_{is}v_{it}\right] = \sum_{s=2}^{T_d-1}\sum_{t=s}^{T_d}(a_{22}^{s-1,t-1})^2 Var\left(\boldsymbol{\xi}_{is}v_{it}\right) \\
&= \sum_{t=2}^{T_d}(a_{22}^{t-1,t-1})^2 Var\left(\boldsymbol{\xi}_{it}v_{it}\right) + \sum_{s=2}^{T_d-1}\sum_{t=s+1}^{T_d}(a_{22}^{s-1,t-1})^2 Var\left(\boldsymbol{\xi}_{is}v_{it}\right) \\
&= O(1) + O\left((\log T)^2\right).
\end{aligned}$$

Finally, using $a_{23}^{t-1,1} = O(\log T/T)$ for all t, and the definition of $\boldsymbol{\Xi}_{3i}$, we have

$$Var(\mathbf{s}_{5i}) = Var\left(\sum_{t=2}^{T_d} a_{23}^{t-1,1}\boldsymbol{\xi}_{it}v_{iT}\right) = \sum_{t=2}^{T_d}(a_{23}^{t-1,1})^2 Var\left(\boldsymbol{\xi}_{it}v_{iT}\right) = O\left(\frac{(\log T)^2}{T}\right),$$

$$Var(\mathbf{s}_{6i}) = Var\left(\boldsymbol{\xi}_{iT}v_{iT}\right) = O(1).$$

For the covariances, we have

$$\begin{aligned}
& Cov(\mathbf{s}_{1i},\mathbf{s}_{2i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{3i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{4i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{6i}) = Cov(\mathbf{s}_{2i},\mathbf{s}_{3i}) \\
= \; & Cov(\mathbf{s}_{2i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{2i},\mathbf{s}_{6i}) = Cov(\mathbf{s}_{3i},\mathbf{s}_{4i}) = Cov(\mathbf{s}_{4i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{4i},\mathbf{s}_{6i}) = \mathbf{0},
\end{aligned}$$

$$\begin{aligned}
Cov(\mathbf{s}_{2i},\mathbf{s}_{4i}) &= E\left[\left(\sum_{t=2}^{T_d} a_{12}^{1,t-1}\boldsymbol{\xi}_{i1}v_{it}\right)\left(\sum_{t=2}^{T_d} a_{22}^{t-1,t-1}\boldsymbol{\xi}_{it}v_{it} + \sum_{s=2}^{T_d-1}\sum_{t=s+1}^{T_d} a_{22}^{s-1,t-1}\boldsymbol{\xi}_{is}v_{it}\right)'\right] \\
&= \sum_{t_1=2}^{T_d}\sum_{t_2=2}^{T_d} a_{12}^{1,t_1-1}a_{22}^{t_2-1,t_2-1}E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it_2}'v_{it_1}v_{it_2}) \\
&\quad + \sum_{t_1=2}^{T_d}\sum_{s=2}^{T_d-1}\sum_{t_2=s+1}^{T_d} a_{12}^{1,t_1-1}a_{22}^{s-1,t_2-1}E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{is}'v_{it_1}v_{it_2}) \\
&= \sum_{t=2}^{T_d} a_{12}^{1,t-1}a_{22}^{t-1,t-1}E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it}'v_{it}^2) + \sum_{s=2}^{T_d-1}\sum_{t=s+1}^{T_d} a_{12}^{1,t-1}a_{22}^{s-1,t-1}E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{is}'v_{it}^2) \\
&= O\left(\frac{(\log T)^2}{T}\right) + O\left((\log T)^2\right),
\end{aligned}$$

$$Cov(\mathbf{s}_{3i},\mathbf{s}_{5i}) = \sum_{t=2}^{T_d} a_{13}^{11}a_{23}^{t-1,1}E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{it}'v_{iT}^2) = O\left(\frac{(\log T)^2}{T}\right),$$

$$Cov(\mathbf{s}_{3i},\mathbf{s}_{6i}) = a_{13}^{11} E(\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{iT}' v_{iT}^2) = O\left(\frac{\log T}{T}\right),$$

$$Cov(\mathbf{s}_{5i},\mathbf{s}_{6i}) = \sum_{t=2}^{T_d} a_{23}^{t-1,1} E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{iT}' v_{iT}^2) = O(\log T).$$

Therefore, for FE model, we have $Var(\mathbf{s}_i) = O\left((\log T)^2\right)$, and

$$Var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Xi}_i' \mathbf{A}_T \mathbf{v}_i\right) = \frac{1}{NT}\sum_{i=1}^N Var(\mathbf{s}_i) = O\left(\frac{(\log T)^2}{T}\right)$$

Hence, it follows that $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \boldsymbol{\Xi}_i'\mathbf{A}_T\mathbf{v}_i = O_p\left(\log T/\sqrt{T}\right) = o_p(1)$.

2.D.4 Trend model

$$Var(\mathbf{s}_{1i}) \;=\; Var(\sum_{s=1}^2\sum_{t=1}^2 \boldsymbol{\xi}_{it} v_{is}) = Var(\boldsymbol{\xi}_{i1} v_{i1}) + Var(\boldsymbol{\xi}_{i2} v_{i2}) = O(1).$$

Using $a_{12}^{s,t-1} = O\left(\frac{\log T}{T}\right), s =, 2$ for all t and $a_{13}^{jk} = O\left(\frac{\log T}{T}\right)$, we have

$$Var(\mathbf{s}_{2i}) \;=\; Var\left(\sum_{s=1}^2\sum_{t=3}^{T_{2d}} a_{12}^{j,t-1}\boldsymbol{\xi}_{i1} v_{it}\right) = O\left(\frac{(\log T)^2}{T}\right),$$

$$Var(\mathbf{s}_{3i}) \;=\; Var\left(\sum_{s=1}^2\sum_{t=1}^2 a_{13}^{st}\boldsymbol{\xi}_{it} v_{iT_s+1}\right) = \sum_{s=1}^2\sum_{t=1}^2 (a_{13}^{st})^2 Var\left(\boldsymbol{\xi}_{it} v_{iT_s+1}\right) = O\left(\frac{(\log T)^2}{T^2}\right).$$

Similarly, using $a_{22}^{t-1,t-1} = O(1/(t+1)) + O(1/(T-t))$ and $a_{22}^{s-1,t-1} = O(\log T/T)$ for all $s \neq t$, we have

$$\begin{aligned}
Var(\mathbf{s}_{4i}) &= Var\left[\sum_{s=3}^{T_d-1}\sum_{t=s}^{T_d} a_{22}^{s-1,t-1}\boldsymbol{\xi}_{is} v_{it}\right] = \sum_{s=3}^{T_d-1}\sum_{t=s}^{T_d} (a_{22}^{s-1,t-1})^2 Var\left(\boldsymbol{\xi}_{is} v_{it}\right) \\
&= \sum_{t=3}^{T_d} (a_{22}^{t-1,t-1})^2 Var\left(\boldsymbol{\xi}_{it} v_{it}\right) + \sum_{s=3}^{T_d-1}\sum_{t=s+1}^{T_d} (a_{22}^{s-1,t-1})^2 Var\left(\boldsymbol{\xi}_{is} v_{it}\right) \\
&= O(1) + O\left((\log T)^2\right).
\end{aligned}$$

Finally, using $a_{23}^{t-1,s} = O(\log T/T), s=1,2$ for all t, and the definition of $\mathbf{\Xi}_{3i}$, we have

$$\begin{aligned} Var(\mathbf{s}_{5i}) &= Var\left(\sum_{s=1}^{2}\sum_{t=3}^{T_d} a_{23}^{t-1,s}\boldsymbol{\xi}_{it}v_{iT_s+1}\right) \\ &= \sum_{s=1}^{2}\sum_{t=2}^{T_d}(a_{23}^{t-1,s})^2 Var\left(\boldsymbol{\xi}_{it}v_{iT_s+1}\right) = O\left(\frac{(\log T)^2}{T}\right), \end{aligned}$$

$$Var(\mathbf{s}_{6i}) = Var\left(\sum_{s=1}^{2}\sum_{t=1}^{2}\boldsymbol{\xi}_{iT_t}v_{iT_s}\right) = O(1).$$

For the covariances of trend model,

$$\begin{aligned} & Cov(\mathbf{s}_{1i},\mathbf{s}_{2i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{3i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{4i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{1i},\mathbf{s}_{6i}) = Cov(\mathbf{s}_{2i},\mathbf{s}_{3i}) \\ = \; & Cov(\mathbf{s}_{2i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{2i},\mathbf{s}_{6i}) = Cov(\mathbf{s}_{3i},\mathbf{s}_{4i}) = Cov(\mathbf{s}_{4i},\mathbf{s}_{5i}) = Cov(\mathbf{s}_{4i},\mathbf{s}_{6i}) = \mathbf{0}, \end{aligned}$$

$$\begin{aligned} Cov(\mathbf{s}_{2i},\mathbf{s}_{4i}) &= E\left[\left(\sum_{s=1}^{2}\sum_{t=2}^{T_d} a_{12}^{s,t-1}\boldsymbol{\xi}_{is}v_{it}\right)\left(\sum_{t=3}^{T_d} a_{22}^{t-1,t-1}\boldsymbol{\xi}_{it}v_{it} + \sum_{s=3}^{T_d-1}\sum_{t=s+1}^{T_d} a_{22}^{s-1,t-1}\boldsymbol{\xi}_{is}v_{it}\right)'\right] \\ &= \sum_{s=1}^{2}\sum_{t_1=3}^{T_d}\sum_{t_2=3}^{T_d} a_{12}^{1,t_1-1}a_{22}^{t_2-1,t_2-1}E(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it_2}'v_{it_1}v_{it_2}) \\ &+ \sum_{s_1=1}^{2}\sum_{t_1=3}^{T_d}\sum_{s_2=3}^{T_d-1}\sum_{t_2=s_2+1}^{T_d} a_{12}^{1,t_1-1}a_{22}^{s_2-1,t_2-1}E(\boldsymbol{\xi}_{is_1}\boldsymbol{\xi}_{is_2}'v_{it_1}v_{it_2}) \\ &= \sum_{s=1}^{2}\sum_{t=3}^{T_d} a_{12}^{s,t-1}a_{22}^{t-1,t-1}E(\boldsymbol{\xi}_{is}\boldsymbol{\xi}_{it}'v_{it}^2) + \sum_{s_1=1}^{2}\sum_{s_2=2}^{T_d-1}\sum_{t=s_2+1}^{T_d} a_{12}^{s_1,t-1}a_{22}^{s_2-1,t-1}E(\boldsymbol{\xi}_{is_1}\boldsymbol{\xi}_{is_2}'v_{it}^2) \\ &= O\left(\frac{(\log T)^2}{T}\right) + O\left((\log T)^2\right), \end{aligned}$$

$$Cov(\mathbf{s}_{3i},\mathbf{s}_{5i}) = \sum_{s_1=1}^{2}\sum_{s_2=1}^{2}\sum_{t=3}^{T_d} a_{13}^{s_1s_21}a_{23}^{t-1,s_2}E(\boldsymbol{\xi}_{is_1}\boldsymbol{\xi}_{it}'v_{iT_{s1}+1}^2) = O\left(\frac{(\log T)^2}{T}\right),$$

$$Cov(\mathbf{s}_{3i},\mathbf{s}_{6i}) = \sum_{s_1=1}^{2}\sum_{s_2=1}^{2} a_{13}^{s_1s_2}E(\boldsymbol{\xi}_{is_1}\boldsymbol{\xi}_{iT_{s2}+1}'v_{iT_{s_1}+1}^2) = O\left(\frac{\log T}{T}\right),$$

$$Cov(\mathbf{s}_{5i}, \mathbf{s}_{6i}) = \sum_{s=1}^{2}\sum_{t=3}^{T_d} a_{23}^{t-1,s} E(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{iT_s+1}' v_{iT_s+1}^2) = O(\log T).$$

Therefore, for trend model, we have $var(\mathbf{s}_i) = O\left((\log T)^2\right)$, and

$$var\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{A}_T\mathbf{v}_i\right) = \frac{1}{NT}\sum_{i=1}^{N} var(\mathbf{s}_i) = O\left(\frac{(\log T)^2}{T}\right)$$

Hence, it follows that $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{A}_T\mathbf{v}_i = O_p\left(\log T/\sqrt{T}\right) = o_p(1)$.

2.E Proof of Theorems 2 and 3

We first provide a proof of Theorem 2. Using (2.74) and Assumption 4, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}' + O_p\left(\frac{1}{T}\right) \to^p \boldsymbol{\Gamma}_0. \tag{2.76}$$

Next, we have the following decomposition

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} + \sqrt{\frac{N}{T}}\bar{\mathbf{h}}_N$$

where $\bar{\mathbf{h}}_N = N^{-1}\sum_{i=1}^{N}\mathbf{h}_i$. Hence, using Assumption 4, as $N, T \to \infty$ with $N/T \to \kappa, (0 < \kappa < \infty)$, we obtain

$$\begin{aligned}
\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{FE} - \boldsymbol{\delta}) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{v}_i \\
&= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} \\
&\quad + \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i\right)^{-1}\sqrt{\frac{N}{T}}\bar{\mathbf{h}}_N \\
&\to^d \mathcal{N}(\sqrt{\kappa}\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{h}}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^{-1})
\end{aligned}$$

where $\bar{\mathbf{h}} = \text{plim}_{N,T\to\infty}\bar{\mathbf{h}}_N$. Theorem 3 can be proved by noting that $\widehat{\boldsymbol{\Gamma}}_0$ and $\widehat{\bar{\mathbf{h}}}$ are consistent estimators of $\boldsymbol{\Gamma}_0$ and $\bar{\mathbf{h}}$ with large N and T.

2.F Proof of Theorem 4

Using Lemma 1(a), (2.76) and Assumption 4, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i + O_p\left(\frac{\log T}{T}\right) \to^p \mathbf{\Gamma}_0. \quad (2.77)$$

Also, using Lemma 1(c) and Assumption 4, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} + o_p(1) \to^d \mathcal{N}(\mathbf{0},\mathbf{\Omega}). \quad (2.78)$$

Combining (2.77) and (2.78), we obtain the result.

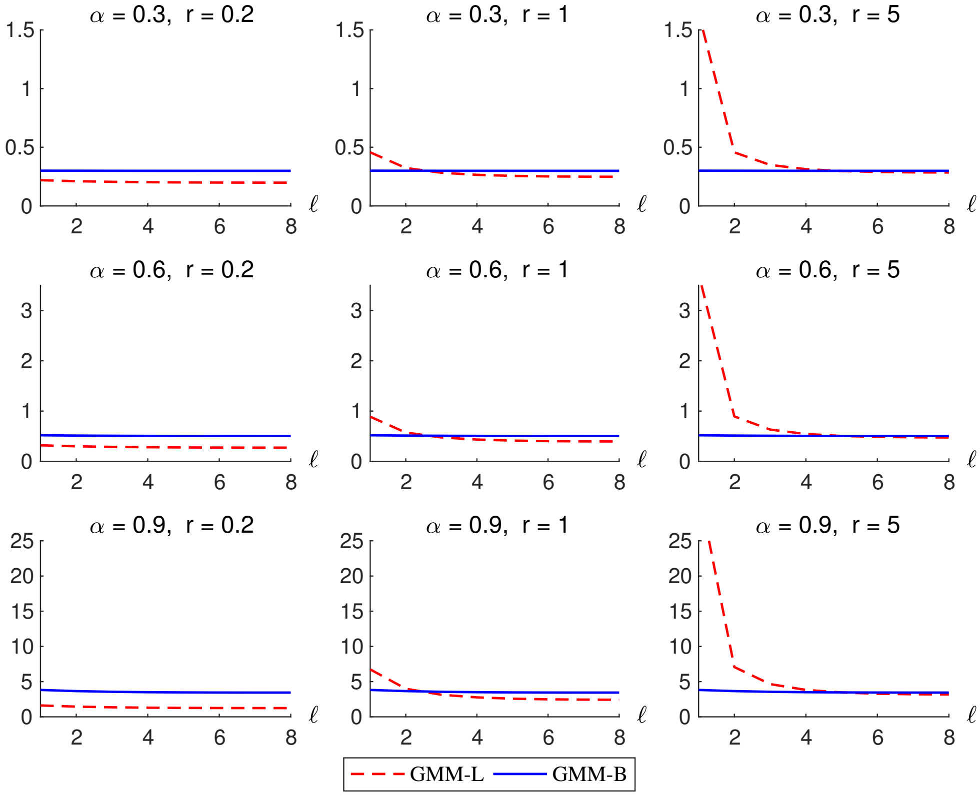

Figure 2.1: Asymptotic variance of GMM estimators with various instruments lag length ℓ ($T = 10$)

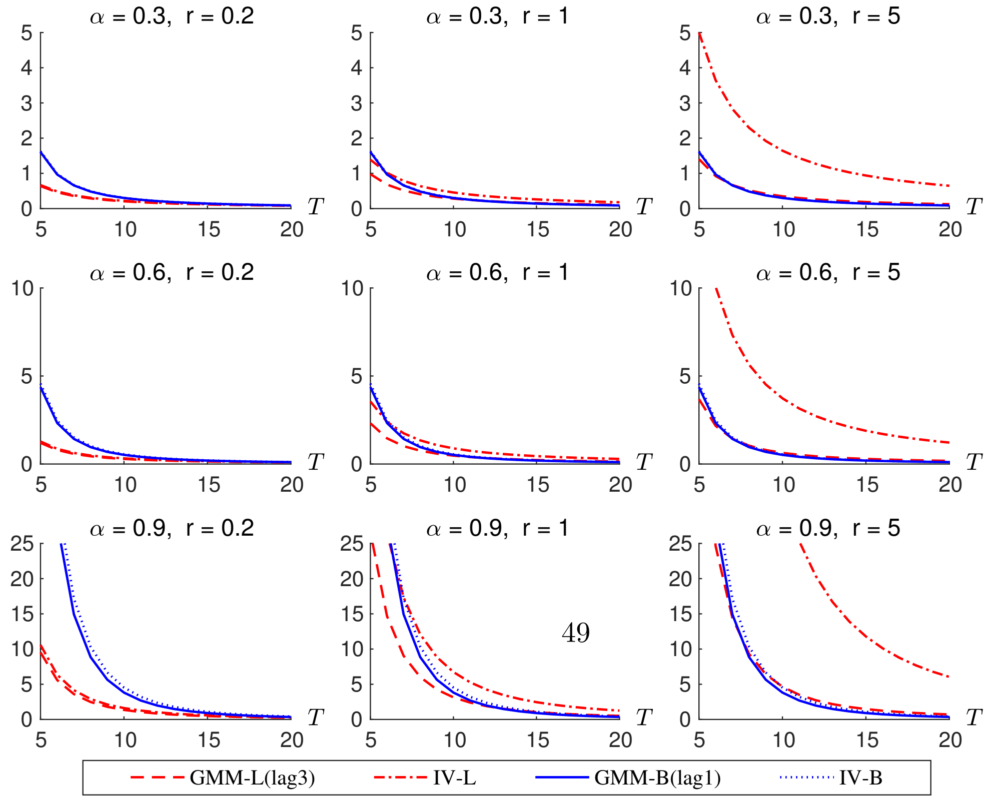

Figure 2.2: Asymptotic variance of IV/GMM estimators with various T

Table 2.1: Fixed effects model: $\alpha = 0.4,\ \ \beta = 1.0$

$N = 50,\ \ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.113	0.003	-0.042	-0.036	0.000	-0.013	-0.035	0.034	0.004	-0.085	-0.039	0.000	-0.014	-0.025
IQR	0.054	0.254	0.156	0.086	0.100	0.097	0.088	0.140	0.579	0.361	0.223	0.302	0.275	0.252
MAE	0.113	0.126	0.080	0.052	0.050	0.048	0.050	0.075	0.285	0.183	0.112	0.150	0.137	0.129
Size	0.818	0.021	0.057	0.086	0.055	0.061	0.090	0.072	0.017	0.041	0.064	0.055	0.056	0.058
	$T = 25$													
Bias	-0.045	0.001	-0.016	-0.011	0.000	-0.004	-0.009	0.021	0.007	-0.023	-0.008	-0.001	-0.003	-0.001
IQR	0.031	0.131	0.090	0.045	0.043	0.041	0.039	0.084	0.243	0.168	0.111	0.116	0.110	0.105
MAE	0.045	0.065	0.045	0.024	0.021	0.020	0.020	0.044	0.120	0.086	0.056	0.058	0.055	0.052
Size	0.490	0.045	0.062	0.070	0.063	0.057	0.066	0.077	0.035	0.055	0.066	0.064	0.067	0.062
	$T = 50$													
Bias	-0.022	-0.002	-0.010	-0.004	0.000	-0.001	-0.003	0.014	0.000	-0.012	0.001	0.002	0.002	0.002
IQR	0.022	0.083	0.063	0.031	0.026	0.026	0.025	0.054	0.144	0.105	0.068	0.068	0.066	0.064
MAE	0.022	0.042	0.032	0.016	0.013	0.013	0.013	0.029	0.071	0.053	0.034	0.034	0.033	0.033
Size	0.281	0.045	0.057	0.075	0.066	0.062	0.063	0.073	0.044	0.056	0.056	0.062	0.061	0.056
	$T = 100$													
Bias	-0.011	0.001	-0.004	-0.002	0.001	0.000	-0.001	0.006	0.001	-0.003	0.000	0.001	0.002	0.001
IQR	0.015	0.058	0.043	0.020	0.017	0.017	0.017	0.038	0.090	0.072	0.045	0.043	0.043	0.042
MAE	0.011	0.029	0.022	0.010	0.008	0.008	0.008	0.020	0.045	0.035	0.023	0.022	0.021	0.021
Size	0.167	0.053	0.058	0.060	0.056	0.055	0.059	0.070	0.049	0.050	0.056	0.057	0.058	0.055

$N = 100,\ \ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.114	0.001	-0.027	-0.021	0.001	-0.005	-0.017	0.038	0.009	-0.053	-0.021	0.005	0.000	-0.014
IQR	0.039	0.177	0.130	0.070	0.070	0.069	0.066	0.100	0.402	0.302	0.168	0.213	0.207	0.192
MAE	0.114	0.089	0.066	0.037	0.035	0.035	0.036	0.055	0.201	0.157	0.086	0.109	0.104	0.097
Size	0.987	0.027	0.055	0.073	0.061	0.061	0.073	0.094	0.026	0.045	0.062	0.055	0.060	0.058
	$T = 25$													
Bias	-0.044	0.002	-0.011	-0.005	0.000	-0.001	-0.004	0.022	0.003	-0.018	-0.003	0.001	0.001	0.002
IQR	0.022	0.090	0.073	0.033	0.030	0.029	0.029	0.057	0.169	0.135	0.077	0.079	0.076	0.075
MAE	0.044	0.045	0.036	0.017	0.015	0.015	0.014	0.034	0.085	0.068	0.038	0.039	0.039	0.037
Size	0.749	0.038	0.058	0.055	0.046	0.049	0.051	0.083	0.040	0.055	0.053	0.060	0.060	0.059
	$T = 50$													
Bias	-0.022	-0.002	-0.006	-0.003	0.000	-0.001	-0.002	0.013	-0.001	-0.007	0.000	0.001	0.001	0.002
IQR	0.016	0.062	0.051	0.024	0.018	0.018	0.018	0.037	0.104	0.084	0.049	0.047	0.046	0.045
MAE	0.022	0.031	0.025	0.012	0.009	0.009	0.009	0.021	0.051	0.041	0.025	0.023	0.023	0.023
Size	0.461	0.050	0.060	0.058	0.063	0.062	0.064	0.072	0.045	0.055	0.052	0.049	0.051	0.050
	$T = 100$													
Bias	-0.011	0.001	-0.003	-0.001	0.000	0.000	-0.001	0.006	0.001	-0.002	0.000	0.000	0.000	0.000
IQR	0.011	0.041	0.034	0.016	0.012	0.012	0.012	0.028	0.064	0.056	0.034	0.031	0.031	0.031
MAE	0.011	0.021	0.017	0.008	0.006	0.006	0.006	0.015	0.032	0.027	0.017	0.016	0.016	0.015
Size	0.266	0.056	0.054	0.053	0.050	0.051	0.053	0.061	0.051	0.049	0.049	0.048	0.051	0.052

$N = 250,\ \ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.115	0.001	-0.017	-0.009	0.001	-0.003	-0.008	0.035	0.003	-0.028	-0.011	0.003	0.000	-0.007
IQR	0.024	0.111	0.092	0.045	0.046	0.045	0.044	0.061	0.249	0.209	0.102	0.130	0.124	0.117
MAE	0.115	0.055	0.047	0.023	0.023	0.022	0.023	0.040	0.124	0.107	0.052	0.065	0.062	0.059
Size	1.000	0.035	0.040	0.058	0.057	0.058	0.060	0.106	0.039	0.037	0.047	0.045	0.043	0.047
	$T = 25$													
Bias	-0.044	0.000	-0.006	-0.003	0.000	-0.001	-0.002	0.022	-0.001	-0.008	-0.001	-0.001	-0.001	0.000
IQR	0.014	0.057	0.048	0.022	0.018	0.018	0.018	0.036	0.107	0.092	0.050	0.048	0.048	0.048
MAE	0.044	0.028	0.025	0.011	0.009	0.009	0.009	0.024	0.053	0.047	0.025	0.024	0.024	0.024
Size	0.982	0.053	0.053	0.055	0.048	0.051	0.055	0.129	0.056	0.054	0.054	0.057	0.050	0.047
	$T = 50$													
Bias	-0.022	-0.001	-0.003	-0.002	0.000	-0.001	-0.001	0.012	-0.002	-0.004	-0.001	0.000	0.000	0.001
IQR	0.010	0.039	0.035	0.015	0.012	0.012	0.012	0.025	0.065	0.058	0.032	0.029	0.029	0.029
MAE	0.022	0.019	0.017	0.008	0.006	0.006	0.006	0.015	0.033	0.030	0.016	0.015	0.015	0.014
Size	0.840	0.046	0.053	0.056	0.060	0.055	0.055	0.105	0.050	0.054	0.050	0.052	0.050	0.053
	$T = 100$													
Bias	-0.011	0.000	-0.001	0.000	0.000	0.000	0.000	0.006	0.000	-0.001	0.000	0.000	0.000	0.000
IQR	0.007	0.025	0.023	0.010	0.008	0.007	0.007	0.018	0.042	0.037	0.020	0.019	0.018	0.018
MAE	0.011	0.012	0.011	0.005	0.004	0.004	0.004	0.010	0.021	0.019	0.010	0.009	0.009	0.009
Size	0.549	0.050	0.050	0.052	0.059	0.063	0.061	0.076	0.049	0.046	0.055	0.052	0.051	0.051

Table 2.1:

Table 2.1(cont.): Fixed effects model: $\alpha = 0.4, \ \ \beta = 1.0$

$N = 50, \ \ \sigma^2_\eta = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.113	0.010	-0.054	-0.036	0.000	-0.013	-0.035	0.034	0.039	-0.112	-0.044	0.000	-0.014	-0.025
IQR	0.054	0.389	0.172	0.089	0.100	0.097	0.088	0.140	0.929	0.439	0.228	0.302	0.275	0.252
MAE	0.113	0.197	0.091	0.052	0.050	0.048	0.050	0.075	0.467	0.223	0.118	0.150	0.137	0.129
Size	0.818	0.012	0.054	0.086	0.055	0.061	0.090	0.072	0.009	0.044	0.068	0.055	0.056	0.058
	$T = 25$													
Bias	-0.045	0.007	-0.022	-0.011	0.000	-0.004	-0.009	0.021	0.014	-0.032	-0.010	-0.001	-0.003	-0.001
IQR	0.031	0.217	0.103	0.046	0.043	0.041	0.039	0.084	0.417	0.201	0.114	0.116	0.110	0.105
MAE	0.045	0.108	0.053	0.024	0.021	0.020	0.020	0.044	0.205	0.104	0.057	0.058	0.055	0.052
Size	0.490	0.035	0.066	0.068	0.063	0.057	0.066	0.077	0.027	0.061	0.070	0.064	0.067	0.062
	$T = 50$													
Bias	-0.022	-0.004	-0.014	-0.004	0.000	-0.001	-0.003	0.014	-0.003	-0.020	0.000	0.002	0.002	0.002
IQR	0.022	0.137	0.074	0.031	0.026	0.026	0.025	0.054	0.241	0.123	0.068	0.068	0.066	0.064
MAE	0.022	0.067	0.037	0.016	0.013	0.013	0.013	0.029	0.117	0.064	0.034	0.034	0.033	0.033
Size	0.281	0.039	0.063	0.074	0.066	0.062	0.063	0.073	0.037	0.068	0.061	0.062	0.061	0.056
	$T = 100$													
Bias	-0.011	0.002	-0.004	-0.002	0.001	0.000	-0.001	0.006	0.003	-0.005	0.000	0.001	0.002	0.001
IQR	0.015	0.091	0.050	0.020	0.017	0.017	0.017	0.038	0.149	0.084	0.047	0.043	0.043	0.042
MAE	0.011	0.045	0.024	0.010	0.008	0.008	0.008	0.020	0.075	0.042	0.023	0.022	0.021	0.021
Size	0.167	0.045	0.062	0.060	0.056	0.055	0.059	0.070	0.042	0.051	0.057	0.057	0.058	0.055

$N = 100, \ \ \sigma^2_\eta = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.114	0.012	-0.040	-0.022	0.001	-0.005	-0.017	0.038	0.030	-0.079	-0.025	0.005	0.000	-0.014
IQR	0.039	0.291	0.164	0.072	0.070	0.069	0.066	0.100	0.698	0.391	0.178	0.213	0.207	0.192
MAE	0.114	0.145	0.082	0.037	0.035	0.035	0.036	0.055	0.351	0.199	0.090	0.109	0.104	0.097
Size	0.987	0.016	0.050	0.073	0.061	0.061	0.073	0.094	0.017	0.047	0.068	0.055	0.060	0.058
	$T = 25$													
Bias	-0.044	0.004	-0.016	-0.005	0.000	-0.001	-0.004	0.022	0.013	-0.025	-0.003	0.001	0.001	0.002
IQR	0.022	0.149	0.091	0.034	0.030	0.029	0.029	0.057	0.286	0.177	0.079	0.079	0.076	0.075
MAE	0.044	0.073	0.045	0.018	0.015	0.015	0.014	0.034	0.141	0.090	0.039	0.039	0.039	0.037
Size	0.749	0.038	0.052	0.056	0.046	0.049	0.051	0.083	0.034	0.057	0.052	0.060	0.060	0.059
	$T = 50$													
Bias	-0.022	-0.002	-0.007	-0.003	0.000	-0.001	-0.002	0.013	-0.003	-0.013	0.000	0.001	0.001	0.002
IQR	0.016	0.096	0.064	0.024	0.018	0.018	0.018	0.037	0.167	0.108	0.051	0.047	0.046	0.045
MAE	0.022	0.048	0.031	0.012	0.009	0.009	0.009	0.021	0.084	0.054	0.025	0.023	0.023	0.023
Size	0.461	0.048	0.054	0.055	0.063	0.062	0.064	0.072	0.044	0.049	0.050	0.049	0.051	0.050
	$T = 100$													
Bias	-0.011	0.001	-0.003	-0.001	0.000	0.000	-0.001	0.006	0.001	-0.006	-0.001	0.000	0.000	0.000
IQR	0.011	0.066	0.044	0.016	0.012	0.012	0.012	0.028	0.107	0.072	0.035	0.031	0.031	0.031
MAE	0.011	0.033	0.022	0.008	0.006	0.006	0.006	0.015	0.052	0.036	0.017	0.016	0.016	0.015
Size	0.266	0.053	0.061	0.053	0.050	0.051	0.053	0.061	0.056	0.053	0.049	0.048	0.051	0.052

