ESSAYS ON DYNAMIC PANEL DATA MODELS WITH INTERACTIVE EFFECTS

YAN-TING CHEN

PhD
University of York
Economics
August, 2021

ABSTRACT

This thesis investigates estimation and inferential methods for dynamic panel data models with multifactor error structure. Chapter 2 reviews the existing estimation methods for short T dynamic panel data models and compare their finite sample behaviour by means of Monte Carlo simulation. Chapter 3 investigates the speed of US firms that reverted back to its long-run equilibrium of target leverage ratio. These firms' behaviours are highly heterogeneous and unobserved common shocks are likely to influence the speed of adjustment of these firms. To ascertain these economic activities, we develop a robust method against cross-sectional heteroskedasticity by employing the approach proposed by Hayakawa et al. (2021). The estimation results confirm that the managers partially adjust the leverage quickly toward the target leverage. Chapter 4 extends the quasi-maximum likelihood (QML) estimator for short T dynamic panel data models with interactive effect proposed by Hayakawa et al. (2021) to panel vector autoregression (VAR) models with interactive effects. In chapter 5, we apply the model averaging method of Kuersteiner and Okui (2010) to the Instrumental Variable (IV) estimator of Norkutė et al. (2021) for the dynamic heterogeneous panel data model with a multifactor error structure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

First I would like to thank my supervisor and co-supervisor, Takashi Yamagata and Vanessa Smith, that without your tremendous patience, kind support, and valuable advice I would likely not have been able to advance quickly. I will always remember the feeling of meeting with my supervisor and co-supervisor. Their dedication to work and support for students set a great example of what kind of people I want to become.

I would like to thank the third member of my advisory panel, Jia Chen, for providing invaluable advice and support. I would also like to thank Yongcheol Shin and Kazuhiko Hayakawa for his helpful comments on my work. Especially, I would like to thank Sebastian Kripfganz, for his supervision in University of Exeter when I was a master student.

I want to thank all my friends and fellow PhD students for their friendship. In particular, I would like to thank my friends in the department, Rui Lin, Chaowen Zheng, Jamilya Junussova, Hui-Lan Chang, Fabio Profumo, Shuyi Lou, Jiaxin Peng and Min Lin. I am so lucky to have such lovely friends.

Last but not least, I would like to dedicate this thesis to my parents and Chia-Yu Wu. I could not have done this without them.

DECLARATION

I declare that this thesis is a presentation of original work and I am the sole author of the three self-contained chapters. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

An earlier version of Chapter 2 was presented at the Research Student Workshop at the University of York in April 2019.

An earlier version of Chapter 3 was presented at the 2021 Asian Meeting (virtual) of the Econometric Society.

An earlier version of Chapter 4 was presented at he Research Student Workshop at the University of York in April 2019..

Chapter 1

Introduction

Over the past few decades, the number of empirical research using panel data sets in economics and finance has grown dramatically and has been occupying a central place. Panel data combines cross-section and time series data, which is useful for research in labor economics, health economics, empirical finance, macroeconomics, among many others. There are mainly two statistical advantages to using panel data set in empirical studies. The first one is the gain of higher estimation precision by having a larger number of observations than pure cross-sectional or time-series data set. The second one is the ability to control potentially endogenous unobserved effects. We often categorise panel data sets into micro (or short) panels, macro (or long) panels and large panels. The macro panels have a large number of time observations (T), while the number of cross-section units (N) can be small. The micro panels have relatively small T but large N. Finally, N and T of large panels are both sizable.

In this thesis, we focus on studying estimation methods for analysing micro panels and large panels. In particular, we investigate the estimation and inferential methods for linear dynamic panel data models with unobserved additive and interactive effects. The dynamic panel data models can capture the dynamic behavioral relationships of the cross-sections, which are very useful for forecasting and various counter-factual analyses, such as impulse response analysis. Since the appearance of two important research papers, Pesaran (2006) and Bai (2009), the econometric methods for the models with interactive effects have been one of the central themes in the literature. Despite of the important contributions by Chudik and Pesaran (2015), Moon and Weidner (2017) and Norkutė et al. (2021), the methodological developments for dynamic panel data models with interactive effects have not been sufficient and more studies are required.

Chapter 2 compares some conventional estimators and recent developed estimator for short T dynamic panel data model. In particular, we consider three recently developed estimators, namely transformed maximum likelihood (TML) estimator, bias-corrected method of moments (BMM) estimator and double filter instrument variable (DFIV) estimator, along with the Instrument Variable (IV) estimator of

Anderson and Hsiao (1982, 1981) and the Generalized Method of Moments (GMM) estimators of Arellano and Bover (1995), Arellano and Bond (1991) and Blundell and Bond (1998). The distinctive feature of the TML estimator of Hsiao et al. (2002) is that it allows the initial value to depend on fixed effects and does not require the initial values to have the same mean across i (Hayakawa and Pesaran (2015a)). The BMM estimator of Chudik et al. (2020) uses the differenced dependent variable as an instrument, and exploits quadratic moment conditions. The advantage of this estimator is that it can allow more general conditions, such as the deviations of initial value from their long run means. The DFIV estimator of Hayakawa et al. (2019) uses the forward demeaning to eliminate the individual effects and the backward demeaning is applied to the instruments. In Monte Carlo simulation, we compare the finite sample behaviour of the TML, BMM and DFIV estimators under different initial conditions. The results show that the TML and BMM estimators mostly outperform the DFIV estimator.

In Chapter 3, we extend the quasi maximum likelihood (QML) estimator for short dynamic panel data models with interactive effects proposed by Hayakawa et al. (2021) to the case where the errors are cross-sectionally heteroskedastic, using a similar approach discussed in Hayakawa and Pesaran (2015a). Extending the error to cross-sectionally heteroskedastic is important in empirical studies, because heteroskedasticity is commonly observed, and it is standard practice to control for its impact on estimation and inference. (Hansen (2020)). We apply the developed method to a annual panel data set which consists of 16,502 US firms over the period from 1960 to 2017 to empirically asses the trade off theory (Graham and Harvey (2001)). We find that the speed of adjustment (SOA) of US firms is between 74% and 86% from 1960 to 1999, whilst it decreases around 32% from 2000 to 2007, followed by around 60% after 2008. The value of SOA we have found is higher than that in other existing research (Ozkan (2001), Fama and French (2002), Kayhan and Titman (2007), Flannery and Rangan (2006), Lemmon et al. (2008) and Dang et al. (2014)), which suggests the importance of controlling unobserved interactive effects.

Chapter 4 extends the QML estimator of Hayakawa et al. (2021) for the estimation of short panel vector autoregressive (VAR) models with interactive effects. Holtz-Eakin et al. (1988) and Binder et al. (2005), among others, consider estimation and inference for cross-sectionally independent short panel VAR model. The finite sample evidence provided by Juodis (2018) shows that the TML estimator outperforms the GMM based estimators in short panel VAR models.. A few studies focus on the estimation of panel VAR models with cross section dependence (Mutl (2009) and Huang (2008)), however, these methods are not asymptotically justified for the models with interactive effects. In the Monte Carlo results, we show that the proposed QML estimator performs reasonably well.

Based on Kuersteiner and Okui (2010), in Chapter 5 we proposes a method to choose a set of weights to average instrumental variable (IV) estimators proposed by Norkutė et al. (2021) for large dynamic panel data models with interactive effects. Norkutė et al. (2021) provides IV estimators that use the lagged defactored covariates in the model as IVs. Therefore, this IV estimator does not need to search for instruments outside the model. However, when T increases, the number of valid instruments rises. Norkutė et al. (2021) is silent about how to choose a set of instruments to avoid the problems due to having too many instruments or weak instruments. To tackle this problem, we take the approach proposed by Kuersteiner and Okui (2010). In particular, they propose a procedure to choose a set of weights to average the IV estimators, which are computed using different subset of available IVs. The weights are chosen so that the mean squared error of the average of the IV estimators is theoretically minimised. The estimator is called model average two stage least square (2SLS) estimator, and we apply this approach to the IV estimator of Norkutė et al. (2021). The empirical evidence shows that the proposed model average 2SLS estimator can reduce the bias but not MSE.

The rest of this thesis is organized as follows. In the second chapter, we review the existing methods for short dynamic panel data models with additive effects, then in the third chapter we extend the quasi maximum likelihood (QML) estimator for short dynamic panel data models with interactive effects proposed by Hayakawa et al. (2021) to the case where the errors are cross-sectionally heteroskedastic, using a similar approach discussed in Hayakawa and Pesaran (2015a). Then, in chapter 4, we extend the QML estimator of Hayakawa et al. (2021) for the estimation of panel vector autoregressive (VAR) models with interactive effects. Finally, in Chapter 5, based on Kuersteiner and Okui (2010), we propose a method to choose a set of instrumental variables (IVs) which minimise the theoretical mean squared errors of the IV estimator (proposed by Norkutė et al. (2021)), for large dynamic panel data models with interactive effects. Chapter 6 concludes the thesis.

Chapter 2

Short T dynamic panel data models: A survey

2.1 Introduction

Panel data combines cross-section and time series data. In recent years, there has been a dramatic proliferation of available panel data sets, which are useful for research in labor economics, health economics, empirical finance, macroeconomics, among many others. There are mainly two statistical advantages to using panel data set in empirical studies. The first one is the gain of higher estimation precision by having a larger number of observations than pure cross-sectional or time-series data set. The second one is the ability to control potentially endogenous unobserved effects. We often categorise panel data sets into micro panels and macro panels. The macro panels have a large number of time observations (T), while the number of cross-section units (N) can be small. The micro panels have relatively large N but small T. In this chapter, we focused on micro panels. The study for short T panels is important because, for example, household panel data has observations over several years only for each cross-section unit.

If panel data models contain lagged dependent variables, we call this model as dynamic panel data model. The dynamic panel data models can capture the dynamic behavioral relationships of the cross-sections. It is well known that the least square based estimator is inconsistent for dynamic panel data models when T is fixed. The within group estimator eliminates the individual effects by first differencing the variables, but the bias arises due to the correlation between the within transformed regressors and the error terms. Nickell (1981) analyses the bias of within group estimators, and he provides that it is $O(1/T)$.

To deal with this endogenous problem, instruments variable (IV) estimators/ Generalized Method of Moments (GMM) estimators have been proposed. Anderson and Hsiao (1982) propose a just identified IV estimator for the first differenced model.

Arellano and Bond (1991) propose a GMM estimator which exploits all the available instruments, which is more efficient than the IV estimator of Anderson and Hsiao (1982). Blundell and Bond (1998) argue that the first differenced GMM estimator of Arellano and Bond (1991) suffers from weak instrument problem when the model is persistent and/or mean variance dominates the idiosyncratic variance. To overcome these problems, they propose to extend the first differenced GMM estimator by combining the lagged level instruments, which is called system GMM estimator.

Bun and Windmeijer (2010) show that when the variance ratio of individual effects to idiosyncratic errors is large, the system GMM estimators also can face the weak instrument problem.

Besides these traditional GMM estimators, Chudik and Pesaran (2017) develop a novel GMM estimator by using a self-instrumenting target variable and a quadratic moment condition. The instruments have maximum correlation with target variables and do not have the weak instrument problem. Moreover, this GMM estimator is robust under flexible initial conditions. Hayakawa et al. (2019) propose a double filter instruments variable estimator (DFIV), which uses the forward filter to eliminate individual effects and a backward filter as instruments. Hayakawa et al. (2019) also show that the DFIV estimator has the same asymptotic distribution as the bias-corrected fixed effect (FE) estimator as N and T tend to infinity.

As an alternative to GMM estimators, Hsiao et al. (2002) propose the transformed maximum likelihood (TML) method. Under suitable assumptions on the initial value process, the TML estimator is shown to be consistent and more efficient than other conventional GMM estimators.

However, the procedure can fail to achieve the global maximum. Bun et al. (2017) have found that taking non negative variance constraint in transformed likelihood estimator performed better than the unconstrained transformed likelihood estimator.

Although many estimators have been provided, experiments under the same setting for a comparative assessment of the performance of these estimators have been under-researched. This is what we provide in this chapter.

The rest of this chapter is structured as follows. Section 2.2 analyses the bias of classical estimators. Section 2.3 reviews some GMM estimators and ML estimators. Section 2.4 investigates finite sample performance of the estimators discussed in 2.2-2.3 by using Monte Carlo simulation. Section 2.4 presents concluding remarks.

2.2 Bias of Conventional Estimators

It is well known that the least square estimator suffers from serious bias in dynamic panel data models when T is fixed. Nickell (1981) analysed the bias of the within group estimator for dynamic panel data models when the number of time observations, T, is fixed.

For simplicity, let us consider the following panel AR(1) model

$$y_{i,t} = \phi y_{i,t-1} + \alpha_i + u_{i,t}, \ |\phi| < 1; \ i = 1, 2, \ldots, N; \ t = 1, 2, \ldots, T, \tag{2.1}$$

where $y_{i,0}$ are observed and α_i are unobserved unit-specific effects, u_{it} is independently, identically distributed ($i.i.d$) with mean 0 and variance σ_u^2, with $0 < \sigma_u^2 < \infty$.

The standard method to eliminate individual effects α_i is to subtract the time mean of equation itself (within transformation). The within model can be expressed as

$$\tilde{y}_{i,t} = \phi \tilde{y}_{i,t-1} + \tilde{u}_{i,t}, \tag{2.2}$$

where $\tilde{y}_{i,t} = y_{i,t} - \bar{y}_i$, $\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$ and $\tilde{u}_{i,t} = u_{i,t} - \bar{u}_i$ in which $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{i,t}$, $\bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}$. The least square (LS) estimator is given by

$$\begin{aligned}\hat{\phi}_{LS} =& \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t-1} - \bar{y}_{i,-1})^2\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(y_{i,t} - \bar{y}_i\right)\right) \\ =&\phi + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(u_{i,t} - \bar{u}_i\right)/NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2/NT}.\end{aligned} \tag{2.3}$$

Consider the numerator. Taking probability limits as N tends to infinity, we have [1]

$$\begin{aligned}&\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(u_{i,t} - \bar{u}_{i,-1}\right) = \\ &\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}u_{i,t} - \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\bar{u}_i \\ &- \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \bar{y}_{i,-1}u_{i,t} + \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \bar{y}_{i,-1}\bar{u}_i.\end{aligned} \tag{2.4}$$

By weak law of large numbers, the first term of equation (2.4) is given by

$$\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}u_{i,t} = E\left(y_{i,t-1}u_{i,t}\right) = 0. \tag{2.5}$$

For the second term of equation (2.4), we have

$$\begin{aligned}&\operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t-1}\bar{u}_i = \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\bar{u}_i\sum_{t=1}^{T} y_{i,t-1} \\ &= \operatorname*{plim}_{N\to\infty} \frac{1}{NT}\sum_{i=1}^{N} T\bar{y}_{i,-1}\bar{u}_i = \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_i.\end{aligned} \tag{2.6}$$

[1]See the derivation in Appendix A at the end of the thesis.

For the third term of equation (2.4), we have

$$\begin{aligned}&\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}u_{i,t}=\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\bar{y}_{i,-1}\sum_{t=1}^{T}u_{i,t}\\&=\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_{i}.\end{aligned}\tag{2.7}$$

For the fourth term of equation (2.4), we have

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{y}_{i,-1}\bar{u}_{i}=\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_{i}.\tag{2.8}$$

Therefore

$$\operatorname*{plim}_{N\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)\left(u_{i,t}-\bar{u}_{i,-1}\right)=-\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_{i},\tag{2.9}$$

but as

$$\sum_{t=1}^{T}y_{i,t-1}=\frac{1-\phi^{T-1}}{1-\phi}u_{i1}+\frac{1-\phi^{T-2}}{1-\phi}u_{i2}+..+u_{i,T-1}+\frac{1-\phi^{T}}{1-\phi}y_{i0}+\frac{T-1-T\phi+\phi^{T}}{(1-\phi)^{2}}\alpha_{i}\tag{2.10}$$

we can easily derive that

$$-\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{y}_{i,-1}\bar{u}_{i}=-\frac{\sigma_u^2}{T^2}\frac{T-1-T\phi+\phi^{T}}{(1-\phi)^{2}}.\tag{2.11}$$

Therefore the LS estimator $\hat{\phi}_{LS}$ has an $O(T^{-1})$ bias when N tends to infinity (see Hsiao (2014) for more detailed derivation).

2.3 Existing Estimation Methods

2.3.1 Conventional IV/GMM estimation

Due to the correlation between $y_{i,t-1}$ and $u_{i.}$, least square estimator is biased on dynamic panel data models when T is fixed. Therefore, IV/GMM approach play an important role in estimation of short T dynamic panel data models.

We now consider three IV/GMM estimators which are widely used in empirical research. Firstly, we consider the AH-IV estimator by Anderson and Hsiao (1982, 1981).

Taking first difference of the model (2.1), we have

$$\Delta y_{i,t}=\phi\Delta y_{i,t-1}+\Delta u_{i,t},\ |\phi|<1;\ i=1,\ldots,N;t=2,\ldots,T,\tag{2.12}$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$, $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ and $\Delta u_{i,t} = u_{i,t} - u_{i,t-1}$.

Anderson and Hsiao (1982) proposes two valid instruments, $y_{i,t-2}$ and $\Delta y_{i,t-2}$. Therefore, the AH-IV estimator is based on the following moment conditions

$$\begin{aligned} &E\left(y_{i,t-2}\Delta u_{i,t}\right) = 0, \\ &E\left(\Delta y_{i,t-2}\Delta u_{i,t}\right) = 0, \quad t = 3, 4, \ldots, T. \end{aligned} \tag{2.13}$$

The AH-IV estimator is given by

$$\hat{\phi}_{AH-IV} = \left(\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{i,t-2}\Delta y_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{i,t-2}\Delta y_{i,t}\right). \tag{2.14}$$

However, Arellano (1989) show that the AH-IV estimator suffers from a weak instrument problem when the autoregressive coefficient tends to unity.

Arellano and Bond (1991) propose a more efficient AB-GMM estimator that uses all the available instruments but require stronger conditions on individual effects and the initial values. AB-GMM estimator require $E(y_{i,s}\Delta u_{i,t}) = 0,\ for\ i = 1, \ldots, N, s = 0, 1, \ldots, t-2,\ and\ t = 2, 3, \ldots, T$.

The moment conditions of the AB-GMM estimator can be expressed as

$$E\left(y_{i,s}\Delta u_{i,t}\right) = 0,\ for\ s = 0, 1, \ldots, t-2;\ t = 2, \ldots, T. \tag{2.15}$$

Stacking the moment conditions as a vector form, we have

$$E\left(\boldsymbol{Z}_i^{'}\Delta\boldsymbol{u}_i\right) = \boldsymbol{0}, \tag{2.16}$$

where $\Delta\boldsymbol{u}_i = \left(\Delta u_{i,2}, \Delta u_{i,3}, \ldots, \Delta u_{i,T}\right)^{'}$ and the instrument

$$\boldsymbol{Z}_i = \begin{bmatrix} y_{i,0} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i,0} & y_{i,1} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & y_{i,0} & y_{i,1} & y_{i,2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i,0} & \cdots & y_{i,T-2} \end{bmatrix} \tag{2.17}$$

is a $(T-1) \times \frac{T(T-1)}{2}$ block diagonal matrix.

Under the above moment condition, we can build the one-step AB-GMM estimator

$$\begin{aligned} \hat{\phi}_{AB-GMM}^{(1)} = &\left(\left(N^{-1}\sum_{i=1}^{N}\Delta\boldsymbol{y}_{i,-1}^{'}\boldsymbol{Z}_i\right)\boldsymbol{V}^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i^{'}\Delta\boldsymbol{y}_{i,-1}\right)\right)^{-1} \\ &\left(N^{-1}\sum_{i=1}^{N}\Delta\boldsymbol{y}_{i,-1}^{'}\boldsymbol{Z}_i\right)\boldsymbol{V}^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i^{'}\Delta\boldsymbol{y}_i\right), \end{aligned} \tag{2.18}$$

where $\Delta \boldsymbol{y}_i = (\Delta y_{i,1}, \ldots, \Delta y_{i,T})'$, $\Delta \boldsymbol{y}_{i,-1} = (\Delta y_{i,0}, \ldots, \Delta y_{i,T-1})'$. The weighting matrix $\boldsymbol{V}^{-1}$ is ~~the inverse of the covariance matrix~~ $\cancel{E\left(\boldsymbol{Z}_i' \Delta \boldsymbol{u}_i \Delta \boldsymbol{u}_i' \boldsymbol{Z}_i\right)}$, ~~as~~

$$\boldsymbol{V}^{-1} = \left(N^{-1} \sum_{i=1}^{N} E\left(\boldsymbol{Z}_i' \Delta \boldsymbol{u}_i \Delta \boldsymbol{u}_i' \boldsymbol{Z}_i\right)\right)^{-1} = \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \boldsymbol{D} \boldsymbol{D}' \boldsymbol{Z}_i\right)^{-1}, \tag{2.19}$$

where $\boldsymbol{D}$ is the $(T-1) \times T$ first difference matrix as

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & & -1 & 1 \end{bmatrix}. \tag{2.20}$$

Then, based on the one-step AB-GMM estimator, we can get the one-step AB-GMM residuals $\Delta \hat{\boldsymbol{u}}_i$.

Using the one-step AB-GMM residual $\hat{\boldsymbol{u}}_i$, we can build a two-step AB-GMM weight matrix as

$$\hat{\boldsymbol{V}}_2^{-1} = \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \hat{\boldsymbol{u}}_i \Delta \hat{\boldsymbol{u}}_i' \boldsymbol{Z}_i\right)^{-1}. \tag{2.21}$$

Then, the two-step AB-GMM estimator is given by

$$\begin{aligned} \hat{\phi}^{(2)}_{AB-GMM} = &\left(\left(N^{-1} \sum_{i=1}^{N} \Delta \boldsymbol{y}_{i,-1}' \boldsymbol{Z}_i\right) \hat{\boldsymbol{V}}_2^{-1} \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_{i,-1}\right)\right)^{-1} \\ &\left(N^{-1} \sum_{i=1}^{N} \Delta \boldsymbol{y}_{i,-1}' \boldsymbol{Z}_i\right) \hat{\boldsymbol{V}}_2^{-1} \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{Z}_i' \Delta \boldsymbol{y}_i\right). \end{aligned} \tag{2.22}$$

Although AB-GMM estimator is consistent in short T dynamic panel data models, this estimator needs distributional assumptions on individual effects. Also, when the autoregressive parameter is close to unity or the variance ratio $(\frac{\sigma_\alpha^2}{\sigma_u^2})$ increases, there is a large finite sample bias in the AB-GMM estimator.

Arellano and Bover (1995) and Blundell and Bond (1998) provide the system GMM (SYS-GMM) estimator by using lags of the first differenced dependent variables as instruments for the level equation[2]. Also, Arellano and Bover (1995) demonstrate the forward orthogonal transformation can be seen as the first difference transformation to remove fixed effects plus a GLS transformation to eliminate the serial correlation. The SYS-GMM estimator require $E(y_{i,s} \Delta u_{i,t}) = 0$ *and* $E(\Delta y_{i,t-1}(\alpha_i +$

[2]More detail can be found in Appendix A

$u_{i,t})) = 0, for\ i = 1, \ldots, N, s = 0, 1, \ldots, t-2,\ and\ t = 2, 3, \ldots, T$. The SYS-GMM estimator uses extra moment conditions:

$$E\left[\Delta y_{i,t-1}\left(u_{i,t} + \alpha_i\right)\right] = 0,\ \ for\ t = 2, 3, \ldots, T. \tag{2.23}$$

This moment condition relies on the covariance stationarity of the initial value. The equation (2.1) can be expressed as

$$\Delta y_{i,t} = \phi^{t-1}\Delta y_{i,1} + \sum_{s=0}^{t-2}\phi^s \Delta u_{i,t-s}\ for\ t = 2, \ldots, T. \tag{2.24}$$

If we want to guarantee $E\left(\Delta y_{i,t}\alpha_i\right)$=0, we need to assume $E\left(\Delta y_{i,1}\alpha_i\right)$ =0. Therefore, we need to restrict $E\left[\left(y_{i,0} - \alpha_i/(1-\phi)\right)\alpha_i\right] = 0$ which means that the initial value is stationary.

Based on the moment conditions (2.15) and (2.23), the moment restrictions can be written as a vector form

$$E\left(\boldsymbol{Z}_i^{*'}\boldsymbol{u}_i^*\right) = \boldsymbol{0},\ \ for\ i = 1, \ldots, N, \tag{2.25}$$

where $\boldsymbol{u}_i^* = \left(\Delta\boldsymbol{u}_i, \boldsymbol{u}_i + \alpha_i\boldsymbol{\iota}_{T-1}\right)$ is a $2(T-1)\times 1$ vector , $\boldsymbol{u}_i = \left(u_{i,2}, \ldots, u_{i,T}\right)^{'}$, $\boldsymbol{\iota}_{T-1}$ is the $(T-1)$ vector of ones and

$$\boldsymbol{Z}_i^* = \begin{bmatrix} \boldsymbol{Z}_i & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i,1} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \Delta y_{i,T-1} \end{bmatrix}. \tag{2.26}$$

Then, under the moment condition (2.25), the one-step SYS-GMM estimator is given by

$$\begin{aligned} \hat{\phi}^{(1)}_{SYS-GMM} = & \left(\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{y}_{i,-1}^{*'}\boldsymbol{Z}_i^*\right)\boldsymbol{V}^{*-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i^{*'}\boldsymbol{y}_{i,-1}^*\right)\right)^{-1} \\ & \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{y}_{i,-1}^{*'}\boldsymbol{Z}_i^*\right)\boldsymbol{V}^{*-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i^{*'}\boldsymbol{y}_i^*\right), \end{aligned} \tag{2.27}$$

where $\boldsymbol{y}_i^{*'} = \left(\Delta\boldsymbol{y}_i^{'}\,\boldsymbol{y}_i^{'}\right)^{'}$, $\boldsymbol{y}_{i,-1}^{*'} = \left(\Delta\boldsymbol{y}_{i,-1}^{'}\,\boldsymbol{y}_{i,-1}^{'}\right)^{'}$ and

$$\boldsymbol{V}^{*-1} = \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i^{*'}\boldsymbol{H}\boldsymbol{Z}_i^*\right)^{-1}, \tag{2.28}$$

where $\boldsymbol{H}$ is $2(T-1) \times 2(T-1)$ weight matrix, as

$$\begin{bmatrix} 2 & -1 & \cdots & \cdots & \cdots & 0 \\ -1 & \ddots & & & & \vdots \\ \vdots & & 2 & & & \vdots \\ \vdots & & & 1 & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 1 \end{bmatrix}. \tag{2.29}$$

When the autoregressive parameter tends to one, the SYS-GMM estimators are not consistent because the instrument variables will not be correlated to the explanatory variables[3]. Bun and Windmeijer (2010) also argue that when the variance ratio is large the SYS-GMM estimator faces the weak instrument problem. Therefore, the ML based approach can be attractive under such circumstances (Hsiao et al. (2002)) [4].

2.3.2 Bias-corrected method of moments (BMM) estimation

Chudik and Pesaran (2017) provide a novel bias-corrected method of moments (BMM) estimator. The advantage of the BMM estimator is that we do not require additional assumptions on individual effects and initial values[5]. Therefore, this estimator can be widely applicable.

Assumption 1 $|\phi| < 1$*, and it is assumed that* ϕ *in a compact set.*

Assumption 2 *The error term* $u_{i,t}$ *is serially and cross-sectionally independently distributed, with* $E(u_{i,t}) = 0$ *and* $E(u_{i,t}^2) = \sigma_{i,t}^2$*, such that* $0 < c_1 < \sigma_{i,t}^2 < c_2$ *where* c_1 *and* c_2 *are positive constants. Also, it is assumed that* $\bar{\sigma}_t^2 \equiv N^{-1} \sum_{i=1}^{N} \sigma_{i,t}^2 \to \bar{\sigma}_t^2$ *as* $N \to \infty$*, and* $\sup_{i,t} E|u_{i,t}|^{4+\epsilon} < c_2$ *for some* $\epsilon > 0$*.*

Let observations $y_{i,t}$ be from arbitrary past, where $t = -m_i + 1, \ldots, T$. Then, from equation (2.12), we can express $\Delta y_{i,1}$ as

$$\Delta y_{i,1} = b_i - (1-\phi) \sum_{\ell=0}^{m_i - 1} \phi^{\ell} u_{i,-\ell} + u_{i,1}, \tag{2.30}$$

where

$$b_i = -\phi^{m_i} (1-\phi) \left(y_{i,-m_i} - \frac{\alpha_i}{1-\phi} \right) \tag{2.31}$$

Assumption 3 *It is assumed that* $E(b_i^2) = \sigma_{bi}^2$ *and* $\bar{\sigma}_{bi}^2 = N^{-1} \sum_{i=1}^{2} \sigma_{bi}^2 \to \bar{\sigma}^2$ *as* $N \to \infty$*, and* $\sup_i E|b_i|^{4+\epsilon} < c_2$ *for some* $\epsilon > 0$*. It follows the condition,* $E(b_i \Delta u_{i,t}) = 0$*, for* $t = 2, 3, \ldots, T$ *and* $i = 1, \ldots, N$*.*

---

[3]See Appendix A for discussions of the weak instruments problem of the AB-GMM estimator and the relationship between the SYS-GMM estimator and the AB-GMM estimator.

[4]In the Monte Carlo results of Hsiao et al. (2002), the ML based estimator outperforms the GMM based estimator.

[5]Chudik et al. (2020) propose a new BMM type estimator by augmenting the AH estimator.

Chudik and Pesaran (2017) use endogenous regressors $\Delta y_{i,t-1}$ to construct the moment conditions. This moment condition is

$$E\left(\Delta u_{i,t}\Delta y_{i,t-1}\right) = -\sigma_{i,t-1}^2, \quad t = 2, 3, \ldots, T-1. \tag{2.32}$$

Note $E(\Delta u_{i,t})^2 = \sigma_{i,t-1}^2 + \sigma_{i,t}^2$ and $E(\Delta u_{i,t-1}\Delta y_{i,t}) = -\sigma_{i,t}^2$. Then, we have

$$\sigma_{i,t-1}^2 = E(\Delta u_{i,t})^2 + E(\Delta u_{i,t+1}\Delta y_{i,t}) \tag{2.33}$$

Therefore, we obtain the quadratic moment condition,

$$E(\Delta u_{i,t}\Delta y_{i,t-1}) + E(\Delta u_{i,t})^2 + E(\Delta u_{i,t+1}\Delta y_{i,t}) = 0, \quad t = 2, 3, \ldots, T-1. \tag{2.34}$$

Averaging the moment conditions (2.34) over t, then substituting the model (2.12) into $\Delta u_{i,t}$ and $\Delta u_{i,t+1}$, we have

$$E\left(M_{i,T}(\phi)\right) = 0, \quad i = 1, 2, \ldots, N, \tag{2.35}$$

where

$$M_{i,T}(\phi) = \frac{1}{T-2}\sum_{t=2}^{T-1}\left(\left(\Delta y_{i,t} - \phi\Delta y_{i,t-1}\right)\Delta y_{i,t-1} + \left(\Delta y_{i,t} - \phi\Delta y_{i,t-1}\right)^2 + \left(\Delta y_{i,t+1} - \phi\Delta y_{i,t}\right)\Delta y_{i,t}\right). \tag{2.36}$$

Then, the BMM estimator can be obtained by

$$\hat{\phi}_{BMM} = arg\min_{\phi\in\Theta}\left\|\bar{M}_{NT}(\phi)\right\|, \tag{2.37}$$

where $\Theta \subset (-1, 1]$ is a compact set and $\bar{M}_{NT}(\phi) = \frac{1}{N}\sum_{i=1}^{N} M_{i,T}(\phi)$.