$N = 250, \ \ \sigma^2_\eta = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.115	0.006	-0.023	-0.009	0.001	-0.003	-0.008	0.035	0.010	-0.048	-0.014	0.003	0.000	-0.007
IQR	0.024	0.175	0.121	0.046	0.046	0.045	0.044	0.061	0.436	0.288	0.112	0.130	0.124	0.117
MAE	0.115	0.088	0.061	0.023	0.023	0.022	0.023	0.040	0.219	0.147	0.056	0.065	0.062	0.059
Size	1.000	0.019	0.038	0.057	0.057	0.058	0.060	0.106	0.019	0.032	0.047	0.045	0.043	0.047
	$T = 25$													
Bias	-0.044	0.000	-0.008	-0.003	0.000	-0.001	-0.002	0.022	-0.002	-0.014	-0.002	-0.001	-0.001	0.000
IQR	0.014	0.094	0.066	0.022	0.018	0.018	0.018	0.036	0.180	0.131	0.052	0.048	0.048	0.048
MAE	0.044	0.047	0.034	0.011	0.009	0.009	0.009	0.024	0.090	0.065	0.026	0.024	0.024	0.024
Size	0.982	0.042	0.054	0.054	0.048	0.051	0.055	0.129	0.042	0.058	0.057	0.057	0.050	0.047
	$T = 50$													
Bias	-0.022	-0.001	-0.006	-0.002	0.000	-0.001	-0.001	0.012	-0.005	-0.008	-0.001	0.000	0.000	0.001
IQR	0.010	0.059	0.049	0.015	0.012	0.012	0.012	0.025	0.106	0.082	0.033	0.029	0.029	0.029
MAE	0.022	0.030	0.024	0.008	0.006	0.006	0.006	0.015	0.052	0.042	0.016	0.015	0.015	0.014
Size	0.840	0.042	0.053	0.055	0.060	0.055	0.055	0.105	0.039	0.060	0.047	0.052	0.050	0.053
	$T = 100$													
Bias	-0.011	0.001	-0.002	0.000	0.000	0.000	0.000	0.006	0.001	-0.002	0.000	0.000	0.000	0.000
IQR	0.007	0.040	0.032	0.010	0.008	0.007	0.007	0.018	0.066	0.054	0.021	0.019	0.018	0.018
MAE	0.011	0.020	0.016	0.005	0.004	0.004	0.004	0.010	0.033	0.026	0.011	0.009	0.009	0.009
Size	0.549	0.051	0.058	0.051	0.059	0.063	0.061	0.076	0.048	0.050	0.049	0.052	0.051	0.051

Table 2.2: Fixed effects model: $\alpha = 0.8,\ \beta = 1.0$

$N = 50,\ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.036	-0.189	-0.109	0.004	-0.053	-0.101	0.006	0.034	-0.164	-0.111	0.013	-0.074	-0.122
IQR	0.049	0.641	0.243	0.112	0.188	0.143	0.112	0.137	0.745	0.407	0.254	0.429	0.335	0.279
MAE	0.150	0.320	0.191	0.110	0.094	0.079	0.102	0.068	0.370	0.224	0.151	0.210	0.168	0.159
Size	0.995	0.001	0.106	0.243	0.029	0.085	0.215	0.062	0.005	0.038	0.097	0.031	0.057	0.094
	$T = 25$													
Bias	-0.054	0.062	-0.072	-0.032	-0.002	-0.010	-0.022	0.021	0.046	-0.030	-0.023	0.000	-0.008	-0.016
IQR	0.024	0.387	0.115	0.048	0.049	0.045	0.040	0.084	0.293	0.184	0.123	0.122	0.116	0.112
MAE	0.054	0.205	0.075	0.035	0.025	0.024	0.025	0.043	0.150	0.095	0.062	0.061	0.058	0.057
Size	0.908	0.012	0.067	0.142	0.057	0.064	0.124	0.081	0.014	0.054	0.068	0.052	0.061	0.064
	$T = 50$													
Bias	-0.026	0.061	-0.034	-0.012	-0.001	-0.003	-0.007	0.017	0.012	-0.004	-0.003	0.002	0.001	0.001
IQR	0.015	0.289	0.075	0.029	0.023	0.022	0.021	0.055	0.156	0.113	0.075	0.066	0.065	0.065
MAE	0.026	0.157	0.042	0.017	0.012	0.011	0.011	0.029	0.078	0.056	0.037	0.033	0.033	0.032
Size	0.664	0.030	0.059	0.086	0.056	0.063	0.080	0.079	0.016	0.060	0.059	0.063	0.061	0.062
	$T = 100$													
Bias	-0.012	0.044	-0.015	-0.005	0.000	-0.001	-0.002	0.009	-0.001	0.005	0.000	0.001	0.001	0.002
IQR	0.010	0.207	0.048	0.019	0.013	0.012	0.012	0.039	0.089	0.072	0.049	0.042	0.041	0.040
MAE	0.012	0.113	0.024	0.010	0.006	0.006	0.006	0.020	0.045	0.036	0.025	0.021	0.020	0.020
Size	0.400	0.072	0.063	0.083	0.056	0.059	0.064	0.071	0.012	0.044	0.048	0.057	0.059	0.056

$N = 100,\ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.041	-0.174	-0.077	0.001	-0.033	-0.068	0.006	0.046	-0.177	-0.097	0.009	-0.041	-0.093
IQR	0.034	0.662	0.257	0.097	0.134	0.110	0.093	0.098	0.758	0.374	0.210	0.299	0.254	0.222
MAE	0.150	0.323	0.178	0.078	0.067	0.060	0.072	0.050	0.381	0.204	0.126	0.151	0.126	0.126
Size	1.000	0.001	0.088	0.186	0.044	0.068	0.165	0.054	0.004	0.036	0.086	0.035	0.060	0.090
	$T = 25$													
Bias	-0.054	0.057	-0.064	-0.020	-0.002	-0.006	-0.013	0.022	0.038	-0.026	-0.011	0.002	-0.001	-0.008
IQR	0.016	0.342	0.114	0.039	0.033	0.031	0.029	0.055	0.235	0.143	0.087	0.089	0.084	0.078
MAE	0.054	0.178	0.069	0.024	0.017	0.016	0.017	0.033	0.124	0.073	0.045	0.044	0.041	0.041
Size	0.995	0.010	0.062	0.096	0.044	0.051	0.082	0.083	0.018	0.041	0.065	0.057	0.056	0.059
	$T = 50$													
Bias	-0.025	0.040	-0.029	-0.008	0.000	-0.001	-0.004	0.016	0.009	-0.004	-0.002	0.001	0.001	0.000
IQR	0.011	0.237	0.067	0.023	0.017	0.016	0.016	0.038	0.105	0.078	0.053	0.046	0.045	0.045
MAE	0.025	0.126	0.037	0.013	0.008	0.008	0.008	0.022	0.053	0.039	0.026	0.023	0.023	0.022
Size	0.910	0.033	0.050	0.082	0.060	0.063	0.075	0.085	0.021	0.046	0.058	0.052	0.055	0.052
	$T = 100$													
Bias	-0.012	0.020	-0.015	-0.003	0.000	-0.001	-0.001	0.009	0.000	0.002	-0.001	0.000	0.000	0.000
IQR	0.007	0.163	0.046	0.015	0.009	0.009	0.009	0.028	0.058	0.052	0.037	0.031	0.031	0.031
MAE	0.012	0.083	0.024	0.008	0.004	0.004	0.004	0.015	0.029	0.026	0.018	0.015	0.015	0.015
Size	0.668	0.060	0.059	0.069	0.050	0.046	0.051	0.075	0.021	0.054	0.050	0.051	0.052	0.052

$N = 250,\ \sigma_\eta^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.038	-0.157	-0.044	0.002	-0.016	-0.037	0.004	0.047	-0.161	-0.054	0.002	-0.024	-0.054
IQR	0.022	0.572	0.217	0.068	0.084	0.077	0.069	0.060	0.654	0.271	0.145	0.179	0.168	0.155
MAE	0.150	0.286	0.166	0.049	0.042	0.040	0.044	0.031	0.327	0.176	0.081	0.090	0.086	0.086
Size	1.000	0.001	0.073	0.118	0.056	0.056	0.108	0.041	0.003	0.038	0.074	0.050	0.052	0.065
	$T = 25$													
Bias	-0.054	0.029	-0.051	-0.011	0.000	-0.002	-0.005	0.021	0.019	-0.025	-0.009	-0.001	-0.003	-0.005
IQR	0.011	0.274	0.099	0.028	0.021	0.020	0.020	0.035	0.176	0.096	0.058	0.055	0.051	0.050
MAE	0.054	0.140	0.060	0.015	0.011	0.010	0.010	0.024	0.092	0.048	0.030	0.028	0.026	0.026
Size	1.000	0.014	0.043	0.082	0.049	0.048	0.062	0.130	0.012	0.038	0.062	0.052	0.049	0.051
	$T = 50$													
Bias	-0.025	0.014	-0.023	-0.004	0.000	-0.001	-0.002	0.015	0.001	-0.005	-0.002	0.000	0.000	0.000
IQR	0.007	0.182	0.062	0.017	0.011	0.010	0.010	0.025	0.067	0.051	0.035	0.030	0.029	0.029
MAE	0.025	0.089	0.035	0.009	0.005	0.005	0.005	0.016	0.034	0.026	0.018	0.015	0.015	0.015
Size	1.000	0.031	0.052	0.060	0.048	0.060	0.057	0.140	0.023	0.044	0.051	0.051	0.052	0.052
	$T = 100$													
Bias	-0.012	0.004	-0.013	-0.002	0.000	0.000	-0.001	0.009	0.001	0.001	0.001	0.000	0.001	0.000
IQR	0.004	0.112	0.042	0.010	0.006	0.006	0.006	0.017	0.036	0.035	0.023	0.019	0.018	0.018
MAE	0.012	0.053	0.022	0.005	0.003	0.003	0.003	0.011	0.018	0.017	0.011	0.009	0.009	0.009
Size	0.958	0.059	0.059	0.058	0.054	0.048	0.051	0.099	0.025	0.044	0.050	0.056	0.055	0.056

Table 2.2:

Table 2.2(cont.): Fixed effects model: $\alpha = 0.8$, $\beta = 1.0$

$N = 50$, $\sigma_\eta^2 = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.054	-0.199	-0.109	0.004	-0.053	-0.101	0.006	0.037	-0.120	-0.115	0.013	-0.074	-0.122
IQR	0.049	0.651	0.260	0.112	0.188	0.143	0.112	0.137	1.000	0.597	0.269	0.429	0.335	0.279
MAE	0.150	0.325	0.202	0.110	0.094	0.079	0.102	0.068	0.500	0.299	0.161	0.210	0.168	0.159
Size	0.995	0.000	0.079	0.244	0.029	0.085	0.215	0.062	0.005	0.037	0.104	0.031	0.057	0.094
	$T = 25$													
Bias	-0.054	0.067	-0.077	-0.032	-0.002	-0.010	-0.022	0.021	0.045	-0.032	-0.025	0.000	-0.008	-0.016
IQR	0.024	0.380	0.121	0.049	0.049	0.045	0.040	0.084	0.459	0.287	0.131	0.122	0.116	0.112
MAE	0.054	0.199	0.081	0.034	0.025	0.024	0.025	0.043	0.237	0.143	0.067	0.061	0.058	0.057
Size	0.908	0.015	0.056	0.140	0.057	0.064	0.124	0.081	0.007	0.053	0.075	0.052	0.061	0.064
	$T = 50$													
Bias	-0.026	0.062	-0.036	-0.012	-0.001	-0.003	-0.007	0.017	0.019	-0.003	-0.004	0.002	0.001	0.001
IQR	0.015	0.289	0.075	0.029	0.023	0.022	0.021	0.055	0.267	0.184	0.077	0.066	0.065	0.065
MAE	0.026	0.160	0.044	0.017	0.012	0.011	0.011	0.029	0.134	0.093	0.039	0.033	0.033	0.032
Size	0.664	0.030	0.057	0.089	0.056	0.063	0.080	0.079	0.012	0.060	0.061	0.063	0.061	0.062
	$T = 100$													
Bias	-0.012	0.051	-0.015	-0.005	0.000	-0.001	-0.002	0.009	0.010	0.004	0.000	0.001	0.001	0.002
IQR	0.010	0.218	0.048	0.019	0.013	0.012	0.012	0.039	0.163	0.113	0.053	0.042	0.041	0.040
MAE	0.012	0.117	0.024	0.010	0.006	0.006	0.006	0.020	0.081	0.057	0.026	0.021	0.020	0.020
Size	0.400	0.065	0.058	0.074	0.056	0.059	0.064	0.071	0.012	0.047	0.050	0.057	0.059	0.056

$N = 100$, $\sigma_\eta^2 = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.031	-0.189	-0.078	0.001	-0.033	-0.068	0.006	0.038	-0.169	-0.110	0.009	-0.041	-0.093
IQR	0.034	0.671	0.271	0.097	0.134	0.110	0.093	0.098	0.907	0.515	0.229	0.299	0.254	0.222
MAE	0.150	0.334	0.195	0.080	0.067	0.060	0.072	0.050	0.453	0.270	0.140	0.151	0.126	0.126
Size	1.000	0.001	0.085	0.188	0.044	0.068	0.165	0.054	0.006	0.034	0.091	0.035	0.060	0.090
	$T = 25$													
Bias	-0.054	0.063	-0.071	-0.021	-0.002	-0.006	-0.013	0.022	0.051	-0.030	-0.017	0.002	-0.001	-0.008
IQR	0.016	0.344	0.122	0.039	0.033	0.031	0.029	0.055	0.366	0.237	0.094	0.089	0.084	0.078
MAE	0.054	0.184	0.075	0.025	0.017	0.016	0.017	0.033	0.185	0.119	0.046	0.044	0.041	0.041
Size	0.995	0.012	0.054	0.099	0.044	0.051	0.082	0.083	0.013	0.038	0.067	0.057	0.056	0.059
	$T = 50$													
Bias	-0.025	0.046	-0.034	-0.008	0.000	-0.001	-0.004	0.016	0.022	-0.003	-0.004	0.001	0.001	0.000
IQR	0.011	0.245	0.072	0.024	0.017	0.016	0.016	0.038	0.187	0.138	0.057	0.046	0.045	0.045
MAE	0.025	0.133	0.041	0.013	0.008	0.008	0.008	0.022	0.097	0.068	0.029	0.023	0.023	0.022
Size	0.910	0.032	0.043	0.080	0.060	0.063	0.075	0.085	0.017	0.046	0.053	0.052	0.055	0.052
	$T = 100$													
Bias	-0.012	0.024	-0.016	-0.003	0.000	-0.001	-0.001	0.009	0.006	0.000	-0.001	0.000	0.000	0.000
IQR	0.007	0.177	0.048	0.015	0.009	0.009	0.009	0.028	0.111	0.086	0.039	0.031	0.031	0.031
MAE	0.012	0.091	0.025	0.008	0.004	0.004	0.004	0.015	0.056	0.043	0.020	0.015	0.015	0.015
Size	0.668	0.057	0.050	0.069	0.050	0.046	0.051	0.075	0.019	0.054	0.056	0.051	0.052	0.052

$N = 250$, $\sigma_\eta^2 = 5$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.150	0.039	-0.192	-0.047	0.002	-0.016	-0.037	0.004	0.065	-0.192	-0.064	0.002	-0.024	-0.054
IQR	0.022	0.580	0.248	0.069	0.084	0.077	0.069	0.060	0.742	0.395	0.166	0.179	0.168	0.155
MAE	0.150	0.286	0.197	0.051	0.042	0.040	0.044	0.031	0.378	0.230	0.091	0.090	0.086	0.086
Size	1.000	0.001	0.065	0.125	0.056	0.056	0.108	0.041	0.006	0.032	0.080	0.050	0.052	0.065
	$T = 25$													
Bias	-0.054	0.034	-0.068	-0.011	0.000	-0.002	-0.005	0.021	0.030	-0.036	-0.010	-0.001	-0.003	-0.005
IQR	0.011	0.286	0.117	0.028	0.021	0.020	0.020	0.035	0.253	0.162	0.064	0.055	0.051	0.050
MAE	0.054	0.151	0.075	0.016	0.011	0.010	0.010	0.024	0.133	0.084	0.032	0.028	0.026	0.026
Size	1.000	0.013	0.040	0.086	0.049	0.048	0.062	0.130	0.010	0.031	0.063	0.052	0.049	0.051
	$T = 50$													
Bias	-0.025	0.017	-0.030	-0.004	0.000	-0.001	-0.002	0.015	0.006	-0.010	-0.002	0.000	0.000	0.000
IQR	0.007	0.190	0.072	0.017	0.011	0.010	0.010	0.025	0.123	0.090	0.039	0.030	0.029	0.029
MAE	0.025	0.096	0.039	0.009	0.005	0.005	0.005	0.016	0.062	0.045	0.020	0.015	0.015	0.015
Size	1.000	0.031	0.045	0.060	0.048	0.060	0.057	0.140	0.021	0.035	0.059	0.051	0.052	0.052
	$T = 100$													
Bias	-0.012	0.006	-0.016	-0.001	0.000	0.000	-0.001	0.009	0.004	0.001	0.000	0.000	0.001	0.000
IQR	0.004	0.118	0.045	0.010	0.006	0.006	0.006	0.017	0.064	0.061	0.025	0.019	0.018	0.018
MAE	0.012	0.057	0.025	0.005	0.003	0.003	0.003	0.011	0.032	0.031	0.012	0.009	0.009	0.009
Size	0.958	0.057	0.058	0.060	0.054	0.048	0.051	0.099	0.026	0.045	0.052	0.056	0.055	0.056

Table 2.3: Trend model: $\alpha = 0.4$, $\beta = 1.0$,

$N = 50$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.247	-0.025	-0.071	-0.113	-0.024	-0.129	-0.231	0.009	-0.034	-0.175	-0.178	-0.139	-0.167	-0.170
IQR	0.062	0.636	0.173	0.123	0.510	0.261	0.193	0.160	3.425	0.757	0.465	2.720	0.963	0.613
MAE	0.247	0.321	0.099	0.115	0.259	0.160	0.231	0.080	1.724	0.386	0.260	1.389	0.500	0.331
Size	1.000	0.007	0.083	0.242	0.014	0.117	0.369	0.063	0.003	0.054	0.079	0.003	0.038	0.066
	$T = 25$													
Bias	-0.092	0.022	-0.031	-0.035	-0.002	-0.020	-0.041	0.034	0.075	-0.079	-0.052	0.010	-0.022	-0.031
IQR	0.032	0.737	0.096	0.057	0.092	0.076	0.063	0.089	2.306	0.316	0.182	0.336	0.266	0.201
MAE	0.092	0.367	0.053	0.039	0.046	0.041	0.045	0.049	1.162	0.165	0.096	0.168	0.133	0.102
Size	0.958	0.001	0.067	0.123	0.052	0.073	0.140	0.094	0.001	0.051	0.064	0.050	0.056	0.059
	$T = 50$													
Bias	-0.045	0.046	-0.017	-0.014	0.001	-0.005	-0.012	0.024	0.118	-0.035	-0.012	0.001	-0.002	-0.003
IQR	0.022	0.740	0.072	0.037	0.040	0.038	0.035	0.057	1.811	0.180	0.094	0.113	0.099	0.091
MAE	0.045	0.377	0.037	0.020	0.020	0.019	0.019	0.033	0.907	0.089	0.047	0.057	0.050	0.046
Size	0.758	0.001	0.078	0.090	0.051	0.055	0.086	0.095	0.003	0.065	0.055	0.060	0.056	0.058
	$T = 100$													
Bias	-0.021	0.077	-0.008	-0.006	0.000	-0.002	-0.005	0.012	0.169	-0.016	-0.003	0.001	0.001	0.001
IQR	0.015	0.666	0.052	0.022	0.022	0.021	0.021	0.040	1.359	0.112	0.058	0.059	0.056	0.054
MAE	0.021	0.344	0.026	0.012	0.011	0.010	0.011	0.021	0.717	0.057	0.029	0.029	0.028	0.027
Size	0.456	0.007	0.060	0.071	0.055	0.055	0.063	0.082	0.007	0.060	0.050	0.056	0.050	0.056

$N = 100$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.245	-0.038	-0.054	-0.077	-0.009	-0.087	-0.168	0.013	-0.145	-0.182	-0.190	-0.024	-0.179	-0.198
IQR	0.045	0.605	0.141	0.104	0.373	0.214	0.160	0.120	3.100	0.674	0.455	2.103	0.855	0.565
MAE	0.245	0.303	0.079	0.081	0.186	0.124	0.169	0.061	1.551	0.351	0.253	1.056	0.457	0.321
Size	1.000	0.004	0.061	0.165	0.013	0.077	0.275	0.056	0.003	0.040	0.079	0.004	0.040	0.070
	$T = 25$													
Bias	-0.090	0.025	-0.026	-0.023	0.001	-0.010	-0.024	0.035	0.129	-0.069	-0.040	0.010	-0.008	-0.026
IQR	0.023	0.707	0.085	0.047	0.063	0.056	0.050	0.060	2.173	0.284	0.151	0.235	0.195	0.163
MAE	0.090	0.344	0.047	0.029	0.032	0.029	0.031	0.040	1.076	0.154	0.081	0.118	0.099	0.081
Size	0.999	0.002	0.058	0.083	0.051	0.053	0.095	0.114	0.001	0.055	0.059	0.044	0.052	0.057
	$T = 50$													
Bias	-0.044	0.061	-0.014	-0.008	0.001	-0.002	-0.006	0.023	0.144	-0.033	-0.010	0.000	-0.002	-0.002
IQR	0.016	0.689	0.060	0.029	0.029	0.027	0.026	0.038	1.690	0.152	0.075	0.083	0.077	0.073
MAE	0.044	0.357	0.034	0.016	0.014	0.013	0.014	0.026	0.870	0.084	0.038	0.042	0.039	0.036
Size	0.956	0.001	0.050	0.083	0.058	0.060	0.074	0.113	0.001	0.046	0.052	0.060	0.052	0.054
	$T = 100$													
Bias	-0.022	0.085	-0.008	-0.003	0.000	-0.001	-0.002	0.012	0.187	-0.014	-0.003	0.000	0.000	0.000
IQR	0.011	0.668	0.048	0.018	0.015	0.015	0.014	0.029	1.393	0.102	0.043	0.042	0.039	0.039
MAE	0.022	0.344	0.024	0.009	0.008	0.007	0.007	0.017	0.707	0.050	0.021	0.021	0.019	0.019
Size	0.750	0.003	0.059	0.061	0.058	0.055	0.064	0.097	0.002	0.054	0.052	0.058	0.055	0.054

$N = 250$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.247	-0.019	-0.035	-0.048	-0.002	-0.045	-0.099	0.011	-0.059	-0.144	-0.167	-0.060	-0.150	-0.188
IQR	0.027	0.488	0.115	0.084	0.225	0.161	0.134	0.073	2.367	0.577	0.417	1.323	0.756	0.531
MAE	0.247	0.245	0.063	0.055	0.113	0.083	0.104	0.037	1.179	0.298	0.237	0.664	0.388	0.296
Size	1.000	0.003	0.040	0.113	0.020	0.065	0.198	0.050	0.001	0.035	0.069	0.010	0.025	0.059
	$T = 25$													
Bias	-0.090	0.015	-0.017	-0.013	0.001	-0.004	-0.011	0.034	0.073	-0.053	-0.028	0.006	-0.002	-0.014
IQR	0.014	0.571	0.071	0.037	0.039	0.036	0.035	0.039	1.775	0.247	0.123	0.149	0.129	0.119
MAE	0.090	0.288	0.037	0.020	0.020	0.019	0.020	0.034	0.896	0.125	0.064	0.075	0.064	0.061
Size	1.000	0.001	0.060	0.091	0.050	0.054	0.077	0.221	0.001	0.053	0.066	0.051	0.052	0.053
	$T = 50$													
Bias	-0.044	0.041	-0.010	-0.005	-0.001	-0.002	-0.004	0.022	0.106	-0.024	-0.006	0.000	-0.001	-0.002
IQR	0.010	0.645	0.054	0.019	0.018	0.018	0.017	0.026	1.544	0.140	0.053	0.052	0.050	0.047
MAE	0.044	0.322	0.027	0.010	0.009	0.009	0.009	0.022	0.789	0.069	0.027	0.026	0.025	0.024
Size	1.000	0.001	0.059	0.058	0.052	0.050	0.053	0.217	0.001	0.057	0.051	0.046	0.050	0.049
	$T = 100$													
Bias	-0.021	0.071	-0.007	-0.001	0.000	0.000	-0.001	0.012	0.150	-0.013	-0.001	0.001	0.000	0.001
IQR	0.007	0.675	0.040	0.012	0.010	0.009	0.009	0.018	1.404	0.085	0.028	0.025	0.024	0.025
MAE	0.021	0.341	0.021	0.006	0.005	0.005	0.005	0.013	0.700	0.045	0.014	0.013	0.012	0.012
Size	0.987	0.003	0.063	0.063	0.052	0.054	0.058	0.152	0.003	0.058	0.050	0.051	0.048	0.048

Table 2.3:

Table 2.3(cont.): Trend model: $\alpha = 0.4$, $\beta = 1.0$,

$N = 50$, $\sigma^2_\eta = 5$, $\sigma^2_\lambda = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.247	-0.007	-0.106	-0.115	-0.024	-0.129	-0.231	0.009	0.054	-0.196	-0.190	-0.139	-0.167	-0.170
IQR	0.062	0.632	0.207	0.125	0.510	0.261	0.193	0.160	3.398	0.782	0.472	2.720	0.963	0.613
MAE	0.247	0.320	0.124	0.116	0.259	0.160	0.231	0.080	1.665	0.417	0.269	1.389	0.500	0.331
Size	1.000	0.005	0.084	0.249	0.014	0.117	0.369	0.063	0.002	0.054	0.077	0.003	0.038	0.066
	$T = 25$													
Bias	-0.092	0.025	-0.033	-0.036	-0.002	-0.020	-0.041	0.034	0.084	-0.079	-0.055	0.010	-0.022	-0.031
IQR	0.032	0.717	0.103	0.057	0.092	0.076	0.063	0.089	2.242	0.330	0.191	0.336	0.266	0.201
MAE	0.092	0.362	0.055	0.040	0.046	0.041	0.045	0.049	1.120	0.170	0.102	0.168	0.133	0.102
Size	0.958	0.001	0.067	0.120	0.052	0.073	0.140	0.094	0.001	0.053	0.067	0.050	0.056	0.059
	$T = 50$													
Bias	-0.045	0.046	-0.017	-0.014	0.001	-0.005	-0.012	0.024	0.128	-0.039	-0.014	0.001	-0.002	-0.003
IQR	0.022	0.733	0.075	0.038	0.040	0.038	0.035	0.057	1.801	0.180	0.098	0.113	0.099	0.091
MAE	0.045	0.372	0.038	0.021	0.020	0.019	0.019	0.033	0.910	0.092	0.050	0.057	0.050	0.046
Size	0.758	0.002	0.081	0.094	0.051	0.055	0.086	0.095	0.002	0.066	0.055	0.060	0.056	0.058
	$T = 100$													
Bias	-0.021	0.078	-0.009	-0.006	0.000	-0.002	-0.005	0.012	0.168	-0.018	-0.004	0.001	0.001	0.001
IQR	0.015	0.656	0.053	0.023	0.022	0.021	0.021	0.040	1.349	0.111	0.058	0.059	0.056	0.054
MAE	0.021	0.343	0.026	0.012	0.011	0.010	0.011	0.021	0.706	0.058	0.030	0.029	0.028	0.027
Size	0.456	0.007	0.058	0.072	0.055	0.055	0.063	0.082	0.007	0.062	0.054	0.056	0.050	0.056

$N = 100$, $\sigma^2_\eta = 5$, $\sigma^2_\lambda = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.245	-0.023	-0.078	-0.081	-0.009	-0.087	-0.168	0.013	-0.101	-0.228	-0.195	-0.024	-0.179	-0.198
IQR	0.045	0.632	0.171	0.108	0.373	0.214	0.160	0.120	3.055	0.716	0.466	2.103	0.855	0.565
MAE	0.245	0.320	0.100	0.083	0.186	0.124	0.169	0.061	1.500	0.385	0.261	1.056	0.457	0.321
Size	1.000	0.002	0.067	0.173	0.013	0.077	0.275	0.056	0.003	0.038	0.090	0.004	0.040	0.070
	$T = 25$													
Bias	-0.090	0.029	-0.030	-0.025	0.001	-0.010	-0.024	0.035	0.111	-0.079	-0.050	0.010	-0.008	-0.026
IQR	0.023	0.705	0.095	0.049	0.063	0.056	0.050	0.060	2.207	0.295	0.167	0.235	0.195	0.163
MAE	0.090	0.354	0.052	0.031	0.032	0.029	0.031	0.040	1.086	0.164	0.089	0.118	0.099	0.081
Size	0.999	0.002	0.062	0.092	0.051	0.053	0.095	0.114	0.001	0.057	0.061	0.044	0.052	0.057
	$T = 50$													
Bias	-0.044	0.060	-0.016	-0.009	0.001	-0.002	-0.006	0.023	0.151	-0.035	-0.012	0.000	-0.002	-0.002
IQR	0.016	0.682	0.064	0.031	0.029	0.027	0.026	0.038	1.682	0.157	0.082	0.083	0.077	0.073
MAE	0.044	0.350	0.035	0.016	0.014	0.013	0.014	0.026	0.844	0.083	0.041	0.042	0.039	0.036
Size	0.956	0.002	0.050	0.079	0.058	0.060	0.074	0.113	0.001	0.045	0.055	0.060	0.052	0.054
	$T = 100$													
Bias	-0.022	0.083	-0.009	-0.003	0.000	-0.001	-0.002	0.012	0.182	-0.016	-0.003	0.000	0.000	0.000
IQR	0.011	0.676	0.049	0.018	0.015	0.015	0.014	0.029	1.394	0.102	0.044	0.042	0.039	0.039
MAE	0.022	0.344	0.025	0.009	0.008	0.007	0.007	0.017	0.703	0.051	0.022	0.021	0.019	0.019
Size	0.750	0.003	0.062	0.060	0.058	0.055	0.064	0.097	0.002	0.052	0.051	0.058	0.055	0.054

$N = 250$, $\sigma^2_\eta = 5$, $\sigma^2_\lambda = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.247	-0.009	-0.057	-0.053	-0.002	-0.045	-0.099	0.011	0.012	-0.205	-0.199	-0.060	-0.150	-0.188
IQR	0.027	0.534	0.145	0.090	0.225	0.161	0.134	0.073	2.544	0.670	0.423	1.323	0.756	0.531
MAE	0.247	0.267	0.084	0.059	0.113	0.083	0.104	0.037	1.266	0.354	0.261	0.664	0.388	0.296
Size	1.000	0.001	0.048	0.123	0.020	0.065	0.198	0.050	0.001	0.035	0.075	0.010	0.025	0.059
	$T = 25$													
Bias	-0.090	0.024	-0.023	-0.015	0.001	-0.004	-0.011	0.034	0.094	-0.072	-0.038	0.006	-0.002	-0.014
IQR	0.014	0.572	0.083	0.039	0.039	0.036	0.035	0.039	1.775	0.265	0.142	0.149	0.129	0.119
MAE	0.090	0.288	0.043	0.022	0.020	0.019	0.020	0.034	0.889	0.143	0.076	0.075	0.064	0.061
Size	1.000	0.001	0.063	0.096	0.050	0.054	0.077	0.221	0.001	0.053	0.067	0.051	0.052	0.053
	$T = 50$													
Bias	-0.044	0.043	-0.012	-0.005	-0.001	-0.002	-0.004	0.022	0.112	-0.030	-0.008	0.000	-0.001	-0.002
IQR	0.010	0.643	0.058	0.021	0.018	0.018	0.017	0.026	1.536	0.147	0.058	0.052	0.050	0.047
MAE	0.044	0.320	0.031	0.011	0.009	0.009	0.009	0.022	0.778	0.077	0.030	0.026	0.025	0.024
Size	1.000	0.001	0.060	0.061	0.052	0.050	0.053	0.217	0.002	0.056	0.058	0.046	0.050	0.049
	$T = 100$													
Bias	-0.021	0.073	-0.008	-0.001	0.000	0.000	-0.001	0.012	0.147	-0.016	-0.001	0.001	0.000	0.001
IQR	0.007	0.682	0.043	0.012	0.010	0.009	0.009	0.018	1.412	0.091	0.030	0.025	0.024	0.025
MAE	0.021	0.342	0.022	0.006	0.005	0.005	0.005	0.013	0.702	0.047	0.015	0.013	0.012	0.012
Size	0.987	0.003	0.059	0.063	0.052	0.054	0.058	0.152	0.003	0.056	0.052	0.051	0.048	0.048