As noted in Chudik and Pesaran (2017), the BMM estimator is less restrictive in terms of the initial condition.

The asymptotic properties of estimator are as follows:

Theorem 1 *: Suppose that Assumptions (1)-(3) hold, and consider the BMM estimator $\hat{\phi}$ as T be fixed and $N \to \infty$,*

$$\sqrt{N}\left(\hat{\phi}_{BMM} - \phi\right) \xrightarrow{d} N\left(0, \sigma_{BMM}^2\right), \tag{2.38}$$

where $\sigma_{BMM}^2 = \bar{B}^{-2}S$.

The variance term σ^2_{BMM} can be estimated as

$$\begin{aligned}
&\hat{\sigma}^2_{BMM} = \hat{\bar{B}}^{-2}\hat{S}, \\
&\textit{where} \\
&\hat{\bar{B}} = \sum_{i=1}^{N}\left(D_{1i} + D_{2i} + 2\hat{H}_i\right), \; \hat{S} = \frac{\sum_{i=1}^{N}\hat{V}_i^2}{N}, \\
&\textit{with} \\
&D_{1i} = \frac{\sum_{t=2}^{T-1}\Delta y^2_{i,t-1}}{T-2}, \; D_{2i} = \frac{\sum_{t=2}^{T-1}\Delta y^2_{i,t}}{T-2}, \; \hat{H}_i = \frac{\sum_{t=2}^{T-1}\Delta\hat{u}_{i,t}\Delta y^2_{i,t-1}}{T-2}, \\
&\Delta\hat{u}_{i,t} = \Delta y_{i,t} - \hat{\phi}_{BMM}\Delta y_{i,t-1}, \\
&\hat{V}_i = -\sum_{t=2}^{T-1}\left(\Delta\hat{u}_{i,t}\Delta y_{i,t-1} + \Delta\hat{u}^2_{i,t} + \Delta\hat{u}_{i,t+1}\Delta y_{i,t}\right).
\end{aligned} \tag{2.39}$$

2.3.3 Double filter IV/GMM estimation

Next, we consider the double filter IV (DFIV) estimator proposed by Moon and Phillips (2000), Hayakawa (2009) and Hayakawa et al. (2019). We make the following assumption

Assumption 4 *(a) The unobserved individual effects α_i is i.i.d. with $E(\alpha_i) = 0$ and $Var(\alpha_i) = \sigma^2_\alpha$;(b)*

$$y_{i,0} = \frac{\alpha_i}{1-\phi} + e_{i,0}, \; \textit{where } e_{i,0} \sim iid\left(0, \frac{\sigma^2_u}{1-\phi}\right) \tag{2.40}$$

Stacking T observations for each i yields

$$\boldsymbol{y}_i = \phi\boldsymbol{y}_{i,-1} + \alpha_i\boldsymbol{\iota}_T + \boldsymbol{u}_i, \tag{2.41}$$

where $\boldsymbol{y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,T})'$, $\boldsymbol{y}_{i,-1} = (y_{i,0}, y_{i,1}, \ldots, y_{i,T-1})'$, $\boldsymbol{u}_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,T})'$ and $\boldsymbol{\iota}_T$ is $T \times 1$ vector of one. Define $(T-1) \times T$ forward orthogonal deviations matrix as

$$\boldsymbol{F} = diag(c_1, c_2, \ldots c_{T-1})\begin{bmatrix} 1 & \frac{-1}{T-1} & \cdots & \cdots & \frac{-1}{T-1} \\ \vdots & 1 & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}, \tag{2.42}$$

where $c_t = \sqrt{(T-t)(T-t+1)}$. Multiplying (2.41) by $\boldsymbol{F}$, the model to be estimated becomes

$$\dot{\boldsymbol{y}}_i = \dot{\boldsymbol{y}}_{i,-1} + \dot{\boldsymbol{u}}_i, \quad i = 1, \ldots, N. \tag{2.43}$$

where $\dot{\boldsymbol{y}}_i = \boldsymbol{F}\boldsymbol{y}_i = (\dot{y}_{i,1}, \dot{y}_{i,2}, \ldots, \dot{y}_{i,T-1})'$, $\dot{\boldsymbol{y}}_{i,-1} = \boldsymbol{F}\boldsymbol{y}_{i,-1} = (\dot{y}_{i,0}, \dot{y}_{i,1}, \ldots, \dot{y}_{i,T-2})'$ and $\dot{\boldsymbol{u}}_i = \boldsymbol{F}\boldsymbol{u}_i = (\dot{u}_{i,1}, \dot{u}_{i,2}, \ldots, \dot{u}_{i,T-1})'$ Then, we use the variables deviated from past means as instruments. We define

$$\boldsymbol{B} = diag(c_{T-1}, \ldots, c_2, c_1) \begin{bmatrix} -1 & 1 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 1 & 0 & \cdots & \cdots & 0 & \\ \vdots & \vdots & & & \ddots & \ddots & & \vdots \\ \frac{-1}{T-3} & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & 1 & 0 & 0 & \\ \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} & 1 & 0 & \\ \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & 1 & \end{bmatrix}, \tag{2.44}$$

Then, we define the instruments as

$$\ddot{\boldsymbol{y}}_i = \boldsymbol{B}\boldsymbol{y}_i = (\ddot{y}_{i,2}, \ldots, \ddot{y}_{i,T})' \tag{2.45}$$

where

$$\ddot{y}_{i,t} = c_{T-t+1}\left(y_{i,t} - \frac{y_{i,t-1} + \cdots + y_{i,1}}{t-1}\right), \quad t = 2, \ldots, T \tag{2.46}$$

Hence, we have a moment condition $E(\ddot{y}_{i,s}\dot{u}_{i,t}) = 0$ for $2 \leq s \leq t \leq T-1$. The DFIV estimator is

$$\hat{\phi}_{DFIV} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} \ddot{y}_{i,t}\dot{y}_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} \ddot{y}_{i,t}\dot{y}_{i,t}\right). \tag{2.47}$$

Theorem 2 *Asymptotic variances of DFIV estimator with fixed T and large N asymptotics is given by*

$$Avar\left(\hat{\phi}_{DFIV}\right) = \left(1-\phi^2\right)\left(\sum_{t=2}^{T-1} c_{T-t+1}^2 A_t\right)\left(\sum_{t=2}^{T-1} \xi_t c_{T-t+1}\left(1 - \frac{\phi\psi_{t-1}}{t-1}\right)^2\right)^{-2},$$

with

$$\xi_t = c_t\left(1 - \frac{\phi\psi_{T-t}}{T-t}\right), \ \psi_t = \frac{1-\phi^t}{1-\phi},$$

$$A_t = \left(1 - \frac{2\phi\psi_{t-1}}{t-1} + \frac{1}{(t-1)^2}\left(\frac{(t-1)(1+\phi)}{1-\phi} - \frac{2\phi\left(1-\phi^{t-1}\right)}{(1-\phi)^2}\right)\right). \tag{2.48}$$

2.3.4 Maximum Likelihood Estimation

Apart from the GMM type estimators, Hsiao et al. (2002) provide the TML estimation for short T dynamic panel data models. As we know, the TML estimator is more efficient than the GMM type estimators in general (Hsiao et al. (2002), Hayakawa and Pesaran (2015a) and Kruiniger (2013)). However, the initial condition is important for the ML type estimator. It is still unclear whether the TML estimator has good performance in different initial conditions.

To begin with, we consider the transformed model (2.12). By recursive substitution, we have (2.12) and let $m_i = m$, we have

$$\begin{aligned} \Delta y_{i,1} &= \phi^m \Delta y_{i,-m+1} + \sum_{j=0}^{m-1} \phi^j \Delta u_{i,1-j} \\ &= \phi^m \Delta y_{i,-m+1} + \nu_{i,1}. \end{aligned} \tag{2.49}$$

There are two assumptions for initial values, $\Delta y_{i,1}$:
Assumption 5 *(i) $|\phi| < 1$ and $m \to \infty$, the process has been going on. Then we have*

$$\begin{aligned} &E\left(\Delta y_{i,1}\right) = \lim_{m\to\infty} \phi^m E\left(\Delta y_{i,-m+1}\right) + \lim_{m\to\infty} E\left(\sum_{j=0}^{m-1} \phi^j \Delta u_{i,1-j}\right), \\ &Var\left(\Delta y_{i,1}\right) = \frac{2\sigma_u^2}{1+\phi}, \\ &Cov\left(\nu_{i,1}, \Delta u_{i,2}\right) = -E\left(u_{i,1}^2\right) = -\sigma_u^2 \; and \; Cov\left(\nu_{i,1}, \Delta u_{i,t}\right) = 0, \\ &for\ t = 3, 4, \ldots T; i = 1, \ldots N. \end{aligned} \tag{2.50}$$

(ii) If the process has started from a finite past period, we have

$$\begin{aligned} &E\left(\Delta y_{i,1}\right) = b, \\ &Var\left(\Delta y_{i,1}\right) = c\sigma_u^2, where\, c > 0, Cov\left(\nu_{i,1}, \Delta u_{i,2}\right) = -\sigma_u^2\, and \\ &Cov\left(\nu_{i,1}, \Delta u_{i,t}\right) = 0\ for\, t = 3, 4, \ldots, T,\ i = 1, 2, \ldots, N. \end{aligned} \tag{2.51}$$

Let $\Delta \boldsymbol{y}_i = \left(\Delta y_{i,1}, \ldots, \Delta y_{i,T}\right)'$ and $\Delta \boldsymbol{u}_i^* = \left(\Delta y_{i,1} - b^*, \Delta u_{i,2}, \ldots, \Delta u_{i,T}\right)'$, where $b = 0$ under infinite past starting point and $b = b^*$ which is an unknown parameter under finite past starting point. Note that Assumption 5 (ii) restricts that the expected $\Delta y_{i,1}$ are the same across all individuals, but it does not require that the initial value $y_{i,-m+1}$ have the same mean across all individuals. The covariance matrix of $\boldsymbol{u}_i^*$ is given by

$$\boldsymbol{\Omega} = \sigma_u^2 \begin{bmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} = \sigma_u^2 \boldsymbol{\Omega}^*, \tag{2.52}$$

where $\omega = (1/\sigma_u^2)\, Var\,(\Delta y_{i,1}) = \frac{2}{1+\phi}$. Assume $u_{i,t}$ is independent normal, the joint probability distribution function of $\Delta \boldsymbol{y}_i$ is

$$\prod_{i=1}^{N} (2\pi)^{-T/2} \left|\boldsymbol{\Omega}\right|^{-1/2} exp\left\{ -\frac{1}{2} \Delta \boldsymbol{u}_i^{*\prime} \boldsymbol{\Omega}^{-1} \Delta \boldsymbol{u}_i^{*} \right\}, \tag{2.53}$$

The TML estimator ϕ can be obtained by maximising the following log-likelihood function,

$$\begin{aligned} &\log L(\phi, \sigma_u^2, b) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln\left|\boldsymbol{\Omega}\right| - \frac{1}{2} \sum_{i=1}^{N} \Delta \boldsymbol{u}_i^{*\prime} \boldsymbol{\Omega}^{-1} \Delta \boldsymbol{u}_i^{*}, \\ &(\hat{\phi}, \hat{\sigma}_u^2, \hat{b}) = arg \max \log L(\phi, \sigma_u^2, b). \end{aligned} \tag{2.54}$$

Under the first and the second order conditions with a fixed number of parameters, the transformed MLE is consistent and asymptotically normally distributed.

Hsiao et al. (2002) recommend using the AH-IV estimator $\hat{\phi}_{AH-IV}$ as an initial (consistent) estimator. σ_u^2 can be estimated by

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} \left(\Delta y_{i,t} - \hat{\phi} \Delta y_{i,t-1} \right)^2}{2N(T-2)}. \tag{2.55}$$

Under Assumption 4 (i), an initial estimate of ω can be obtained by $\frac{2}{1+\hat{\phi}_{AH-IV}}$. Under Assumption 4 (ii), we can estimate ω by following procedure:
(i) Estimate ω by

$$\hat{\omega} = \frac{\sum_{i=1}^{N} (\Delta y_{i,1} - \hat{b})^2}{(N-1)\hat{\sigma}_u^2}, \tag{2.56}$$

where $\hat{\sigma}_u^2$ is given by (2.55) and $\hat{b} = \frac{\sum_{i=1}^{N} \Delta y_{i,1}}{N}$ is a consistent estimator of b.
(ii) Using $\hat{\omega}$ and $\hat{b}$ in minimum distance estimator of

$$\sum_{i=1}^{N} \Delta \boldsymbol{u}_i^{*\prime} \Omega^{*-1} \Delta \boldsymbol{u}_i^{*} \tag{2.57}$$

to obtain minimum distance estimator of ϕ.
Repeat the process (i) and (ii) until converge, we can obtain consistent $\hat{\omega}$.

Hayakawa and Pesaran (2015a) extends the transformed likelihood estimation to the case where the errors are cross-sectionally heteroskedastic. In Hayakawa and Pesaran (2015a), they allow $E(u_{i,t}) = 0$ and $E(u_{i,t}^2) = \sigma_{ui}^2$ such that $0 < \sigma_{ui}^2 < K < \infty$, for $i = 1, \ldots, N$ and $t = 1, 2, \ldots, T$. Thus, the covariance matrix of $\boldsymbol{u}_i^*$ is given by

$$\boldsymbol{\Omega}(\omega_i) = \sigma_{ui}^2 \begin{bmatrix} \omega_i & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} = \sigma_{ui}^2 \boldsymbol{\Omega}^*(\omega_i), \tag{2.58}$$

where $\omega_i > 0$ is a free parameter. Then, the log-likelihood function is give by

$$\begin{aligned} logL(\boldsymbol{\theta}_N) = &- \frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_{ui}^2 - \frac{1}{2}\sum_{i=1}^{N}\ln[1+T(\omega_i-1)] - \\ &\frac{1}{2}\sum_{i=1}^{N}\frac{1}{\sigma_{ui}^2}\Delta\boldsymbol{u}_i^{*'}\boldsymbol{\Omega}(\omega_i)^{-1}\Delta\boldsymbol{u}_i^{*'}, \end{aligned} \tag{2.59}$$

where $\boldsymbol{\theta}_N = (b, \phi, \omega_1, \ldots, \omega_N, \sigma_{u1}^2, \ldots, \sigma_{uN}^2)'$.

Therefore, the transformed likelihood estimation encounters the incidental parameters problem when sample size N increases. By using mis-specified model where the error variances are assumed to be homoskedastic, we can construct a pseudo log-likelihood function

$$\begin{aligned} logL_p(\boldsymbol{\theta_p}) = &- \frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln\sigma_u^2 - \frac{N}{2}\ln[1+T(\omega-1)] - \\ &\frac{1}{2\sigma_u^2}\sum_{i=1}^{N}\Delta\boldsymbol{u}_i^{*'}\boldsymbol{\Omega}(\omega)^{-1}\Delta\boldsymbol{u}_i^{*'}, \end{aligned} \tag{2.60}$$

where $\boldsymbol{\theta}_p = (b, \phi, \omega, \sigma_u^2)'$. By minimizing log-likelihood function (2.60), we obtain $\hat{\theta}_p$. To give the assumption of the relationship between the true value $\boldsymbol{\theta}_{0N} = (b_0, \phi_0, \omega_{01}, \ldots, \omega_{0N}, \sigma_{01}^2, \ldots, \sigma_{0N}^2)'$ and the pseudo true value $\boldsymbol{\theta}_{0p} = (b_{0p}, \phi_{0p}, \omega_{0p}, \sigma_{0p}^2)'$ below

Assumption 6 *The average true value as*

$$\bar{\sigma}_{0N}^2 = N^{-1}\sum_{i=1}^{N}\sigma_{0i}^2, \quad and \quad \bar{\omega}_{0N} = \frac{N^{-1}\sum_{i=1}^{N}\omega_{0i}\sigma_{0i}^2}{N^{-1}\sum_{i=1}^{N}\sigma_{0i}^2}, \tag{2.61}$$

as $N \to \infty$

$$\bar{\sigma}_0^2 = \lim_{N\to\infty}\bar{\sigma}_{0N}^2 \quad and \quad \bar{\omega}_0 = \frac{\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\omega_{0i}\sigma_{0i}^2}{\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_{0i}^2}. \tag{2.62}$$

If $|\sigma_{0i}|$ *and* $|\omega_{0i}|$ *are finite and bounded away from zero, the above assumption is satisfied.*

On theorem 2 of Hayakawa and Pesaran (2015a), they show that the pseudo estimator is consistent with the average true value when N is large, $\boldsymbol{\theta}_{0p} = (b_0, \phi_0, \bar{\sigma}_0^2, \bar{\omega}_0)$.