Table 2.4: Trend model: $\alpha = 0.8$, $\beta = 1.0$,

$N = 50$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.365	-0.091	-0.390	-0.301	-0.184	-0.370	-0.488	-0.099	0.125	-0.234	-0.379	-0.559	-0.406	-0.355
IQR	0.063	1.181	0.322	0.171	1.251	0.369	0.249	0.153	3.577	0.805	0.467	3.756	0.885	0.583
MAE	0.365	0.569	0.390	0.301	0.671	0.373	0.488	0.109	1.801	0.434	0.386	1.968	0.529	0.405
Size	1.000	0.004	0.275	0.682	0.010	0.303	0.782	0.162	0.002	0.050	0.207	0.004	0.067	0.138
	$T = 25$													
Bias	-0.122	0.008	-0.084	-0.085	0.001	-0.086	-0.106	0.013	0.077	-0.131	-0.133	0.028	-0.156	-0.131
IQR	0.029	0.662	0.110	0.065	0.399	0.113	0.077	0.086	2.043	0.320	0.191	1.011	0.316	0.222
MAE	0.122	0.330	0.085	0.085	0.198	0.091	0.106	0.044	1.024	0.179	0.142	0.507	0.195	0.145
Size	1.000	0.001	0.120	0.422	0.002	0.168	0.492	0.071	0.002	0.064	0.159	0.004	0.107	0.135
	$T = 50$													
Bias	-0.054	0.022	-0.037	-0.036	0.002	-0.025	-0.036	0.024	0.115	-0.070	-0.055	0.002	-0.037	-0.039
IQR	0.017	0.577	0.068	0.039	0.080	0.051	0.039	0.057	1.129	0.165	0.108	0.187	0.132	0.106
MAE	0.054	0.288	0.041	0.036	0.040	0.032	0.037	0.032	0.582	0.091	0.066	0.094	0.070	0.060
Size	0.995	0.001	0.111	0.272	0.034	0.093	0.256	0.094	0.007	0.079	0.122	0.036	0.065	0.090
	$T = 100$													
Bias	-0.025	0.059	-0.019	-0.015	0.000	-0.006	-0.011	0.016	0.113	-0.029	-0.018	0.000	-0.004	-0.006
IQR	0.010	0.447	0.044	0.023	0.027	0.023	0.020	0.039	0.699	0.090	0.059	0.065	0.057	0.056
MAE	0.025	0.225	0.025	0.017	0.013	0.012	0.013	0.023	0.372	0.048	0.032	0.033	0.029	0.028
Size	0.918	0.006	0.088	0.157	0.049	0.072	0.138	0.092	0.007	0.074	0.075	0.057	0.067	0.067

$N = 100$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.362	-0.091	-0.352	-0.250	-0.187	-0.315	-0.428	-0.098	-0.002	-0.318	-0.446	-0.577	-0.498	-0.435
IQR	0.046	1.174	0.316	0.166	1.203	0.324	0.227	0.118	3.279	0.756	0.444	3.752	0.847	0.570
MAE	0.362	0.567	0.352	0.250	0.623	0.320	0.428	0.100	1.640	0.442	0.446	1.928	0.578	0.451
Size	1.000	0.000	0.236	0.572	0.007	0.254	0.735	0.245	0.000	0.052	0.275	0.001	0.086	0.172
	$T = 25$													
Bias	-0.122	-0.009	-0.073	-0.075	0.003	-0.073	-0.091	0.014	0.109	-0.141	-0.149	0.025	-0.148	-0.156
IQR	0.020	0.735	0.103	0.060	0.280	0.113	0.074	0.060	1.923	0.304	0.172	0.755	0.295	0.195
MAE	0.122	0.367	0.076	0.075	0.140	0.079	0.091	0.032	0.971	0.174	0.151	0.375	0.181	0.159
Size	1.000	0.000	0.103	0.350	0.004	0.149	0.413	0.065	0.003	0.074	0.195	0.004	0.099	0.169
	$T = 50$													
Bias	-0.054	0.042	-0.034	-0.032	0.000	-0.015	-0.028	0.023	0.122	-0.072	-0.059	0.004	-0.022	-0.039
IQR	0.012	0.592	0.060	0.036	0.060	0.043	0.034	0.039	1.145	0.145	0.096	0.140	0.108	0.091
MAE	0.054	0.299	0.039	0.032	0.030	0.024	0.028	0.026	0.575	0.090	0.066	0.070	0.056	0.052
Size	1.000	0.001	0.077	0.231	0.050	0.071	0.210	0.116	0.004	0.073	0.137	0.052	0.057	0.101
	$T = 100$													
Bias	-0.025	0.055	-0.016	-0.011	0.000	-0.003	-0.007	0.015	0.104	-0.026	-0.016	-0.001	-0.003	-0.006
IQR	0.007	0.431	0.041	0.020	0.018	0.017	0.015	0.029	0.651	0.080	0.046	0.045	0.042	0.040
MAE	0.025	0.215	0.022	0.013	0.009	0.009	0.009	0.018	0.348	0.042	0.026	0.023	0.021	0.020
Size	0.998	0.003	0.064	0.114	0.051	0.063	0.092	0.126	0.009	0.069	0.074	0.062	0.062	0.062

$N = 250$, $\sigma_\eta^2 = 1$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.365	-0.070	-0.322	-0.211	-0.159	-0.258	-0.354	-0.100	-0.032	-0.405	-0.484	-0.504	-0.566	-0.535
IQR	0.029	0.952	0.296	0.154	1.178	0.317	0.212	0.070	2.496	0.659	0.426	3.560	0.817	0.531
MAE	0.365	0.475	0.322	0.211	0.607	0.270	0.354	0.100	1.249	0.463	0.484	1.858	0.607	0.536
Size	1.000	0.002	0.235	0.499	0.007	0.211	0.620	0.483	0.002	0.071	0.340	0.002	0.105	0.259
	$T = 25$													
Bias	-0.121	-0.011	-0.059	-0.062	-0.004	-0.059	-0.076	0.013	0.076	-0.144	-0.154	0.001	-0.137	-0.159
IQR	0.013	0.713	0.093	0.059	0.176	0.099	0.066	0.038	1.746	0.258	0.176	0.469	0.257	0.174
MAE	0.121	0.355	0.063	0.062	0.088	0.065	0.076	0.021	0.866	0.165	0.155	0.235	0.161	0.160
Size	1.000	0.001	0.088	0.307	0.015	0.128	0.353	0.071	0.003	0.071	0.241	0.015	0.097	0.224
	$T = 50$													
Bias	-0.054	0.034	-0.027	-0.024	-0.001	-0.009	-0.018	0.022	0.095	-0.056	-0.051	-0.001	-0.017	-0.031
IQR	0.008	0.570	0.055	0.030	0.035	0.030	0.026	0.025	1.094	0.129	0.081	0.085	0.074	0.064
MAE	0.054	0.288	0.033	0.025	0.017	0.016	0.019	0.022	0.561	0.073	0.056	0.043	0.038	0.038
Size	1.000	0.001	0.079	0.167	0.043	0.060	0.146	0.215	0.005	0.067	0.139	0.046	0.055	0.090
	$T = 100$													
Bias	-0.025	0.051	-0.014	-0.007	0.000	-0.001	-0.003	0.015	0.096	-0.024	-0.011	0.001	-0.001	-0.003
IQR	0.005	0.458	0.035	0.015	0.012	0.010	0.010	0.018	0.735	0.067	0.035	0.029	0.028	0.027
MAE	0.025	0.240	0.020	0.009	0.006	0.005	0.006	0.015	0.388	0.038	0.019	0.015	0.014	0.014
Size	1.000	0.001	0.064	0.092	0.059	0.050	0.071	0.217	0.007	0.060	0.065	0.050	0.048	0.054

Table 2.4(cont.): Trend model: $\alpha = 0.8$, $\beta = 1.0$,

$N = 50$, $\sigma_\eta^2 = 5$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.365	-0.057	-0.407	-0.302	-0.184	-0.370	-0.488	-0.099	0.106	-0.146	-0.376	-0.559	-0.406	-0.355
IQR	0.063	1.170	0.335	0.173	1.251	0.369	0.249	0.153	3.291	0.832	0.475	3.756	0.885	0.583
MAE	0.365	0.580	0.408	0.302	0.671	0.373	0.488	0.109	1.639	0.421	0.383	1.968	0.529	0.405
Size	1.000	0.004	0.280	0.683	0.010	0.303	0.782	0.162	0.001	0.042	0.198	0.004	0.067	0.138
	$T = 25$													
Bias	-0.122	0.006	-0.103	-0.087	0.001	-0.086	-0.106	0.013	0.097	-0.100	-0.137	0.028	-0.156	-0.131
IQR	0.029	0.644	0.122	0.065	0.399	0.113	0.077	0.086	1.972	0.363	0.191	1.011	0.316	0.222
MAE	0.122	0.323	0.103	0.087	0.198	0.091	0.106	0.044	0.985	0.192	0.146	0.507	0.195	0.145
Size	1.000	0.001	0.134	0.431	0.002	0.168	0.492	0.071	0.002	0.050	0.165	0.004	0.107	0.135
	$T = 50$													
Bias	-0.054	0.022	-0.044	-0.037	0.002	-0.025	-0.036	0.024	0.112	-0.075	-0.061	0.002	-0.037	-0.039
IQR	0.017	0.568	0.070	0.040	0.080	0.051	0.039	0.057	1.165	0.177	0.109	0.187	0.132	0.106
MAE	0.054	0.284	0.046	0.037	0.040	0.032	0.037	0.032	0.593	0.102	0.072	0.094	0.070	0.060
Size	0.995	0.001	0.112	0.279	0.034	0.093	0.256	0.094	0.006	0.073	0.115	0.036	0.065	0.090
	$T = 100$													
Bias	-0.025	0.063	-0.020	-0.015	0.000	-0.006	-0.011	0.016	0.115	-0.032	-0.021	0.000	-0.004	-0.006
IQR	0.010	0.440	0.046	0.023	0.027	0.023	0.020	0.039	0.689	0.093	0.061	0.065	0.057	0.056
MAE	0.025	0.224	0.026	0.017	0.013	0.012	0.013	0.023	0.361	0.051	0.034	0.033	0.029	0.028
Size	0.918	0.006	0.095	0.155	0.049	0.072	0.138	0.092	0.008	0.074	0.073	0.057	0.067	0.067

$N = 100$, $\sigma_\eta^2 = 5$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.362	-0.079	-0.380	-0.249	-0.187	-0.315	-0.428	-0.098	-0.064	-0.212	-0.441	-0.577	-0.498	-0.435
IQR	0.046	1.105	0.332	0.164	1.203	0.324	0.227	0.118	3.218	0.821	0.449	3.752	0.847	0.570
MAE	0.362	0.567	0.380	0.249	0.623	0.320	0.428	0.100	1.608	0.432	0.444	1.928	0.578	0.451
Size	1.000	0.000	0.237	0.569	0.007	0.254	0.735	0.245	0.000	0.032	0.275	0.001	0.086	0.172
	$T = 25$													
Bias	-0.122	-0.007	-0.093	-0.077	0.003	-0.073	-0.091	0.014	0.100	-0.135	-0.159	0.025	-0.148	-0.156
IQR	0.020	0.707	0.118	0.062	0.280	0.113	0.074	0.060	1.870	0.337	0.180	0.755	0.295	0.195
MAE	0.122	0.353	0.094	0.077	0.140	0.079	0.091	0.032	0.955	0.188	0.161	0.375	0.181	0.159
Size	1.000	0.000	0.118	0.372	0.004	0.149	0.413	0.065	0.002	0.062	0.203	0.004	0.099	0.169
	$T = 50$													
Bias	-0.054	0.043	-0.040	-0.033	0.000	-0.015	-0.028	0.023	0.117	-0.081	-0.068	0.004	-0.022	-0.039
IQR	0.012	0.583	0.065	0.037	0.060	0.043	0.034	0.039	1.156	0.163	0.104	0.140	0.108	0.091
MAE	0.054	0.296	0.043	0.034	0.030	0.024	0.028	0.026	0.582	0.099	0.073	0.070	0.056	0.052
Size	1.000	0.001	0.085	0.250	0.050	0.071	0.210	0.116	0.004	0.069	0.153	0.052	0.057	0.101
	$T = 100$													
Bias	-0.025	0.054	-0.018	-0.012	0.000	-0.003	-0.007	0.015	0.106	-0.030	-0.020	-0.001	-0.003	-0.006
IQR	0.007	0.430	0.043	0.021	0.018	0.017	0.015	0.029	0.653	0.084	0.050	0.045	0.042	0.040
MAE	0.025	0.219	0.024	0.013	0.009	0.009	0.009	0.018	0.347	0.045	0.029	0.023	0.021	0.020
Size	0.998	0.003	0.068	0.115	0.051	0.063	0.092	0.126	0.009	0.064	0.077	0.062	0.062	0.062

$N = 250$, $\sigma_\eta^2 = 5$, $\sigma_\lambda^2 = 1$

	α							β						
	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3	FE	IV LEV	GMM LEV1	GMM LEV3	IV BOD	GMM BOD1	GMM BOD3
	$T = 10$													
Bias	-0.365	-0.065	-0.361	-0.213	-0.159	-0.258	-0.354	-0.100	-0.028	-0.311	-0.508	-0.504	-0.566	-0.535
IQR	0.029	0.949	0.318	0.153	1.178	0.317	0.212	0.070	2.545	0.770	0.426	3.560	0.817	0.531
MAE	0.365	0.469	0.361	0.213	0.607	0.270	0.354	0.100	1.269	0.455	0.509	1.858	0.607	0.536
Size	1.000	0.002	0.228	0.500	0.007	0.211	0.620	0.483	0.001	0.042	0.351	0.002	0.105	0.259
	$T = 25$													
Bias	-0.121	0.003	-0.080	-0.068	-0.004	-0.059	-0.076	0.013	0.088	-0.175	-0.175	0.001	-0.137	-0.159
IQR	0.013	0.692	0.111	0.061	0.176	0.099	0.066	0.038	1.698	0.314	0.180	0.469	0.257	0.174
MAE	0.121	0.346	0.082	0.068	0.088	0.065	0.076	0.021	0.845	0.197	0.175	0.235	0.161	0.160
Size	1.000	0.000	0.082	0.335	0.015	0.128	0.353	0.071	0.002	0.070	0.261	0.015	0.097	0.224
	$T = 50$													
Bias	-0.054	0.039	-0.037	-0.028	-0.001	-0.009	-0.018	0.022	0.099	-0.079	-0.064	-0.001	-0.017	-0.031
IQR	0.008	0.572	0.065	0.033	0.035	0.030	0.026	0.025	1.121	0.152	0.092	0.085	0.074	0.064
MAE	0.054	0.287	0.042	0.028	0.017	0.016	0.019	0.022	0.564	0.091	0.067	0.043	0.038	0.038
Size	1.000	0.001	0.079	0.185	0.043	0.060	0.146	0.215	0.005	0.066	0.151	0.046	0.055	0.090
	$T = 100$													
Bias	-0.025	0.050	-0.018	-0.008	0.000	-0.001	-0.003	0.015	0.096	-0.030	-0.013	0.001	-0.001	-0.003
IQR	0.005	0.457	0.040	0.017	0.012	0.010	0.010	0.018	0.739	0.075	0.042	0.029	0.028	0.027
MAE	0.025	0.239	0.024	0.010	0.006	0.005	0.006	0.015	0.386	0.044	0.023	0.015	0.014	0.014
Size	1.000	0.001	0.061	0.104	0.059	0.050	0.071	0.217	0.007	0.058	0.076	0.050	0.048	0.054

Chapter 3

Further Results on the Weak Instruments Problem of the System GMM Estimator in Dynamic Panel Data Models

[1]

In this chapter, we investigate the weak instruments problem of GMM estimator for dynamic panel data models. Bun and Windmeijer [2010] demonstrates that the system GMM estimator combining models in first-differences and levels suffers from the weak instruments problem when the variance ratio of individual effects to the disturbances is large, which is mainly due to the model in levels. In this chapter, we alternatively consider first-difference and level models transformed by forward GLS transformation, and demonstrate that this transformation yields a higher concentration parameter compared with original models. This indicates that this transformation yields stronger instruments. Monte Carlo simulation results reveal that the system GMM estimator for the transformed model, called the forward system GMM estimator, performs better than the conventional system GMM estimator for models in first-differences and levels, and the performance of the new system GMM estimator is reasonably well even when the variance ratio is large.

[1] This is a joint work with Kazuhiko Hayakawa.

3.1 Introduction

Dynamic panel data models have been used in a wide range of economic applications, including labor economics, macro economics, finance and so on. Since the standard fixed effects estimator is inconsistent when T is small and N is large where T and N denote the sample sizes of time series and cross section[Nickell, 1981], alternative estimation methods have been proposed. Among others, an instrumental variables approach have been discussed extensively since the work of Anderson and Hsiao [1981]. Thereafter, the generalized method of moments (GMM) estimators have been proposed by Holtz-Eakin et al. [1988], Arellano and Bond [1991], Arellano and Bover [1995], Ahn and Schmidt [1995], Blundell and Bond [1998] and so on. Holtz-Eakin et al. [1988] and Arellano and Bond [1991] suggest to remove the fixed effect from the model by taking first-differences and estimate those first difference model, denoted as DIF model, using lagged level variables as instruments. Arellano and Bover [1995] suggest to consider level models, denoted as LEV model, and suggest to lagged first-differenced variable as instruments. They also suggest to consider a system model(SYS model) by combining DIF and LEV models, and the corresponding GMM estimator is known as the system GMM estimator, which is now a standard GMM estimator used in empirical studies. Blundell and Bond [1998] and Blundell et al. [2000] investigate the weak instruments problem of these GMM estimators and demonstrated that the GMM estimator for DIF models suffers from the weak instruments problem when the persistency of data is strong and/or the variance ratio of fixed effects to the error is large[2]. They also demonstrate that the the moment condition for LEV models with first-differenced instruments remain informative even when persistency is strong by showing that the reduced form coefficient has non zero probability limit. Bun and Windmeijer [2010] investigate the weak instruments problem in more detail by deriving the concentration parameter which is a formal measure of strength of instruments. They consider DIF and LEV models and demonstrate that the the weak instruments problem arises for both models when persistency is strong and/or the variance ratio of fixed effects to the errors is large. What is important in their result is that while the concentration parameter for DIF model remain positive even when the variance ratio tends to infinity, that for LEV model converges to zero. This implies that the weak instruments problem associated with the large variance ratio is more evident for LEV models than DIF models.

In this chapter, we further investigate the weak instruments problem associated with GMM estimation of dynamic panel data models. Although Bun and Windmeijer [2010] consider DIF and LEV models, we consider alternative models. Specifically, as an alternative to DIF model, we consider models in forward orthogonal deviations(FOD model) suggested by Arellano and Bover [1995], which is obtained by upper triangular GLS

[2]For finite sample properties of these GMM estimators, see Bun and Kiviet [2006], Hayakawa [2007, 2010].

transformation to DIF models[3]. Also, as an alternative to the LEV model, we consider a model introduced by Hayakawa [2010, 2015] where upper triangular GLS transformation is applied to LEV models. We call this transformed model the forward random effect (FRE) model since the OLS estimator of that transformed model becomes the random effect model. In this chapter, we investigate these two alternative FOD and FRE models and derive their concentration parameters. Since it is difficult to derive the concentration parameter in panel regression model, which can be seen as a system, following Blundell and Bond [1998], Blundell et al. [2000], Bun and Windmeijer [2010], we consider a cross-section regression at each period. Consequently, we demonstrate that these two alternative FOD and FRE models yields larger or the same concentration parameters than DIF and LEV models for all periods despite the same instruments are used. Also, it is shown that the concentration parameter for LEV models tends to zero as the variance ratio of fixed effects to the errors gets larger, whereas this is not the case for FRE model; the concentration parameter for the FRE model remains positive even if the variance ratio gets larger. Thus, the upper triangular GLS transformation mitigates the weak instruments problem associated with the large variance ratio for FRE models. This property affects the behavior of the system GMM estimator. Since the system GMM estimator can be written as a weight sum of GMM estimators for DIF and LEV models, it could suffer from the weak instruments problem when the variance ratio is large, which is due to LEV models. However, if the FRE model is used instead of original LEV model in the system, the corresponding system GMM estimator, which is called the *forward system GMM estimator*, is less affected by the variance ratio. We confirm these properties by extensive Monte Carlo simulation and show that the original system GMM estimator for DIF and LEV models are vulnerable to the variance ratio while it is not the case for the forward system GMM estimator for FOD and FRE models. Monte Carlo simulation is carried out to investigate the finite sample behavior of several GMM estimators. Consequently, we find that the forward system GMM estimator combining FOD and FRE models is affected little by the variance ratio and tends to perform best in terms of bias and accuracy of inference.

The rest of chapter is organized as follows. In section 2, we introduce the model and GMM estimators, In section 3, we derive the concentration parameter for four models. In section 4, we conduct a Monte Carlo simulation to investigate the finite sample properties of the GMM estimator and finally in section 5, we conclude. Throughout the paper, we use the notation $T_j = T - j$.

[3]Note that the GMM estimators for DIF and FOD models are numerically equivalent if all past variables are used as instruments[Arellano and Bover, 1995]. For further comparison of DIF and FOD models, see Hayakawa [2009b] and Hsiao and Zhou [2015].

3.2 Model and GMM estimator

We consider an AR(1) dynamic panel data model with fixed effects given by

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (i = 1, 2, \ldots, N; t = 1, 2, \ldots, T) \tag{3.1}$$

where $|\alpha| < 1$, $v_{it} \sim iid(0, \sigma_v^2)$, $\eta_i \sim iid(0, \sigma_\eta^2)$ and v_{it} and η_i are uncorrelated. Further, we assume that the initial conditions satisfy

$$y_{i0} = \frac{\eta_i}{1-\alpha} + w_{i0}, \quad (i = 1, ..., N)$$

where $w_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i,-j}$. With this initial condition, y_{it} becomes a covariance stationary process. These assumptions are used to simplify the theoretical derivation. Indeed, some of them are not necessary for consistency of GMM estimators. For example, we can allow for heteroskedasticity for v_{it} such that $Var(v_{it}) = \sigma_{it}^2$. For more empirically relevant model, see Section 3.4. In a matrix form, (3.1) can be written as

$$\mathbf{y}_i = \alpha \mathbf{y}_{i,-1} + \eta_i \boldsymbol{\iota}_T + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{3.2}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, ..., y_{i,T-1})'$, $\boldsymbol{\iota}_T = (1, ..., 1)'$ and $\mathbf{v}_i = (v_{i1}, ..., v_{iT})'$.

First, we introduce two GMM estimators considered by Bun and Windmeijer [2010] and then introduce alternative GMM estimators.

3.2.1 First difference (DIF) Model

The first-difference model is given by

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it}, \qquad (i = 1, ..., N; t = 2, ..., T) \tag{3.3}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and $\Delta v_{it} = v_{it} - v_{i,t-1}$. Since $y_{is}, (s = 0, ..., t-2)$ are uncorrelated with Δv_{it}, but correlated with $\Delta y_{i,t-1}$, they can be used as instruments. Specifically, we have moment conditions $E(\mathbf{z}_{it}^D \Delta v_{it}) = \mathbf{0}$ for $t = 2, ..., T$ where $\mathbf{z}_{it}^D = (y_{i,t-1-\ell}, \cdots, y_{i,t-2})', (\ell = 1, ..., t-1)$[4]. Note that ℓ denotes the number of instruments used in each period. For example, if $\ell = t-1$, all past variables are used as instruments while if $\ell = \min\{3, t-1\}$, most recent three-lagged variables are used in each period at the maximum.

In a matrix form, taking the first-differences is equivalent to multiplying the matrix

[4] If the model (3.1) contains a global constant such that $y_{it} = \mu + \alpha y_{i,t-1} + \eta_i + v_{it}$, Han and Kim [2014] suggest to include a constant term in the instruments such that $\mathbf{z}_{it}^D = (1, y_{i,t-1-\ell}, \cdots, y_{i,t-2})'$.

$\mathbf{D}_T$ to (3.2) where $T_1 \times T$ matrix $\mathbf{D}_T$ is defined as

$$\mathbf{D}_T = \begin{bmatrix} -1 & 1 & 0 & & & 0 \\ 0 & -1 & 1 & & & \\ & & & \ddots & \ddots & \\ 0 & & & & -1 & 1 \end{bmatrix}.$$

The first-difference model (3.3) can be written in a matrix form as follows:

$$\Delta\mathbf{y}_i = \alpha\Delta\mathbf{y}_{i,-1} + \Delta\mathbf{v}_i, \qquad (i = 1, ..., N)$$

where $\Delta\mathbf{y}_i = \mathbf{D}_T\mathbf{y}_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, $\Delta\mathbf{y}_{i,-1} = \mathbf{D}_T\mathbf{y}_{i,-1} = (\Delta y_{i1}, ..., \Delta y_{i,T-1})'$, and $\Delta\mathbf{v}_i = \mathbf{D}_T\mathbf{v}_i = (\Delta v_{i2}, ..., \Delta v_{iT})'$. The moment conditions can be written as

$$E(\mathbf{Z}_i^{D\prime}\Delta\mathbf{v}_i) = \mathbf{0}$$

where $\mathbf{Z}_i^D = \text{diag}(\mathbf{z}_{i2}^{D\prime}, ..., \mathbf{z}_{iT}^{D\prime})$. The GMM estimator using $\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{Z}_i^D\right)^{-1}$ as the weighting matrix is given by

$$\widetilde{\alpha}^D = \frac{\left(\sum_{i=1}^N \Delta\mathbf{y}_{i,-1}'\mathbf{Z}_i^D\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{Z}_i^D\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\Delta\mathbf{y}_i\right)}{\left(\sum_{i=1}^N \Delta\mathbf{y}_{i,-1}'\mathbf{Z}_i^D\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{Z}_i^D\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\Delta\mathbf{y}_{i,-1}\right)}. \tag{3.4}$$

Similarly, the GMM estimator using $\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i^D\right)^{-1}$ as the weighting matrix is given by

$$\widehat{\alpha}^D = \frac{\left(\sum_{i=1}^N \Delta\mathbf{y}_{i,-1}'\mathbf{Z}_i^D\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i^D\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\Delta\mathbf{y}_i\right)}{\left(\sum_{i=1}^N \Delta\mathbf{y}_{i,-1}'\mathbf{Z}_i^D\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i^D\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{D\prime}\Delta\mathbf{y}_{i,-1}\right)} \tag{3.5}$$

where $\mathbf{D}_T\mathbf{D}_T'$ is a matrix with 2's on the main diagonal, -1's on the first sub-diagonal and 0's otherwise. Since $Var(\mathbf{Z}_i^{D\prime}\Delta\mathbf{v}_i) = \sigma_v^2 E(\mathbf{Z}_i^{D\prime}\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i^D)$, $\widehat{\alpha}^D$ is an asymptotically efficient one-step GMM estimator.

3.2.2 Level (LEV) model

We consider model in levels

$$y_{it} = \alpha y_{i,t-1} + u_{it}, \qquad (i = 1, ..., N; t = 2, ..., T) \tag{3.6}$$

where $u_{it} = \eta_i + v_{it}$. Under the assumption of mean-stationarity, since $\Delta y_{is}, (s = 1, ..., t-1)$ are uncorrelated with $u_{it} = \eta_i + v_{it}$, but correlated with $y_{i,t-1}$, they can be used as instruments. Specifically, we have moment conditions $E(\mathbf{z}_{it}^L u_{it}) = \mathbf{0}$ where $\mathbf{z}_{it}^L = (\Delta y_{i,t-\ell}, \cdots, \Delta y_{i,t-1})', (\ell = 1, ..., t-1)$.

In a matrix form, the model (3.6) can be written as

$$\dot{\mathbf{y}}_i = \alpha \dot{\mathbf{y}}_{i,-1} + \dot{\mathbf{u}}_i, \qquad (i = 1, ..., N) \tag{3.7}$$

where $\dot{\mathbf{y}}_i = \mathbf{L}_T \mathbf{y}_i = (y_{i2}, ..., y_{iT})'$, $\dot{\mathbf{y}}_{i,-1} = \mathbf{L}_T \mathbf{y}_{i,-1} = (y_{i1}, ..., y_{i,T-1})'$, $\dot{\mathbf{u}}_i = \mathbf{L}_T \mathbf{u}_i = (u_{i2}, ..., u_{iT})'$ and $\mathbf{L}_T = (\mathbf{0}_T, \mathbf{I}_{T_1})$ with $\mathbf{0}_T = (0, ..., 0)'$. The moment conditions can be written as

$$E(\mathbf{Z}_i^{L\prime} \dot{\mathbf{u}}_i) = \mathbf{0}$$

where $\mathbf{Z}_i^L = \text{diag}(\mathbf{z}_{i2}^{L\prime}, ..., \mathbf{z}_{iT}^{L\prime})$. The GMM estimator using $\left(\sum_{i=1}^N \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L\right)^{-1}$ as the weighting matrix is given by

$$\widetilde{\alpha}^L = \frac{\left(\sum_{i=1}^N \dot{\mathbf{y}}_{i,-1}' \mathbf{Z}_i^L\right) \left(\sum_{i=1}^N \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L\right)^{-1} \left(\sum_{i=1}^N \mathbf{Z}_i^{L\prime} \dot{\mathbf{y}}_i\right)}{\left(\sum_{i=1}^N \dot{\mathbf{y}}_{i,-1}' \mathbf{Z}_i^L\right) \left(\sum_{i=1}^N \mathbf{Z}_i^{L\prime} \mathbf{Z}_i^L\right)^{-1} \left(\sum_{i=1}^N \mathbf{Z}_i^{L\prime} \dot{\mathbf{y}}_{i,-1}\right)} \tag{3.8}$$

Note that different from first difference model, there is no asymptotically efficient one-step GMM estimator.

The GMM estimators $\widetilde{\alpha}^D$ and $\widetilde{\alpha}^L$ are considered in Bun and Windmeijer [2010]. Next, we consider alternative models that are variants of DIF and LEV models..

3.2.3 Forward orthogonal deviation (FOD) Model

In DIF model, since $Var(\mathbf{D}_T \mathbf{v}_i) = \sigma_v^2 \mathbf{D}_T \mathbf{D}_T'$, the error in DIF model is serially correlated. To correct for the serial correlation, we use the following transformation matrix, which is a GLS transformation in DIF model:

$$\mathbf{F}_T = \left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2} \mathbf{D}_T,$$

where $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2}$ is the *upper* triangular Cholesky factorization of $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1}$. The specific form of $\mathbf{F}_T, (T_1 \times T)$ is given by the following:[5]

$$\mathbf{F}_T = \begin{bmatrix} \sqrt{\frac{T_1}{T}} & & & \mathbf{O} \\ & \sqrt{\frac{T_2}{T_1}} & & \\ & & \ddots & \\ \mathbf{O} & & & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

Note that $\mathbf{F}_T' \mathbf{F}_T = \mathbf{I}_T - \boldsymbol{\iota}_T \boldsymbol{\iota}_T'/T$ and $\mathbf{F}_T \mathbf{F}_T' = \mathbf{I}_{T_1}$. Multiplying $\mathbf{F}_T$ to (3.2), we have

$$\mathbf{y}_i^* = \alpha \mathbf{y}_{i,-1}^* + \mathbf{v}_i^*, \qquad (i = 1, ..., N) \tag{3.9}$$

[5] This form is derived in Arellano and Bover [1995].

where $\mathbf{y}_i^* = \mathbf{F}_T\mathbf{y}_i = (y_{i1}^*, ..., y_{i,T-1}^*)'$, $\mathbf{y}_{i,-1}^* = \mathbf{F}_T\mathbf{y}_{i,-1} = (y_{i0}^*, ..., y_{i,T-2}^*)'$, $\mathbf{v}_i^* = \mathbf{F}_T\mathbf{v}_i = (v_{i1}^*, ..., v_{i,T-1}^*)'$ and

$$\begin{aligned} y_{it}^* &= c_t\left[y_{it} - \frac{y_{i,t+1} + \cdots + y_{iT}}{T-t}\right], \qquad y_{i,t-1}^* = c_t\left[y_{i,t-1} - \frac{y_{it} + \cdots + y_{i,T-1}}{T-t}\right], \\ v_{it}^* &= c_t\left[v_{it} - \frac{v_{i,t+1} + \cdots + v_{iT}}{T-t}\right], \qquad c_t = \sqrt{\frac{T-t}{T-t+1}}. \end{aligned}$$

The tth row of (3.9) can be written as

$$y_{it}^* = \alpha y_{i,t-1}^* + v_{it}^*, \qquad (i = 1, ..., N;\ t = 1, ..., T-1). \tag{3.10}$$

Note that if $Var(v_{it}) = \sigma_v^2$ and $Cov(v_{it}, v_{is}) = 0$ for $s \neq t$, then $Var(v_{it}^*) = \sigma_v^2$ and $Cov(v_{it}^*, v_{is}^*) = 0$ for $t \neq s$. Note also that a pooled OLS estimator of (3.10) becomes identical to the fixed effects estimator.

Since $y_{is}, (s = 0, ..., t-1)$ are uncorrelated with v_{it}^*, but correlated with $y_{i,t-1}^*$, they can be used as instruments. Specifically, we have moment conditions $E(\mathbf{z}_{it}^F v_{it}^*) = \mathbf{0}$ for $t = 1, ..., T-1$ where $\mathbf{z}_{it}^F = (y_{i,t-\ell}, \cdots, y_{i,t-1})', (\ell = 1, ..., t)$. In a matrix form, the moment conditions can be written as

$$E(\mathbf{Z}_i^{F\prime}\mathbf{v}_i^*) = \mathbf{0} \tag{3.11}$$

where $\mathbf{Z}_i^F = \mathrm{diag}(\mathbf{z}_{i1}^{F\prime}, ..., \mathbf{z}_{i,T-1}^{F\prime})$. The GMM estimator using $\left(\sum_{i=1}^N \mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F\right)^{-1}$ as the weighting matrix is given by

$$\widetilde{\alpha}^F = \frac{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime}\mathbf{Z}_i^F\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{F\prime}\mathbf{y}_i^*\right)}{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime}\mathbf{Z}_i^F\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{F\prime}\mathbf{y}_{i,-1}^*\right)}. \tag{3.12}$$

Note that since $Var(\mathbf{Z}_i^{F\prime}\mathbf{v}_i^*) = \sigma_v^2 E(\mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F)$, $\widehat{\alpha}^F$ is an asymptotically efficient one-step GMM estimator when v_{it} is *iid*.