Hayakawa and Pesaran (2015a) use a mis-specified model to provide the relationship between the true value and the pseudo true value. Hayakawa and Pesaran (2015a) show that the quasi(pseudo) ML estimators are consistent under mis-specification.

2.4 Comparison of Finite Sample Behaviour of the Estimators

In this section, we use Monte Carlo simulations to investigate the finite sample performance of the TML estimator, the BMM estimator and the DFIV estimator in difference scenarios. In this exercise, we examine the behaviour under different initial conditions. We report the bias, root mean square error (RMSE) and size of the t-test for each estimator.

2.4.1 Monte Carlo design

In this Monte Carlo simulations, the $y_{i,t}$ are generated as

$$y_{i,t} = \phi y_{i,t-1} + \alpha_i + u_{i,t},\ u_{i,t} \sim U(-0.25, 0.25),\ t = 1, 2, \ldots T;\ i = 1, \ldots, N.$$

$$\alpha_i = \sum_{t=1}^{T} \gamma^t u_{i,t} + \pi_i,\ \pi_i \overset{iid}{\sim} N(0,1).$$

The processes of initial value $y_{i,0}$, as

$$y_{i,0} = \mu_i + \eta\pi_i + v_i,\ v_i \overset{iid}{\sim} N(0,1), \tag{2.63}$$

where $\mu_i = \frac{\alpha_i}{1-\phi}$ is long run means. If $\gamma \neq 0$, the individual effects are uncorrelated with errors $u_{i,t}$. The AB-GMM estimator and SYS-GMM estimator are not satisfied when $\gamma \neq 0$. If $\eta \neq 0$, the deviations of starting values from the long-run means affect initial value. The SYS-GMM estimator is not satisfied when $\eta \neq 0$.

Follow above setting, we consider four cases, as

1. $\gamma = 0.8$, $\eta = 1$.
2. $\gamma = 0.8$, $\eta = 0$.
3. $\gamma = 0$, $\eta = 0$.
4. $\gamma = 0$, $\eta = 1$.

The sample size $T = \{5, 10\}$, $N = \{100, 200, 500, 1000\}$ and the parameter, $\gamma = \{0.4, 0.8\}$. The number of replications is 2000.

2.4.2 Monte Carlo results

In the Monte Carlo simulation results, we investigate the behaviour of the TML estimator, BMM estimator and DFIV estimator in finite sample.

Table 2.1 reports the bias, RMSE and size of the t-test for the TML estimator, the BMM estimator and the DFIV estimator. We allow individual effects to be correlated with errors $u_{i,t}$ and the deviations of initial values from their long-run means. As we can see that the bias and the RMSE of the TML estimator is small in the Case 1. The size of the t-test based on the TML estimator is close to the nominal value of 0.05. As ϕ increases from 0.5 to 0.8, the TML estimator still performs better in terms of the bias, RMSE and size. In the Case 1, the bias of the BMM estimator is small in finite sample. The RMSE of the BMM estimator is slightly larger than that of the TML estimator in Case 1. As we can see that the size of the t-test based on the BMM estimator is close to 5% for all values of N considered. Compared to the TML estimator and the BMM estimator, the RMSE of the DFIV estimator is slightly larger. The size of the t-test based on the DFIV estimator is close to the norminal value of 0.05.

In Table 2.2, we restrict the individual effects uncorrelated with the deviations of initial values from their long-run means, whilst the individual effects are correlated with errors. We can see that the bias and the RMSE of the TML estimator are smaller that those in the Case 2. The reported size is close to 5% for all the values of N considered. The bias of the BMM estimator is small in Case 2. As expected, the RMSE of the BMM esimator is slightly larger than that of the TML estimator. Also, the size of the t-test based on the BMM estimator is close to the nominal value of 0.05. The bias of the DFIV estimator is small when both N and T are small. The size of the t-test based on the DFIV estimator is close to the nominal value of 0.05.

In Table 2.3, we restrict the individual effects uncorrelated with the deviations of initial values from their long-run means and the individual effects uncorrelated with errors. In this case, the bias and the RMSE of the TML estimator are small and the size of the TML estimator is close to 5%. The RMSE of the DFIV estimator is larger than the TML estimator, and the size is close to the nominal value of 0.05.

Table 2.4 shows the results for the case in which the individual effects are uncorrelated with errors but correlated with the deviations of initial values from their long-run means. As to be expected, the RMSE of the DFIV estimator is larger than those of the BMM estimator and the TML estimator in Case 4.

From table 2.1 to 2.4, we can see that the performance of the TML estimator the BMM and the DFIV estimator is robust to the different initial conditions.

Table 2.1 Case 1, Bias, RMSE and Size. ($\gamma = 0.8$, $\eta = 1$)

Results for $\phi = 0.5$.

	Bias(×100)				RMSE(×100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	0.04	0.00	-0.03	0.01	1.37	0.76	0.60	0.42	7.70	5.40	5.35	5.15
10	-0.03	0.02	0.00	0.00	1.05	0.59	0.46	0.33	6.75	4.60	5.75	5.25
BMM												
5	0.04	-0.01	-0.03	0.00	1.54	0.84	0.68	0.48	6.80	5.15	4.80	4.95
10	-0.04	0.03	0.01	-0.01	1.41	0.84	0.65	0.46	5.20	5.30	5.30	4.60
DFIV												
5	0.01	0.05	0.03	0.03	4.18	2.46	1.91	1.35	5.05	3.70	5.25	4.60
10	-0.09	-0.05	0.01	0.01	4.26	2.36	1.85	1.34	5.25	4.55	5.00	3.90

Results for $\phi = 0.8$.

	Bias(×100)				RMSE(×100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	0.04	0.00	-0.07	0.01	2.26	1.29	1.01	0.70	7.35	5.65	5.60	5.00
10	-0.04	0.03	-0.01	-0.01	1.21	0.68	0.52	0.37	7.45	4.85	5.55	5.85
BMM												
5	0.04	0.00	-0.06	0.01	2.48	1.44	1.13	0.80	5.70	5.95	6.10	5.35
10	-0.07	0.05	-0.01	-0.01	1.68	0.97	0.75	0.52	5.45	5.60	5.15	4.45
DFIV												
5	0.05	-0.01	0.02	-0.02	3.22	1.90	1.46	1.05	3.60	3.55	4.45	4.95
10	0.02	-0.01	0.02	-0.01	1.88	1.06	0.84	0.57	4.20	4.60	4.45	4.30

$y_{i,t} = \phi y_{i,t-1} + \alpha_i + u_{i,t},\ u_{i,t} \sim U(-0.25, 0.25),\quad t = 1, 2, \ldots T;\ i = 1, \ldots, N.\ \alpha_i = \sum_{t=1}^{T} \gamma^t u_{i,t} + \pi_i,\ \pi_i \overset{iid}{\sim} N(0,1)$. The processes of initial value $y_{i,0}$, as $y_{i,0} = \mu_i + \eta\pi_i + v_i,\ v_i \overset{iid}{\sim} N(0,1)$, where $\mu_i = \frac{\alpha_i}{1-\phi}$ is long run means.

Table 2.2 Case 2, Bias, RMSE and Size. ($\gamma = 0.8$, $\eta = 0$)

Results for $\phi = 0.5$.

	Bias(×100)				RMSE(×100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	-0.01	-0.03	-0.05	0.01	1.92	1.06	0.83	0.59	6.45	4.95	5.00	5.70
10	-0.03	0.00	0.01	0.01	1.42	0.82	0.62	0.45	6.05	5.25	5.10	5.20
BMM												
5	-0.01	-0.05	-0.05	0.00	2.19	1.19	0.94	0.68	6.00	4.70	4.65	4.70
10	-0.02	0.01	0.02	-0.01	2.03	1.19	0.91	0.63	4.90	5.15	4.75	5.05
DFIV												
5	0.02	0.01	0.04	-0.02	4.24	2.45	1.89	1.38	5.65	4.45	4.75	4.40
10	-0.04	-0.06	-0.01	-0.03	4.18	2.42	1.85	1.29	5.30	4.00	5.00	4.45

Results for $\phi = 0.8$.

	Bias(×100)				RMSE(×100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	-0.03	-0.05	-0.09	0.03	3.24	1.81	1.43	0.99	6.60	4.90	5.60	5.60
10	-0.06	-0.01	0.00	0.02	1.66	0.95	0.74	0.53	6.10	5.60	5.40	5.90
BMM												
5	0.05	-0.06	-0.09	0.03	3.76	2.11	1.66	1.18	5.35	4.75	5.40	5.70
10	-0.05	0.03	0.00	0.00	2.48	1.49	1.12	0.78	4.95	5.65	5.00	4.90
DFIV												
5	0.04	-0.02	0.02	-0.03	2.42	1.41	1.08	0.79	4.60	4.10	4.75	5.00
10	0.01	0.00	0.01	-0.01	1.36	0.79	0.62	0.42	4.65	4.60	5.10	4.60

See the note to Table 2.1.

Table 2.3 Case 3, Bias, RMSE and Size. ($\gamma = 0$, $\eta = 0$)

Results for $\phi = 0.5$.

	Bias($\times$100)				RMSE($\times$100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	-0.01	-0.03	-0.05	0.01	1.92	1.05	0.83	0.59	6.45	4.95	4.95	5.65
10	-0.03	-0.01	0.01	0.01	1.42	0.82	0.62	0.45	6.00	5.45	5.10	5.20
BMM												
5	-0.01	-0.05	-0.05	0.00	2.19	1.19	0.94	0.68	6.00	4.70	4.65	4.70
10	-0.02	0.01	0.02	-0.01	2.03	1.19	0.91	0.63	4.90	5.15	4.75	5.05
DFIV												
5	0.02	0.01	0.04	-0.02	4.24	2.45	1.89	1.38	5.65	4.45	4.75	4.40
10	-0.04	-0.06	-0.01	-0.03	4.18	2.42	1.85	1.29	5.30	4.00	5.00	4.45

Results for $\phi = 0.8$.

	Bias($\times$100)				RMSE($\times$100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	-0.03	-0.05	-0.09	0.03	3.24	1.81	1.43	0.99	6.65	4.90	5.50	5.50
10	-0.06	-0.01	0.00	0.02	1.66	0.95	0.74	0.53	5.95	5.65	5.35	5.90
BMM												
5	0.05	-0.06	-0.09	0.03	3.76	2.11	1.66	1.18	5.35	4.75	5.40	5.70
10	-0.05	0.03	0.00	0.00	2.48	1.49	1.12	0.78	4.95	5.65	5.00	4.90
DFIV												
5	0.04	-0.02	0.02	-0.03	2.42	1.41	1.08	0.79	4.60	4.10	4.75	5.00
10	0.01	0.00	0.01	-0.01	1.36	0.79	0.62	0.42	4.65	4.60	5.10	4.60

See the note to Table 2.1.

Table 2.4 Case 4, Bias, RMSE and Size. ($\gamma = 0$, $\eta = 1$)

Results for $\phi = 0.5$.												
	Bias($\times$100)				RMSE($\times$100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	0.04	0.00	-0.03	0.01	1.37	0.77	0.60	0.42	7.55	5.40	5.35	5.15
10	-0.03	0.03	0.00	0.00	1.05	0.59	0.46	0.33	6.80	4.70	5.65	5.35
BMM												
5	0.04	-0.01	-0.03	0.00	1.54	0.84	0.68	0.48	6.80	5.15	4.80	4.95
10	-0.04	0.03	0.01	-0.01	1.41	0.84	0.65	0.46	5.20	5.30	5.30	4.60
DFIV												
5	0.01	0.05	0.03	0.03	4.18	2.46	1.91	1.35	5.05	3.70	5.25	4.60
10	-0.09	-0.05	0.01	0.01	4.26	2.36	1.85	1.34	5.25	4.55	5.00	3.90

Results for $\phi = 0.8$.												
	Bias($\times$100)				RMSE($\times$100)				Size			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
TML												
5	0.04	0.00	-0.07	0.01	2.26	1.29	1.01	0.70	7.30	5.70	5.60	4.95
10	-0.04	0.03	0.00	0.00	1.21	0.68	0.52	0.37	7.55	4.80	5.50	5.80
BMM												
5	0.04	0.00	-0.06	0.01	2.48	1.44	1.13	0.80	5.70	5.95	6.10	5.35
10	-0.07	0.05	-0.01	-0.01	1.68	0.97	0.75	0.52	5.45	5.60	5.15	4.45
DFIV												
5	0.05	-0.01	0.02	-0.02	3.60	1.90	1.46	1.05	3.60	3.55	4.45	4.95
10	0.02	-0.01	0.02	-0.01	1.88	1.06	0.84	0.57	4.20	4.60	4.45	4.30

See the note to Table 2.1.

2.5 Concluding remarks

This chapter compares the performance of different estimators for short T dynamic panel data models, namely, the TML estimator, the BMM estimator and the DFIV estimator. In the Monte Carlo experiment, we particularly investigate effects of different initial conditions for these estimators. We find that in terms of the bias and the RMSE, the TML estimator and the BMM estimator perform better for all the designs. The size of the t-tests based on the TML estimator and the BMM estimator is close to the nominal level. The DFIV estimator is the least efficient in our design.

Chapter 3

The speed of adjustment to the target capital structure in the US

3.1 Introduction

The capital structure irrelevance theory (Modigliani and Miller (1958)) is one of the most important theories of corporate finance. The idea of this theory is that the value of a company is not influenced by its capital structure. The assumption of capital structure irrelevance theory is based on the perfectly efficient markets. Although this assumption is quite strong in the real world[1], it gives the clear explanation of financing strategy. After Modigliani and Miller (1958) developed the capital structure irrelevance theory, many researchers relaxed the assumption of the perfectly efficient markets by including corporate tax, bankruptcy cost and agency-related costs (Kraus and Litzenberger (1973), Scott Jr (1976) and Kim (1978)). Much of the extant literature focuses on two preeminent capital structure theories. One is the static trade-off model (Graham and Harvey (2001)), in which firms choose optimal leverage to balance the costs and benefits. The other is pecking order theory by Myers and Majluf (1984), in which the most preferred way of financing by firms is self financing by retained earning, then by debt, and the least preferred is by issuing new equity. The above two main theories are successful in explaining firms' heterogeneous capital structure.

Due to the shocks, firms may temporarily deviate from their target leverage. To investigate whether or not firms adjust their leverage towards target leverage is important for examining trade off theory (Ross et al. (2014), Arioglu and Tuan (2014) and Drobetz and Wanzenried (2006)). Also, to investigate how fast firms adjust their leverage towards their target leverage would help us to understand the behavior of firm management while considering financing policy. However, the target leverage is not observed in practice. This issue increases the difficulty of estimation of the speed of adjustment (SOA). In empirical literature, we often assume the

[1] Modigliani and Miller (1963) developed the theory by including taxes.

target leverage is associate with firms characteristics (e.g., growth opportunities, tangibility, size, profitability and non-debt tax shields). But, the explanatory power of these characteristics are very low (Westerlund et al. (2021) and Lemmon et al. (2008)). Therefore, Westerlund et al. (2021) indicate that controlling for unobserved heterogeneity is important in estimating the SOA.