3.2.4 Forward random effect (FRE) model

The FOD model above is obtained by upper GLS transformation for DIF models to account for serial correlation in Δv_{it}. The same strategy can be applied to LEV models. Since $Var(\dot{\mathbf{u}}_i) = \sigma_\eta^2 \boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}' + \sigma_v^2 \mathbf{I}_{T_1} = \sigma_v^2 \boldsymbol{\Sigma}_{u,T_1}$ where $\boldsymbol{\Sigma}_{u,T_1} = r\boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}' + \mathbf{I}_{T_1}$ and $r = \sigma_\eta^2/\sigma_v^2$, the error term in LEV model $\dot{\mathbf{u}}_i$ is serially correlated . To correct for this serial correlation, we can use the GLS principle by multiplying $\boldsymbol{\Sigma}_{u,T_1}^{-1/2}$ to the model (3.7), which is used to obtain so-called the random effects estimator. However, although an alternative expression of $\boldsymbol{\Sigma}_{T_1}^{-1/2}$ is derived by Wansbeek and Kapteyn [1982, 1983], it cannot be used in this context, because it is not upper triangular and past variables cannot be used as instruments. Alternatively, we can use an upper triangular form of

$\boldsymbol{\Sigma}_{u,T_1}^{-1/2}$ derived in Hayakawa [2010, 2015]. The specific form of $\boldsymbol{\Sigma}_{u,T_1}^{-1/2}, (T_1 \times T_1)$ is given by

$$\begin{aligned}
\mathbf{R}_{T_1} &= \boldsymbol{\Sigma}_{u,T_1}^{-1/2} \\
&= \operatorname{diag}\left(\sqrt{\frac{T_1-1+\frac{1}{r}}{T_1+\frac{1}{r}}}, \sqrt{\frac{T_1-2+\frac{1}{r}}{T_1-1+\frac{1}{r}}}, ..., \sqrt{\frac{1+\frac{1}{r}}{2+\frac{1}{r}}}, \sqrt{\frac{\frac{1}{r}}{1+\frac{1}{r}}}\right) \\
&\times \begin{bmatrix} 1 & \frac{-1}{T_1-1+\frac{1}{r}} & \frac{-1}{T_1-1+\frac{1}{r}} & & \cdots & \frac{-1}{T_1-1+\frac{1}{r}} \\ 0 & 1 & \frac{-1}{T_1-2+\frac{1}{r}} & & \cdots & \frac{-1}{T_1-2+\frac{1}{r}} \\ 0 & 0 & 1 & & & \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ & & & & 1 & \frac{-1}{1+\frac{1}{r}} \\ 0 & 0 & & \cdots & 0 & 1 \end{bmatrix}.
\end{aligned} \tag{3.13}$$

By multiplying $\mathbf{R}_{T_1}$ given by (3.13) to the model (3.7), we have

$$\mathbf{y}_i^+ = \alpha \mathbf{y}_{i,-1}^+ + \mathbf{u}_i^+ \qquad (i = 1, ..., N) \tag{3.14}$$

where $\mathbf{y}_i^+ = \mathbf{R}_{T_1}\dot{\mathbf{y}}_i = (y_{i2}^+, ..., y_{iT}^+)'$, $\mathbf{y}_{i,-1}^+ = \mathbf{R}_{T_1}\dot{\mathbf{y}}_{i,-1} = (y_{i1}^+, ..., y_{i,T-1}^+)'$, $\mathbf{u}_i^+ = \mathbf{R}_{T_1}\dot{\mathbf{u}}_i = (u_{i2}^+, ..., u_{iT}^+)'$. Note that due to the upper triangular nature of $\mathbf{R}_{T_1}$, u_{it}^+ is composed with future errors $u_{is}, (s = t+1, ..., T)$ and past errors $u_{is}, (s = 1, ..., t-1)$ are not included. This structure is particularly important when estimating with past variables as instruments.

The tth row of (3.14) can be written as[6]

$$y_{it}^+ = \alpha y_{i,t-1}^+ + u_{it}^+, \qquad (i = 1, ..., N;\ t = 2, ..., T) \tag{3.15}$$

where

$$\begin{aligned}
y_{it}^+ &= \begin{cases} d_t\left(y_{it} - \frac{y_{i,t+1}+\cdots+y_{iT}}{T-t+\frac{1}{r}}\right) & (i = 1, ..., N;\ t = 2, ..., T-1) \\ d_T y_{iT} & (i = 1, ..., N;\ t = T), \end{cases} \\
y_{i,t-1}^+ &= \begin{cases} d_t\left(y_{i,t-1} - \frac{y_{it}+\cdots+y_{i,T-1}}{T-t+\frac{1}{r}}\right) & (i = 1, ..., N;\ t = 2, ..., T-1) \\ d_T y_{i,T-1} & (i = 1, ..., N;\ t = T), \end{cases} \\
u_{it}^+ &= \begin{cases} d_t\left(u_{it} - \frac{u_{i,t+1}+\cdots+u_{iT}}{T-t+\frac{1}{r}}\right) = k_t\eta_i + v_{it}^+ & (i = 1, ..., N;\ t = 2, ..., T-1) \\ d_T u_{iT} & (i = 1, ..., N;\ t = T), \end{cases}
\end{aligned}$$

[6] See Hayakawa [2015]

where

$$d_t^2 = \frac{T-t+\frac{1}{r}}{T-t+1+\frac{1}{r}}, \quad k_t = \frac{d_t}{r(T-t+\frac{1}{r})} = \frac{1}{r\sqrt{T-t+1+\frac{1}{r}}\sqrt{T-t+\frac{1}{r}}},$$

$$v_{it}^{+} = d_t\left(v_{it} - \frac{v_{i,t+1}+\cdots+v_{iT}}{T-t+\frac{1}{r}}\right).$$

Note that if $Var(v_{it}) = \sigma_v^2$ and $Cov(v_{it}, v_{is}) = 0$ for $s \neq t$, then $Var(u_{it}^{+}) = \sigma_v^2$ and $Cov(u_{it}^{+}, u_{is}^{+}) = 0$ for $t \neq s$. Since a pooled OLS estimator of (3.15) becomes identical to random effect estimator, we call the model (3.15) the *forward random effect* model. Under the assumption of mean-stationarity, since $\Delta y_{is}, (s = 1, ..., t-1)$ are uncorrelated with u_{it}^{+}, but correlated with $y_{i,t-1}$, they can be used as instruments. Specifically, we have moment conditions $E(\mathbf{z}_{it}^{R} u_{it}^{+}) = \mathbf{0}$ where $\mathbf{z}_{it}^{R} = (\Delta y_{i,t-\ell}, \cdots, \Delta y_{i,t-1})', (\ell = 1, ..., t-1)$.

In a matrix form, the moment conditions can be written as

$$E(\mathbf{Z}_i^{R\prime} \mathbf{u}_i^{+}) = \mathbf{0}$$

where $\mathbf{Z}_i^R = \text{diag}(\mathbf{z}_{i2}^{R\prime}, ..., \mathbf{z}_{iT}^{R\prime})$. The GMM estimator using $\left(\sum_{i=1}^{N} \mathbf{Z}_i^{R\prime} \mathbf{Z}_i^{R}\right)^{-1}$ as the weighting matrix is given by

$$\widetilde{\alpha}^R = \frac{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{+\prime} \mathbf{Z}_i^R\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{R\prime}\mathbf{Z}_i^R\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{R\prime}\mathbf{y}_i^{+}\right)}{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{+\prime} \mathbf{Z}_i^R\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{R\prime}\mathbf{Z}_i^R\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{R\prime}\mathbf{y}_{i,-1}^{+}\right)}. \tag{3.16}$$

Note that the GMM estimator $\widetilde{\alpha}^R$ is infeasible since the transformation matrix $\mathbf{R}_{T_1}$ contains unknown parameter. A feasible estimator that allows for heteroskedasticity in v_{it} is considered in Section 3.4.

3.2.5 Unified model

To provide a systematic analysis, we reformulate the above four models (3.3),(3.6), (3.9) and (3.15) in a unified model. For $t = 2, ..., T$, let $(\widetilde{y}_{it}, \widetilde{y}_{i,t-1}, \widetilde{u}_{it}, \widetilde{\mathbf{z}}_{it})$ denote $(\Delta y_{it}, \Delta y_{i,t-1}, \Delta u_{it}, \mathbf{z}_{it}^{D})$ for DIF model, $(\dot{y}_{it}, \dot{y}_{i,t-1}, \dot{u}_{it}, \mathbf{z}_{it}^{L})$ for LEV model, $(y_{i,t-1}^{*}, y_{i,t-2}^{*}, u_{i,t-1}^{*}, \mathbf{z}_{i,t-1}^{F})$ for FOD model, and $(y_{it}^{+}, y_{i,t-1}^{+}, u_{it}^{+}, \mathbf{z}_{it}^{R})$ for FRE model, respectively. The cross section regression for the unified model at the tth period is given by

$$\widetilde{y}_{it} = \alpha \widetilde{y}_{i,t-1} + \widetilde{v}_{it}, \qquad (i = 1, ..., N) \tag{3.17}$$

with the moment condition $E(\widetilde{\mathbf{z}}_{it}\widetilde{u}_{it}) = \mathbf{0}$. In a matrix form, we have $E(\widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{u}}_i) = \mathbf{0}$. The GMM estimators (3.4), (3.8), (3.12) and (3.16) can be written as

$$\widetilde{\alpha} = \frac{\left(\sum_{i=1}^N \widetilde{\mathbf{y}}_{i,-1}'\widetilde{\mathbf{Z}}_i\right)\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{Z}}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{y}}_i^+\right)}{\left(\sum_{i=1}^N \widetilde{\mathbf{y}}_{i,-1}'\widetilde{\mathbf{Z}}_i\right)\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{Z}}_i\right)^{-1}\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{y}}_{i,-1}\right)}.$$

Note that, using the fact that $\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{Z}}_i\right)$ is a block-diagonal matrix, $\widetilde{\alpha}$ can be rewritten as

$$\widetilde{\alpha} = \frac{\sum_{t=2}^T \widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_t}{\sum_{t=2}^T \widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_{t-1}} = \sum_{t=2}^T w_t\widetilde{\alpha}_t$$

where $\widetilde{\mathbf{y}}_{t-1} = (\widetilde{y}_{1,t-1}, ..., \widetilde{y}_{N,t-1})'$, $\widetilde{\mathbf{Z}}_t = (\widetilde{\mathbf{z}}_{1t}, ..., \widetilde{\mathbf{z}}_{Nt})'$ and $\widetilde{\mathbf{y}}_t = (\widetilde{y}_{1t}, ..., \widetilde{y}_{Nt})'$,

$$\widetilde{\alpha}_t = \frac{\widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_t}{\widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_{t-1}}$$

is a cross-section 2SLS estimator at period t, and

$$w_t = \frac{\widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_{t-1}}{\sum_{t=2}^T \widetilde{\mathbf{y}}_{t-1}'\widetilde{\mathbf{Z}}_t(\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{Z}}_t)^{-1}\widetilde{\mathbf{Z}}_t'\widetilde{\mathbf{y}}_{t-1}}. \tag{3.18}$$

Thus, we can interpret that the GMM estimator $\widetilde{\alpha}$ is weighted average of cross-section two-stage least squares (2SLS) estimator at each period with weight w_t.

3.3 Concentration Parameter

We now assess the strength of instruments in the above four models by deriving the concentration parameter(CP). Since it is difficult to derive the CPs in panel regression model, we consider a cross-section regression at tth period. The same strategy is used in Bun and Windmeijer [2010] as a measure to strength of instrument.

The two-stage cross section regression for the unified model at the tth period is given by

$$\begin{aligned} \widetilde{y}_{it} &= \alpha\widetilde{y}_{i,t-1} + \widetilde{v}_{it}, \qquad (i = 1, ..., N), \\ \widetilde{y}_{i,t-1} &= \boldsymbol{\pi}_t'\widetilde{\mathbf{z}}_{it} + r_{it}. \end{aligned}$$

Then, using $\boldsymbol{\pi}_t = \left[E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\right]^{-1} E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)$ and $Var(r_{it}) = E\left[\left(\widetilde{y}_{i,t-1} - \widetilde{\mathbf{z}}_{it}'\boldsymbol{\pi}_t\right)^2\right]$, the expected concentration parameter is given by

$$CP_t = \frac{\boldsymbol{\pi}_t' E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\boldsymbol{\pi}_t}{Var(r_{it})} = \frac{E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)'\left[E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\right]^{-1}E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)}{E\left(\widetilde{y}_{i,t-1}^2\right) - E\left(\widetilde{\mathbf{z}}_{it}'\widetilde{y}_{i,t-1}\right)'\left[E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\right]^{-1}E\left(\widetilde{\mathbf{z}}_{it}'\widetilde{y}_{i,t-1}\right)}.$$

The following Theorem 1 is the concentration parameter for DIF and LEV models at tth period.

Theorem 6. *The concentration parameter for cross-section regression of DIF model denoted by CP_t^D and LEV model denoted by CP_t^L at period t are given by*

$$CP_t^D = \frac{(1-\alpha)^2[1+r(\ell-1)]}{(1-\alpha^2)+r(1+\alpha)[(\ell+1)-(\ell-1)\alpha]}, \quad (t=2,..,T), \tag{3.19}$$

$$CP_t^L = \frac{(1-\alpha)^2\ell}{(1-\alpha^2)+r(1+\alpha)[(\ell+1)-(\ell-1)\alpha]}, \quad (t=2,..,T). \tag{3.20}$$

This Theorem extends the analysis of Bun and Windmeijer [2010] in such a way that a particular choice of lag length of instruments ℓ is allowed.

The following Theorem 2 is the concentration parameter for FOD and FRE models.

Theorem 7. *The concentration parameter for cross-section regression of FOD model, denoted by CP_t^F, and FRE model, denoted by CP_t^R, at period t are given by*

$$CP_t^F = \frac{\psi_{t-1}^2(1-\alpha)^2[1+r(\ell-1)]}{\left[\psi_{t-1}^2 r+\varphi_{t-1}(1-r)\right](1-\alpha^2)+\varphi_{t-1}r(1+\alpha)[(\ell+1)-(\ell-1)\alpha]}, \quad (t=2,...,T) \tag{3.21}$$

$$CP_t^R = \begin{cases} \dfrac{(1-\alpha)^2\ell}{\left[\frac{(1-\alpha)}{1+\alpha}+\frac{\varphi_t+rq_t^2}{(\psi_t+q_t)^2}\right](1+\alpha)[\ell+1-(\ell-1)\alpha]-\ell(1-\alpha)^2} & (t=2,...,T-1) \\ \dfrac{(1-\alpha)^2\ell}{\left(\frac{(1-\alpha)}{1+\alpha}+r\right)(1+\alpha)[\ell+1-(\ell-1)\alpha]-(1-\alpha)^2\ell} & (t=T) \end{cases} \tag{3.22}$$

where

$$\psi_{t-1} = 1-\frac{\alpha\phi_{T-t+1}}{T-t+1}, \quad \phi_j = \frac{1-\alpha^j}{1-\alpha}, \quad \varphi_{t-1} = \frac{1}{(T-t+1)^2}\sum_{j=1}^{T-t+1}(1-\alpha^j)^2, \quad q_t = \frac{1}{r(T-t)} \tag{3.23}$$

From these theorems, we have several remarks. First, we find that $CP_t^D = CP_t^F$ and $CP_t^L = CP_t^R$ for $t=T$ regardless of r. We also find that $CP_t^D = CP_t^L$ for all t when $r=1$, which is consistent with Bun and Windmeijer [2010]. Second, as documented in Bun and Windmeijer [2010], one of the most important implications for DIF and LEV models is the asymmetric effect of r. In DIF model, even if r gets larger CP_t^D does not converges to zero. Hence even if r is large, the DIF model keeps the strength of instruments to some extent. However, this is not the case for LEV model. From (3.20), it is found that as r gets larger, CP_t^L tends to zero, which indicates that the instruments are very weak when r is large. This led Bun and Windmeijer [2010] to conclude that the system GMM estimator which is a linear combination of DIF and LEV model suffers from the weak instruments problem when r is large. However, for FRE models, from (3.22), we find that the CP_t^R does not tend to zero even when r gets larger, and the FRE model keeps the strength of instruments to some extent even when r is large. Thus, the FRE model is robust to large r compared with the LEV model in terms of strength of instruments and expected to perform better in practice. The reason behind this results lies in the behavior of $E(\widetilde{y}_{i,t-1}^2)$ that appears in the denominator of the concentration parameter. In LEV models, we have $\widetilde{y}_{i,t-1} = y_{i,t-1}$ and $E(\widetilde{y}_{i,t-1}^2) = E(y_{i,t-1}^2) = [r/(1-\alpha)^2 + 1/(1-\alpha^2)]\sigma_v^2$ diverges as r gets larger(with σ_v^2 being fixed). However, for the FRE model, since $\widetilde{y}_{i,t-1} = y_{i,t-1}^+$ and $E(\widetilde{y}_{i,t-1}^2) = E[(y_{i,t-1}^+)^2]$ has the form given in (3.60), $E(\widetilde{y}_{i,t-1}^2)$ does not diverge even when r gets larger. This result can also be explained intuitively in terms of goodness of fit in the first stage regression. To see this, let us define the population R^2 in the first stage regression as follows

$$R_p^2 = \frac{Var(\boldsymbol{\pi}_t'\widetilde{\mathbf{z}}_{it})}{Var(\widetilde{y}_{i,t-1})} = \frac{\boldsymbol{\pi}_t'E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\boldsymbol{\pi}_t}{Var(\widetilde{y}_{i,t-1})} = \frac{E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)'\left[E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\right]^{-1}E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)}{E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right)'\left[E\left(\widetilde{\mathbf{z}}_{it}\widetilde{\mathbf{z}}_{it}'\right)\right]^{-1}E\left(\widetilde{\mathbf{z}}_{it}\widetilde{y}_{i,t-1}\right) + Var(r_{it})}.$$

Then, the concentration parameter can be written as $CP_t = R_p^2/(1-R_p^2)$. Hence, a better goodness of fit leads to stronger instruments. Figure 1 depicts the scatter plot for the first stage regression of LEV and FRE models with $(t,T) = (4,6)$ for $\alpha = 0.2, 0.8$ and $r = 1, 5$. From this figure, we can see that the FRE model has substantially better fit than the DIF model especially when r is large.

Next, we compare the magnitude of four concentration parameters. However, since the analytical form of concentration parameters are complicated, we draw some figures for CP_t^D, CP_t^L, CP_t^F and CP_t^F. Figures 2-8 depict the concentration parameters (3.19), (3.20), (3.21) and (3.22) and weight (3.18) for $\alpha = 0.0, 0.1, 0.2..., 0.9, 0.95$, $r = 0.2, 0.4, ..., 5$, and $T = 6$[7]. From Figures 2 to 5, we find that the CPs become smaller as α approaches one or the variance ratio r gets larger. This tendency is well known in the literature[e.g. Blundell and Bond, 1998, Bun and Windmeijer, 2010]. Also, we can confirm the tendency visually that $CP_t^L \to 0$ as r increases while it is not case for CP_t^D, CP_t^F and CP_t^R. When $t = T$, or $r = 1$, the CPs of DIF and FOD models, and LEV and FRE models are identical. However, for other cases, we find that the CPs of FOD and FRE models are larger than those of DIF and LEV models, respectively. To see this more precisely, we provide Figures 6 to 8 that depicts the CPs and weights at each period t. From these, we find that the CPs of FOD and FRE models are larger than those of DIF and LEV models for all periods except for the last period where the CPs are identical. This means that the instruments of FOD and FRE models are stronger than those of DIF and LEV model. Also, comparing the shape of CPs and weights, it

[7] In a supplement, figures with $T = 12$ are provided.

is observed that the tendency of CPs and weights is very similar. For example, when $r = 5$ and $\alpha = 0.2$, the CP of FOD model becomes largest around $t = 3$ or $t = 4$ and becomes smallest when $t = 1$. This pattern also applies to the weight. The weight becomes largest when $t = 4$ and becomes smallest when $t = 1$. This means that the GMM estimator is constructed in such a way that cross-sectional 2SLS estimator with strong instruments is assigned with a large weight and that with a weak instruments are assigned with a small weight. Note that such a pattern generally applies for other (but not all) cases.

The above analysis is based on cross section regression at each period and limited to a simple AR(1) model. Although it is desirable to consider in a panel regression models with covariates besides the lagged dependent variable, it is not trivial to derive the concentration parameter in such a general context[8]. Therefore, we conduct an extensive Monte Carlo simulation to investigate the performance of GMM estimators.

3.4 GMM estimators for models with covariates

In this section, we introduce the GMM estimators for dynamic panel data models with covariates, which are used in Monte Carlo study in the next section.

Let us consider the following dynamic panel data model with covariates[9]:

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + \eta_i + v_{it} = \mathbf{w}'_{it}\boldsymbol{\delta} + \eta_i + v_{it}. \qquad (i = 1, ...N; t = 1, ..., T) \tag{3.24}$$

where $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}'_{it})'$ and $\boldsymbol{\delta} = (\alpha, \boldsymbol{\beta}')'$. We assume that $|\alpha| < 1$, $\eta_i \sim iid(0, \sigma^2_\eta)$ and the errors satisfy $E(v_{it}) = 0$, $Var(v_{it}) = \sigma^2_{it}$ and are serially and cross-sectionally uncorrelated. The covariate $\mathbf{x}_{it}$ can be either strictly exogenous, weakly exogenous or endogenous. Here, we assume that $\mathbf{x}_{it}$ is endogenous in the sense that $E(\mathbf{x}_{it} v_{is}) \neq \mathbf{0}$ for $t \geq s$ and $E(\mathbf{x}_{it} v_{is}) = \mathbf{0}$ for $t < s$[10]. Also, we assume that $E(y_{it}\eta_i)$ and $E(\mathbf{x}_{it}\eta_i)$ are constant over time when considering the GMM estimator for LEV, FRE and system models[11].

Let us consider the unified model which is an extension of (3.17) to the current setting:

$$\widetilde{y}_{it} = \widetilde{\mathbf{w}}'_{it}\boldsymbol{\delta} + \widetilde{u}_{it}, \qquad (i = 1, ..., N; t = 2, ..., T) \tag{3.25}$$

where instruments $\widetilde{\mathbf{z}}_{it}$ that satisfy the moment condition $E(\widetilde{\mathbf{z}}_{it}\widetilde{u}_{it}) = \mathbf{0}$ are available. In

[8]Even for panel AR(1) model, we cannot derive the concentration parameter as in Rothenberg [1984] since the error variance in the first stage regression is time-varying.

[9]If time effects are included, we can remove them by taking deviations from cross-sectional averages.

[10]It is straightforward to modify the following discussion if covariates are strictly or weakly exogenous.

[11]With this assumption, the fixed effects are removed by taking first differences, and first-differenced variables can be used as instruments in LEV, FRE and system models containing LEV and FRE models.

a matrix form, (3.25) can be written as

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{W}}_i\boldsymbol{\delta} + \widetilde{\mathbf{u}}_i, \qquad (i = 1, ..., N) \tag{3.26}$$

where $\widetilde{\mathbf{y}}_i = (\widetilde{y}_{i2}, ..., \widetilde{y}_{iT})'$, $\widetilde{\mathbf{W}}_i = (\widetilde{\mathbf{w}}_{i2}, ..., \widetilde{\mathbf{w}}_{iT})'$, and $\widetilde{\mathbf{u}}_i = (\widetilde{u}_{i2}, ..., \widetilde{u}_{iT})'$. The moment condition can be written in a matrix form as $E(\widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{u}}_i) = \mathbf{0}$ where $\widetilde{\mathbf{Z}}_i = \text{diag}(\widetilde{\mathbf{z}}_{i2}', ..., \widetilde{\mathbf{z}}_{iT}')$. The one- and two-step GMM estimators for the model (3.26) utilizing the moment condition $E(\widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{u}}_i) = \mathbf{0}$ can be defined as

$$\widehat{\boldsymbol{\delta}}_{1step} = \left(\mathbf{A}_N'\mathbf{V}_N^{-1}\mathbf{A}_N\right)^{-1}\left(\mathbf{A}_N'\mathbf{V}_N^{-1}\mathbf{b}_N\right), \tag{3.27}$$

$$\widehat{\boldsymbol{\delta}}_{2step} = \left(\mathbf{A}_N'\boldsymbol{\Omega}_N^{-1}\mathbf{A}_N\right)^{-1}\left(\mathbf{A}_N'\boldsymbol{\Omega}_N^{-1}\mathbf{b}_N\right) \tag{3.28}$$

where $\mathbf{A}_N = \sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{W}}_i$, $\mathbf{b}_N = \sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\widetilde{\mathbf{y}}_i$, $\boldsymbol{\Omega}_N = \sum_{i=1}^N \mathbf{Z}_i'\widehat{\widetilde{\mathbf{u}}}_i\widehat{\widetilde{\mathbf{u}}}_i'\mathbf{Z}_i$ and $\widehat{\widetilde{\mathbf{u}}}_i = \widetilde{\mathbf{y}}_i - \widetilde{\mathbf{W}}_i\widehat{\boldsymbol{\delta}}_{1step}$.

All the GMM estimator considered in this chapter can be obtained by specifying a form for $(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N)$.

The GMM estimator for DIF model, denoted as "DIF", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\Delta\mathbf{y}_i, \Delta\mathbf{W}_i, \mathbf{Z}_i^D, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i'\mathbf{D}_T\mathbf{D}_T'\widetilde{\mathbf{Z}}_i\right) \tag{3.29}$$

where $\Delta\mathbf{y}_i = \mathbf{D}_T\mathbf{y}_i$, $\Delta\mathbf{W}_i = \mathbf{D}_T\mathbf{W}_i$, $\mathbf{Z}_i^D = \text{diag}(\mathbf{z}_{i2}^{D\prime}, ..., \mathbf{z}_{iT}^{D\prime})$, and
$\mathbf{z}_{it}^D = (y_{i,t-1-\ell}, ..., y_{i,t-2}, x_{i,t-\ell-1}, ..., x_{i,t-2})'$.

The GMM estimator for LEV model, denoted as "LEV", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\dot{\mathbf{y}}_i, \dot{\mathbf{W}}_i, \mathbf{Z}_i^L, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{L\prime}\widetilde{\mathbf{Z}}_i^L\right) \tag{3.30}$$

where $\dot{\mathbf{y}}_i = \mathbf{L}_T\mathbf{y}_i$, $\dot{\mathbf{W}}_i = \mathbf{L}_T\mathbf{W}_i$, $\mathbf{Z}_i^L = \text{diag}(\mathbf{z}_{i2}^{L\prime}, ..., \mathbf{z}_{iT}^{L\prime})$, and
$\mathbf{z}_{it}^L = (\Delta y_{i,t-\ell}, ..., \Delta y_{i,t-1}, \Delta x_{i,t-\ell}, ..., \Delta x_{i,t-1})'$.

The GMM estimator for FOD model, denoted as "FOD", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\mathbf{y}_i^*, \mathbf{W}_i^*, \mathbf{Z}_i^F, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{F\prime}\widetilde{\mathbf{Z}}_i^F\right) \tag{3.31}$$

where $\mathbf{y}_i^* = \mathbf{F}_T\mathbf{y}_i$, $\mathbf{W}_i^* = \mathbf{F}_T\mathbf{W}_i$, $\mathbf{Z}_i^F = \text{diag}(\mathbf{z}_{i1}^{F\prime}, ..., \mathbf{z}_{i,T-1}^{F\prime})$, and
$\mathbf{z}_{it}^F = (y_{i,t-\ell}, ..., y_{i,t-1}, x_{i,t-\ell}, ..., x_{i,t-1})'$.

For FRE model, we need to compute the covariance matrix of $\dot{\mathbf{u}}_i$ to transform the LEV model. If the error term v_{it} is *iid* over i and t, the covariance matrix of $\dot{\mathbf{u}}_i$ is given by $\boldsymbol{\Sigma}_{u,T_1}$ and the upper-triangular Cholesky factorization of $\boldsymbol{\Sigma}_{u,T_1}$ is given by (3.13). However, in practice, v_{it} might be heteroskedastic such that $Var(v_{it}) = \sigma_{it}^2$. In order to allow for such a heteroskedastic case, we propose an estimator for σ_η^2 and

$\bar{\sigma}_t^2 = N^{-1}\sum_{i=1}^N \sigma_{it}^2$. Since all off-diagonal elements of $\mathbf{\Sigma}_{u,T_1}$ are σ_η^2, we can estimate it consistently by

$$\widehat{\sigma}_\eta^2 = \frac{1}{T_1^2 - T_1}\sum_{s=2}^T \sum_{t=2,s\neq t}^T \widehat{\sigma}_{st}$$

where $\widehat{\sigma}_{st} = \frac{1}{N}\sum_{i=1}^N \widehat{u}_{is}\widehat{u}_{it}$ is the (s,t) element of $T_1 \times T_1$ matrix $N^{-1}\sum_{i=1}^N \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$, $\widehat{\mathbf{u}}_i = (\widehat{u}_{i2}, ..., \widehat{u}_{iT})'$, and $\widehat{u}_{it} = y_{it} - \widetilde{\boldsymbol{\delta}}'\mathbf{w}_{it}$, $(i = 1, ..., N; t = 2, ..., T)$. A consistent estimator for $\bar{\sigma}_t^2$ is then given by $\widehat{\sigma}_t^2 = N^{-1}\sum_{i=1}^N \widehat{u}_{it}^2 - \widehat{\sigma}_\eta^2$. Using these, we have a consistent estimator for $\mathbf{\Sigma}_{u,T_1}$ given by $\widehat{\mathbf{\Sigma}}_{u,T_1} = \widehat{\sigma}_\eta^2 \boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}' + \text{diag}(\widehat{\sigma}_2^2, ..., \widehat{\sigma}_T^2)$. The upper triangular Cholesky factorization for $\widehat{\mathbf{\Sigma}}_{u,T_1}^{-1}$ can be obtained numerically in practice.

Then, the GMM estimator for FRE model, denoted as "FRE", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\widehat{\mathbf{y}}_i^+, \widehat{\mathbf{W}}_i^+, \mathbf{Z}_i^R, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{R\prime}\widetilde{\mathbf{Z}}_i^R\right)$$

where $\widehat{\mathbf{y}}_i^+ = \widehat{\mathbf{\Sigma}}_{u,T_1}^{-1/2}\mathbf{y}_i$, $\widehat{\mathbf{W}}_i^+ = \widehat{\mathbf{\Sigma}}_{u,T_1}^{-1/2}\mathbf{W}_i$, $\widehat{\mathbf{\Sigma}}_{u,T_1}^{-1/2}$ is the upper triangular Cholesky factorization of $\widehat{\mathbf{\Sigma}}_{u,T_1}^{-1}$, $\mathbf{Z}_i^R = \text{diag}(\mathbf{z}_{i2}^{R\prime}, ..., \mathbf{z}_{iT}^{R\prime})$, and $\mathbf{z}_{it}^R = (\Delta y_{i,t-\ell}, ..., \Delta y_{i,t-1}, \Delta x_{i,t-\ell}, ..., \Delta x_{i,t-1})'$. For the purpose of comparison, we also consider the FRE model where model is transformed by using the true parameter values for σ_η^2 and $\sigma_t^2, (t = 2, ..., T)$. Such a GMM estimator for FRE model is denoted as "FRE*".

Yet another way to compute GMM estimator for FRE models is to use $\widehat{\mathbf{V}}_{u,N} = N^{-1}\sum_{i=1}^N \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'$ as a consistent estimator for $\mathbf{\Sigma}_{u,T_1}$, and transform the model by $\widehat{\mathbf{V}}_{u,N}^{-1/2}$ instead of $\mathbf{R}_{T_1}$ or $\widehat{\mathbf{\Sigma}}_{u,T_1}^{-1}$ where $\widehat{\mathbf{V}}_{u,N}^{-1/2}$ denotes the upper triangular factorization of $\widehat{\mathbf{V}}_{u,N}^{-1}$. Note that this is exactly the same approach proposed by Keane and Runkle [1992]. The GMM estimator for this alternative FRE model, denoted as "KR", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\widehat{\mathbf{y}}_i^\dagger, \widehat{\mathbf{W}}_i^\dagger, \mathbf{Z}_i^R, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{R\prime}\widetilde{\mathbf{Z}}_i^R\right)$$

where $\widehat{\mathbf{y}}_i^\dagger = \widehat{\mathbf{V}}_{u,N}^{-1/2}\mathbf{y}_i$, $\widehat{\mathbf{W}}_i^\dagger = \widehat{\mathbf{V}}_{u,N}^{-1/2}\mathbf{W}_i$.