Some researchers apply the ordinary least squares (OLS) method to estimate the SOA, which yield estimated values ranging from 9% to 18% per year (Fama and French (2002), Kayhan and Titman (2007), Flannery and Rangan (2006) and Lemmon et al. (2008)). As the OLS estimator tends to be downward biased for dynamic models, other researchers estimate SOA using the generalised method of moments (GMM) or instrumental variable (IV) method. Under the GMM/IV estimation, the SOA is estimated around 17% to 34% per year (Flannery and Rangan (2006), Lemmon et al. (2008), Huang and Ritter (2009) and Dang et al. (2014)). Iliev and Welch (2010) estimated an negative SOA, which implies that leverage is not mean reverting. Dang et al. (2012) and Dang et al. (2014) consider asymmetric capital structure adjustments. They develop dynamic panel threshold models to investigate heterogeneous SOA in different regimes. They find that more constrained firms have higher SOA over the pre-crisis period.

In the recent literature, the importance of controlling unobserved heterogeneity in the model has been emphasised for estimating the SOA . Lemmon et al. (2008) indicate that including the firms' fixed effects and time effects in the model is important for estimating SOA. Westerlund et al. (2021) also recognize this unobserved heterogeneity should not be ignore on estimation of a partial adjustment model. Westerlund et al. (2021) and DeAngelo and Roll (2015) indicate that controlling conventional additive effects in the partial adjustment model is not enough to resolve the endogeneity problem. As suggested by Westerlund et al. (2021), controlling unobserved interactive effects in partial adjustment model is important. However, the aforementioned works do not control heteroskedasticity, which is likely in the firm data models.

In this chapter, we estimate the SOA controlling the unobserved interactive effects and cross-sectional heteroskedasticity using US firms data from 1960 to 2017. We apply the approach proposed by (Hayakawa et al. (2021)) to control the heteroskedasticity in the model, which is new in the literature. The data contain 315,621 firms-year observations that consist of 16,502 firms from 1960 to 2017. We use a rolling window approach to examine the trade off theory because this approach is used to evaluate the stability of coefficients of the model in the sample size. The approach which we take is to use rolling 8 year fixed windows of data to estimate SOA.

Following Hayakawa et al. (2021), we have estimated the number of factor by using the sequential multiple testing likelihood ratio (MTLR) procedure, which provides the evidence of unobserved interactive effects that may affect the estimation of the SOA. In the results, we show that the SOA decrease from around 0.9 to 0.5 from 1960 to 1982. The SOA are fluctuated in the range of 0.2 to 0.8 from 1990 to

2010.

This chapter is organized as follows: Section 2 sets out the partial adjustment model. Section 3 introduces robust QML estimators for the short T dynamic panel data models with interactive effects. This section also introduces the sequential multiple testing likelihood ratio procedure for estimating the number of factors. Section 4 studies the potential determinants of the SOA. Data and empirical results are reported in Section 5, and Section 6 concludes.

3.2 Partial adjustment model

To understand whether firms have target leverage ratio and how quickly they adjust toward them, we use the following standard partial adjustment model of leverage (Flannery and Rangan (2006)):

$$\Delta l_{i,t} = \phi \left(l_{i,t}^{*} - l_{i,t-1} \right) + u_{i,t},\ i = 1, \dots N,\ t = 1, \dots, T. \tag{3.1}$$

where $l_{i,t}$ is the actual leverage ratio and $l_{i,t}^{*}$ is the target leverage ratios for firm i at time t. $u_{i,t}$ is the error term. ϕ is the SOA to the target leverage ratio. The higher the value of ϕ, the faster the adjustment is.

The target leverage is associated with firm characteristics as:

$$l_{i,t}^{*} = \boldsymbol{\delta}' \boldsymbol{x}_{i,t}, \tag{3.2}$$

where $\boldsymbol{x}_{i,t}$ is a $k \times 1$ vector of exogenous variables and $\boldsymbol{\delta}$ is a $k \times 1$ vector of structural parameters. As we can see, the target leverage differs across firms and over time accordingly to the time varying firm characteristics. These firm characteristics include profitability, growth opportunities, firm size, tangibility and non-debt tax shields etc. According to the trade off theory, $\boldsymbol{\delta} \neq \mathbf{0}$, and the variation in $l_{i,t}^{*}$ should be nontrivial.

Substituting (3.2) into (3.1), we have

$$l_{i,t} = \rho l_{i,t-1} + \boldsymbol{\beta}' \boldsymbol{x}_{i,t} + u_{i,t}, \tag{3.3}$$

where $\rho = 1 - \phi$ and $\boldsymbol{\beta} = \phi \boldsymbol{\delta}$.

3.3 Econometric methodology

3.3.1 Estimation method

In much of the literature (See eg. Flannery and Rangan (2006), Lemmon et al. (2008) and Dang et al. (2012)), they find that firms' specific unobserved fixed effects substantially influence estimated SOA. However, most literature ignore that distinct sources of firms specific unobserved effects may varies over times. Also, most empirical work assume that the errors are cross-sectionally homoscedastic. To capture this corss-sectional dependence, we apply estimation method by Hayakawa et al. (2021) and extend it to permit cross-sectionally heteroskedasticity.

From model (3.3), we have

$$l_{i,t} = \rho l_{i,t-1} + \boldsymbol{\beta}'\boldsymbol{x}_{i,t} + u_{i,t},\ \ for\ t = 0, \dots, T;\ i = 1, \dots, N, \tag{3.4}$$

with

$$u_{i,t} = \alpha_i + \boldsymbol{\gamma}_i'\boldsymbol{f}_t + e_{i,t}, \tag{3.5}$$

where α_i denote unit-specific fixed effects and $\boldsymbol{\gamma}_i'\boldsymbol{f}_t$ denote an interactive effects with $\boldsymbol{f}_t$ an $m \times 1$ vector of unobserved common factors, $\boldsymbol{\gamma}_i$ an $m \times 1$ vector of associated factor loading, and $e_{i,t}$ is the idiosyncratic error term.

First, we combine (3.4) and (3.5) and eliminate the individual effects α_i by first differencing. Then, we have

$$\Delta l_{i,t} = \rho \Delta l_{i,t-1} + \boldsymbol{\beta}'\Delta\boldsymbol{x}_{i,t} + \boldsymbol{g}_t'\gamma_i + \Delta e_{i,t},\ \ for\ t = 2, \dots, T;\ i = 1, \dots, N, \tag{3.6}$$

where $\boldsymbol{g}_t = \Delta\boldsymbol{f}_t$ for some $t \geq 2$.

When the process start from some arbitrary point at $t = -S + 1$ with $\Delta l_{i,-S+1}$ as given, we have

$$\Delta l_{i,1} = \rho^S \Delta l_{i,-S+1} + \sum_{j=0}^{S-1} \rho^j \boldsymbol{\beta}'\Delta\boldsymbol{x}_{i,1-j} + \tilde{\boldsymbol{g}}_1'\gamma_i + \sum_{j=0}^{S-1} \rho^j \Delta e_{i,1-j}, \tag{3.7}$$

where $\tilde{\boldsymbol{g}}_1 = \sum_{j=0}^{S-1} \rho^j \boldsymbol{g}_{1-j}$.

Following from Hsiao et al. (2002) and Hayakawa et al. (2021) the initial value $\Delta l_{i,1}$ depends on the $\Delta\boldsymbol{x}_i$ which can be observed. Then, we have $\Delta l_{i,1}$ as

$$\Delta l_{i,1} = b + \boldsymbol{\pi}'\Delta\boldsymbol{x}_i + \xi_{i,1}, \tag{3.8}$$

with

$$\xi_{i,1} = \tilde{\boldsymbol{g}}_1'\gamma_i + v_{i,1}, \tag{3.9}$$

where b is a constant, $\boldsymbol{\pi}$ is a $T \times 1$ vector of constant, $\Delta\boldsymbol{x}_i = \left(\Delta\boldsymbol{x}_{i,1}', \Delta\boldsymbol{x}_{i,2}', \dots, \Delta\boldsymbol{x}_{i,T}'\right)'$ and $v_{i,1}$ is independently distributed across i, such that $v_{i,1} \overset{i.i.d.}{\sim} (0, \omega_i\sigma_i^2)$, and

$$\begin{aligned} Cov\left(v_{i,1}, \Delta e_{i,2}\right) &= -\sigma_i^2 \\ Cov\left(v_{i,1}, \Delta e_{i,t}\right) &= 0,\ \ for\ t = 3, \dots, T \end{aligned} \tag{3.10}$$

Let $\Delta\boldsymbol{l}_i = \left(\Delta l_{i,1}, \Delta l_{i,2}, \dots, \Delta l_{i,T}\right)'$, and the $T \times (Tk + k + 2)$ matrix given by

$$\Delta\boldsymbol{W}_i = \begin{bmatrix} 1 & \Delta\boldsymbol{X}_i' & 0 & 0 \\ 0 & \boldsymbol{0} & \Delta\boldsymbol{X}_{i,2}' & \Delta y_{i,1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \boldsymbol{0} & \Delta\boldsymbol{X}_{i,T}' & \Delta y_{i,T-1} \end{bmatrix} \tag{3.11}$$

Stacking the T observations for each i, the transformed model can be expressed as

$$\begin{aligned}\Delta \boldsymbol{l}_i &= \Delta \boldsymbol{W}_i \boldsymbol{\varphi} + \boldsymbol{\xi}_i \\ \boldsymbol{\xi}_i &= \boldsymbol{G}\boldsymbol{\gamma}_i + \boldsymbol{r}_i,\end{aligned} \tag{3.12}$$

where $\boldsymbol{\varphi} = \left(b, \boldsymbol{\pi}', \boldsymbol{\beta}', \rho\right)'$, $\boldsymbol{G}' = (\tilde{\boldsymbol{g}}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_T)$, $\boldsymbol{r}_i = (v_{i,1}, \Delta e_{i,2}, \ldots, \Delta e_{i,T})'$, and $\boldsymbol{\xi}_i = (\xi_{i,1}, \xi_{i,2}, \ldots, \xi_{i,T})'$.

Consider the transformed model (3.12) and under the heteroskedasticity, we have

$$E\left(\boldsymbol{r}_i \boldsymbol{r}_i'\right) = \sigma_i^2 \begin{bmatrix} \omega_i & -1 & & & 0 \\ -1 & 2 & \ddots & & 0 \\ & & \ddots & & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_i^2 \boldsymbol{\Omega}(\omega_i) \tag{3.13}$$

Since $\boldsymbol{\gamma}_i$ and $\boldsymbol{r}_i$ are independently distributed, we have

$$Var\left(\boldsymbol{\xi}_i\right) = \boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_N\right) = \sigma_i^2 \boldsymbol{\Omega}\left(\omega_i\right) + \boldsymbol{G}\boldsymbol{\Omega}_\gamma \boldsymbol{G}' = \sigma_i^2\left(\boldsymbol{\Omega}\left(\omega_i\right) + \boldsymbol{Q}_i \boldsymbol{Q}_i'\right), \tag{3.14}$$

where $\boldsymbol{Q}_i' = \sigma_i^{-1} \boldsymbol{G}\boldsymbol{\Omega}_\gamma^{1/2}$, $rank(\boldsymbol{Q}_i) = m$, and
$\boldsymbol{\psi}_N = \left(\omega_1, \ldots, \omega_N, \sigma_1^2, \ldots, \sigma_N^2, vec(\boldsymbol{Q}_1)', \ldots, vec(\boldsymbol{Q}_N)'\right)'$.

The quasi-log-likelihood function of the transformed model (3.12) is given by

$$\begin{aligned}\ell\left(\boldsymbol{\theta}_N\right) &= -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln\left|\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_N\right)\right| - \frac{1}{2}\sum_{i=1}^N \boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}\right)\boldsymbol{\Sigma}_\xi\left(\boldsymbol{\psi}_N\right)\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}\right) \\ &= -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln\left(2\sigma_i\right) - \frac{1}{2}\sum_{i=1}^N \ln\left|\boldsymbol{\Omega}\left(\omega_i\right) + \boldsymbol{Q}_i\boldsymbol{Q}_i'\right| - \\ &\quad \frac{1}{2}\sum_{i=1}^N \frac{1}{\sigma_i^2}\boldsymbol{\xi}_i'\left(\boldsymbol{\varphi}\right)\left(\boldsymbol{\Omega}\left(\omega_i\right) + \boldsymbol{Q}_i\boldsymbol{Q}_i'\right)^{-1}\boldsymbol{\xi}_i\left(\boldsymbol{\varphi}\right),\end{aligned} \tag{3.15}$$

where $\boldsymbol{\theta}_N = \left(\boldsymbol{\varphi}', \boldsymbol{\psi}_N'\right)'$ However, to find the optimal solution from the above likelihood function (3.15) is impossible, because the number of parameters increases with N (Neyman and Scott (1948)). To deal with this problem, we follow Hayakawa and Pesaran (2015b) to show that pseudo quasi maximum likelihood estimator of $\boldsymbol{\varphi}$ are consistent under mis-specification.

The pseudo log-likelihood function of the transformed model (3.12) is given by

$$\begin{aligned}\ell_p(\boldsymbol{\theta}) &= -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})| - \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\xi}_i'(\boldsymbol{\varphi})\boldsymbol{\Sigma}_\xi(\boldsymbol{\psi})\boldsymbol{\xi}_i(\boldsymbol{\varphi}) \\ &= -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(2\sigma) - \frac{N}{2}\ln\left|\boldsymbol{\Omega}(\omega) + \boldsymbol{Q}\boldsymbol{Q}'\right| - \\ &\frac{1}{2\sigma^2}\boldsymbol{\xi}_i'(\boldsymbol{\varphi})\left(\boldsymbol{\Omega}(\omega) + \boldsymbol{Q}\boldsymbol{Q}'\right)^{-1}\boldsymbol{\xi}_i(\boldsymbol{\varphi}),\end{aligned} \tag{3.16}$$

Based on the heteroskedastic errors, the pseudo-true value of $\boldsymbol{\theta}$ is $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_*', \boldsymbol{\psi}_*'\right)'$ which is the solution of $\lim_{N\to\infty} E\left(\partial\ell_p(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}\right) = 0$.

Follow Theorem 2 of Hayakawa and Pesaran (2015b), we have the pseudo true value

$$\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_0', \bar{\boldsymbol{\psi}}'\right) = \left(\boldsymbol{\varphi}_0', \bar{\sigma}_0^2, \bar{\omega}_0, vec\left(\bar{\boldsymbol{Q}}_0\right)'\right)' = \bar{\boldsymbol{\theta}}_0, \tag{3.17}$$

where

$$\bar{\sigma}_{N,0}^2 = \frac{\sum_{i=1}^N \sigma_{i0}^2}{N}, \ , \ \bar{\omega}_{N,0} = \frac{N^{-1}\sum_{i=1}^N \omega_{i,0}\sigma_{i0}^2}{N^{-1}\sum_{i=1}^N \sigma_{i,0}^2}, \ and \ \bar{\boldsymbol{Q}}_{N,0} = \frac{\sum_{i=1}^N \frac{1}{\sigma_i}\boldsymbol{G}\boldsymbol{\Omega}_\eta^{1/2}}{N} \tag{3.18}$$

$$\bar{\sigma}_0^2 = \lim_{N\to\infty}\bar{\sigma}_{N,0}^2, \quad \bar{\omega}_0 = \frac{\lim_{N\to\infty}\sum_{i=1}^N \omega_{i,0}\sigma_{i,0}^2}{\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \sigma_{i,0}^2}, \ and \ \bar{\boldsymbol{Q}}_0 = \lim_{N\to\infty}\bar{\boldsymbol{Q}}_{N,0} \tag{3.19}$$

By maximizing pseudo log likelihood function (3.16) respect to $\boldsymbol{\theta}$, we can obtain the pseudo estimator $\hat{\boldsymbol{\theta}}$[2]. Then as $N\to\infty$, the estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normal with

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{A}^{*-1}\boldsymbol{B}^*\boldsymbol{A}^{*-1}\right), \tag{3.20}$$

where

$$\begin{aligned}\boldsymbol{\theta}_* &= \left(\boldsymbol{\varphi}_0', \bar{\sigma}_0^2, \bar{\omega}_0, vec\left(\bar{\boldsymbol{Q}}_0\right)'\right)' \\ \boldsymbol{A}^* &= \lim_{N\to\infty} E\left(-\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) \\ \boldsymbol{B}^* &= \lim_{N\to\infty} E\left(\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\theta}}\frac{\partial\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\theta}'}\right).\end{aligned} \tag{3.21}$$

[2]Following Hayakawa et al. (2021), we use an eigenvalue approach to simplify the computations. See Appendix B for details.