The system GMM estimator combining DIF and LEV models, denoted as "SYS(DIF & LEV)", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\mathbf{y}_i^{DL}, \mathbf{W}_i^{DL}, \mathbf{Z}_i^{DL}, \text{diag}\left(\sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{D\prime}\mathbf{D}_T\mathbf{D}_T'\widetilde{\mathbf{Z}}_i^D, \sum_{i=1}^N \widetilde{\mathbf{Z}}_i^{L1\prime}\widetilde{\mathbf{Z}}_i^{L1}\right)\right)$$

where $\mathbf{y}_i^{DL} = (\Delta\mathbf{y}_i', \dot{\mathbf{y}}_i')'$, $\mathbf{W}_i^{DL} = (\Delta\mathbf{W}_i', \dot{\mathbf{W}}_i')'$ $\mathbf{Z}_i^{DL} = \text{diag}(\mathbf{Z}_i^D, \mathbf{Z}_i^{L1})$ $\mathbf{Z}_i^{L1} = \text{diag}(\mathbf{z}_{i2}^{L1}, ..., \mathbf{z}_{iT}^{L1})$ and $\mathbf{z}_{i2}^{L1} = \text{diag}(\Delta y_{i,t-1}, \Delta x_{i,t-1})'$.

The system GMM estimator combining FOD and LEV models, denoted as "SYS(FOD

& LEV)", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\mathbf{y}_i^{FL}, \mathbf{W}_i^{FL}, \mathbf{Z}_i^{FL}, \sum_{i=1}^{N} \widetilde{\mathbf{Z}}_i^{FL\prime} \widetilde{\mathbf{Z}}_i^{FL}\right)$$

where $\mathbf{y}_i^{FL} = (\mathbf{y}_i^{*\prime}, \dot{\mathbf{y}}_i^{\prime})^{\prime}$, $\mathbf{W}_i^{FL} = (\mathbf{W}_i^{*\prime}, \dot{\mathbf{W}}_i^{\prime})^{\prime}$ $\mathbf{Z}_i^{FL} = \mathrm{diag}(\mathbf{Z}_i^{F}, \mathbf{Z}_i^{L1})$.

The system GMM estimator combining FOD and FRE models, denoted as "SYS(FOD & FRE)", is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\mathbf{y}_i^{FR}, \mathbf{W}_i^{FR}, \mathbf{Z}_i^{FR}, \sum_{i=1}^{N} \widetilde{\mathbf{Z}}_i^{FR\prime} \widetilde{\mathbf{Z}}_i^{FR}\right)$$

where $\mathbf{y}_i^{FR} = (\mathbf{y}_i^{*\prime}, \widehat{\mathbf{y}}_i^{+\prime})^{\prime}$, $\mathbf{W}_i^{FR} = (\mathbf{W}_i^{*\prime}, \widehat{\mathbf{W}}_i^{+\prime})^{\prime}$ and $\mathbf{Z}_i^{FR} = \mathrm{diag}(\mathbf{Z}_i^{F}, \mathbf{Z}_i^{R1})$ and $\mathbf{Z}_i^{R1} = \mathbf{Z}_i^{L1}$. We call this system GMM estimator as *the forward system GMM estimator* since the forward GLS transformation is used for both DIF and LEV models. For the purpose of comparison, we also compute an infeasible version where the true parameter values are used, which is denoted as "SYS(FOD & FRE*)"

The system GMM estimator combining FOD and alternative FRE models, denoted as "SYS(FOD & KR)" is obtained by letting

$$\left(\widetilde{\mathbf{y}}_i, \widetilde{\mathbf{W}}_i, \widetilde{\mathbf{Z}}_i, \mathbf{V}_N\right) = \left(\mathbf{y}_i^{SKR}, \mathbf{W}_i^{SKR}, \mathbf{Z}_i^{SKR}, \sum_{i=1}^{N} \widetilde{\mathbf{Z}}_i^{SKR\prime} \widetilde{\mathbf{Z}}_i^{SKR}\right)$$

where $\mathbf{y}_i^{SKR} = (\mathbf{y}_i^{*\prime}, \widehat{\mathbf{y}}_i^{\dagger\prime})^{\prime}$, $\mathbf{W}_i^{FKR} = (\mathbf{W}_i^{*\prime}, \widehat{\mathbf{W}}_i^{\dagger\prime})^{\prime}$ and $\mathbf{Z}_i^{SKR} = \mathrm{diag}(\mathbf{Z}_i^{F}, \mathbf{Z}_i^{R1})$ and $\mathbf{Z}_i^{R1} = \mathbf{Z}_i^{L1}$.

3.5 Monte Carlo Simulation

In this section, we investigate the finite sample of the GMM estimator discussed in the previous sections. First, we investigate an pure AR(1) model and then consider a dynamic panel data models with an endogenous variable, which is practically relevant.

3.5.1 AR(1) model

First, we consider an AR(1) model. The data are generated as

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (t = 1, ..., T; i = 1, .., N) \tag{3.32}$$

where $v_{it} \sim iid\mathcal{N}(0, \sigma_v^2)$ and $\eta_i \sim iid\mathcal{N}(0, \sigma_\eta^2)$. The initial conditions are generated from stationary distribution

$$y_{i0} = \frac{\eta_i}{1-\alpha} + e_{i0} \tag{3.33}$$

where $e_{i0} \sim iid\mathcal{N}(0, \sigma_v^2/(1-\alpha^2))$. This specification is the same as the one used in theoretical analysis. For parameter values, we consider $\alpha = 0.2, 0.5, 0.8$, $\sigma_\eta^2 = 0.2, 1, 5$ and $\sigma_v^2 = 1$. For the sample size, we consider $T = 6, 12$ and $N = 200$. We report median("Median"), interquartile range("IQR"), and median absolute error(MAE) at each period[12], The number of replications is 5,000. We compare six estimators. Among them, four estimators are $\widetilde{\alpha}^D$ given by (3.4), $\widetilde{\alpha}^L$ given by (3.8), $\widetilde{\alpha}^F$ given by (3.12) and $\widetilde{\alpha}^R$ given by (3.16), and these are denoted as "DIF", "LEV", "FOD" and"FRE*", respectively[13]. Another two estimators are the system GMM estimator combining DIF and LEV models, and FOD and FRE models, and defined as

$$\widetilde{\alpha}^{DL} = \frac{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{DL\prime}\mathbf{Z}_i^{DL}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{DL\prime}\mathbf{Z}_i^{DL}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{DL\prime}\mathbf{y}_i^{DL}\right)}{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{DL\prime}\mathbf{Z}_i^{DL}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{DL\prime}\mathbf{Z}_i^{DL}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{DL\prime}\mathbf{y}_{i,-1}^{DL}\right)},$$

$$\widetilde{\alpha}^{FR} = \frac{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{FR\prime}\mathbf{Z}_i^{FR}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{FR\prime}\mathbf{Z}_i^{FR}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{FR\prime}\mathbf{y}_i^{FR}\right)}{\left(\sum_{i=1}^N \mathbf{y}_{i,-1}^{FR\prime}\mathbf{Z}_i^{FR}\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{FR\prime}\mathbf{Z}_i^{FR}\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{FR\prime}\mathbf{y}_{i,-1}^{FR}\right)}$$

where $\mathbf{y}_i^{DL} = (\Delta\mathbf{y}_i', \dot{\mathbf{y}}_i')'$, $\mathbf{y}_{i,-1}^{DL} = (\Delta\mathbf{y}_{i,-1}', \dot{\mathbf{y}}_{i,-1})'$, $\mathbf{Z}_i^{DL} = \mathrm{diag}(\mathbf{Z}_i^D, \mathbf{Z}_i^L)$, $\mathbf{y}_i^{FR} = (\mathbf{y}_i^{*\prime}, \mathbf{y}_i^{+\prime})'$, $\mathbf{y}_{i,-1}^{FR} = (\mathbf{y}_{i,-1}^{*\prime}, \mathbf{y}_{i,-1}^{+\prime})'$, and $\mathbf{Z}_i^{FR} = \mathrm{diag}(\mathbf{Z}_i^F, \mathbf{Z}_i^R)$. These are denoted as "SYS(DIF & LEV)" and "SYS(FOD & FRE)" in tables, respectively. We set the instruments lag length $\ell = 3$.

The simulation results are reported in Tables 1 to 3. To save space, we only provide the results with $T = 6$[14]. Major implications from the theoretical analysis are (i) the instruments become weaker as α approaches one and/or the variance ratio r increases, (ii) the instruments for FOD and FRE models are stronger than those for DIF and LEV models, (iii) the instruments for LEV model is substantially weak compared with those for FRE model when the variance ratio r is large. Monte Carlo simulation results are consistent with these implications. With regard to (i), from the tables, we observe that MAE increases as α approaches one and/or the variance ratio, or σ_η^2, increases for each period in most of the cases, which is consistent with the theoretical implication. With regard to the point (ii), from the tables, we find that, in terms of MAE, the 2SLS for FOD model outperforms that for DIF model in almost all cases(except for the last period where two 2SLS estimators are identical). With regard to the 2SLS estimators for LEV

[12]Since the number of parameter and instruments are identical at $t = 1$, i.e., just-identified, the moments of 2SLS estimator do not exist and some outliers appeared during the replications. Hence, we use a robust measure such as median.

[13]"*" in "FRE*" and "SYS(FOD & FRE*)" indicates that it uses the true parameter value for the computation of $\mathbf{R}_{T_1}$ and hence it is infeasible. Also, note that in DIF model inefficient weighting matrix is used.

[14]The simulation results with $T = 12$ are provided in a supplementary appendix. The results with $T = 12$ are qualitatively similar to those of $T = 6$.

and FRE models, we find that the 2SLS for FRE model tends to have smaller bias than that for LEV models. However, in terms of dispersion, the results are mixed: in some cases, the 2SLS for FRE model has smaller IQR while in other cases, the 2SLS for LEV has smaller IQR. This is also the case for the GMM estimator where 2SLS estimators are aggregated. This is because the 2SLS estimator for LEV model shrinks to one and tends to have smaller dispersion as the variance ratio r gets larger[15]. However, since the bias of GMM estimator for FRE model is smaller than that for LEV estimator, the GMM estimator for FRE model outperforms that for LEV estimator in terms of MAE. Finally, with regard to the point (iii), we find that the 2SLS estimator for FRE model substantially perform better than that for LEV model. In terms of bias, although the 2SLS for LEV model has an upward bias for all periods, that for FRE model has upward bias for some periods and downward bias in other periods, and this structure makes the GMM estimator for FRE model less biased compared with LEV model. Thus, the Monte Carlo simulation results are consistent with the theoretical implications derived in Section 3.3.

3.5.2 Model with an endogenous variable

In the previous subsection, we considered a simple AR(1) model without covariates and also considered an infeasible estimator for FRE model. However, in practice, we usually include covariates besides the lagged dependent variable and also, the covariance matrix used in the FRE model is unknown. Therefore, in this subsection, we extend the AR(1) model to include an endogenous variable, consider feasible estimator for FRE model and investigate whether results similar to those in the AR(1) case are obtained.

We consider the following model:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it}, \quad (i = 1, ..., N; t = -49, -48, ..., -1, 0, 1, ..., T), \\ x_{it} &= \rho x_{i,t-1} + \tau \eta_i + \theta v_{it} + e_{it}, \end{aligned}$$

where $\eta_i \sim iid(0, \sigma_\eta^2)$ and $e_{it} \sim iid(0, \sigma_e^2)$. For the error term v_{it}, we allow for heteroskedasticity both in cross-section and time-series and generated as $v_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$, where σ_{it}^2 is generated as $\sigma_{it}^2 = \delta_i \tau_t$, in which $\delta_i \sim \mathcal{U}[0.5, 1.5]$ and $\tau_t = 0.5 + (t-1)/(T-1)$ for $t = 1..., T$ and $\tau_t = 0.5$ for $t = -49, ..., -1, 0$. The data are generated as according

[15] The intuition behind this behavior is as follows. The 2SLS estimator tends to have a similar mean to the inconsistent OLS estimator as instruments gets weaker[Han and Schmidt, 2001]. Since the probability limit of the OLS estimator for LEV model $y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}$, denoted by $\widehat{\alpha}_{LEVOLS}$ is given by $\text{plim}_{N\to\infty} \widehat{\alpha}_{LEVOLS} = \alpha + (1-\alpha)\frac{r}{r + \frac{1-\alpha}{1+\alpha}}$, it follows that as $r \to \infty$, $\text{plim}_{N\to\infty} \widehat{\alpha}_{LEVOLS} \to 1$. Thus, the 2SLS estimator shrinks to one as r gets larger.

to the following VAR(1) model:

$$\begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix} \begin{pmatrix} y_{i,t-1} \\ x_{i,t-1} \end{pmatrix} + \begin{pmatrix} (1+\beta\tau)\eta_i \\ \tau\eta_i \end{pmatrix} + \begin{pmatrix} (1+\theta\beta)v_{it} + \beta e_{it} \\ \theta v_{it} + e_{it} \end{pmatrix}$$

with the initial conditions:

$$\begin{pmatrix} y_{i,-50} \\ x_{i,-50} \end{pmatrix} = \begin{pmatrix} 1-\bar{\alpha} & -\beta\bar{\rho} \\ 0 & 1-\bar{\rho} \end{pmatrix}^{-1} \begin{pmatrix} (1+\beta\tau)\eta_i \\ \tau\eta_i \end{pmatrix} + \boldsymbol{\xi}_{i,-50}, \quad and$$

$$\boldsymbol{\xi}_{i,-50} = \sum_{j=0}^{100} \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix}^j \begin{pmatrix} (1+\theta\beta)v_{ij} + \beta e_{ij} \\ \theta v_{ij} + e_{ij} \end{pmatrix}, \tag{3.34}$$

where $v_{ij} \sim \mathcal{N}(0, \sigma^2_{i,-49})$ and $e_{ij} \sim iid\mathcal{N}(0, \sigma^2_e)$ $(j = 0, ..., 100)$.

For the parameter values, we consider $\alpha = 0.2, 0.8$, $\rho = 0.2, 0.8$, $\beta = 1-\alpha$, $\theta = -0.1$, and $\tau = 0.25$. For the values of σ^2_η and σ^2_e, we follow the approach of Bun and Sarafidis [2013]:

$$\begin{aligned}
\sigma^2_\eta &= \frac{c_v^2 \bar{\sigma}_v^2 VR}{\zeta^2}, \\
\sigma^2_e &= \frac{SNR + 1 - c_v^2}{c_e^2}, \\
c_v^2 &= \frac{(1+\alpha\rho)}{(1-\alpha^2)(1-\rho^2)(1-\alpha\rho)} \left[(1+\beta\theta)^2 + \rho^2 - \frac{2\rho(\alpha+\rho)(1+\beta\theta)}{1+\alpha\rho} \right], \\
c_e^2 &= \frac{(1+\alpha\rho)\beta^2}{(1-\alpha^2)(1-\rho^2)(1-\alpha\rho)}, \quad and \\
\zeta^2 &= \frac{\beta\tau + 1 - \rho}{(1-\alpha)(1-\rho)},
\end{aligned}$$

where $\bar{\sigma}_v^2 = \frac{1}{N(T+1)} \sum_{i=1}^N \sum_{t=0}^T E(\sigma^2_{it})$, VR stands for the 'variance ratio' and SNR stands for the 'signal-to-noise ratio': $SNR = (Var(y_{it}|\eta_i) - \sigma_v^2)/\sigma_v^2$[16]. We consider $VR = \{10, 50\}$ and $SNR = \{3, 8\}$. The values of σ^2_η for each design is provided in the table.

Estimators compared here are provided in Section 4. For the lag length of instruments, we set $\ell = 3$. Simulation results of two-step GMM estimators are provided in Tables 4 to 6[17]. To save space, we only report the results with $\alpha = 0.2, 0.8$, $\rho = 0.8$, and $SNR = 3$. Also, the results for estimators using $\widehat{\mathbf{V}}_{u,N}$ are omitted since the performance of those estimators are very similar to those using $\widehat{\boldsymbol{\Sigma}}_{u,T_1}$. Complete results are provided in a supplementary appendix.

[16] Note that $\bar{\sigma}_v^2 = 1$ in the current setup.

[17] For the empirical size, we use corrected standard errors by Windmeijer [2005].

From the table, we find that most of the estimators tend to work well when $\alpha = 0.2$ and $VR = 10$. This result is consistent with the theoretical implication since in such cases, the instruments are sufficiently strong. However, it is not the case when $\alpha = 0.8$ and/or $VR = 50$. In these cases, since instruments tend to be weak, the GMM estimators tend to work poorly. When $\alpha = 0.2$ and $VR = 50$, the LEV, SYS(DIF & LEV) and SYS(FOD & LEV), all contain LEV model, are biased and empirical size is more distorted than other estimators. However, the FRE and SYS(FOR & FRE) perform well. This result is consistent with the theoretical implication that the LEV model is more vulnerable to the large variance ratio than the FRE model. If α is increased from 0.2 to 0.8, we find that the performance of DIF and FOD model gets worsened: both bias and SD gets larger and empirical sizes are more distorted. Among the estimators, SYS(FOD & FRE) tends to perform best in terms of bias and accuracy of inference. With regard to the effect of T, most of the implications for the case $T = 6$ apply to the case with $T = 12$, and overall performance improves as T gets larger. When $T = 12$, SYS(FOD & FRE) tends to perform best among the estimators. Also, with regard to the effect of ρ and SNR, in view of the results provided in a supplement, we find that the performance of all estimators improves if ρ is decreased from 0.8 to 0.2 and SNR is increased from 3 to 8. We now investigate the effect of transformation matrix associated with FRE and SYS(FOD & FRE) models. We compute two estimators for these models: one is using $\widehat{\mathbf{\Sigma}}_{u,T_1} = \widehat{\sigma}_\eta^2 \boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}' + \text{diag}(\widehat{\sigma}_2^2, ..., \widehat{\sigma}_T^2)$ where estimated parameter values are used while the other is $\mathbf{\Sigma}_{u,T_1} = \sigma_\eta^2 \boldsymbol{\iota}_{T_1}\boldsymbol{\iota}_{T_1}' + \text{diag}(\sigma_2^2, ..., \sigma_T^2)$ where the true parameter values for $\sigma_\eta^2, \sigma_2^2, ..., \sigma_T^2$ are used.[18] Estimators using the true covariance matrix is denoted as "FRE*" and "SYS(FOD & FRE*)". Comparing the estimators with/without "*", we find that "FRE" and "SYS(FOD & FRE)" tend to have a large dispersion and RMSE than those with using true covariance matrix. This comes from estimation uncertainty associated with estimation of covariances matrix, and not negligible when $N = 100$. However, as N gets larger, this estimation uncertainty disappears and in fact, when $N = 500$, estimators using estimated covariance matrix and true one perform very similarly.

Summarizing the simulation results, we find that the LEV model or system model containing it tend to perform poorly when VR is large, which is consistent with the theoretical implication, while the FRE model or system model containing it is less affected by the variance ratio and SYS(FOD & FRE) tends to perform best in terms of bias and accuracy of inference. Thus, while the simulation design is somewhat limited, SYS(FOD & FRE) can be an alternative to SYS(DIF & LEV) or SYS(FOD & LEV) estimators.

[18] As an initial estimate for $\boldsymbol{\delta}$ for the computation of residual, we use the double filter IV estimator proposed by Breitung, Hayakawa, and Qi [2016] since it is less biased and more efficient than other estimators.

3.6 Conclusion

In this chapter, we investigated the weak instruments problem of the system GMM estimator in dynamic panel data models. We showed that the the FOD and FRE models have a higher concentration parameter than DIF and LEV model despite the same instruments are used.

We investigated the finite sample performance of various GMM estimators and found that the system GMM estimator combining FOD and FRE models performs well even when the variance ratio of individual effects to the errors is large despite the conventional system GMM estimators do not.

Appendix

3.A Proof of Theorems 1 and 2

First, we provide some expressions that are useful to derive the results. Variables $y_{i,t-1}$, $y^*_{i,t-1}$ and $y^+_{i,t-1}$ can be written as[19]

$$
\begin{aligned}
y_{i,t-1} &= \mu_i + w_{i,t-1}, \qquad (t=2,...,T), \\
y^*_{i,t-1} &= c_t(\psi_t w_{i,t-1} - \tilde{v}_{itT}), \qquad (t=1,...,T-1), \qquad (3.35) \\
y^+_{i,t-1} &= \begin{cases} b_t y^*_{i,t-1} + k_t y_{i,t-1} = b_t c_t (h_t w_{i,t-1} - \tilde{v}_{itT} + q_t \mu_i) & (t=2,...,T-1) \\ k_t y_{i,t-1} = k_t \mu_i + k_t w_{i,t-1} & (t=T) \end{cases} \qquad (3.36)
\end{aligned}
$$

where $\mu_i = \eta_i/(1-\alpha)$, $w_{it} = \sum_{j=0}^{\infty} \alpha^j v_{i,t-j}$,

$$
\begin{aligned}
\tilde{v}_{itT} &= \frac{\phi_{T-t} v_{it} + \cdots + \phi_1 v_{i,T-1}}{T-t}, \qquad c_t = \frac{\sqrt{T-t}}{\sqrt{T-t+1}}, \quad (t=1,...,T-1), \qquad (3.37) \\
b_t &= \frac{\sqrt{(T-t+1)}\sqrt{(T-t)}}{\sqrt{\left(T-t+1+\frac{1}{r}\right)}\sqrt{\left(T-t+\frac{1}{r}\right)}}, \qquad (t=2,...,T-1), \\
k_t &= \frac{1}{r\sqrt{T-t+1+\frac{1}{r}}\sqrt{T-t+\frac{1}{r}}}, \qquad (t=2,...,T), \\
h_t &= (\psi_t + q_t), \qquad q_t = \frac{1}{r(T-t)}, \qquad (t=2,...,T-1).
\end{aligned}
$$

3.A.1 Derivation for DIF model

The concentration parameter for DIF model at period t can be written as

$$
CP^D_t = \frac{E\left(\mathbf{z}^D_{it}\Delta y_{i,t-1}\right)' \left[E\left(\mathbf{z}^D_{it}\mathbf{z}^{D\prime}_{it}\right)\right]^{-1} E\left(\mathbf{z}^D_{it}\Delta y_{i,t-1}\right)}{E\left[(\Delta y_{i,t-1})^2\right] - E\left(\mathbf{z}^D_{it}\Delta y_{i,t-1}\right)' \left[E\left(\mathbf{z}^D_{it}\mathbf{z}^{D\prime}_{it}\right)\right]^{-1} E\left(\mathbf{z}^D_{it}\Delta y_{i,t-1}\right)}, \quad (t=2,...,T) \qquad (3.38)
$$

Although we can use the same approach as Bun and Windmeijer [2010] for the derivation of CP^D_t, we provide another derivation different from them.

[19] See Alvarez and Arellano [2003] and Hayakawa [2015].

Let us define

$$\dot{\mathbf{z}}_{it}^{D}=\begin{bmatrix}\mathbf{z}_{it}^{D}\\ \Delta y_{i,t-1}\end{bmatrix}=\begin{bmatrix}1 & & & & 0\\ & 1 & & & \\ & & \ddots & & \\ 0 & & & 1 & 0\\ 0 & & & -1 & 1\end{bmatrix}\begin{bmatrix}y_{i,t-1-\ell}\\ \vdots\\ y_{i,t-2}\\ y_{i,t-1}\end{bmatrix}=\dot{\mathbf{H}}\mathbf{y}_{i}^{(\ell)}.$$

Then, we have

$$E\left(\dot{\mathbf{z}}_{it}^{D}\dot{\mathbf{z}}_{it}^{D\prime}\right)=\begin{bmatrix}E\left(\mathbf{z}_{it}^{D}\mathbf{z}_{it}^{D\prime}\right) & E\left(\mathbf{z}_{it}^{D}\Delta y_{i,t-1}\right)\\ E\left(\Delta y_{i,t-1}\mathbf{z}_{it}^{D\prime}\right) & E\left[(\Delta y_{i,t-1})^{2}\right]\end{bmatrix}=\dot{\mathbf{H}}E\left(\mathbf{y}_{i}^{(\ell)}\mathbf{y}_{i}^{(\ell)\prime}\right)\dot{\mathbf{H}}'=\mathbf{\Gamma}^{D}=\begin{bmatrix}\mathbf{\Gamma}_{11}^{D} & \boldsymbol{\gamma}_{12}^{D}\\ \boldsymbol{\gamma}_{12}^{D\prime} & \gamma_{22}^{D}\end{bmatrix}$$

and

$$\left[E\left(\dot{\mathbf{z}}_{it}^{D}\dot{\mathbf{z}}_{it}^{D\prime}\right)\right]^{-1}=\dot{\mathbf{H}}'^{-1}\left[E\left(\mathbf{y}_{i}^{(\ell)}\mathbf{y}_{i}^{(\ell)\prime}\right)\right]^{-1}\dot{\mathbf{H}}^{-1}=(\mathbf{\Gamma}^{D})^{-1}=\begin{bmatrix}\mathbf{\Gamma}^{D11} & \boldsymbol{\gamma}^{D12}\\ \boldsymbol{\gamma}^{D12\prime} & \gamma^{D22}\end{bmatrix} \quad (3.39)$$

Now, consider a partitioned matrix $\mathbf{A}$ given by

$$\mathbf{A}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{a}_{12}\\ \mathbf{a}_{12}' & a_{22}\end{bmatrix} \quad (3.40)$$

where $\mathbf{A}_{11}$ is $m\times m$, $\mathbf{a}_{12}$ is $m\times 1$ and a_{22} is a scaler. Then, the inverse formula for partitioned matrix is given by

$$\mathbf{A}^{-1}=\begin{bmatrix}\mathbf{A}_{11} & \mathbf{a}_{12}\\ \mathbf{a}_{12}' & a_{22}\end{bmatrix}^{-1}=\begin{bmatrix}\mathbf{A}_{11}^{-1}+\mathbf{A}_{11}^{-1}\mathbf{a}_{12}b^{-1}\mathbf{a}_{12}'\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{a}_{12}b^{-1}\\ -b^{-1}\mathbf{a}_{12}'\mathbf{A}_{11}^{-1} & b^{-1}\end{bmatrix}=\begin{bmatrix}\mathbf{A}^{11} & \mathbf{a}^{12}\\ \mathbf{a}^{12\prime} & a^{22}\end{bmatrix}$$

where $b=a_{22}-\mathbf{a}_{12}'\mathbf{A}_{11}^{-1}\mathbf{a}_{12}$. Hence we have

$$\mathbf{a}_{12}'\mathbf{A}_{11}^{-1}\mathbf{a}_{12}=a_{22}-b=a_{22}-(a^{22})^{-1}. \quad (3.41)$$

By putting $\mathbf{A}=\mathbf{\Gamma}^{D}$ in (3.40) and using (3.41), we have

$$E\left(\Delta y_{i,t-1}\mathbf{z}_{it}^{D\prime}\right)E\left(\mathbf{z}_{it}^{D}\mathbf{z}_{it}^{D\prime}\right)^{-1}E\left(\mathbf{z}_{it}^{D}\Delta y_{i,t-1}\right)=E\left[(\Delta y_{i,t-1})^{2}\right]-(\gamma^{D22})^{-1}. \quad (3.42)$$

It is easy to show that the first term is given by

$$E[(\Delta y_{i,t-1})^{2}]=\frac{2\sigma_{v}^{2}}{1+\alpha}. \quad (3.43)$$

For the derivation of γ^{D22}, by substituting

$$
\begin{aligned}
\dot{\mathbf{H}}^{-1} &= \begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ 0 & & & 1 & 0 \\ 0 & & & -1 & 1 \end{bmatrix}, \\
E\left(\mathbf{y}_i^{(\ell)}\mathbf{y}_i^{(\ell)\prime}\right)^{-1} &= \frac{1}{\sigma_v^2}\left(\mathbf{\Phi}_\ell^{-1} - \frac{\sigma_\mu^2}{\sigma_v^2+\sigma_\mu^2\boldsymbol{\iota}_\ell'\mathbf{\Phi}_\ell^{-1}\boldsymbol{\iota}_\ell}\mathbf{\Phi}_\ell^{-1}\boldsymbol{\iota}_\ell\boldsymbol{\iota}_\ell'\mathbf{\Phi}_\ell^{-1}\right), \quad \sigma_\mu^2 = \frac{\sigma_\eta^2}{(1-\alpha)^2}, \\
\mathbf{\Phi}_\ell^{-1} &= \begin{bmatrix} 1 & -\alpha & 0 & & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & & \\ & \ddots & \ddots & \ddots & \\ & & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & & 0 & -\alpha & 1 \end{bmatrix}
\end{aligned}
$$

into (3.39), we obtain

$$
\gamma^{D22} = \frac{1}{\sigma_v^2}\left(1 - \frac{\sigma_\eta^2}{\sigma_v^2+\sigma_\eta^2\left(\ell-1+\frac{1+\alpha}{1-\alpha}\right)}\right). \tag{3.44}
$$

Hence, substituting (3.43) and (3.44) into (3.42), we obtain

$$
\begin{aligned}
& E\left(\Delta y_{i,t-1}\mathbf{z}_{it}^{D\prime}\right)E\left(\mathbf{z}_{it}^{D}\mathbf{z}_{it}^{D\prime}\right)^{-1}E\left(\mathbf{z}_{it}^{D}\Delta y_{i,t-1}\right) \\
&\qquad = \frac{(1-\alpha)^4\left[1+r(\ell-1)\right]}{(1-\alpha^2)\left\{(1-r)(1-\alpha)^2+r(1-\alpha)\left[(\ell+1)-(\ell-1)\alpha\right]\right\}}
\end{aligned} \tag{3.45}
$$

The concentration parameter (3.19) is obtained by substituting (3.43) and (3.45) into (3.38).

3.A.2 Derivation for LEV model

The concentration parameter for LEV model at period t can be written as

$$
CP_t^L = \frac{E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)'\left[E\left(\mathbf{z}_{it}^L\mathbf{z}_{it}^{L\prime}\right)\right]^{-1}E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)}{E(y_{i,t-1}^2) - E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)'\left[E\left(\mathbf{z}_{it}^L\mathbf{z}_{it}^{L\prime}\right)\right]^{-1}E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)}, \quad (t=2,...,T). \tag{3.46}
$$

To derive the numerator of (3.46), let us define

$$\dot{\mathbf{z}}_{it}^L = \begin{bmatrix} \mathbf{z}_{it}^L \\ -y_{i,t-1} \end{bmatrix} = \begin{bmatrix} -1 & 1 & & & 0 \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ 0 & & & & -1 \end{bmatrix} \begin{bmatrix} y_{i,t-1-\ell} \\ \vdots \\ y_{i,t-2} \\ y_{i,t-1} \end{bmatrix} = \dot{\mathbf{D}}\mathbf{y}_i^{(\ell)}.$$

Then, we have

$$\begin{aligned} E\left(\dot{\mathbf{z}}_{it}^L \dot{\mathbf{z}}_{it}^{L\prime}\right) &= \begin{bmatrix} E\left(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime}\right) & -E\left(\mathbf{z}_{it}^L \Delta y_{i,t-1}\right) \\ -E\left(y_{i,t-1}\mathbf{z}_{it}^{L\prime}\right) & E(y_{i,t-1}^2) \end{bmatrix} \\ &= \dot{\mathbf{D}} E\left(\mathbf{y}_i^{(\ell)} \mathbf{y}_i^{(\ell)\prime}\right) \dot{\mathbf{D}}' \\ &= \boldsymbol{\Gamma}^L \\ &= \begin{bmatrix} \boldsymbol{\Gamma}_{11}^L & \boldsymbol{\gamma}_{12}^L \\ \boldsymbol{\gamma}_{12}^{L\prime} & \gamma_{22}^L \end{bmatrix} \end{aligned}$$

and

$$\left[E\left(\dot{\mathbf{z}}_{it}^L \dot{\mathbf{z}}_{it}^{L\prime}\right)\right]^{-1} = \dot{\mathbf{D}}'^{-1}\left[E\left(\mathbf{y}_i^{(\ell)} \mathbf{y}_i^{(\ell)\prime}\right)\right]^{-1} \dot{\mathbf{D}}^{-1} = (\boldsymbol{\Gamma}^L)^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{L11} & \mathbf{fl}^{L12} \\ \mathbf{fl}^{L12\prime} & \mathbf{fl}^{L22} \end{bmatrix}.$$

If we put $\mathbf{A} = \boldsymbol{\Gamma}^L$ in (3.40) and use (3.41), we have

$$E\left(y_{i,t-1}\mathbf{z}_{it}^{L\prime}\right) E\left(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime}\right)^{-1} E\left(\mathbf{z}_{it}^L y_{i,t-1}\right) = E(y_{i,t-1}^2) - (\gamma^{L22})^{-1}. \tag{3.47}$$

The first term is given by

$$E(y_{i,t-1}^2) = \sigma_v^2 \left(\frac{1}{1-\alpha^2} + \frac{r}{(1-\alpha)^2}\right). \tag{3.48}$$

For the second term, by using (3.44) and

$$\dot{\mathbf{D}}^{-1} = \begin{bmatrix} -1 & \cdots & -1 \\ & \ddots & \vdots \\ 0 & & -1 \end{bmatrix},$$

we have

$$\gamma^{L22} = \frac{\left(1-\alpha^2\right) + \ell\left(1-\alpha\right)}{\sigma_v^2}. \tag{3.49}$$

Hence, substituting (3.48) and (3.49) into (3.47), we have

$$E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)' \left[E\left(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^L y_{i,t-1}\right) = \frac{\sigma_v^2 \ell}{(1+\alpha)[(\ell+1)-(\ell-1)\alpha]}. \tag{3.50}$$

The concentration parameter (3.20) is obtained by substituting (3.48) and (3.50) into (3.46).

3.A.3 Derivation for FOD model

The concentration parameter for FOD model at period t can be written as

$$
\begin{aligned}
CP_t^F &= \frac{E\left(\mathbf{z}_{it}^F y_{i,t-1}^*\right)' \left[E\left(\mathbf{z}_{it}^F \mathbf{z}_{it}^{F\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^F y_{i,t-1}^*\right)}{E[(y_{i,t-1}^*)^2] - E\left(\mathbf{z}_{it}^F y_{i,t-1}^*\right)' \left[E\left(\mathbf{z}_{it}^F \mathbf{z}_{it}^{F\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^F y_{i,t-1}^*\right)}, \quad (t = 1, ..., T-1) \\
&= \frac{E\left(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*\right)' \left[E\left(\mathbf{z}_{i,t-1}^F \mathbf{z}_{i,t-1}^{F\prime}\right)\right]^{-1} E\left(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*\right)}{E[(y_{i,t-2}^*)^2] - E\left(\mathbf{z}_{it}^F y_{i,t-2}^*\right)' \left[E\left(\mathbf{z}_{i,t-1}^F \mathbf{z}_{i,t-1}^{F\prime}\right)\right]^{-1} E\left(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*\right)}, \quad (3.51) \\
&\qquad (t = 2, ..., T).
\end{aligned}
$$

Note that the range of time index t in FOD model is different from other three models. First, consider the numerator of (3.51). Note that, for DIF and FOD models, we have

$$
E(\mathbf{z}_{it}^D \Delta y_{i,t-1}) = \frac{-\sigma_v^2}{1+\alpha} \begin{bmatrix} \alpha^{\ell-1} \\ \vdots \\ 1 \end{bmatrix}, \qquad (t = 2, ..., T), \tag{3.52}
$$

$$
E(\mathbf{z}_{it}^F y_{i,t-1}^*) = c_t \psi_t \frac{\sigma_v^2}{1-\alpha^2} \begin{bmatrix} \alpha^{\ell-1} \\ \vdots \\ 1 \end{bmatrix}, \qquad (t = 1, ..., T-1). \tag{3.53}
$$

Hence, we have

$$
E(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*) = A_t E(\mathbf{z}_{it}^D \Delta y_{i,t-1}), \qquad (t = 2, ..., T) \tag{3.54}
$$

where $A_t = -c_{t-1}\psi_{t-1}/(1-\alpha)$. Using (3.54) and noting that $\mathbf{z}_{it}^F = \mathbf{z}_{it}^D$, the numerator of (3.51) can be written as

$$
\begin{aligned}
&E\left(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*\right)' \left[E\left(\mathbf{z}_{i,t-1}^F \mathbf{z}_{i,t-1}^{F\prime}\right)\right]^{-1} E\left(\mathbf{z}_{i,t-1}^F y_{i,t-2}^*\right) \\
&\qquad = A_t^2 E\left(\mathbf{z}_{it}^D \Delta y_{i,t-1}\right)' \left[E\left(\mathbf{z}_{it}^D \mathbf{z}_{it}^{D\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^D \Delta y_{i,t-1}\right), \qquad (t = 2, ..., T)
\end{aligned} \tag{3.55}
$$

Also, from (3.35), we have

$$
E[(y_{i,t-2}^*)^2] = \sigma_v^2 c_{t-1}^2 \left[\psi_{t-1}^2 \left(\frac{1}{1-\alpha^2}\right) + \frac{\varphi_{t-1}}{(1-\alpha)^2}\right], \quad (t = 2, ..., T) \tag{3.56}
$$

where c_t is defined in (3.37) and ψ_{t-1} and φ_{t-1} are defined in (3.23). The concentration parameter (3.21) is obtained by substituting (3.55) and (3.56) into (3.51).

3.A.4 Derivation for FRE model

The concentration parameter for FRE model at period t is given by

$$CP_t^R = \frac{E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right)' \left[E\left(\mathbf{z}_{it}^R \mathbf{z}_{it}^{R\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right)}{E(y_{i,t-1}^{+2}) - E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right)' \left[E\left(\mathbf{z}_{it}^R \mathbf{z}_{it}^{R\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right)}, \quad (t = 2, ..., T) \tag{3.57}$$

As in the derivation FOD model, we utilize the results of LEV model. Since

$$\begin{aligned} E\left(\mathbf{z}_{it}^L y_{i,t-1}\right) &= \frac{\sigma_v^2}{1+\alpha} \begin{bmatrix} \alpha^{\ell-1} \\ \vdots \\ 1 \end{bmatrix}, \qquad (t = 2, ..., T), \\ E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right) &= B_t \frac{\sigma_v^2}{1+\alpha} \begin{bmatrix} \alpha^{\ell-1} \\ \vdots \\ 1 \end{bmatrix}, \qquad (t = 2, ..., T) \end{aligned}$$

where

$$B_t = \begin{cases} b_t c_t h_t & (t = 2, ..., T-1) \\ k_t & (t = T) \end{cases},$$

we have

$$E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right) = B_t E\left(\mathbf{z}_{it}^L y_{i,t-1}\right), \quad (t = 2, ..., T). \tag{3.58}$$

Then, using (3.58) and noting that $\mathbf{z}_{it}^L = \mathbf{z}_{it}^R$, the numerator of (3.57) can be written as

$$\begin{aligned} &E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right)' \left[E\left(\mathbf{z}_{it}^R \mathbf{z}_{it}^{R\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^R y_{i,t-1}^+\right) \\ &= B_t^2 E\left(\mathbf{z}_{it}^L y_{i,t-1}\right)' \left[E\left(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^L y_{i,t-1}\right), \quad (t \\ &= 2, ..., T). \end{aligned} \tag{3.59}$$

Also, from (3.36), we have

$$\begin{aligned} E[(y_{i,t-1}^+)^2] &= \begin{cases} \sigma_v^2 b_t^2 c_t^2 \left(\frac{h_t^2}{1-\alpha^2} + \frac{\varphi_t + r q_t^2}{(1-\alpha)^2}\right) & (t = 2, ..., T-1) \\ \sigma_v^2 k_t^2 \left(\frac{1}{1-\alpha^2} + \frac{r}{(1-\alpha)^2}\right) & (t = T) \end{cases}. \\ &= \begin{cases} \sigma_v^2 b_t^2 c_t^2 \left(\frac{h_t^2}{1-\alpha^2} + \frac{\varphi_t + \frac{1}{r(T-t)^2}}{(1-\alpha)^2}\right) & (t = 2, ..., T-1) \\ \sigma_v^2 k_t \left(\frac{k_t}{1-\alpha^2} + \frac{1}{(1-\alpha)^2 \sqrt{T-t+1+\frac{1}{r}}\sqrt{T-t+\frac{1}{r}}}\right) & (t = T) \end{cases}. \end{aligned} \tag{3.60}$$

The concentration parameter (3.22) is obtained by substituting (3.59) and (3.60) into (3.57).

Table 3.1: Table 3.1 Simulation results for AR(1) model : $T = 6,\ N = 200\ ,\ \alpha = 0.2,$ $\sigma_\eta^2 = 0.2,\ 1,\ 5$

$T = 6,\ N = 200,\ \alpha = 0.2,\ \sigma_\eta^2 = 0.2$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.203	0.195	0.181	0.174	0.168	0.177	0.200	0.201	0.196	0.194	0.168	0.192
IQR	0.184	0.188	0.177	0.173	0.170	0.077	0.121	0.127	0.129	0.138	0.170	0.062
MAE	0.091	0.094	0.090	0.089	0.090	0.041	0.060	0.064	0.065	0.069	0.090	0.031
	LEV						FRE*					
Median	0.203	0.201	0.207	0.206	0.207	0.206	0.204	0.198	0.203	0.202	0.207	0.204
IQR	0.161	0.136	0.127	0.119	0.117	0.060	0.158	0.133	0.128	0.122	0.117	0.059
MAE	0.081	0.068	0.063	0.060	0.059	0.031	0.079	0.067	0.064	0.061	0.059	0.030
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.202	0.199	0.195	0.193	0.190	0.194	0.202	0.200	0.199	0.198	0.195	0.199
IQR	0.147	0.127	0.114	0.106	0.103	0.060	0.101	0.099	0.093	0.096	0.100	0.054
MAE	0.073	0.064	0.057	0.053	0.051	0.030	0.051	0.050	0.046	0.048	0.050	0.027

$T = 6,\ N = 200,\ \alpha = 0.2,\ \sigma_\eta^2 = 1$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.201	0.191	0.177	0.169	0.164	0.170	0.200	0.201	0.195	0.193	0.164	0.191
IQR	0.257	0.219	0.196	0.187	0.178	0.087	0.170	0.148	0.145	0.147	0.178	0.072
MAE	0.128	0.110	0.100	0.097	0.096	0.048	0.085	0.075	0.072	0.073	0.096	0.037
	LEV						FRE*					
Median	0.203	0.208	0.218	0.227	0.231	0.224	0.204	0.198	0.200	0.198	0.231	0.207
IQR	0.208	0.171	0.158	0.149	0.147	0.081	0.169	0.145	0.144	0.149	0.147	0.068
MAE	0.103	0.086	0.081	0.079	0.078	0.044	0.085	0.073	0.072	0.074	0.078	0.035
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.204	0.204	0.203	0.205	0.205	0.204	0.201	0.201	0.197	0.196	0.205	0.200
IQR	0.190	0.150	0.130	0.123	0.118	0.070	0.125	0.111	0.100	0.110	0.118	0.064
MAE	0.095	0.075	0.065	0.061	0.059	0.035	0.063	0.055	0.050	0.055	0.059	0.032

$T = 6,\ N = 200,\ \alpha = 0.2,\ \sigma_\eta^2 = 5$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.197	0.186	0.173	0.167	0.162	0.163	0.202	0.204	0.195	0.193	0.162	0.187
IQR	0.488	0.247	0.208	0.191	0.178	0.093	0.318	0.166	0.154	0.151	0.178	0.081
MAE	0.238	0.124	0.107	0.099	0.098	0.055	0.159	0.083	0.076	0.075	0.098	0.041
	LEV						FRE*					
Median	0.205	0.250	0.284	0.315	0.334	0.305	0.204	0.198	0.197	0.183	0.334	0.211
IQR	0.371	0.277	0.237	0.215	0.202	0.139	0.175	0.153	0.151	0.171	0.202	0.080
MAE	0.176	0.147	0.144	0.150	0.158	0.109	0.087	0.077	0.076	0.087	0.158	0.040
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.209	0.232	0.247	0.263	0.275	0.260	0.202	0.200	0.194	0.186	0.234	0.200
IQR	0.336	0.214	0.181	0.169	0.159	0.109	0.156	0.117	0.106	0.119	0.150	0.076
MAE	0.164	0.111	0.100	0.099	0.101	0.068	0.078	0.059	0.053	0.060	0.080	0.038

Table 3.2: Table 3.2 Simulation results for AR(1) model : $T = 6,\ N = 200$, $\alpha = 0.5$, $\sigma_\eta^2 = 0.2,\ 1,\ 5$

$T = 6,\ N = 200,\ \alpha = 0.5,\ \sigma_\eta^2 = 0.2$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.504	0.484	0.457	0.437	0.429	0.445	0.500	0.499	0.491	0.483	0.429	0.485
IQR	0.293	0.287	0.264	0.253	0.252	0.099	0.140	0.151	0.161	0.186	0.252	0.077
MAE	0.145	0.144	0.140	0.136	0.142	0.064	0.070	0.076	0.081	0.095	0.142	0.039
	LEV						FRE*					
Median	0.504	0.503	0.509	0.510	0.510	0.508	0.504	0.497	0.503	0.505	0.510	0.503
IQR	0.182	0.141	0.128	0.114	0.112	0.064	0.194	0.155	0.139	0.124	0.112	0.062
MAE	0.091	0.070	0.064	0.058	0.056	0.033	0.097	0.077	0.069	0.063	0.056	0.032
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.502	0.494	0.492	0.490	0.488	0.492	0.500	0.497	0.496	0.497	0.495	0.498
IQR	0.180	0.137	0.118	0.107	0.102	0.063	0.121	0.114	0.107	0.107	0.100	0.059
MAE	0.090	0.069	0.059	0.053	0.051	0.031	0.060	0.057	0.054	0.053	0.050	0.030

$T = 6,\ N = 200,\ \alpha = 0.5,\ \sigma_\eta^2 = 1$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.499	0.470	0.439	0.422	0.414	0.419	0.501	0.499	0.487	0.479	0.414	0.477
IQR	0.466	0.365	0.313	0.282	0.272	0.119	0.227	0.192	0.191	0.208	0.272	0.096
MAE	0.229	0.186	0.166	0.155	0.155	0.087	0.113	0.096	0.096	0.106	0.155	0.050
	LEV						FRE*					
Median	0.505	0.517	0.530	0.538	0.545	0.535	0.506	0.495	0.499	0.502	0.545	0.514
IQR	0.241	0.178	0.155	0.138	0.135	0.088	0.223	0.187	0.177	0.173	0.135	0.079
MAE	0.117	0.090	0.083	0.079	0.079	0.052	0.112	0.093	0.089	0.086	0.079	0.041
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.503	0.506	0.510	0.515	0.517	0.514	0.500	0.494	0.490	0.492	0.517	0.501
IQR	0.238	0.171	0.143	0.127	0.124	0.083	0.165	0.141	0.130	0.135	0.124	0.076
MAE	0.116	0.085	0.073	0.065	0.063	0.043	0.083	0.071	0.065	0.068	0.063	0.038

$T = 6,\ N = 200,\ \alpha = 0.5,\ \sigma_\eta^2 = 5$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.436	0.457	0.431	0.417	0.409	0.401	0.500	0.499	0.486	0.476	0.409	0.469
IQR	0.933	0.418	0.336	0.294	0.278	0.132	0.458	0.221	0.201	0.217	0.278	0.109
MAE	0.459	0.216	0.182	0.165	0.159	0.103	0.229	0.110	0.102	0.110	0.159	0.059
	LEV						FRE*					
Median	0.534	0.591	0.629	0.653	0.672	0.643	0.506	0.492	0.491	0.472	0.672	0.529
IQR	0.409	0.251	0.199	0.165	0.149	0.125	0.237	0.206	0.211	0.239	0.149	0.100
MAE	0.190	0.153	0.153	0.166	0.177	0.144	0.118	0.103	0.106	0.124	0.177	0.055
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.522	0.566	0.593	0.615	0.628	0.609	0.500	0.494	0.484	0.471	0.583	0.505
IQR	0.376	0.232	0.187	0.157	0.146	0.121	0.212	0.153	0.145	0.161	0.154	0.095
MAE	0.179	0.132	0.127	0.131	0.136	0.113	0.106	0.078	0.074	0.084	0.105	0.048

Table 3.3: Table 3.3 Simulation results for AR(1) model : $T = 6,\ N = 200$, $\alpha = 0.8$, $\sigma_\eta^2 = 0.2,\ 1,\ 5$

$T = 6,\ N = 200,\ \alpha = 0.8,\ \sigma_\eta^2 = 0.2$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.800	0.695	0.610	0.560	0.531	0.576	0.800	0.790	0.764	0.715	0.531	0.747
IQR	0.683	0.591	0.502	0.458	0.434	0.176	0.227	0.251	0.269	0.329	0.434	0.127
MAE	0.330	0.316	0.314	0.312	0.316	0.224	0.114	0.126	0.138	0.181	0.316	0.074
	LEV						FRE*					
Median	0.805	0.807	0.814	0.819	0.817	0.816	0.806	0.798	0.807	0.815	0.817	0.811
IQR	0.204	0.140	0.119	0.106	0.098	0.061	0.271	0.182	0.148	0.123	0.098	0.062
MAE	0.101	0.071	0.061	0.055	0.052	0.034	0.136	0.091	0.074	0.063	0.052	0.033
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.793	0.790	0.796	0.799	0.795	0.797	0.794	0.789	0.793	0.802	0.803	0.799
IQR	0.216	0.146	0.119	0.101	0.096	0.064	0.184	0.154	0.133	0.114	0.095	0.063
MAE	0.103	0.071	0.059	0.051	0.047	0.032	0.091	0.077	0.067	0.057	0.048	0.032

$T = 6,\ N = 200,\ \alpha = 0.8,\ \sigma_\eta^2 = 1$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.652	0.553	0.507	0.472	0.458	0.460	0.799	0.778	0.738	0.679	0.458	0.697
IQR	1.218	0.776	0.592	0.517	0.465	0.227	0.445	0.353	0.341	0.386	0.465	0.172
MAE	0.618	0.452	0.404	0.393	0.380	0.340	0.222	0.177	0.181	0.222	0.380	0.117
	LEV						FRE*					
Median	0.836	0.845	0.860	0.867	0.870	0.863	0.808	0.792	0.806	0.834	0.870	0.843
IQR	0.244	0.156	0.117	0.103	0.091	0.065	0.394	0.279	0.237	0.184	0.091	0.072
MAE	0.126	0.089	0.082	0.079	0.078	0.066	0.197	0.140	0.119	0.095	0.078	0.052
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.812	0.827	0.840	0.847	0.849	0.846	0.785	0.775	0.777	0.804	0.849	0.820
IQR	0.264	0.155	0.118	0.102	0.091	0.070	0.287	0.230	0.194	0.164	0.091	0.079
MAE	0.118	0.084	0.073	0.068	0.064	0.055	0.143	0.115	0.098	0.082	0.064	0.044

$T = 6,\ N = 200,\ \alpha = 0.8,\ \sigma_\eta^2 = 5$

t	1	2	3	4	5	GMM	1	2	3	4	5	GMM
	DIF						FOD					
Median	0.217	0.450	0.449	0.428	0.433	0.403	0.748	0.765	0.725	0.665	0.433	0.659
IQR	1.865	0.865	0.631	0.542	0.474	0.247	0.944	0.425	0.377	0.409	0.474	0.200
MAE	1.082	0.547	0.456	0.430	0.404	0.397	0.471	0.215	0.200	0.237	0.404	0.151
	LEV						FRE*					
Median	0.919	0.930	0.939	0.945	0.947	0.940	0.806	0.768	0.763	0.796	0.947	0.891
IQR	0.237	0.128	0.092	0.078	0.067	0.055	0.487	0.383	0.357	0.326	0.067	0.077
MAE	0.161	0.138	0.142	0.146	0.147	0.140	0.240	0.193	0.184	0.163	0.147	0.094
	SYS(DIF & LEV)						SYS(FOD & FRE*)					
Median	0.908	0.919	0.929	0.936	0.938	0.932	0.771	0.756	0.731	0.741	0.921	0.850
IQR	0.217	0.121	0.089	0.075	0.066	0.056	0.391	0.282	0.251	0.243	0.071	0.097
MAE	0.153	0.130	0.131	0.136	0.138	0.132	0.198	0.145	0.134	0.131	0.122	0.067

Table 3.4: Simulation results for dynamic panel data models with an endogenous variable ($N = 100,\ \rho = 0.8,\ SNR = 3$)

	α				β				α				β			
Model	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size
	$T = 6,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 1.39,\ VR = 10$								$T = 6,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 6.951,\ VR = 50$							
DIF	-0.039	0.099	0.106	0.074	-0.003	0.117	0.117	0.060	-0.049	0.109	0.120	0.076	-0.003	0.118	0.118	0.061
FOD	-0.024	0.092	0.095	0.063	-0.002	0.114	0.114	0.055	-0.027	0.097	0.101	0.061	-0.002	0.114	0.114	0.054
LEV	0.055	0.080	0.097	0.145	0.025	0.129	0.132	0.070	0.157	0.098	0.185	0.439	0.097	0.155	0.182	0.120
FRE*	0.023	0.082	0.086	0.083	0.032	0.136	0.140	0.067	0.042	0.097	0.105	0.114	0.103	0.180	0.208	0.120
FRE	0.018	0.086	0.088	0.087	0.023	0.141	0.143	0.063	0.039	0.105	0.112	0.131	0.085	0.190	0.207	0.116
SYS(DIF & LEV)	0.034	0.069	0.077	0.101	0.007	0.100	0.101	0.058	0.125	0.083	0.150	0.375	0.047	0.112	0.121	0.088
SYS(FOD & LEV)	0.033	0.069	0.076	0.101	0.007	0.099	0.100	0.064	0.120	0.084	0.147	0.355	0.043	0.111	0.119	0.086
SYS(FOD & FRE*)	0.010	0.070	0.071	0.068	-0.001	0.101	0.101	0.051	0.010	0.081	0.082	0.060	0.004	0.108	0.109	0.058
SYS(FOD & FRE)	0.006	0.073	0.074	0.077	-0.001	0.102	0.102	0.052	0.010	0.089	0.090	0.083	0.004	0.115	0.115	0.054
	$T = 6,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 0.64,\ VR = 10$								$T = 6,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 3.18,\ VR = 50$							
DIF	-0.334	0.217	0.399	0.298	-0.134	0.153	0.203	0.133	-0.400	0.225	0.459	0.382	-0.158	0.158	0.224	0.168
FOD	-0.245	0.191	0.310	0.193	-0.091	0.137	0.164	0.090	-0.284	0.202	0.348	0.238	-0.104	0.143	0.177	0.099
LEV	0.070	0.052	0.088	0.406	0.003	0.086	0.086	0.046	0.125	0.035	0.130	0.887	0.013	0.085	0.086	0.049
FRE*	0.051	0.062	0.080	0.231	0.007	0.102	0.102	0.052	0.087	0.063	0.108	0.506	0.034	0.131	0.135	0.055
FRE	-0.008	0.128	0.128	0.159	-0.013	0.144	0.144	0.058	0.007	0.162	0.162	0.343	0.002	0.179	0.179	0.061
SYS(DIF & LEV)	0.043	0.059	0.073	0.203	0.002	0.079	0.079	0.060	0.108	0.041	0.116	0.742	0.019	0.077	0.079	0.058
SYS(FOD & LEV)	0.045	0.059	0.074	0.215	0.001	0.077	0.077	0.062	0.110	0.041	0.117	0.751	0.016	0.075	0.077	0.058
SYS(FOD & FRE*)	0.003	0.073	0.074	0.079	-0.007	0.083	0.084	0.062	-0.005	0.096	0.096	0.116	-0.006	0.094	0.094	0.065
SYS(FOD & FRE)	-0.061	0.125	0.139	0.124	-0.025	0.097	0.100	0.070	-0.091	0.174	0.196	0.260	-0.030	0.113	0.117	0.080
	$T = 12,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 1.39,\ VR = 10$								$T = 12,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 6.95,\ VR = 50$							
DIF	-0.018	0.050	0.053	0.063	-0.005	0.062	0.062	0.049	-0.032	0.059	0.067	0.083	-0.007	0.062	0.062	0.051
FOD	-0.002	0.047	0.047	0.049	-0.007	0.058	0.058	0.058	-0.003	0.048	0.048	0.050	-0.007	0.058	0.058	0.054
LEV	0.080	0.046	0.092	0.427	0.008	0.073	0.073	0.056	0.199	0.058	0.207	0.936	0.051	0.085	0.099	0.109
FRE*	0.014	0.045	0.047	0.070	0.001	0.077	0.077	0.058	0.016	0.048	0.050	0.077	0.008	0.087	0.088	0.064
FRE	0.013	0.047	0.049	0.077	0.000	0.079	0.079	0.060	0.012	0.054	0.055	0.079	0.006	0.098	0.098	0.062
SYS(DIF & LEV)	0.033	0.042	0.054	0.130	0.010	0.060	0.061	0.060	0.138	0.052	0.147	0.791	0.050	0.068	0.085	0.127
SYS(FOD & LEV)	0.035	0.042	0.055	0.144	0.000	0.057	0.057	0.055	0.128	0.051	0.138	0.757	0.020	0.065	0.068	0.070
SYS(FOD & FRE*)	0.003	0.043	0.043	0.048	-0.007	0.056	0.056	0.057	0.002	0.045	0.045	0.046	-0.007	0.057	0.057	0.058
SYS(FOD & FRE)	0.002	0.043	0.043	0.055	-0.007	0.056	0.057	0.060	0.000	0.049	0.049	0.059	-0.004	0.144	0.144	0.061
	$T = 12,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 0.64,\ VR = 10$								$T = 12,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 3.18,\ VR = 50$							
DIF	-0.168	0.118	0.205	0.330	-0.067	0.078	0.103	0.144	-0.260	0.136	0.294	0.515	-0.109	0.087	0.140	0.258
FOD	-0.069	0.074	0.101	0.131	-0.018	0.051	0.054	0.059	-0.076	0.078	0.109	0.142	-0.020	0.052	0.056	0.054
LEV	0.080	0.029	0.086	0.766	0.002	0.048	0.048	0.048	0.133	0.019	0.134	0.997	0.009	0.047	0.048	0.049
FRE*	0.030	0.044	0.054	0.147	0.003	0.065	0.065	0.054	0.043	0.051	0.067	0.226	0.012	0.078	0.079	0.056
FRE	0.020	0.066	0.069	0.264	-0.004	0.070	0.070	0.061	0.029	0.087	0.092	0.350	0.003	0.088	0.088	0.068
SYS(DIF & LEV)	0.047	0.036	0.059	0.298	0.008	0.048	0.049	0.052	0.114	0.023	0.117	0.978	0.028	0.047	0.054	0.091
SYS(FOD & LEV)	0.049	0.035	0.061	0.333	0.004	0.042	0.042	0.056	0.115	0.023	0.117	0.980	0.016	0.041	0.044	0.066
SYS(FOD & FRE*)	-0.015	0.051	0.053	0.052	-0.009	0.046	0.047	0.060	-0.021	0.059	0.063	0.054	-0.010	0.048	0.049	0.055
SYS(FOD & FRE)	-0.017	0.067	0.069	0.143	-0.010	0.047	0.048	0.062	-0.016	0.086	0.088	0.220	-0.009	0.053	0.053	0.060

Table 3.5: Simulation results for dynamic panel data models with an endogenous variable ($N = 250,\ \rho = 0.8,\ SNR = 3$)

Model	α				β				α				β			
	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size
	$T = 6,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 1.39,\ VR = 10$								$T = 6,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 6.95,\ VR = 50$							
DIF	-0.016	0.063	0.065	0.054	-0.001	0.073	0.073	0.049	-0.022	0.073	0.076	0.058	-0.001	0.073	0.073	0.049
FOD	-0.010	0.059	0.059	0.052	-0.001	0.072	0.072	0.051	-0.011	0.063	0.064	0.052	-0.001	0.072	0.072	0.053
LEV	0.021	0.050	0.055	0.082	0.016	0.086	0.087	0.058	0.070	0.067	0.097	0.241	0.082	0.112	0.139	0.165
FRE*	0.012	0.051	0.052	0.063	0.016	0.093	0.094	0.067	0.022	0.060	0.064	0.066	0.061	0.128	0.142	0.100
FRE	0.010	0.051	0.052	0.062	0.013	0.094	0.095	0.065	0.021	0.062	0.065	0.075	0.056	0.130	0.142	0.101
SYS(DIF & LEV)	0.015	0.042	0.044	0.074	0.002	0.064	0.064	0.053	0.056	0.054	0.078	0.211	0.028	0.072	0.077	0.091
SYS(FOD & LEV)	0.014	0.042	0.044	0.070	0.003	0.062	0.062	0.055	0.052	0.053	0.075	0.196	0.025	0.070	0.074	0.088
SYS(FOD & FRE*)	0.008	0.042	0.043	0.058	0.000	0.063	0.063	0.051	0.010	0.049	0.050	0.062	0.003	0.069	0.070	0.056
SYS(FOD & FRE)	0.007	0.043	0.043	0.060	0.000	0.063	0.063	0.053	0.009	0.050	0.051	0.064	0.004	0.069	0.069	0.054
	$T = 6,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 0.64,\ VR = 10$								$T = 6,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 3.18,\ VR = 50$							
DIF	-0.188	0.153	0.242	0.187	-0.080	0.095	0.124	0.100	-0.256	0.170	0.307	0.279	-0.110	0.106	0.152	0.142
FOD	-0.132	0.131	0.186	0.124	-0.051	0.082	0.097	0.067	-0.165	0.145	0.220	0.158	-0.064	0.089	0.110	0.076
LEV	0.039	0.041	0.057	0.273	0.006	0.058	0.059	0.048	0.099	0.037	0.106	0.787	0.020	0.062	0.065	0.052
FRE*	0.030	0.046	0.055	0.174	0.007	0.067	0.067	0.052	0.067	0.056	0.087	0.410	0.029	0.090	0.094	0.059
FRE	0.014	0.063	0.065	0.143	0.002	0.078	0.078	0.048	0.035	0.099	0.105	0.357	0.010	0.112	0.112	0.052
SYS(DIF & LEV)	0.025	0.043	0.050	0.147	0.002	0.050	0.050	0.052	0.083	0.041	0.092	0.622	0.017	0.051	0.054	0.069
SYS(FOD & LEV)	0.025	0.043	0.049	0.155	0.002	0.048	0.048	0.060	0.083	0.041	0.093	0.622	0.016	0.050	0.053	0.082
SYS(FOD & FRE*)	0.009	0.048	0.049	0.091	-0.002	0.051	0.051	0.058	0.008	0.072	0.072	0.121	-0.001	0.060	0.060	0.055
SYS(FOD & FRE)	-0.007	0.062	0.062	0.106	-0.005	0.053	0.054	0.049	-0.019	0.104	0.106	0.224	-0.008	0.065	0.065	0.065
	$T = 12,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 1.39,\ VR = 10$								$T = 12,\ \alpha = 0.2,\ \beta = 0.8,\ \sigma_\eta^2 = 6.95,\ VR = 50$							
DIF	-0.006	0.033	0.034	0.061	-0.003	0.042	0.042	0.057	-0.013	0.040	0.042	0.065	-0.003	0.042	0.042	0.059
FOD	0.000	0.032	0.032	0.056	-0.004	0.039	0.039	0.056	0.001	0.034	0.034	0.055	-0.004	0.039	0.040	0.054
LEV	0.027	0.030	0.040	0.165	0.010	0.050	0.051	0.059	0.083	0.040	0.092	0.611	0.058	0.065	0.087	0.186
FRE*	0.007	0.030	0.031	0.056	0.001	0.055	0.055	0.056	0.008	0.032	0.033	0.058	0.005	0.063	0.063	0.057
FRE	0.007	0.031	0.031	0.059	0.000	0.055	0.055	0.058	0.008	0.033	0.034	0.059	0.005	0.065	0.065	0.063
SYS(DIF & LEV)	0.014	0.027	0.030	0.091	0.001	0.041	0.041	0.054	0.059	0.035	0.069	0.454	0.021	0.046	0.051	0.092
SYS(FOD & LEV)	0.012	0.026	0.028	0.084	-0.001	0.037	0.037	0.061	0.051	0.033	0.061	0.398	0.008	0.042	0.042	0.075
SYS(FOD & FRE*)	0.004	0.027	0.027	0.056	-0.003	0.037	0.037	0.065	0.004	0.030	0.031	0.051	-0.002	0.039	0.039	0.060
SYS(FOD & FRE)	0.003	0.028	0.028	0.062	-0.002	0.037	0.038	0.062	0.002	0.031	0.031	0.054	-0.003	0.039	0.039	0.062
	$T = 12,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 0.64,\ VR = 10$								$T = 12,\ \alpha = 0.8,\ \beta = 0.2,\ \sigma_\eta^2 = 3.18,\ VR = 50$							
DIF	-0.080	0.074	0.109	0.167	-0.033	0.047	0.058	0.097	-0.162	0.103	0.191	0.348	-0.071	0.060	0.093	0.216
FOD	-0.034	0.052	0.062	0.097	-0.009	0.032	0.033	0.056	-0.038	0.056	0.068	0.105	-0.010	0.032	0.034	0.053
LEV	0.045	0.024	0.051	0.495	0.006	0.034	0.034	0.048	0.104	0.022	0.106	0.981	0.017	0.035	0.039	0.072
FRE*	0.018	0.033	0.038	0.094	0.004	0.044	0.044	0.048	0.030	0.044	0.054	0.174	0.014	0.055	0.057	0.060
FRE	0.015	0.040	0.043	0.171	0.001	0.045	0.045	0.048	0.029	0.056	0.063	0.274	0.010	0.058	0.059	0.068
SYS(DIF & LEV)	0.027	0.027	0.038	0.208	0.004	0.031	0.031	0.055	0.087	0.025	0.091	0.911	0.021	0.032	0.038	0.106
SYS(FOD & LEV)	0.026	0.026	0.037	0.216	0.003	0.027	0.027	0.062	0.087	0.025	0.090	0.912	0.016	0.027	0.031	0.099
SYS(FOD & FRE*)	0.002	0.031	0.031	0.059	-0.002	0.029	0.029	0.060	-0.001	0.043	0.043	0.067	-0.003	0.031	0.031	0.057
SYS(FOD & FRE)	0.001	0.036	0.036	0.102	-0.002	0.029	0.029	0.061	-0.001	0.052	0.052	0.145	-0.002	0.032	0.032	0.063