3.3.2 Estimating the number of factor

For estimating the number of factors, we follow Hayakawa et al. (2021) by using a sequential multiple testing likelihood ratio procedure. The LR statistics for testing $H_0 : m = m_0$ against $H_1 : m = m_{max}$, for $m_0 = \{0, 1, 2, \ldots, m_{max} - 1\}$ and $m_{max} = T - 2$, are give by

$$LR_N(m_{max}, m_0) = 2\left[\ell_N\left(\hat{\boldsymbol{\theta}}_{m_{max}}\right) - \ell_N\left(\hat{\boldsymbol{\theta}}_{m0}\right)\right] \xrightarrow{d} \chi^2(r_0),\ \ for\ m_0 = 0, 1, 2, \ldots, T-3, \tag{3.22}$$

where $\hat{\boldsymbol{\theta}}_m = argmax_{\phi_m}\ell_N(\boldsymbol{\theta}_m)$ and $r_0 = T(T+1)/2 - 3 - (Tm_0 - m_0(m_0 - 1)/2)$. The testing procedure as follow,

Table 3.1 Sequential multiple testing likelihood ratio procedure

$\hat{m} = 0$	if $LR_N(m_{max}, m_0 = 0) < \chi^2_{r_0}[p/N(T-2)]$
$\hat{m} = 1$	if $LR_N(m_{max}, m_0 = 0) \geq \chi^2_{r_0}[p/N(T-2)]$ and $LR_N(m_{max}, m_0 = 1) < \chi^2_{r_0}[p/N(T-2)]$
$\hat{m} = 2$	if $LR_N(m_{max}, m_0 = 0) \geq \chi^2_{r_0}[p/N(T-2)]$; $LR_N(m_{max}, m_0 = 1) \geq \chi^2_{r_0}[p/N(T-2)]$ and $LR_N(m_{max}, m_0 = 2) < \chi^2_{r_0}[p/N(T-2)]$

The parameter p can be viewed as the nominal size of test.

3.4 Firm specific characteristics

- Growth Opportunities

Myers (1977) explains that the firms' value is the present value of options to make further investments. In some situations, the firm financed with risky debt will give up investment opportunities because it could make a positive net contribution to the market value of the firm. Therefore, the debt ratio of a firm is inversely related to the growth opportunities. Dang et al. (2012) also indicates that most of high growth firms are young and they may limit internal fund (debt). Most of low growth firms are mature and they may seek to maintain a high leverage ratio. In this chapter, we use market value of assets to the book value of assets as a proxy for growth opportunities (Dang et al. (2012) and Flannery and Rangan (2006)).

- Tangibility

In Hall (2012) and Halling et al. (2016), tangibility is the ratio of property, plant, and equipment to total assets. Tangible assets are easy to collateralize and thus they reduce the agency costs of debt. Ozkan (2001) indicates that firms with greater tangible assets have a higher debt capacity. Therefore, the relationship between tangibility and leverage ratio is positive. However, Rajan and Zingales (1995) indicates that the relationship on tangibility varies across the different estimators.

Halling et al. (2016) also find that the relationship between asset tangibility and leverage varies substantially among countries.

- Size

In general, large firms are mature with high tangibility, more diversification and profitability. Therefore, large firms enjoy easier access to the capital market (Ferri and Jones (1979) and Ozkan (2001)). The credit rating of small size firms is generally low, so they prefer lower leverage ratio to avoid liquidation. We use the natural log of total assets, measured in year-2000 dollars, as a proxy for size.

- Profitability

In pecking order theory (Myers (1984) and Myers and Majluf (1984)), firms first prefer internal financing (retained earning), then debt, and lastly raising equity. Therefore, highly profitable firms prefer internal financing over external finance. Following Dang et al. (2012) and Ozkan (2001), we use the ratio of the earnings before interest, tax and depreciation (EBITD) to total assets as a proxy for profitability.

- Non-debt tax shields

Firms can benefit from non-debt tax shields when they own large fixed assets. Firms can also benefit from debt tax shields. DeAngelo and Masulis (1980) indicates that non-debt tax shields can be a substitute for interest expenses. Therefore, there is an inverse relationship between the leverage ratio and non-debt tax shields. Following Ozkan (2001) and Dang et al. (2012), we use the ratio of annual depreciation expense to total assets as a proxy for non-debt tax shields. However, firms with higher depreciation ratios generally have relatively low growth opportunities. Therefore, the depreciation ratio may also be a proxy for growth opportunities (Barclay and Smith Jr (1995) and Krishnaswami and Subramaniam (1999)). Following the argument about growth opportunities, higher depreciation ratios generally have relatively low growth opportunities. Thus, this implies that there is a positive relation between the leverage ratio and non-debt tax shields (Ozkan (2001)).

3.5 Data and Empirical Results

3.5.1 Data

First, we collect balance sheet annual data for US firms from the CRSP/Compustat database. The sample consists of 596,198 firms' annual observations from 1960 to 2017. Then, following empirical studies by (Dang et al. (2012) and Ozkan (2001)), we procreate data as follows. First, we exclude utilities (SIC codes: 4900-4999) and financial firms (SIC codes: 6000-6999) because these industries have different accounting considerations. Second, in order to satisfy the order conditions for identification (Hayakawa et al. (2021)), we retain the firms with at least five years of observations. Third, we remove the observations that include missing data. Finally, all of the variables are winsorized at the 1st and 99st percentiles to avoid the impact of extreme outliers (Flannery and Rangan (2006) and Dang et al. (2012)). This leaves us with data preprocessing, the panel data set with 315,621 firms-year observations that consist of 16,502 firms from 1960 to 2017. In Table 3.2, we present the summary statistics for the variable.

Table 3.2 Variable Definition

Variable	Definitions
Leverage	The ratio of total debt to total assets
Growth Opportunities	The ratio of total liabilities plus the market value of equity to total assets
Tangibility	The ratio of property, plant, and equipment to total assets
Size	The natural log of total assets, measured in year-2000 dollars
Profitability	The ratio of earnings before interest and taxes to total assets
Non-debt tax shields	the ratio of depreciation to total assets

The data set is a panel of US firms collected from the CRSP/Compustat database over the period 1960-2017. Following Flannery and Rangan (2006) and Dang et al. (2014), we collect data from CRSP/Compustat database and the Data items used in target leverage model are as follows. Leverage: $((dltt + dlc)/at)$. Growth Opportunities: $((dltt + dlc + pstkl + csho \times prcc)/at)$. Tangibility: $(ppent/at)$. Size: $((ln(at \times CPI2000/CPI))$, where CPI is the consumer price index). Non-debt tax shields: (dp/at).

3.5.2 Empirical results

In this section, we report the empirical results based on the annual US samples from 1960 to 2017.

Figure 3.2 shows that leverage ratio of US firms increase from about 18.5% to 27.8% from 1960 to 1970. In Figure 3.2, we can see that leverage ratio of US firms increased from 25.3% to 28.3% from 1968 to 1975. From 1971 to 2010, the leverage ratio of US firms fluctuated between 24.3% and 30.6%. After 2010, we can see the leverage ratio of US firms increased from about 24.3% to 29.3% The leverage ratio in the period of Global Financial Crisis (2007–2009) increased from around 25.53% (2007) to 27.55% (2008), and then decreased to 25.50% in 2009.

The Figure 3.2 shows the result of the rolling window approach. The solid line graphs in the figures represent the estimated coefficients. These graphs show the two standard deviation bands (upper and lower bands of doted lines) that confirms the coefficients statistical significance. From the figure 3.2, We can see the SOA is fluctuated from 1967 to 2017.

Byoun (2008) find that the SOA is around 33% when firms have above-target debt with a financial surplus and about 20% when firms have below-target debt with a financial deficit, but that the SOA is substantially decreased when firms have a financial deficit with above-target debt or when they have a financial surplus with below-target debt. Byoun (2008) also indict that firms make the most significant adjustments toward the target when they have above-target debt with a financial surplus. Firms therefore appear to face lower adjustment costs in reducing debt from the above-target debt level than in issuing debt with the below-target debt level, or the costs of keeping above-target debt may be much higher than those of keeping below-target debt.

In figure 3.2, we can see the SOA decrease from around 0.9 to 0.5 from 1967 to 1987. Therefore, from the results above it is deduced that the average firms may face a financial deficit with above-target debt in 1980s . As we know that the U.S. economy experienced a deep recession between 1980 and 1982.

From 1987 to 1990, the SOA increase from around 0.5 to 0.8. Between 1987 and 1990 the average firms may have above-target debt with a financial surplus. Again, the SOA decrease from around 0.8 to 0.4 between 1990 and 1996. In this period, the average firms may face a financial deficit with above-target debt. Between 2004 and 2008, the SOA decrease from around 0.8 to 0.3.

As we know that the financial crisis of 2007–2008 was the sharp decline in economic activity. The SOA substantially decreased because the average firms face a financial deficit with above-target debt in financial crisis.

Then, the SOA increase from around 0.3 to 0.8 between 2007 and 2012. From 2013 to 2016, the SOA decrease from around 0.8 to 0.5. As we know that the 2015–2016 stock market selloff was the period of decline in the value of stock prices globally that occurred between June 2015 to June 2016. The average firms' face a financial deficit with above-target debt. In our empirical result, we confirm that the SOA conditional on the required external capital changes as measured by a financial deficit or financial surplus (Byoun (2008)). Also, we find that there are some unobserved factors that influence the partial adjustment models in the US firms.

Figure 3.1 Mean of leverage ratio from 1960 to 2017

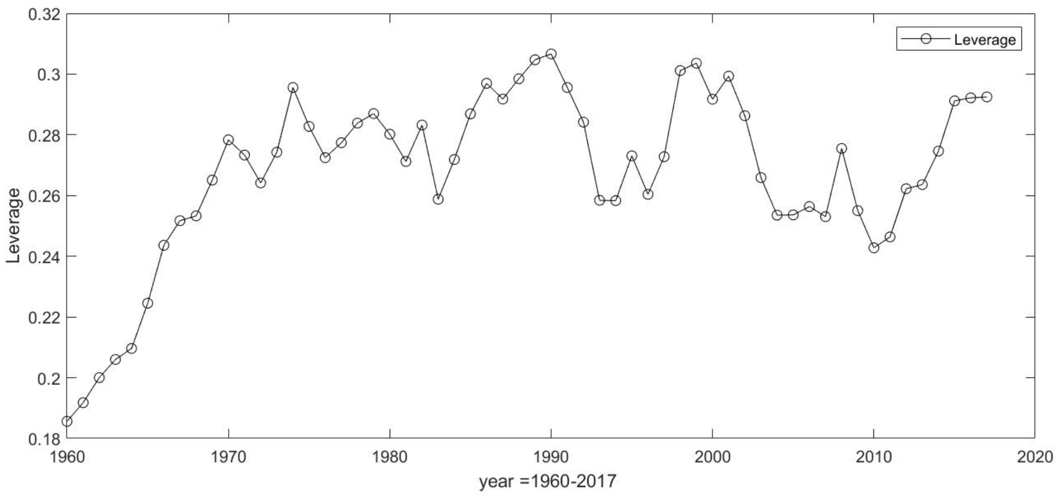
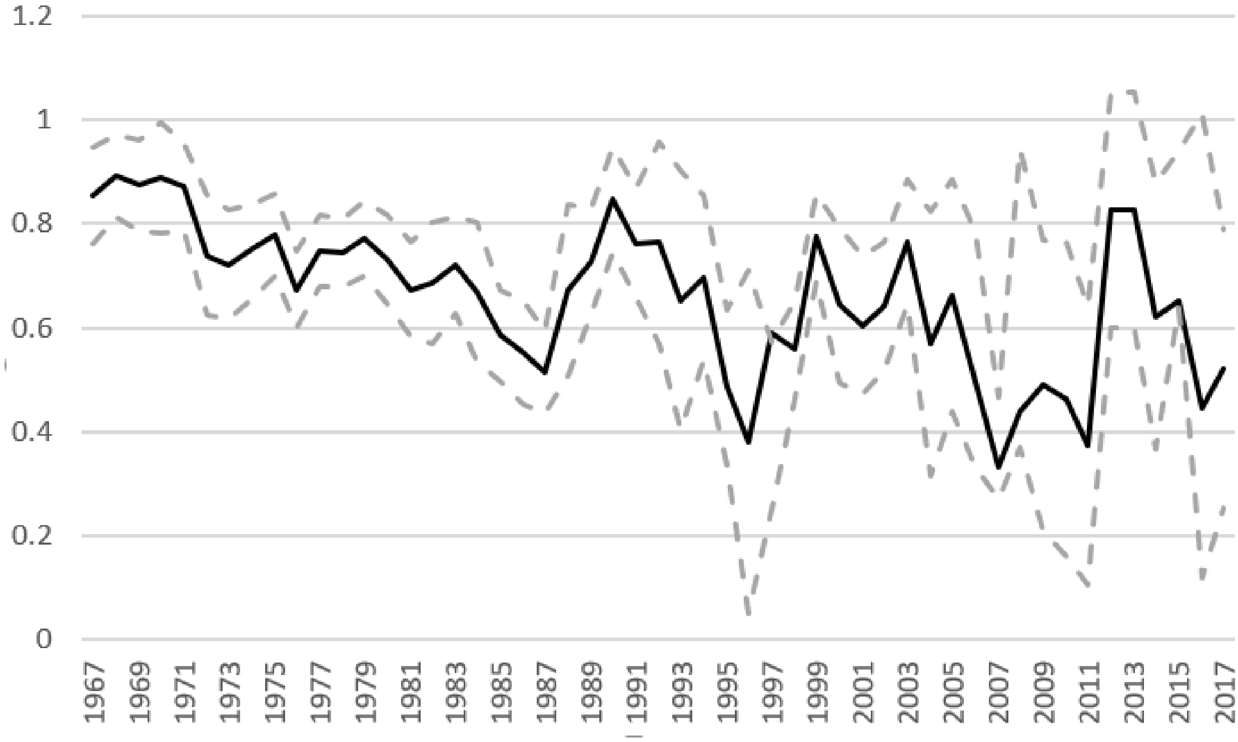


Figure 3.2 Rolling window estimate: the speed of adjustment of US firms



The data set is a panel of US firms collected from the CRSP/Compustat database over the period between 1960-2017. The approach which we take is to use rolling 8 year fixed window.

3.6 Conclusion

This chapter investigates the SOA on US firms from 1960 to 2017. This chapter also extends the short T quasi maximum likelihood estimator by Hayakawa et al. (2021) to the case where the errors are cross-sectionally heteroskedastic. As we consider the partial adjustment model with multi-factors error structure and cross-sectionally heteroskedasticity, we find that the quasi maximum likelihood estimator can be used to well estimate the SOA.

Using a large firm- year observations, we have some significant findings. First, we find that there are some unobserved factors that influence the partial adjustment models in the US firms. Second, we find evidence that average managers partially adjust leverage toward their target leverage fluctuated in the range of 0.3 to 0.8 from 1967 to 2017. The SOA conditional on the required external capital changes as measured by a financial deficit or financial surplus. Based on empirical results, we confirm that firms follow the trade-off theory. In this chapter we consider the partial adjustment model with multi-factor error structure and cross-sectionally heteroskedasticity. However, we do not consider asymmetric capital structure adjustments. It would be interesting to extend the model to the asymmetric partial adjustment model in future research.