Table 3.6: Simulation results for dynamic panel data models with an endogenous variable ($N = 500,\ \rho = 0.8,\ SNR = 3$)

	α				β				α				β			
Model	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size	Bias	SD	RMSE	Size
	$T=6,\ \alpha=0.2,\ \beta=0.8,\ \sigma_\eta^2=1.39,\ VR=10$								$T=6,\ \alpha=0.2,\ \beta=0.8,\ \sigma_\eta^2=6.95,\ VR=50$							
DIF	-0.007	0.045	0.046	0.052	-0.001	0.051	0.051	0.046	-0.010	0.053	0.053	0.062	-0.001	0.051	0.051	0.049
FOD	-0.004	0.042	0.042	0.051	-0.001	0.050	0.050	0.046	-0.004	0.045	0.045	0.049	-0.001	0.050	0.050	0.044
LEV	0.010	0.035	0.036	0.064	0.007	0.060	0.060	0.047	0.032	0.045	0.056	0.136	0.051	0.087	0.100	0.107
FRE*	0.007	0.035	0.035	0.051	0.005	0.064	0.064	0.042	0.012	0.042	0.043	0.065	0.031	0.094	0.099	0.061
FRE	0.006	0.035	0.035	0.050	0.003	0.065	0.065	0.039	0.011	0.042	0.044	0.062	0.028	0.095	0.099	0.061
SYS(DIF & LEV)	0.007	0.027	0.028	0.059	0.001	0.045	0.045	0.048	0.023	0.033	0.040	0.106	0.014	0.051	0.053	0.075
SYS(FOD & LEV)	0.006	0.028	0.028	0.058	0.001	0.044	0.044	0.054	0.021	0.033	0.039	0.099	0.013	0.049	0.051	0.081
SYS(FOD & FRE*)	0.005	0.028	0.028	0.059	0.000	0.044	0.044	0.053	0.007	0.032	0.033	0.064	0.002	0.048	0.048	0.053
SYS(FOD & FRE)	0.005	0.028	0.029	0.055	0.000	0.044	0.044	0.052	0.006	0.032	0.033	0.062	0.002	0.048	0.048	0.055
	$T=6,\ \alpha=0.8,\ \beta=0.2,\ \sigma_\eta^2=0.64,\ VR=10$								$T=6,\ \alpha=0.8,\ \beta=0.2,\ \sigma_\eta^2=3.18,\ VR=50$							
DIF	-0.096	0.113	0.149	0.125	-0.042	0.068	0.080	0.079	-0.149	0.134	0.200	0.188	-0.066	0.078	0.102	0.114
FOD	-0.067	0.097	0.118	0.098	-0.027	0.059	0.065	0.064	-0.089	0.108	0.140	0.113	-0.036	0.063	0.073	0.067
LEV	0.022	0.030	0.037	0.175	0.002	0.042	0.043	0.047	0.071	0.034	0.078	0.639	0.017	0.051	0.054	0.073
FRE*	0.018	0.033	0.037	0.117	0.002	0.047	0.047	0.043	0.047	0.045	0.065	0.323	0.018	0.067	0.069	0.051
FRE	0.013	0.039	0.041	0.121	0.000	0.051	0.051	0.038	0.037	0.063	0.073	0.331	0.010	0.075	0.076	0.053
SYS(DIF & LEV)	0.014	0.029	0.033	0.103	0.001	0.036	0.036	0.053	0.056	0.036	0.066	0.464	0.014	0.040	0.042	0.089
SYS(FOD & LEV)	0.014	0.029	0.033	0.100	0.001	0.035	0.035	0.057	0.056	0.036	0.067	0.461	0.014	0.039	0.041	0.088
SYS(FOD & FRE*)	0.008	0.031	0.032	0.082	-0.001	0.036	0.036	0.054	0.014	0.048	0.050	0.116	0.002	0.043	0.043	0.062
SYS(FOD & FRE)	0.004	0.034	0.035	0.086	-0.002	0.037	0.037	0.055	0.008	0.062	0.062	0.206	0.001	0.045	0.045	0.067
	$T=12,\ \alpha=0.2,\ \beta=0.8,\ \sigma_\eta^2=1.39,\ VR=10$								$T=12,\ \alpha=0.2,\ \beta=0.8,\ \sigma_\eta^2=6.95,\ VR=50$							
DIF	-0.002	0.023	0.023	0.051	-0.002	0.029	0.029	0.047	-0.006	0.028	0.029	0.051	-0.002	0.029	0.029	0.046
FOD	0.000	0.022	0.022	0.050	-0.002	0.027	0.027	0.046	0.000	0.023	0.023	0.048	-0.002	0.027	0.027	0.049
LEV	0.012	0.020	0.024	0.092	0.007	0.037	0.038	0.067	0.034	0.026	0.043	0.300	0.041	0.053	0.067	0.185
FRE*	0.004	0.021	0.021	0.052	0.001	0.040	0.040	0.058	0.004	0.022	0.023	0.053	0.004	0.046	0.046	0.054
FRE	0.004	0.021	0.021	0.052	0.001	0.040	0.040	0.059	0.004	0.022	0.023	0.055	0.004	0.046	0.047	0.057
SYS(DIF & LEV)	0.006	0.018	0.020	0.067	0.000	0.029	0.029	0.055	0.023	0.021	0.032	0.219	0.009	0.032	0.033	0.070
SYS(FOD & LEV)	0.005	0.017	0.017	0.060	0.000	0.025	0.025	0.052	0.019	0.019	0.027	0.185	0.004	0.028	0.028	0.073
SYS(FOD & FRE*)	0.003	0.018	0.018	0.046	-0.001	0.025	0.025	0.048	0.002	0.020	0.020	0.045	0.000	0.026	0.026	0.051
SYS(FOD & FRE)	0.002	0.018	0.018	0.050	0.000	0.025	0.025	0.049	0.001	0.020	0.020	0.047	-0.001	0.026	0.026	0.050
	$T=12,\ \alpha=0.8,\ \beta=0.2,\ \sigma_\eta^2=0.64,\ VR=10$								$T=12,\ \alpha=0.8,\ \beta=0.2,\ \sigma_\eta^2=3.178,\ VR=50$							
DIF	-0.040	0.048	0.063	0.111	-0.017	0.030	0.035	0.069	-0.093	0.072	0.118	0.235	-0.041	0.040	0.058	0.159
FOD	-0.019	0.037	0.041	0.076	-0.004	0.022	0.022	0.052	-0.021	0.040	0.045	0.080	-0.005	0.022	0.023	0.050
LEV	0.024	0.019	0.030	0.300	0.005	0.026	0.026	0.067	0.072	0.021	0.075	0.911	0.020	0.030	0.036	0.135
FRE*	0.010	0.024	0.026	0.082	0.003	0.032	0.032	0.053	0.019	0.034	0.039	0.119	0.010	0.041	0.042	0.058
FRE	0.008	0.027	0.028	0.119	0.002	0.032	0.032	0.056	0.018	0.038	0.042	0.173	0.009	0.042	0.043	0.068
SYS(DIF & LEV)	0.014	0.019	0.023	0.130	0.002	0.022	0.022	0.054	0.057	0.023	0.062	0.751	0.017	0.024	0.029	0.134
SYS(FOD & LEV)	0.013	0.018	0.022	0.138	0.002	0.019	0.020	0.061	0.056	0.023	0.061	0.748	0.015	0.021	0.026	0.174
SYS(FOD & FRE*)	0.004	0.020	0.020	0.066	0.000	0.020	0.020	0.052	0.003	0.028	0.029	0.063	0.001	0.022	0.022	0.053
SYS(FOD & FRE)	0.003	0.021	0.021	0.079	0.000	0.020	0.020	0.051	0.002	0.031	0.031	0.092	0.000	0.022	0.022	0.056

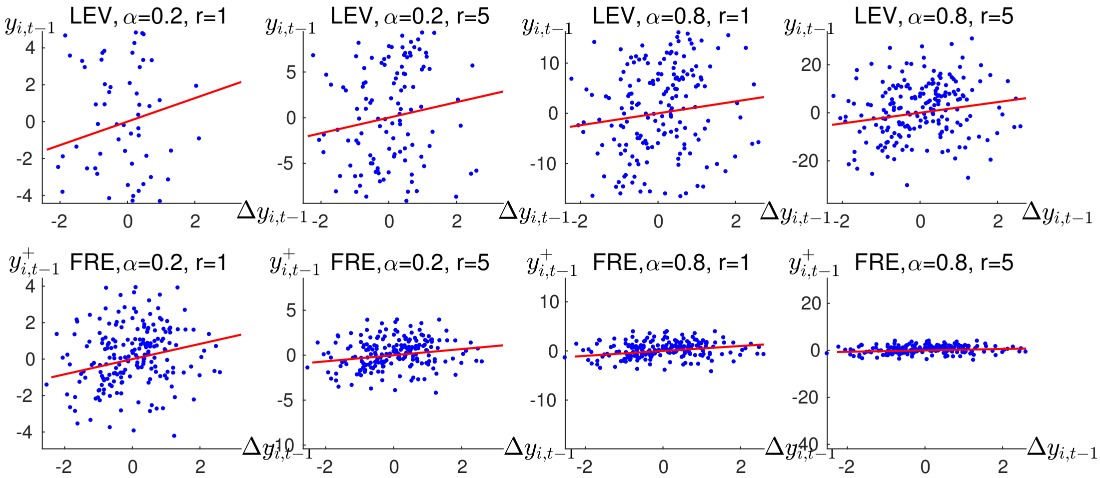

Figure 3.1: Scatter plot of the first stage regression for LEV and FRE models at $(t, T) = (4, 6)$

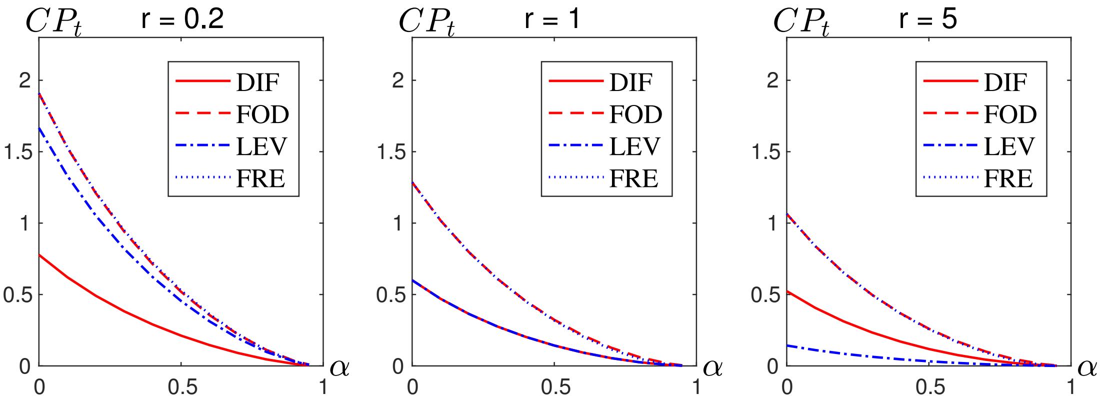

Figure 3.2: CP for cross section at $(t, T) = (4, 6)$

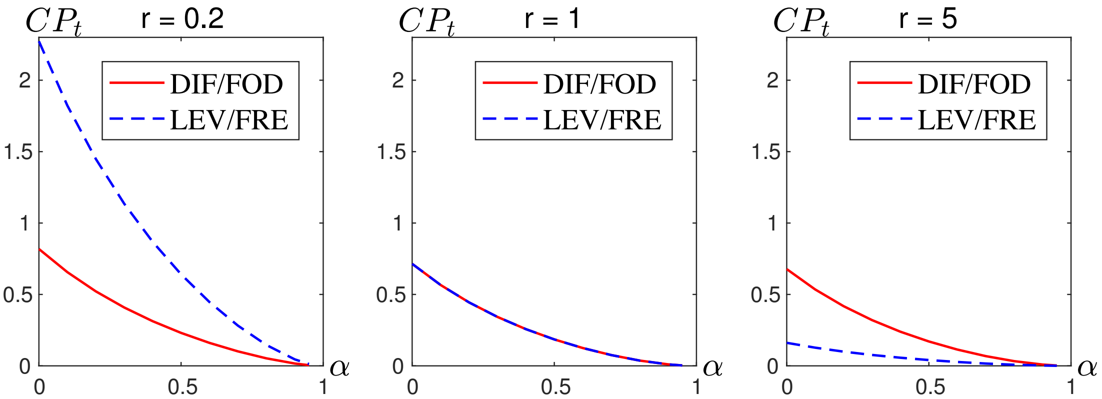

Figure 3.3: CP for cross section at $(t, T) = (6, 6)$

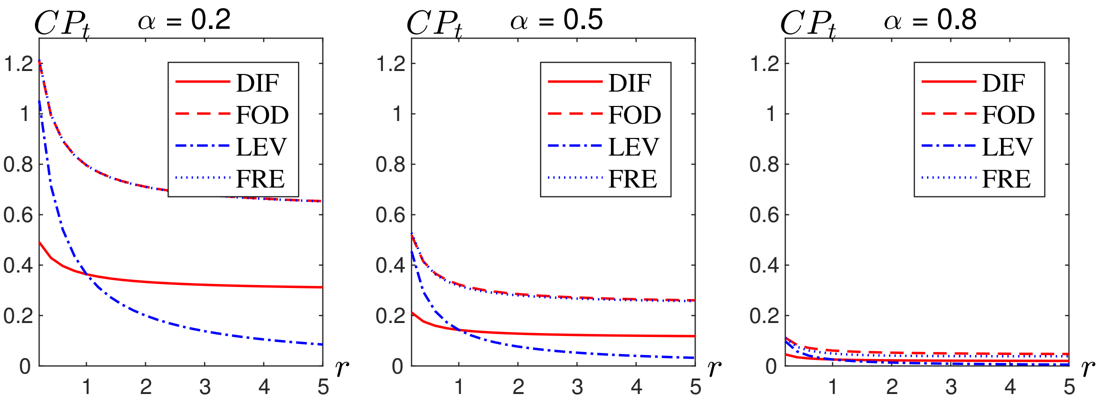

Figure 3.4: CP for cross section at $(t, T) = (4, 6)$

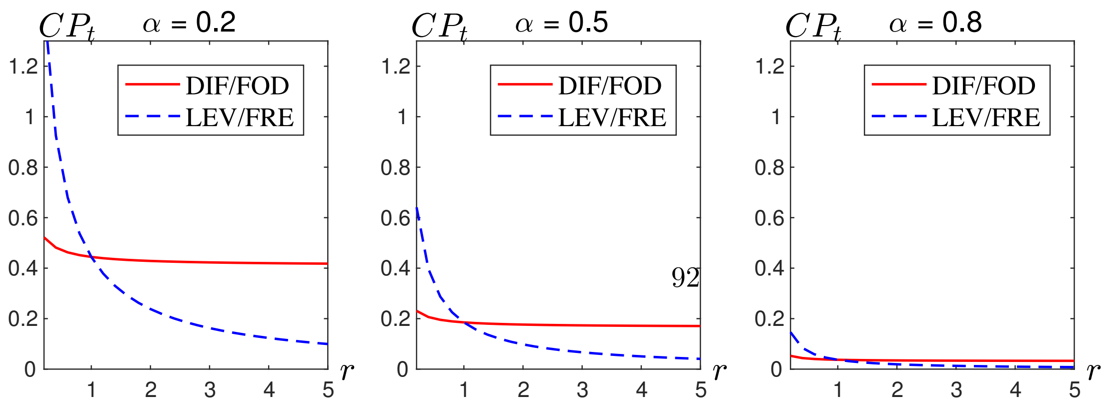

Figure 3.5: CP for cross section at $(t, T) = (6, 6)$

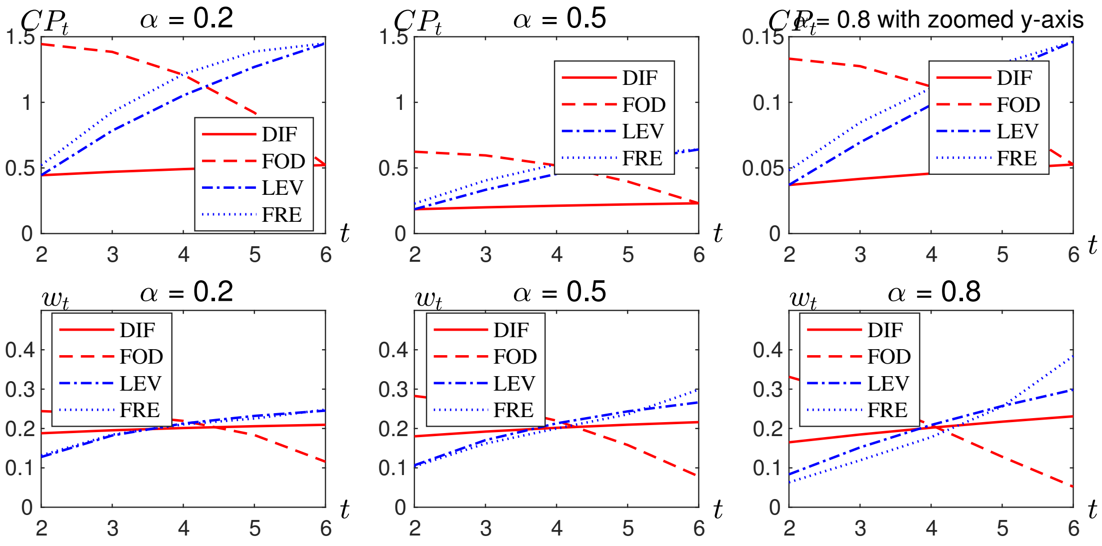

Figure 3.6: CP and weight for $r = 0.2$ and $T = 6$

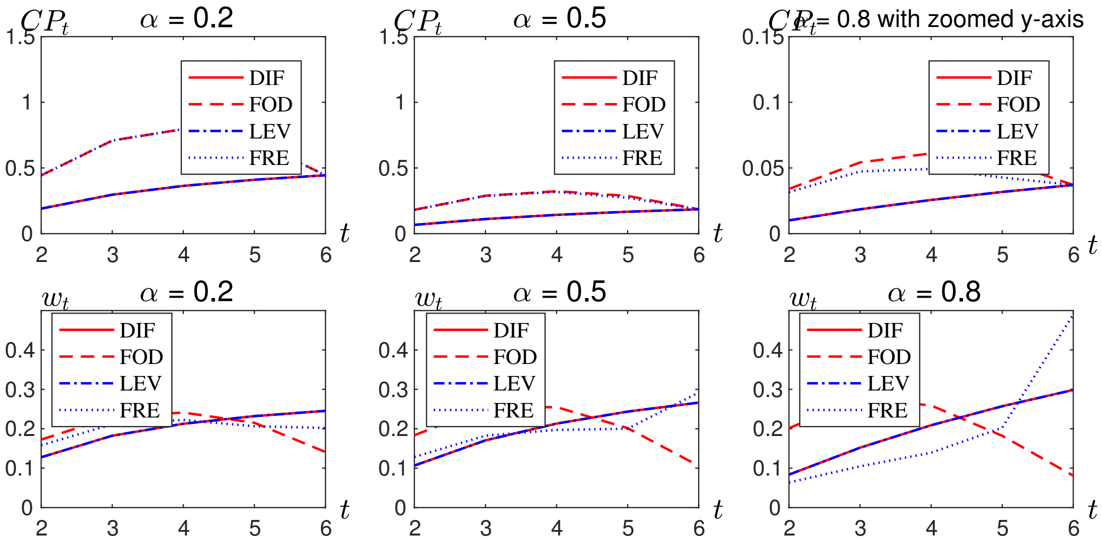

Figure 3.7: CP and weight for $r = 1$ and $T = 6$

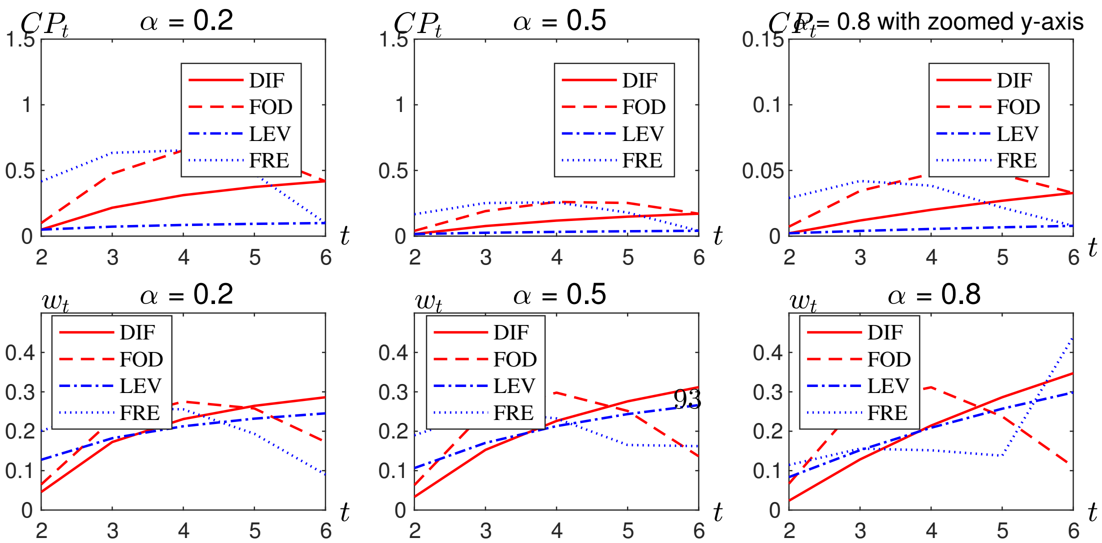

Figure 3.8: CP and weight for $r = 5$ and $T = 6$

Chapter 4

Higher-Order Bias-Corrected Estimation of Heterogeneous Dynamic Panel Data Models

In this chapter, we propose a bias-corrected mean-group estimator for dynamic heterogeneous panel data models. Specifically, we propose to use the second-order bias-corrected estimator by Kiviet and Phillips [2012]when constructing the mean-group estimator. Monte Carlo simulation is conducted and it is confirmed that the proposed estimator has smaller bias than the conentional and first-order bias-corrected mean-group estimators when the length of panel data is not so large.

4.1 Introduction

Panel data analysis has become a standard approach in empirical economic studies thanks to the availability of many panel data sets, and several panel data models have been proposed in the literature. Among those models, we focus on the dynamic panel data models. While linear dynamic panel data models are most widely used, the choice of estimation method changes depending on the sample size of panel data. When the length of panel data, denoted as T, is small and the number of cross-sectional units, denoted as N, is large, we usually allow for cross-sectional heterogeneity in the intercept while other coefficients are assumed to be homogeneous. Such a model is typically estimated by the generalized method of moments (GMM) estimators Arellano and Bond [1991]; Arellano and Bover [1995]; Blundell and Bond [1998]. However, when T is large,@articlekiviet2012higher, title=Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models, author=Kiviet, Jan F and Phillips, Garry DA, journal=Computational Statistics & Data Analysis, volume=56, number=11, pages=3705–3729, year=2012, publisher=Elsevier it is possible to allow for cross-sectional heterogeneity both in the intercept and coefficients. Pesaran and Smith [1995] demonstrate that the GMM estimators assuming homogeneous coefficients become inconsistent if the coefficients are in fact heterogeneous. As a solution to this problem, they proposed the mean-group (MG) estimator where estimated coefficients obtained from time-series regression are averaged over cross-sectional units.

While this MG estimator is consistent when T is large, it is biased in finite samples since OLS estimator of each cross-section is biased. To address this problem, Pesaran and Zhao [1999] proposed to use the first-order bias-corrected estimator by Kiviet and Phillips [1993] which accounts for the bias up to $O(T^{-1})$. Pesaran and Zhao [1999] confirmed through extensive Monte Carlo simulation that this bias-correction is effective in reducing the bias of MG estimator. However, unfortunately, this bias-correction is not always effective especially when persistency is strong. To address this problem, this chapter proposes to use the second-order bias-corrected estimator by Kiviet and Phillips [2012] which accounts for the bias up to $O(T^{-2})$. We conduct Monte Carlo simulation to investigate the effectiveness of second-order bias-correction. Consequently, it is shown that second-order bias-correction successfully reduce the bias of MG estimator compared with naive and first-order bias-corrected estimator.

This chapter is organized as following. In the next section, we introduce the model and MG estimator. In Section 3, first-order and second-order bias-corrected estimators are introduced. In Section 4, Monte Carlo simulation is carried out to investigate the performance of estimators. Finally, in Section 5, we conclude.

4.2 Model and mean-group estimator

Let us consider a dynamic heterogeneous panel data model as follows:

$$y_{it} = \alpha_i + \lambda_i y_{i,t-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \varepsilon_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \tag{4.1}$$

where i denotes the index of cross-section unit, and t denotes the time period. $\mathbf{x}_{it}$ is a $k \times 1$ vector of strictly exogenous variable. We assume that the error term ε_{it} is serially and cross-sectionally uncorrelated with zero mean and variance σ_i^2. The joint process $\{y_{it}, \mathbf{x}_{it}\}$ is assumed to be stable and has started a long time ago. We assume that the parameters λ_i and $\boldsymbol{\beta}_i$ are random and can be characterized as

$$\lambda_i = \lambda + \eta_{1i}, \quad \boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_{2i},$$

where η_{1i} and $\boldsymbol{\eta}_{2i}$ are assumed to have zero means and following covariance matrix:

$$\boldsymbol{\Omega} = \begin{bmatrix} Var(\lambda_i) & Cov(\lambda_i, \boldsymbol{\beta}_i) \\ Cov(\boldsymbol{\beta}_i, \lambda_i) & Var(\boldsymbol{\beta}_i) \end{bmatrix}.$$

The long-run coefficient is defined as

$$\boldsymbol{\theta} = E(\boldsymbol{\theta}_i) = E\left(\frac{\boldsymbol{\beta}_i}{1 - \lambda_i}\right).$$

To estimate parameters, rewrite the model (4.1) as follows:

$$y_{it} = \boldsymbol{\delta}_i' \mathbf{w}_{it} + \varepsilon_{it}$$

where $\mathbf{w}_{it} = (1, y_{i,t-1}, \mathbf{x}_{it}')'$, and $\boldsymbol{\delta}_i = (\alpha_i, \lambda_i, \boldsymbol{\beta}_i')'$. The mean group estimator of $\boldsymbol{\delta}$ based on OLS estimator is defined as

$$\widehat{\boldsymbol{\delta}}_{OLS,MG} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\delta}}_{OLS,i} \tag{4.2}$$

where $\widehat{\boldsymbol{\delta}}_{OLS,i} = \left(\widehat{\alpha}_{OLS,i}, \widehat{\lambda}_{OLS,i}, \widehat{\boldsymbol{\beta}}_{OLS,i}'\right)' = (\mathbf{W}_i'\mathbf{W}_i)^{-1} \mathbf{W}_i'\mathbf{y}_i$ with $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$ and $\mathbf{W}_i = (\mathbf{w}_{i1}, \ldots, \mathbf{w}_{iT})'$. The mean group estimator of long-run coefficient $\boldsymbol{\theta}$ is:

$$\widehat{\boldsymbol{\theta}}_{OLS,MG} = \frac{1}{N} \sum_{i=1}^{N} \frac{\widehat{\boldsymbol{\beta}}_{OLS,i}}{1 - \widehat{\lambda}_{OLS,i}}. \tag{4.3}$$

In the time-series literature, it is well known that $\widehat{\boldsymbol{\delta}}_{OLS,i}$ is biased in finite samples. This problem applies to the MG estimator as well since taking an average over cross-sectional units has no effect on the bias. Hence, to reduce the bias of MG estimator of $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, we need to reduce the bias of estimates for each i. In the next section, we consider two approaches to reduce the bias.

4.3 Bias-corrected estimators

As mentioned above, the OLS estimator of dynamic regression model is biased in finite samples and several approaches to reduce the bias have been proposed. We now introduce two approaches.

4.3.1 First-order bias-correction

Pesaran and Zhao [1999] applies one of the bias-reduction methods, namely, first-order bias-corrected estimator by Kiviet and Phillips [1993], which accounts for the bias up to $O(T^{-1})$, to address the finite sample bias problem of the MG estimator. Specifically, the first-order bias-corrected MG estimator is defined as

$$\widehat{\boldsymbol{\delta}}_{BC1,MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\delta}}_{BC1,i} \tag{4.4}$$

where $\widehat{\boldsymbol{\delta}}_{BC1,i}$ is the bias-corrected estimator by Kiviet and Phillips (1993), which is defined as

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}_{BC1,i} &= \left(\widehat{\alpha}_{BC1,i}, \widehat{\lambda}_{BC1,i}, \widehat{\boldsymbol{\beta}}'_{BC1,i}\right)' = \widehat{\boldsymbol{\delta}}_{OLS,i} - \widehat{\mathbf{b}}^{(1)}_{\delta,i} \\
\widehat{\mathbf{b}}^{(1)}_{\delta,i} &= \left(\widehat{b}^{(1)}_{\alpha,i}, \widehat{b}^{(1)}_{\lambda,i}, \widehat{\mathbf{b}}^{(1)\prime}_{\beta,i}\right)', \\
&= -\widehat{\sigma}^2_i(\widehat{\mathbf{D}})^{-1}\left(\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}(\widehat{\mathbf{D}})^{-1}\mathbf{e}_1 + tr[\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}(\widehat{\mathbf{D}})^{-1}]\mathbf{e}_1 + 2\widehat{\sigma}^2_i\mathbf{e}'_1(\widehat{\mathbf{D}})^{-1}\mathbf{e}_1 tr(\mathbf{C}\mathbf{C}'\mathbf{C})\mathbf{e}_1\right), \\
\widehat{\mathbf{D}} &= \widehat{\mathbf{Z}}'\widehat{\mathbf{Z}} + \widehat{\sigma}^2_i tr(\widehat{\mathbf{C}}'\widehat{\mathbf{C}})\mathbf{e}_1\mathbf{e}'_1, \qquad \widehat{\mathbf{Z}} = \left[y_0\widehat{\mathbf{F}} + \widehat{\mathbf{C}}\mathbf{X}_i\widehat{\boldsymbol{\beta}}_{OLS,i}, \mathbf{X}_i\right], \qquad (4.5) \\
\widehat{\sigma}^2_i &= \frac{(\mathbf{y}_i - \mathbf{W}_i\widehat{\boldsymbol{\delta}}_{OLS,i})'(\mathbf{y}_i - \mathbf{W}_i\widehat{\boldsymbol{\delta}}_{OLS,i})}{T-k-1} \qquad (4.6) \\
\widehat{\mathbf{F}} &= \mathbf{F}(\widehat{\lambda}_{OLS,i}), \qquad \widehat{\mathbf{C}} = \mathbf{C}(\widehat{\lambda}_{OLS,i}), \qquad (4.7) \\
\mathbf{F}(\lambda_i) &= \begin{bmatrix} 1 \\ \lambda_i \\ \lambda_i^2 \\ \vdots \\ \lambda_i^{T-1} \end{bmatrix}, \qquad \mathbf{C}(\lambda_i) = \begin{bmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & 0 & & & & \\ \lambda_i & 1 & 0 & & & & \\ \lambda_i^2 & \lambda_i & 1 & \ddots & & & \\ \vdots & & & & & & \vdots \\ \lambda_i^{T-2} & \cdot & \cdot & \cdot & \lambda_i & 1 & 0 \end{bmatrix}
\end{aligned}$$

and $\mathbf{e}_1 = (1, 0, ..., 0)'$. The first-order bias-corrected estimators for long-run coefficient can be obtained as

$$\widehat{\boldsymbol{\theta}}_{BC1,MG} \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{\boldsymbol{\beta}}_{BC1,i}}{1-\widehat{\lambda}_{BC1,i}}. \tag{4.8}$$

4.3.2 Second-order bias-correction

Recently, Kiviet and Phillips [2012] derived the finite sample bias of OLS estimator of dynamic regression models up to $O(T^{-2})$, which is more accurate than that of Kiviet and Phillips [1993]. This chapter suggests to use this second-order bias-corrected estimator to further reduce the bias of MG estimator. This extension is particularly important since it is often the case that the time-series length of panel data is only moderate.

The second-order bias corrected mean group estimator is defined as

$$\widehat{\boldsymbol{\delta}}_{BC2,MG} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\delta}}_{BC2,i} \tag{4.9}$$

where $\widehat{\boldsymbol{\delta}}_{BC2,i}$ is the bias-corrected estimator by Kiviet and Phillips (2012) defined as

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}_{BC2,i} &= \widehat{\boldsymbol{\delta}}_{OLS,i} - \widehat{\mathbf{b}}^{(2)}_{\delta,i} \\
\widehat{\mathbf{b}}^{(2)}_{\delta,i} &= \left(\widehat{b}^{(2)}_{\alpha,i}, \widehat{b}^{(2)}_{\lambda,i}, \widehat{\mathbf{b}}^{(2)\prime}_{\beta,i}\right)', \\
&= -\widehat{\sigma}_i^2[tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}})\mathbf{q}_1 + \mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1] + \widehat{\sigma}_i^4\{[-2q_{11}tr(\mathbf{G}\mathbf{G}'\mathbf{C}) + 2q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\mathbf{C}\widehat{\mathbf{Z}}) \\
&\quad +2q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\mathbf{C}'\widehat{\mathbf{Z}})] - 2q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}) \qquad (4.10) \\
&\quad -q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}})tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}) + 4\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1 + 2\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\mathbf{C}'\widehat{\mathbf{Z}}\mathbf{q}_1 \\
&\quad -4\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1 - 2\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}'\widehat{\mathbf{Z}}\mathbf{q}_1 - \mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}\widehat{\mathbf{Z}}\mathbf{q}_1 tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}) \\
&\quad -2\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1 tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}})]\mathbf{q}_1 - [q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}) + \mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}\widehat{\mathbf{Z}}\mathbf{q}_1]\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1 \\
&\quad -2[q_{11}tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}) + \mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1]\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{q}_1 + 2q_{11}\mathbf{Q}\widehat{\mathbf{Z}}'[\mathbf{G}\mathbf{G}'\mathbf{C} + \mathbf{C}\mathbf{G}\mathbf{G}' \\
&\quad +\mathbf{G}\mathbf{G}'\mathbf{C}']\widehat{\mathbf{Z}}\mathbf{q}_1 - 2q_{11}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{Q}\widehat{\mathbf{Z}}'[\mathbf{C} + \mathbf{C}']\widehat{\mathbf{Z}}\mathbf{q}_1 - 2q_{11}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{q}_1\} \\
&\quad +\widehat{\sigma}_i^6\{[8q_{11}^2 tr(\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{G}'\mathbf{C} - 2q_{11}^2 tr(\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{G}')tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}) \\
&\quad -4q_{11}^2 tr(\mathbf{G}\mathbf{G}'\mathbf{C})tr(\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}) - 12q_{11}(\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{q}_1)tr(\mathbf{G}\mathbf{G}'\mathbf{C}) \\
&\quad -8q_{11}(\mathbf{q}_1'\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1)tr(\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{G}')]\mathbf{q}_1 - 2q_{11}^2 tr(\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{G}')\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{C}\widehat{\mathbf{Z}}\mathbf{q}_1 \\
&\quad -8q_{11}^2 tr(\mathbf{G}\mathbf{G}'\mathbf{C})\mathbf{Q}\widehat{\mathbf{Z}}'\mathbf{G}\mathbf{G}'\widehat{\mathbf{Z}}\mathbf{q}_1\} - \widehat{\sigma}_i^8[12q_{11}^3 tr(\mathbf{G}\mathbf{G}'\mathbf{C})tr(\mathbf{G}\mathbf{G}'\mathbf{G}\mathbf{G}')\mathbf{q}_1]
\end{aligned}$$

where $\widehat{\sigma}_i^2$, $\widehat{\mathbf{Z}}$, $\widehat{\mathbf{C}}$ and $\widehat{\mathbf{D}}$ are defined as (4.5), (4.6) and (4.7), respectively,

$$\mathbf{Q} = \widehat{\mathbf{D}}^{-1}, \quad \mathbf{q}_1 = \widehat{\mathbf{D}}^{-1}\mathbf{e}_1, \quad q_{11} = \mathbf{e}_1'\widehat{\mathbf{D}}^{-1}\mathbf{e}_1, \quad \mathbf{G} = \boldsymbol{\Lambda}^{-1}(\lambda_{OLS,i})(\mathbf{I}_T, \mathbf{0})\boldsymbol{\Omega};$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\Lambda}^{-1}(\lambda_i) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \lambda_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \lambda_i^{T-1} & \cdots & \lambda_i & 1 \end{bmatrix}$$

and ω is variance of start-up value $y_{i0} \sim N(\bar{y}_{i0}, \omega^2\sigma_i^2)$[1]. The second-order bias-corrected estimators of long-run coefficients can be obtained by

$$\widehat{\boldsymbol{\theta}}_{BC2,MG} = \frac{1}{N}\sum_{i=1}^{N} \frac{\widehat{\boldsymbol{\beta}}_{BC2,i}}{1-\widehat{\lambda}_{BC2,i}}. \tag{4.11}$$

4.3.3 Bias-correction of the Long-run coefficient

Since the estimates of long-run coefficients based on bias-corrected estimator does not always yield desirable performance as demonstrated by Pesaran and Zhao [1999], they proposed a direct bias-correction of long-run coefficient. Specifically, they proposed the following bias-corrected estimator for long-run coefficient[2]:

$$\widehat{\boldsymbol{\theta}}_{DBC1,MG} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_{OLS,i} - \widehat{\mathbf{b}}_{\theta,i}^{(1)}\right), \tag{4.12}$$

$$\widehat{\boldsymbol{\theta}}_{DBC2,MG} = \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\theta}}_{OLS,i} - \widehat{\mathbf{b}}_{\theta,i}^{(2)}\right) \tag{4.13}$$

where for $h = 1, 2$

$$\widehat{\mathbf{b}}_{\theta i}^{(h)} = \frac{\left(1-\widehat{\lambda}_{OLS,i}-\widehat{b}_{\lambda,i}^{(h)}\right)\left(\widehat{\mathbf{b}}_{\beta,i}^{(h)} + \widehat{\boldsymbol{\theta}}_{OLS,i}\widehat{b}_{\lambda,i}^{(h)}\right) + \widehat{Cov}\left(\widehat{\lambda}_{OLS,i}, \widehat{\boldsymbol{\beta}}_{OLS,i}\right) + \widehat{\boldsymbol{\theta}}_{OLS,i}\widehat{Var}\left(\widehat{\lambda}_{OLS,i}\right)}{\left(1-\widehat{\lambda}_{OLS,i}-\widehat{\mathbf{b}}_{\lambda,i}^{(h)}\right)^2}.$$

4.4 Monte Carlo Simulation

In this section, we investigate to what extent the higher-order bias-correction is effective to reduce the bias of the MG estimator.

[1]In the simulation study in the next section, we simply set $\omega = 0$.
[2]Although they proposed two types of direct bias correction methods, we only consider DBC_1 in their notation since it performs better than DBC_2.

4.4.1 Design

We use the same design as Pesaran and Zhao (1997). Specifically, y_{it} and x_{it} are generated as

$$\begin{aligned} y_{it} &= \alpha_i + \lambda y_{i,t-1} + (1-\lambda)\theta_i x_{it} + \varepsilon_{it}, \qquad (i = 1, ..., N, t = 1, ..., T) \\ x_{it} &= \mu_i(1-\rho) + \rho x_{i,t-1} + u_{it} \end{aligned}$$

where α_i, μ_i and long-run coefficient θ_i are independently generated as $\alpha_i \sim N(1,1)$, $\mu_i \sim N(1,1)$ and $\theta_i \sim N(1,1)$. Note that $\beta = E[(1-\lambda)\theta_i] = (1-\lambda)$. The idiosyncratic term ε_{it} and u_{it} are independently generated as $\varepsilon_{it} \sim N(0,1)$, $u_{it} \sim N(0, \sigma_{u,i}^2)$ where[3]

$$\sigma_{u,i}^2 = \frac{(\sigma_s^2 - \frac{\lambda^2}{1-\lambda^2})(1+\lambda)(1-\rho^2)(1-\lambda\rho))}{\theta_i^2(1-\lambda)(1+\lambda\rho)}$$

with $\sigma_s^2 = 2, 8$. For the sample sizes, we consider $T = 15, 20, 25, 50$ and $N = 20, 100$. For parameter values, we consider $\lambda = 0.4, 0.8$, $\rho = 0.6, 0.95$ and $\sigma_s^2 = 2, 8$. For the estimation of λ and β, we consider three estimators, the standard MG estimator (4.2), denoted as "OLS", and two bias-corrected estimators (4.4) and (4.9), which are denoted as "KP1" and "KP2", respectively. For the long-run coefficient θ, we consider five estimators. The first three estimators are (4.3), (4.8), and (4.11), which are denoted as "OLS", "KP1" and "KP2", respectively. The last two are direct-bias-corrected estimators (4.12) and (4.13), which are denoted as "DBC1" and "DBC2", respectively. The first 50 periods are discarded to reduce the effect of initial conditions. The median bias (BIAS) and median absolute error (MAE) based on 20,000 replications are reported.

4.4.2 Result

The simulation results are provided in Tables 1 to Table 4. Theoretical result implies that the bias of KP1 for λ and β should be smaller than that of OLS, and the bias of KP2 should be smaller than that of KP1. From the tables, we can confirm that this is the case. For λ and β, KP2 has the smallest bias in almost all cases while OLS is most biased. This reduction in bias is also reflected in MAE. Indeed, KP2 has the smallest MAE in almost all cases among the three estimators. However, for the long-run coefficient θ, this result does not readily apply. In some cases, the estimate for θ based on KP2 is more biased than that based on KP1. Moreover, the estimates for θ based on bias-corrected estimators are sometimes severely biased. Note that this result is also reported in Pesaran and Zhao [1999]. As a solution to this problem, we consider DBC1 and DBC2. From the tables, it is observed that DBC2 which is based on second-order bias-corrected estimator has the smallest bias in many cases. These results suggest that

[3]For the derivation of $\sigma_{u,i}^2$, see Pesaran and Zhao (1999).

second-order bias-correction is useful in reducing the bias of MG estimators.

4.5 Conclusion

In this chapter, we considered estimation of heterogeneous dynamic panel data models by means of MG estimator. Since MG estimator is biased in finite samples, we proposed to use second-order bias-corrected estimator by Kiviet and Phillips [2012] as a bias-reduction method. Monte Carlo simulation results revealed that the second-order bias-correction is useful in reducing the bias especially when T is not large.

Table 4.1: Simulation results ($\lambda = 0.4, \sigma_s^2 = 2$)

$\lambda = 0.4,\ \sigma_s^2 = 2,\ \rho = 0.6$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.099	-0.074	-0.059	-0.029	-0.100	-0.075	-0.059	-0.029	0.099	0.074	0.059	0.029	0.100	0.075	0.059	0.029
KP1	-0.024	-0.013	-0.008	-0.002	-0.024	-0.014	-0.009	-0.002	0.037	0.028	0.023	0.015	0.025	0.016	0.012	0.007
KP2	-0.015	-0.007	-0.004	-0.001	-0.015	-0.008	-0.005	-0.001	0.035	0.028	0.023	0.015	0.019	0.013	0.011	0.007
	β								β							
OLS	0.028	0.023	0.019	0.009	0.033	0.026	0.021	0.012	0.106	0.103	0.100	0.096	0.053	0.049	0.047	0.043
KP1	0.004	0.001	0.000	-0.003	0.008	0.005	0.002	0.001	0.102	0.099	0.097	0.094	0.046	0.045	0.043	0.042
KP2	0.002	-0.001	-0.001	-0.003	0.006	0.003	0.001	0.000	0.102	0.099	0.097	0.094	0.046	0.045	0.043	0.042
	θ								θ							
OLS	-0.032	-0.034	-0.026	-0.017	-0.023	-0.025	-0.024	-0.013	0.177	0.165	0.161	0.155	0.083	0.077	0.073	0.070
KP1	0.074	0.042	0.031	0.009	0.090	0.052	0.035	0.013	0.207	0.180	0.170	0.159	0.122	0.090	0.080	0.071
KP2	0.093	0.054	0.038	0.010	0.110	0.064	0.041	0.015	0.218	0.184	0.172	0.160	0.139	0.095	0.082	0.071
DBC1	-0.020	-0.017	-0.009	-0.007	-0.012	-0.008	-0.007	-0.002	0.178	0.168	0.162	0.156	0.082	0.077	0.073	0.070
DBC2	-0.008	-0.010	-0.005	-0.006	0.002	-0.001	-0.002	-0.001	0.179	0.169	0.163	0.156	0.082	0.077	0.073	0.070

$\lambda = 0.4,\ \sigma_s^2 = 2,\ \rho = 0.95$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.184	-0.142	-0.114	-0.059	-0.185	-0.142	-0.115	-0.059	0.184	0.142	0.114	0.059	0.185	0.142	0.115	0.059
KP1	-0.067	-0.041	-0.027	-0.007	-0.068	-0.041	-0.027	-0.007	0.069	0.047	0.036	0.021	0.068	0.041	0.028	0.010
KP2	-0.049	-0.029	-0.018	-0.004	-0.049	-0.029	-0.018	-0.004	0.058	0.042	0.033	0.020	0.049	0.029	0.020	0.010
	β								β							
OLS	0.099	0.080	0.071	0.041	0.106	0.090	0.077	0.046	0.178	0.151	0.138	0.110	0.113	0.095	0.083	0.058
KP1	0.030	0.016	0.012	0.001	0.038	0.025	0.018	0.006	0.156	0.132	0.122	0.101	0.076	0.063	0.056	0.045
KP2	0.021	0.009	0.007	-0.001	0.029	0.019	0.013	0.004	0.155	0.132	0.121	0.101	0.074	0.061	0.055	0.045
	θ								θ							
OLS	-0.050	-0.044	-0.033	-0.016	-0.035	-0.031	-0.026	-0.010	0.246	0.208	0.191	0.161	0.120	0.098	0.087	0.073
KP1	0.046	0.023	0.017	0.002	0.078	0.041	0.026	0.008	0.304	0.229	0.201	0.164	0.185	0.113	0.093	0.074
KP2	0.074	0.043	0.028	0.004	0.109	0.061	0.037	0.010	0.339	0.241	0.204	0.164	0.227	0.127	0.097	0.074
DBC1	-0.034	-0.027	-0.017	-0.008	-0.021	-0.015	-0.010	-0.002	0.249	0.210	0.193	0.163	0.120	0.097	0.087	0.073
DBC2	-0.021	-0.019	-0.012	-0.006	-0.007	-0.006	-0.004	-0.001	0.255	0.212	0.194	0.163	0.122	0.098	0.088	0.073

Table 4.2: Simulation results ($\lambda = 0.4, \sigma_s^2 = 8$)

$\lambda = 0.4,\ \sigma_s^2 = 8,\ \rho = 0.6$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.040	-0.030	-0.023	-0.011	-0.041	-0.030	-0.023	-0.011	0.040	0.030	0.023	0.012	0.041	0.030	0.023	0.011
KP1	-0.004	-0.002	-0.001	0.000	-0.005	-0.003	-0.001	0.000	0.021	0.016	0.014	0.009	0.010	0.008	0.006	0.004
KP2	-0.002	-0.001	0.000	0.000	-0.002	-0.001	-0.001	0.000	0.020	0.016	0.014	0.009	0.009	0.007	0.006	0.004
	β								β							
OLS	0.012	0.008	0.007	0.003	0.014	0.010	0.007	0.004	0.094	0.096	0.093	0.093	0.044	0.043	0.042	0.041
KP1	0.000	-0.001	-0.001	-0.002	0.002	0.001	-0.001	0.000	0.092	0.093	0.092	0.092	0.041	0.042	0.041	0.041
KP2	0.000	-0.002	-0.001	-0.002	0.002	0.001	-0.001	0.000	0.092	0.093	0.092	0.092	0.041	0.042	0.041	0.041
	θ								θ							
OLS	-0.012	-0.014	-0.012	-0.008	-0.009	-0.011	-0.011	-0.006	0.156	0.156	0.153	0.153	0.071	0.071	0.069	0.068
KP1	0.034	0.017	0.012	0.002	0.039	0.021	0.013	0.005	0.167	0.162	0.157	0.155	0.081	0.074	0.070	0.069
KP2	0.039	0.019	0.013	0.002	0.043	0.023	0.015	0.005	0.168	0.162	0.157	0.155	0.082	0.075	0.071	0.069
DBC1	-0.005	-0.006	-0.004	-0.004	-0.002	-0.002	-0.003	-0.001	0.156	0.158	0.155	0.154	0.071	0.071	0.069	0.068
DBC2	-0.001	-0.005	-0.003	-0.004	0.002	0.000	-0.002	-0.001	0.157	0.158	0.154	0.154	0.072	0.071	0.069	0.068

$\lambda = 0.4,\ \sigma_s^2 = 8,\ \rho = 0.95$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.138	-0.104	-0.082	-0.041	-0.138	-0.104	-0.083	-0.041	0.138	0.104	0.082	0.041	0.138	0.104	0.083	0.041
KP1	-0.042	-0.024	-0.015	-0.004	-0.042	-0.024	-0.016	-0.004	0.048	0.035	0.027	0.016	0.042	0.025	0.017	0.008
KP2	-0.028	-0.016	-0.009	-0.002	-0.028	-0.016	-0.010	-0.002	0.043	0.033	0.026	0.016	0.029	0.018	0.013	0.008
	β								β							
OLS	0.074	0.061	0.053	0.029	0.079	0.066	0.055	0.032	0.128	0.118	0.113	0.101	0.084	0.071	0.064	0.049
KP1	0.019	0.011	0.007	-0.001	0.024	0.015	0.010	0.003	0.112	0.104	0.102	0.095	0.054	0.049	0.046	0.043
KP2	0.013	0.006	0.004	-0.002	0.017	0.011	0.007	0.002	0.111	0.104	0.101	0.094	0.052	0.048	0.046	0.043
	θ								θ							
OLS	-0.035	-0.031	-0.022	-0.012	-0.026	-0.024	-0.019	-0.008	0.180	0.167	0.162	0.153	0.084	0.077	0.073	0.069
KP1	0.048	0.024	0.016	0.001	0.062	0.032	0.019	0.005	0.206	0.178	0.167	0.156	0.114	0.084	0.076	0.069
KP2	0.067	0.036	0.022	0.002	0.084	0.044	0.026	0.006	0.220	0.182	0.169	0.156	0.134	0.089	0.078	0.069
DBC1	-0.020	-0.016	-0.009	-0.006	-0.014	-0.009	-0.006	-0.002	0.181	0.169	0.163	0.155	0.084	0.077	0.073	0.069
DBC2	-0.010	-0.010	-0.005	-0.005	-0.002	-0.003	-0.002	-0.001	0.183	0.170	0.164	0.155	0.085	0.077	0.073	0.069

Table 4.3: Simulation results ($\lambda = 0.8, \sigma_s^2 = 2$)

$\lambda = 0.8,\ \sigma_s^2 = 2,\ \rho = 0.6$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.254	-0.193	-0.154	-0.075	-0.255	-0.193	-0.155	-0.076	0.254	0.193	0.154	0.075	0.255	0.193	0.155	0.076
KP1	-0.120	-0.077	-0.053	-0.015	-0.121	-0.078	-0.054	-0.016	0.120	0.077	0.053	0.019	0.121	0.078	0.054	0.016
KP2	-0.086	-0.050	-0.032	-0.007	-0.087	-0.051	-0.033	-0.007	0.087	0.053	0.037	0.017	0.087	0.051	0.033	0.009
	β								β							
OLS	0.016	0.015	0.015	0.009	0.020	0.020	0.018	0.012	0.102	0.084	0.074	0.052	0.048	0.040	0.036	0.025
KP1	0.002	0.001	0.001	-0.002	0.007	0.006	0.004	0.002	0.098	0.079	0.071	0.050	0.045	0.037	0.032	0.022
KP2	-0.001	-0.001	-0.001	-0.003	0.004	0.004	0.002	0.001	0.097	0.079	0.070	0.050	0.044	0.036	0.032	0.022
	θ								θ							
OLS	-0.350	-0.266	-0.218	-0.138	-0.283	-0.212	-0.180	-0.114	0.607	0.475	0.403	0.264	0.491	0.338	0.256	0.142
KP1	-0.305	-0.084	0.095	0.159	-0.297	-0.022	0.160	0.215	1.390	1.188	0.955	0.367	1.409	1.162	0.963	0.272
KP2	-0.396	-0.226	-0.022	0.236	-0.385	-0.156	0.033	0.300	1.420	1.302	1.150	0.461	1.444	1.309	1.143	0.407
DBC1	-0.292	-0.206	-0.159	-0.068	-0.105	-0.128	-0.101	-0.040	0.655	0.479	0.403	0.265	1.167	0.409	0.255	0.124
DBC2	-0.217	-0.134	-0.097	-0.040	-0.066	-0.046	-0.029	-0.010	0.671	0.513	0.428	0.271	0.616	0.362	0.255	0.126

$\lambda = 0.8,\ \sigma_s^2 = 2,\ \rho = 0.95$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.303	-0.235	-0.191	-0.097	-0.304	-0.236	-0.192	-0.098	0.303	0.235	0.191	0.097	0.304	0.236	0.192	0.098
KP1	-0.156	-0.105	-0.076	-0.025	-0.158	-0.106	-0.077	-0.026	0.156	0.105	0.076	0.026	0.158	0.106	0.077	0.026
KP2	-0.119	-0.074	-0.050	-0.013	-0.120	-0.075	-0.051	-0.013	0.119	0.075	0.051	0.020	0.120	0.075	0.051	0.014
	β								β							
OLS	0.083	0.075	0.072	0.046	0.087	0.080	0.074	0.053	0.388	0.293	0.241	0.127	0.186	0.145	0.118	0.068
KP1	0.040	0.028	0.024	0.007	0.042	0.033	0.027	0.013	0.351	0.262	0.212	0.110	0.164	0.123	0.097	0.051
KP2	0.029	0.017	0.014	0.001	0.032	0.023	0.018	0.008	0.347	0.259	0.208	0.108	0.161	0.120	0.095	0.050
	θ								θ							
OLS	-0.293	-0.218	-0.172	-0.116	-0.235	-0.165	-0.140	-0.077	1.481	1.097	0.877	0.482	1.141	0.729	0.514	0.231
KP1	-0.252	-0.105	0.037	0.053	-0.173	-0.064	0.088	0.105	3.562	2.786	2.089	0.654	3.624	2.721	1.991	0.384
KP2	-0.325	-0.156	-0.077	0.112	-0.267	-0.135	-0.034	0.178	3.634	3.025	2.519	0.803	3.606	3.049	2.407	0.573
DBC1	-0.231	-0.167	-0.141	-0.075	-0.113	-0.122	-0.100	-0.035	1.597	1.161	0.933	0.506	2.845	0.935	0.566	0.238
DBC2	-0.189	-0.116	-0.095	-0.055	-0.074	-0.045	-0.040	-0.014	1.758	1.293	1.024	0.520	1.620	0.936	0.616	0.247

Table 4.4: Simulation results ($\lambda = 0.8, \sigma_s^2 = 8$)

$\lambda = 0.8,\ \sigma_s^2 = 8,\ \rho = 0.6$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.119	-0.086	-0.066	-0.030	-0.120	-0.087	-0.067	-0.031	0.119	0.086	0.066	0.030	0.120	0.087	0.067	0.031
KP1	-0.037	-0.020	-0.013	-0.003	-0.038	-0.021	-0.013	-0.003	0.039	0.025	0.019	0.010	0.038	0.021	0.013	0.005
KP2	-0.020	-0.009	-0.005	0.000	-0.021	-0.010	-0.005	-0.001	0.032	0.023	0.018	0.010	0.022	0.012	0.009	0.005
	β								β							
OLS	0.008	0.007	0.007	0.004	0.009	0.009	0.008	0.005	0.036	0.034	0.033	0.032	0.017	0.017	0.016	0.015
KP1	0.000	0.000	0.000	-0.001	0.002	0.001	0.001	0.000	0.035	0.033	0.032	0.031	0.016	0.015	0.015	0.014
KP2	-0.001	-0.001	-0.001	-0.001	0.001	0.001	0.000	0.000	0.035	0.033	0.032	0.031	0.016	0.015	0.014	0.014
	θ								θ							
OLS	-0.148	-0.100	-0.082	-0.054	-0.107	-0.071	-0.063	-0.046	0.312	0.238	0.206	0.164	0.241	0.152	0.114	0.078
KP1	0.005	0.136	0.177	0.073	0.023	0.190	0.222	0.086	0.653	0.496	0.370	0.189	0.646	0.483	0.353	0.106
KP2	-0.050	0.128	0.199	0.091	-0.016	0.175	0.240	0.106	0.699	0.562	0.438	0.195	0.712	0.550	0.427	0.120
DBC1	-0.129	-0.070	-0.043	-0.015	0.112	-0.028	-0.026	-0.007	0.326	0.228	0.196	0.165	0.730	0.175	0.107	0.074
DBC2	-0.061	-0.019	-0.008	-0.005	0.049	0.021	0.014	0.003	0.303	0.232	0.204	0.166	0.297	0.147	0.106	0.075

$\lambda = 0.8,\ \sigma_s^2 = 8,\ \rho = 0.95$

	BIAS								MAE							
N	20				100				20				100			
T	15	20	25	50	15	20	25	50	15	20	25	50	15	20	25	50
	λ								λ							
OLS	-0.256	-0.192	-0.151	-0.072	-0.256	-0.192	-0.152	-0.072	0.256	0.192	0.151	0.072	0.256	0.192	0.152	0.072
KP1	-0.122	-0.078	-0.053	-0.015	-0.123	-0.078	-0.054	-0.016	0.122	0.078	0.053	0.018	0.123	0.078	0.054	0.016
KP2	-0.088	-0.051	-0.031	-0.006	-0.089	-0.052	-0.032	-0.007	0.089	0.053	0.035	0.015	0.089	0.052	0.032	0.008
	β								β							
OLS	0.063	0.058	0.054	0.036	0.067	0.062	0.057	0.038	0.092	0.077	0.069	0.049	0.067	0.062	0.057	0.038
KP1	0.025	0.019	0.015	0.005	0.029	0.022	0.018	0.008	0.075	0.060	0.053	0.038	0.039	0.031	0.027	0.018
KP2	0.017	0.011	0.009	0.002	0.021	0.015	0.011	0.004	0.073	0.059	0.052	0.038	0.036	0.029	0.025	0.017
	θ								θ							
OLS	-0.251	-0.185	-0.141	-0.075	-0.207	-0.148	-0.114	-0.063	0.439	0.333	0.274	0.180	0.341	0.229	0.166	0.091
KP1	-0.189	-0.016	0.074	0.063	-0.183	0.009	0.125	0.084	0.895	0.674	0.506	0.213	0.870	0.659	0.469	0.119
KP2	-0.300	-0.094	0.049	0.114	-0.259	-0.062	0.068	0.138	0.947	0.789	0.638	0.248	0.952	0.782	0.631	0.176
DBC1	-0.212	-0.145	-0.102	-0.036	-0.079	-0.091	-0.069	-0.024	0.464	0.331	0.270	0.181	0.741	0.266	0.159	0.083
DBC2	-0.143	-0.097	-0.058	-0.020	-0.042	-0.033	-0.018	-0.007	0.459	0.344	0.282	0.184	0.391	0.227	0.154	0.084

Bibliography

S. C. Ahn and P. Schmidt. Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1):5–27, 1995.

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

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.

M. Arellano. *Panel Data Econometrics.* Oxford University Press, Oxford, 2003.

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

M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

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

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

R. Blundell, S. Bond, and F. Windmeijer. Estimation in dynamic panel data models: Improving on the performance of the standard gmm estimator. In B. H. Baltagi, editor, *Nonstationary Panels, Panel Cointegration and Dynamic Panels*, volume 15 of *Advances in Econometrics*, pages 53–91. JAI Press, Amsterdam, 2000.

J. Breitung and K. Hayakawa. Bias-corrected estimators for various dynamic panel data models. 2015.

J. Breitung, K. Hayakawa, and M. Qi. Double filter instrumental variable estimation of panel data models with weakly exogenous variables. mimeo, 2016.

M. J. G. Bun and M. A. Carree. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics*, 23(2):200–210, 2005.

M. J. G. Bun and J. F. Kiviet. The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models. *Journal of Econometrics*, 132(2):409–444, 2006.

M. J. G. Bun and V. Sarafidis. Dynamic panel data models. In B. Baltagi, editor, *Oxford Handbook on Panel Data*. Oxford University Press, 2013.

M. J. G. Bun and F. Windmeijer. The weak instrument problem of the system gmm estimator in dynamic panel data models. *Econometrics Journal*, 13(1):95–126, 2010.

Y. Chen. A new algorithm for computing the determinant and the inverse of a pentadiagonal toeplitz matrix. *Engineering*, 5(05):25, 2013.

M. Dow. Explicit inverses of toeplitz and associated matrices. *ANZIAM Journal*, 44(E): E185–E215, 2003.

M. El-Mikkawy and A. Karawia. Inversion of general tridiagonal matrices. *Applied Mathematics Letters*, 19(8):712–720, 2006.

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.

C. Han and H. Kim. The role of constant instruments in dynamic panel estimation. *Economics Letters*, 124(3):500–503, 2014.

C. Han and P. Schmidt. The asymptotic distribution of the instrumental variable estimators when the instruments are not correlated with the regressors. *Economics Letters*, 74(1):61–66, 2001.

K. Hayakawa. Efficient gmm estimation of dynamic panel data models where large heterogeneity may be present. 2006.

K. Hayakawa. Small sample bias properties of the system gmm estimator in dynamic panel data models. *Economics Letters*, 95(1):32–38, 2007.

K. Hayakawa. A simple efficient instrumental variable estimator in panel ar(p) models when both n and t are large. *Econometric Theory*, 25(3):873–890, 2009a.

K. Hayakawa. First-difference or forward orthogonal deviation: Which transformation to use in dynamic panel data models?: A simulation study. *Economics Bulletin*, 29 (3):2014–2023, 2009b.

K. Hayakawa. The effects of dynamic feedbacks on ls and mm estimator accuracy in panel data models: Some additional results. *Journal of Econometrics*, 159(1):202–208, 2010.

K. Hayakawa. The asymptotic properties of the system gmm estimator in dynamic panel data models when both n and t are large. *Econometric Theory*, 31(3):647–667, 2015.

K. Hayakawa. Improved gmm estimation of panel var models. *Computational Statistics & Data Analysis*, 2016.

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

C. Hsiao and Q. Zhou. First difference or forward demeaning: Implications for the method of moments estimators. Working paper, 2015.

M. P. Keane and D. E. Runkle. On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *Journal of Business and Economic Statistics*, 10(1):1–9, 1992.

J. F. Kiviet and G. D. A. Phillips. Alternative bias approximations in regressions with a lagged-dependent variable. *Econometric Theory*, 9(1):62–80, 1993.

Jan F Kiviet and Garry DA Phillips. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. *Computational Statistics & Data Analysis*, 56(11):3705–3729, 2012.

Yoon-Jin Lee, Ryo Okui, and Mototsugu Shintani. Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes. *Available at SSRN 2337445*, 2013.

H. R. Moon and P. C. B. Phillips. Estimation of autoregressive roots near unity using panel data. *Econometric Theory*, 16(6):927–997, 2000.

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

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

M. H. Pesaran and R. Smith. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics,*, 68(1):79–113, 1995.

M. H. Pesaran and Z. Zhao. Bias reduction in estimating long-run relationships from dynamic heterogeneous panels. In C. Hsiao, K. Lahiri, L. F. Lee, and M. H. Pesaran, editors, *Analysis of Panels and Limited Dependent Variable Models*, chapter 12. Cambridge University Press, Cambridge, 1999.

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

P. C. B. Phillips and D. Sul. Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics*, 127(1): 162–188, 2007.

D. Roodman. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics*, 71(1):135–158, 2009.

T. J. Rothenberg. Approximating the distributions of econometric estimators and test statistics. In *Handbook of econometrics*, volume 2, chapter 15, pages 881–935. Elsevier, 1984.

T. Wansbeek and A. Kapteyn. A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics*, A11: 2105–2112, 1982.

T. Wansbeek and A. Kapteyn. A note on spectral decomposition and maximum likelihood estimation in models with balanced data. *Statistics and Probability Letters*, 1: 213–215, 1983.

T. J. Wansbeek and T. Knaap. Estimating a dynamic panel data model with heterogeneous trends. *Annales d'Economie et de Statistique*, 55-56:331–349, 1999.

J. Westerlund, H. Karabiyik, and P. Narayan. Testing for predictability in panels with general predictors. *Journal of Applied Econometrics*, 2016.

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