It is also interesting to investigate SOA by heterogeneous dynamic partial adjustment capital structure models. More specifically, consider the following target leverage ratio as

$$l_{i,t}^{*} = \alpha^{*} + \alpha_i^{*} + \boldsymbol{\delta}_i^{'}\boldsymbol{x}_{i,t}, \tag{3.23}$$

where α^{*} is constant term and α_i^{*} is the unobserved unit- specific fixed effects. The leverage ratio $l_{i,t}$ adjusts to its target according to the rule

$$\Delta l_{i,t} = \phi_i \left(l_{i,t}^{*} - l_{i,t-1}\right) + u_{i,t},\ i = 1, \ldots N,\ t = 1, \ldots, T. \tag{3.24}$$

Substituting (3.23) into (3.24), we have

$$l_{i,t} = \alpha + \alpha_i + \rho_i l_{i,t-1} + \boldsymbol{\beta}_i^{'}\boldsymbol{x}_{i,t-1} + u_{i,t}, \tag{3.25}$$

with

$$u_{i,t} = \boldsymbol{\gamma}_i^{'}\boldsymbol{f}_t + e_{i,t}, \tag{3.26}$$

where $\alpha = \alpha^{*}\phi$, $\rho_i = 1 - \phi_i$, $\boldsymbol{\beta}_i = \phi_i\boldsymbol{\delta}_i$.

Chapter 4

Estimation in short panel vector autoregressions with error cross-sectional dependence

4.1 Introduction

In the last several decades there has been a tremendous interest in the study of panel data models with error cross-sectional dependence. Cross-sectional dependence is a typical feature of many empirical datasets. Ignoring it can lead to inconsistent estimates and misleading inferences. Developing tools to detect and account for cross-sectional dependence has therefore, not surprisingly, been at the forefront of econometric research.

The surge of popularity towards cross-sectional dependence in mainstream research has been for large N and large T panel data models. This has motivated researchers to acknowledge its value in big data environments. The study of the error cross-sectional dependence on short T panel data is also important because the use of such data can overcome aggregation problems (see Sarafidis and Wansbeek (2012)). Intuitively, the characteristics of social data are usually not independent but rather interrelated. Therefore, allowing for error cross-sectional dependence is more realistic in empirical studies. For extensive surveys of cross-sectional dependence (see Sarafidis and Wansbeek (2012) and Chudik and Pesaran (2013)).

The two most common ways of accounting for cross-sectional dependence within the literature is either through a spatial error structure or a multi-factor error structure. In the former case, the spatial dependence is captured through a weight matrix which typically characterises the location and/or distance (physical or economical) between units. In the latter case, dependence is characterised through a set of factors and their loadings. In this chapter, we focus on a dynamic multivariate panel data model, namely the panel vector autoregressive (VAR) model with multi-factor error structure.

Early theorization of the estimation and testing of short T panel VAR model with nonstationary individual effects can be traced back to Holtz-Eakin et al. (1988). Their panel VAR model is applied to investigate the dynamic relationship between wages and hours worked in American males. The results show that lagged hours is important in the hours equation.

Besides the methods of Holtz-Eakin et al. (1988), who use instrumental variables within the context of quasi-differenced autoregressive equations, Binder et al. (2005) first proposed the Quasi Maximum Likelihood (QML) estimation method for panel VAR models. In their study, they compare the random effect (RE) and fixed effect (FE) QML estimator and also show that the RE QML estimator is more efficient than the FE QML estimator. Binder et al. (2005) also compares the performance of the QML estimator and the Generalised Method of Moments (GMM) estimator. Their simulation results indicate that the performance of the QML estimator is better than that of the GMM estimator when individual effects have large variations. They also propose a unit root test within the context of the short T panel VAR model. When N tends to infinity, the unit root t statistic is asymptotically distributed as a standard normal variate. Their panel unit root test assumes slope homogeneity given the small T dimension. To allow for a panel unit root test with slope heterogeneity, the model would require large N and large T. They also provide a cointegration test to identify whether the variables cointegrated. Juodis (2018) extend the QML estimation method to panel VAR models with additional strictly exogenous regressors and possible cross-sectional heteroskedasticity. He shows that the likelihood based estimator outperforms the GMM based estimator in terms of bias and root mean square error. He further demonstrates that the multivariate transformed likelihood estimator is only consistent under a restricted parameter set and that the unrestricted QML estimator can not be globally identified for $T = 2$.

The above estimation methods have focused on panel VAR models without cross-sectional dependence. It should be noted, however, that few studies have considered panel VAR models with cross section dependence. For multivariable models with a spatial error term, Mutl (2009) provides a three steps estimation approach for short T panel VAR model with spatial dependence. For the panel VAR model with interactive effects, Huang (2008) provides an estimation method for non-stationary panel VAR model for large N and large T data set. He uses three steps to deal with the unobserved factor error structure. In the first step, Huang (2008) uses ordinary least squares (OLS) estimation ignoring the cross-sectional dependence. In the second step, he performs factor analysis on the residual. In the third step, he uses the factor augmented FM method to re-estimate.

Hayakawa et al. (2021) develops a quasi maximum likelihood estimator for short T dynamic panel data models with interactive effect. This chapter extends the transformed maximum likelihood approach for estimation of dynamic panel data model by Hsiao et al. (2002) to the case where the errors have a multi-factor structure. Hayakawa et al. (2021) also propose a sequential multiple testing likelihood ratio (MTLR) procedure for estimating the number of factors.

The purpose of this study is to extend the univariate short T dynamic panel data model under cross section dependence to the multivariate setting. In particular it considers a panel VAR model with cross-sectional dependence and adopts the quasi maximum likelihood (QML) approach to estimation. The results are expected to lead to a better understanding of cross-sectional dependence and the link across units in an unrestricted fashion.

The rest of this chapter is structured as follow. In Section 4.2, the panel VAR model and underlying assumptions are presented. The QML estimator is described in Section 4.3. Section 4.4 discusses identification of the parameters. In Section 4.5, we derive the asymptotic distribution of the QML estimator. Section 4.6 describes the Monte Carlo simulation setting and provides the corresponding results. Section 4.7 provides some concluding remarks.

Notations: Denote small positive constants by ϵ. $E_0(.)$ denotes expectations taken under the true probability measure. $\overset{p}{\to}$ denotes convergence in probability. Denote almost sure convergence by $\overset{a.s.}{\to}$. $\overset{d}{\to}$ denotes convergence in distribution for fixed T and as $N \to \infty$. $int(\mathbf{\Theta})$ denotes interior of the set $\mathbf{\Theta}$. $\mathbb{R}^{\kappa}$ denotes $\kappa-$dimensional Euclidean space.

4.2 Short T panel VAR model with interactive effects

First, we consider panel VAR model as,

$$\boldsymbol{z}_{i,t} = \mathbf{\Phi}\boldsymbol{z}_{i,t-1} + (\boldsymbol{I}_m - \mathbf{\Phi})\boldsymbol{a}_i + \boldsymbol{\xi}_{i,t}, \quad for\ t = 1, \dots, T;\ i = 1, \dots, N, \tag{4.1}$$

where $\mathbf{\Phi}$ is an $m \times m$ matrix of slope coefficients, $\boldsymbol{\xi}_{i,t}$ is defined as below and it is assumed that $\boldsymbol{z}_{i,0}$ is observable. We consider the panel VAR model for short T under the stationary assumption. The multi-factor structure of reduced form VAR can then be written as

$$\boldsymbol{\xi}_{i,t} = \mathbf{\Gamma}_i \boldsymbol{f}_t + \boldsymbol{\varepsilon}_{i,t}, \tag{4.2}$$

where $\mathbf{\Gamma}_i$ is an $m \times r$ matrix of factor loadings, $\boldsymbol{f}_t$ is an $r \times 1$ vector of unobserved common factor, and $\boldsymbol{\varepsilon}_{i,t}$ is an $m \times 1$ vector of idiosyncratic errors [1]. Let $\tilde{\boldsymbol{z}}_{i,t} = \boldsymbol{z}_{i,t} - \boldsymbol{a}_i$ and note that the model can also be written as

$$(\boldsymbol{I}_m - \mathbf{\Phi}\mathrm{L})\, \tilde{\boldsymbol{z}}_{i,t} = \boldsymbol{\xi}_{i,t}, \ \ for\ t = 2, 3, \dots, T,$$

with

$$\Delta \boldsymbol{z}_{i,1} = -\left(\boldsymbol{I}_m - \mathbf{\Phi}\right) \tilde{\boldsymbol{z}}_{i,0} + \boldsymbol{\xi}_{i,1}.$$

[1] Huang (2008) also proposed the estimation method for panel VAR model with multi-factor structure error. The main difference between Huang (2008) and this study is that our model includes individual and time effects and the estimation method is different. Also, our study focuses on short T panel VAR model but Huang (2008) considers panel VAR with large N and large T.

When T is fixed, it is necessary to consider the initialization of the $\boldsymbol{z}_{it}$ process for estimation and inference, which is reflected in the assumptions that will follow.[2] In the multi-factor error structure term, we treat the unobserved factor as fixed parameters and the associated factor loadings as random, which are independent of the idiosyncratic errors. Bai (2013) in Section 3.2 also provides an estimation method for a panel VAR model with a multi-factor error structure. The main difference between the model of Bai (2013) and our model is that we use a different method to overcome the incidental parameter problem. Moon and Weidner (2017) in their Example 1 provides an estimation method for the panel VAR model with a multi-factor error structure as well. However neither of these provides a thorough analysis of the multivariate structure.

We begin by stating the main assumptions which are similar to those of Hsiao et al. (2002), Binder et al. (2005) and Hayakawa et al. (2021)

Assumption 1 The idiosyncratic error $\boldsymbol{\varepsilon}_{i,t}$ are distributed independently for all i and t, with $E\left(\boldsymbol{\varepsilon}_{i,t}\right) = \mathbf{0}$, and variance $Var\left(\boldsymbol{\varepsilon}_{i,t}\right) = \boldsymbol{\Sigma}$. where $\boldsymbol{\Sigma}$ is a $m \times m$ symmetric positive define matrix, $\boldsymbol{\varepsilon}_{i,t}$ have finite fourth moment.

Assumption 2 The factor loadings $\boldsymbol{\Gamma}_i$ have finite fourth moments and are distributed independently of the idiosyncratic errors $\boldsymbol{\varepsilon}_{j,t}$ and common factors $\boldsymbol{f}_t$, for all i, j and t. The covariance matrix of $\boldsymbol{\Gamma}_i$ is finite and positive definite.

Assumption 3 The vector of individual effects, $\boldsymbol{a}_i$ can be correlated with $\boldsymbol{z}_{j,t}$, $\boldsymbol{\Gamma}_j$ and $\boldsymbol{\varepsilon}_{j,t}$ for all i, j, t.

Assumption 4 The factor loadings, $\boldsymbol{\Gamma}_i$, for $i = 1, \ldots, N$ is distributed independently of unobserved factors $\boldsymbol{f}_t$, for all i and t, and $\boldsymbol{a}_i$ are $i.i.d.$ with zero means and a finite $rm \times rm$ covariance matrix, as given bellow

$$Vec(\boldsymbol{\Gamma}_i) \overset{iid}{\sim} \left(\mathbf{0}, \boldsymbol{\Sigma}_\Gamma\right). \tag{4.3}$$

Assumption 5 The initial deviations, $\tilde{\boldsymbol{z}}_{i,0}$, are $i.i.d.$ across i, with zero means and constant non-singular variance, $E(\tilde{\boldsymbol{z}}_{i,0}\tilde{\boldsymbol{z}}_{i,0}^{'}) = \boldsymbol{\Psi}_{\tilde{\boldsymbol{z}}_0}$.

Under Assumption 5 and stationarity, the process $\boldsymbol{z}_{it}$ can either start from an infinite past or a finite past.

Due to the incidental parameter problem in fixed effects panel VAR models, we take the first difference of equation (4.1) and (4.2) to eliminate the individual effects. This transformation avoids the incidental parameters problem yielding the following

[2]Cheng and Zhou (2018) show that the QMLE is inconsistent for short T dynamic panel data model when treating initial value as fixed constants. When treated as a random variable, the QMLE is consistent for N tend to infinity.

expression

$$\Delta \boldsymbol{z}_{i,t} = \boldsymbol{\Phi} \Delta \boldsymbol{z}_{i,t-1} + \boldsymbol{\Gamma}_i \boldsymbol{g}_t + \Delta \boldsymbol{\varepsilon}_{i,t}, \ \ for \ t = 2, \ldots, T; \ i = 1, \ldots, N, \tag{4.4}$$

where $\boldsymbol{g}_t = \Delta \boldsymbol{f}_t$.

To obtain a consistent QML estimator one needs to work with the unconditional joint probability distribution of $(\Delta \boldsymbol{z}_{i,1}, \Delta \boldsymbol{z}_{i,2}, ..., \Delta \boldsymbol{z}_{i,T})$, or the distribution of $(\Delta \boldsymbol{z}_{i,2}, \Delta \boldsymbol{z}_{i,3}, ..., \Delta \boldsymbol{z}_{i,T})$ conditional on $\Delta \boldsymbol{z}_{i,1}$, and ensure that these distributions are free of the incidental parameters problem. This will clearly be the case if the unconditional distribution of $\Delta \boldsymbol{z}_{i,1}$ does not depend on any incidental parameters. Thus, we further consider the following assumption:

Assumption 6 The following moment restrictions are satisfied:

$$E(\boldsymbol{\kappa}_{i,0} \boldsymbol{\varepsilon}'_{i,1}) = \mathbf{0},$$
$$and$$
$$E(\boldsymbol{\kappa}_{i,0} \Delta \boldsymbol{\varepsilon}'_{i,t}) = \mathbf{0} \ \ for \ t = 2, 3, \ldots, T,$$

where $\boldsymbol{\kappa}_{i,0} = (\boldsymbol{I}_m - \boldsymbol{\Phi}) \tilde{\boldsymbol{z}}_{i,0}$.

Combining Assumption 6 with Assumptions 1-5 and using definition of $\Delta \boldsymbol{z}_{i,1}$ we now have

$$\Delta \boldsymbol{z}_{i,1} \overset{iid}{\sim} (\mathbf{0}, \boldsymbol{\Psi}), \tag{4.5}$$

where $\boldsymbol{\Psi} = (\boldsymbol{I}_m - \boldsymbol{\Phi}) \boldsymbol{\Psi}_{\tilde{z}_0} (\boldsymbol{I}_m - \boldsymbol{\Phi}') + \boldsymbol{\Sigma}$, $\text{Cov}(\Delta \boldsymbol{z}_{i,1}, \Delta \boldsymbol{\varepsilon}_{i,2}) = -\boldsymbol{\Sigma}$ and $Cov\,(\Delta \boldsymbol{z}_{i,1}, \Delta \boldsymbol{\varepsilon}_{i,t}) = \mathbf{0}$, for $t = 3, 4, \ldots, T$, $i = 1, \ldots, N$.

Let $\Delta \boldsymbol{Z}_i = \left(\Delta \boldsymbol{z}'_{i,1}, \Delta \boldsymbol{z}'_{i,2}, \ldots, \Delta \boldsymbol{z}'_{i,T}\right)'$ and $\Delta \boldsymbol{Z}_{i,-1} = \left(\mathbf{0}'_m, \Delta \boldsymbol{z}'_{i,1}, \Delta \boldsymbol{z}'_{i,2}, \ldots, \Delta \boldsymbol{z}'_{i,T-1}\right)'$ which are $Tm \times 1$ matrices, $\boldsymbol{G} = (\boldsymbol{f}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_T)'$ a $T \times r$ matrix, $\boldsymbol{e}_i = \left(\boldsymbol{\varepsilon}'_{i,1}, \Delta \boldsymbol{\varepsilon}'_{i,2}, \ldots, \Delta \boldsymbol{\varepsilon}'_{i,T}\right)'$ a $Tm \times 1$ vector, and $\boldsymbol{\chi}_i = \boldsymbol{G}^* vec\,(\boldsymbol{\Gamma}_i) + \boldsymbol{e}_i$ with $\boldsymbol{G}^* = (\boldsymbol{G} \otimes \boldsymbol{I}_m)$ which is a $Tm \times rm$ matrix. It follows that

$$\boldsymbol{\chi}_i = \boldsymbol{R}\,(\boldsymbol{\Phi}) \Delta \boldsymbol{Z}_i, \tag{4.6}$$

where $\boldsymbol{R}(\boldsymbol{\Phi})$ is given by the $Tm \times Tm$ matrix

$$\boldsymbol{R}(\boldsymbol{\Phi}) = \begin{bmatrix} \boldsymbol{I}_m & \ldots & \ldots & \mathbf{0} \\ -\boldsymbol{\Phi} & \boldsymbol{I}_m & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots & -\boldsymbol{\Phi} & \boldsymbol{I}_m \end{bmatrix}. \tag{4.7}$$

.

4.3 Quasi maximum likelihood estimation

Multiplying $\boldsymbol{R}^{-1}(\boldsymbol{\Phi})$ on both sides of equation (4.6), we have

$$\Delta \boldsymbol{Z}_i = \boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i. \tag{4.8}$$

We define the variance of $\boldsymbol{\Delta Z_i}$ as

$$\boldsymbol{\Sigma}_{R\chi} = Var(\Delta \boldsymbol{Z}_i) = E\left(\left(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i\right)\left(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i\right)^{'}\right) = \boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})\boldsymbol{R}^{'-1}(\boldsymbol{\Phi}), \tag{4.9}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi}) &= \boldsymbol{G}^{*}\boldsymbol{\Sigma}_{\Gamma}\boldsymbol{G}^{*'} + \boldsymbol{\Sigma}_{\boldsymbol{e}},\\ &= \boldsymbol{Q}\boldsymbol{Q}^{'} + \boldsymbol{\Sigma}_{e},\end{aligned} \tag{4.10}$$

$$\boldsymbol{\Sigma}_e = E\left(\boldsymbol{e}_i\boldsymbol{e}_i^{'}\right) = \begin{bmatrix} \boldsymbol{\Psi} & -\boldsymbol{\Sigma} & & 0\\ -\boldsymbol{\Sigma} & 2\boldsymbol{\Sigma} & & 0\\ \vdots & & \ddots & \vdots\\ 0 & & -\boldsymbol{\Sigma} & 2\boldsymbol{\Sigma}\end{bmatrix} \tag{4.11}$$

and $\boldsymbol{Q} = \boldsymbol{G}^{*}\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}^{\frac{1}{2}}$ is a $Tm \times rm$ matrix. We further define $\boldsymbol{\varphi} = \left(Vec(\boldsymbol{\Phi})^{'}\right)^{'}$ which is a $m^2 \times 1$ vector and $\boldsymbol{\phi} = \left(Vech(\boldsymbol{\Psi})^{'}, Vech(\boldsymbol{\Sigma})^{'}, Vec(\boldsymbol{Q})^{'}\right)^{'}$ which is a $(Trm^2 + m(m+1)) \times 1$ vector. Our main interest is in the parameters $\boldsymbol{\varphi}$, and therefore we treat the interactive effects as nuisance parameters. Because $\boldsymbol{Q}\boldsymbol{Q}^{'}$ is of reduced rank, $rank(\boldsymbol{Q}\boldsymbol{Q}^{'}) = rm < Tm$, it is not possible to identify $\boldsymbol{Q}$ without additional restrictions. This is because for any orthonormal $rm \times rm$ matrix $\boldsymbol{\Lambda}$, $\boldsymbol{Q}\boldsymbol{Q}^{'} = \boldsymbol{Q}^{*}\boldsymbol{Q}^{*'}$ where $\boldsymbol{Q}^{*} = \boldsymbol{Q}\boldsymbol{\Lambda}$. To avoid such non-trivial identification $\frac{rm(rm-1)}{2}$ restrictions need to be imposed on $\boldsymbol{Q}$ in line also with the usual $(rm)^2$ restrictions typically imposed on $Var(\boldsymbol{G}^{*}vec(\boldsymbol{\Gamma}_i))$ in factor analysis. The number of non-redundant parameters of $\boldsymbol{Q}$ is then given by $Trm^2 - \frac{rm(rm-1)}{2}$ (see also Hayashi et al. (2007)).

The quasi-log-likelihood function of the transformed model (4.8) can be expressed as

$$\ell_N(\boldsymbol{\theta}) = \ell_N(\boldsymbol{\varphi}, \boldsymbol{\phi}) \propto -\frac{N}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{R\chi}}| - \frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i\right)^{'}\boldsymbol{\Sigma}_{R\chi}^{-1}\left(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i\right), \tag{4.12}$$

with the unknown parameters $\boldsymbol{\theta} = \left(\boldsymbol{\varphi}^{'}, \boldsymbol{\phi}^{'}\right)^{'}$ collected in the $\left((Trm^2 - \frac{rm(rm-1)}{2}) + \frac{m(m+1)}{2} + m^2\right) \times$ 1 vector.

4.3.1 An eigenvalue approach for the computation of the QML estimator

Here we consider an eigenvalue approach that simplifies the computation of the QML estimator. Consider the log-likelihood given in (4.12) without any restrictions on $\boldsymbol{Q}$ which can be further written as follows.

The determinant of $\boldsymbol{\Sigma}_{R\chi}$ is given by

$$\begin{aligned} |\boldsymbol{\Sigma}_{R\chi}| &= |\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\,(\boldsymbol{\Sigma}_{\chi})\,\boldsymbol{R}^{'-1}(\boldsymbol{\Phi})| \\ &= |\boldsymbol{R}^{-1}(\boldsymbol{\Phi})(\boldsymbol{G}^{*}\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}\boldsymbol{G}^{*'} + \boldsymbol{\Sigma}_{\boldsymbol{e}})\boldsymbol{R}^{'-1}(\boldsymbol{\Phi})| \\ &= |\boldsymbol{R}^{-1}(\boldsymbol{\Phi})||\boldsymbol{G}^{*}\Sigma_{\Gamma}\boldsymbol{G}^{*'} + \boldsymbol{\Sigma}_{e}||\boldsymbol{R}^{'-1}(\boldsymbol{\Phi})| \\ &= |\boldsymbol{R}(\boldsymbol{\Phi})|^{-1}|\boldsymbol{G}^{*}\Sigma_{\Gamma}\boldsymbol{G}^{*'} + \boldsymbol{\Sigma}_{e}||\boldsymbol{R}^{'}(\boldsymbol{\Phi})|^{-1} \\ &= |\boldsymbol{G}^{*}\Sigma_{\Gamma}\boldsymbol{G}^{*'} + \boldsymbol{\Sigma}_{e}| \\ &= |\boldsymbol{Q}\boldsymbol{Q}^{'} + \boldsymbol{\Sigma}_{e}|, \end{aligned} \tag{4.13}$$

where $|\boldsymbol{R}^{-1}(\boldsymbol{\Phi})| = |\boldsymbol{R}(\boldsymbol{\Phi})|^{-1} = 1$.

The quasi-log-likelihood function in (4.12) can then be written as

$$\begin{aligned} \ell_N(\boldsymbol{\theta}) &\propto -\frac{N}{2}\log|\boldsymbol{\Sigma}_e + \boldsymbol{Q}\boldsymbol{Q}^{'}| - \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i)^{'}(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\Sigma}_{\chi}\boldsymbol{R}^{'-1}(\boldsymbol{\Phi}))^{-1}(\boldsymbol{R}^{-1}(\boldsymbol{\Phi})\boldsymbol{\chi}_i) \\ &= -\frac{N}{2}\log|\boldsymbol{\Sigma}_e + \boldsymbol{Q}\boldsymbol{Q}^{'}| - \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\chi}_i^{'}\,(\boldsymbol{\Sigma}_{\chi})^{-1}\,\boldsymbol{\chi}_i. \end{aligned} \tag{4.14}$$

Recall $\boldsymbol{Q}\boldsymbol{Q}^{'}$ is rank deficient, $rank(\boldsymbol{Q}\boldsymbol{Q}^{'}) = rm < Tm$, and $\boldsymbol{\Omega}$ is positive definite. We decompose

$$|\boldsymbol{\Sigma}_e + \boldsymbol{Q}\boldsymbol{Q}^{'}| = |\boldsymbol{\Sigma}_e||\boldsymbol{I}_{rm} + \boldsymbol{Q}^{'}\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}.| \tag{4.15}$$

Thus, equation (4.15) can be written as

$$|\boldsymbol{\Sigma}_e + \boldsymbol{Q}\boldsymbol{Q}^{'}| = |\boldsymbol{\Sigma}_e||\boldsymbol{A}|, \tag{4.16}$$

where $\boldsymbol{A} = \boldsymbol{I}_{rm} + \boldsymbol{Q}^{'}\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}$ is a non-singular matrix. By the Woodbury matrix identity, we have

$$\begin{aligned} (\boldsymbol{\Sigma}_{\chi})^{-1} &= \left(\boldsymbol{\Sigma}_e + \boldsymbol{Q}\boldsymbol{Q}^{'}\right)^{-1} \\ &= \boldsymbol{\Sigma}_e^{-1} - \boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\left(\boldsymbol{I}_{rm} + \boldsymbol{Q}^{'}\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^{'}\boldsymbol{\Sigma}_e^{-1} \\ &= \boldsymbol{\Sigma}_e^{-1} - \boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\boldsymbol{A}^{-1}\boldsymbol{Q}^{'}\boldsymbol{\Sigma}_e^{-1} \end{aligned} \tag{4.17}$$

Using the above results, the quasi-log-likelihood function can then be written as

$$\ell_N(\boldsymbol{\theta}) \propto -\frac{N}{2}\log|\boldsymbol{\Sigma}_e| - \frac{N}{2}log|\boldsymbol{A}| - \frac{N}{2}\left[Tr\left(\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-1}\right) - Tr\left(\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\boldsymbol{A}^{-1}\boldsymbol{Q}'\boldsymbol{\Sigma}_e^{-1}\right)\right], \tag{4.18}$$

where $\boldsymbol{C}_N(\boldsymbol{\varphi}) = \frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\chi}_i\boldsymbol{\chi}_i'\right)$.

For analytical convenience, we next define $\boldsymbol{P} = \boldsymbol{\Sigma}_e^{-\frac{1}{2}}\boldsymbol{Q}\boldsymbol{A}^{-\frac{1}{2}}$ where $rank\left(\boldsymbol{P}\right) = rm$. We then have

$$\boldsymbol{I}_{rm} - \boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_{rm} - \boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{Q}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\boldsymbol{A}^{-\frac{1}{2}}, \tag{4.19}$$

From equation (4.16), we know

$$\boldsymbol{Q}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q} = \boldsymbol{A} - \boldsymbol{I}_{rm}. \tag{4.20}$$

Thus, we can rewrite

$$\begin{aligned}\boldsymbol{I}_{rm} - \boldsymbol{P}'\boldsymbol{P} &= \boldsymbol{I}_{rm} - \boldsymbol{A}^{-\frac{1}{2}}\left(\boldsymbol{A} - \boldsymbol{I}_{rm}\right)\boldsymbol{A}^{-\frac{1}{2}} \\ &= \boldsymbol{I}_{rm} - \boldsymbol{I}_{rm} + \boldsymbol{A}^{-1}.\end{aligned} \tag{4.21}$$

Therefore, from the above equations we have

$$\boldsymbol{A}^{-1} = \boldsymbol{I}_{rm} - \boldsymbol{P}'\boldsymbol{P}. \tag{4.22}$$

Also from the quasi-log-likelihood function (4.18), we have

$$Tr\left(\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-1}\right) = Tr\left(\boldsymbol{\Sigma}_e^{-\frac{1}{2}}\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-\frac{1}{2}}\right) = Tr\left(\boldsymbol{D}_N(\boldsymbol{\theta})\right), \tag{4.23}$$

where we denote

$$\boldsymbol{D}_N\left(\boldsymbol{\theta}\right) = \boldsymbol{\Sigma}_e^{-\frac{1}{2}}\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-\frac{1}{2}}. \tag{4.24}$$

Hence, we have

$$Tr\left(\boldsymbol{C}_N(\boldsymbol{\varphi})\boldsymbol{\Sigma}_e^{-1}\boldsymbol{Q}\boldsymbol{A}^{-1}\boldsymbol{Q}'\boldsymbol{\Sigma}_e^{-1}\right) = Tr\left(\boldsymbol{P}'\boldsymbol{D}_N(\boldsymbol{\theta})\boldsymbol{P}\right). \tag{4.25}$$

From the above results, the quasi-log-likelihood function can then be written as

$$\ell_N(\boldsymbol{\theta}) \propto -\frac{N}{2}\log|\boldsymbol{\Sigma}_e| + \frac{N}{2}log|\boldsymbol{I}_{rm} - \boldsymbol{P}'\boldsymbol{P}| - \frac{N}{2}\left[Tr\left(\boldsymbol{D}_N(\boldsymbol{\theta})\right) - Tr\left(\boldsymbol{P}'\boldsymbol{D}_N(\boldsymbol{\theta})\boldsymbol{P}\right)\right]. \tag{4.26}$$

In line with the discussion in Section 4.3, $\boldsymbol{P}$ is not identified without additional restrictions. It is easily seen that the value of $\ell_N(\boldsymbol{\theta})$ is invariant to the orthonomal transformation of $\boldsymbol{P}$, identification of $\boldsymbol{P}$ is not possible without restrictions (Bai

(2009) and Bai and Ng (2002)). For example, $\tilde{\boldsymbol{P}} = \boldsymbol{P}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is an arbitrary $rm \times rm$ invertible matrix and $\boldsymbol{\Lambda}^{'}\boldsymbol{\Lambda} = \boldsymbol{I}_{rm}$. Using $\tilde{\boldsymbol{P}}$, the likelihood function remains unchanged. Denoting $\boldsymbol{P} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{rm})$, where $\boldsymbol{p}_t$ is the t^{th} column of $\boldsymbol{P}$ (a $Tm \times 1$ vector of unknown parameters) we impose the following $\frac{rm\,(rm-1)}{2}$ orthogonality conditions

$$\boldsymbol{p}_t^{'}\boldsymbol{p}_s = 0, \;\; for\ all\ s \neq t = 1, 2, \ldots, rm. \tag{4.27}$$

Using these restrictions, the quasi likelihood function can be expressed as

$$\ell_N(\boldsymbol{\theta}) \propto -\frac{N}{2} log|\boldsymbol{\Sigma}_e| + \frac{N}{2}\sum_{t=1}^{rm} log\left(1 - \boldsymbol{p}_t^{'}\boldsymbol{p}_t\right) + \frac{N}{2}\sum_{t=1}^{rm}\boldsymbol{p}_t^{'}\boldsymbol{D}_N(\boldsymbol{\theta})\boldsymbol{p}_t - \frac{N}{2}Tr\left(\boldsymbol{D}_N(\boldsymbol{\theta})\right). \tag{4.28}$$

Taking the first derivative with respect to $\boldsymbol{p}_t$ and setting this to zero, yields

$$\boldsymbol{D}_N(\boldsymbol{\theta})\hat{\boldsymbol{p}}_t = \left(\frac{1}{1 - \hat{\boldsymbol{p}}_t^{'}\hat{\boldsymbol{p}}_t}\right)\hat{\boldsymbol{p}}_t, for\ t = 1, 2, \ldots, rm, \tag{4.29}$$

where $\hat{\boldsymbol{p}}_t$ is quasi-maximum likelihood estimator of $\boldsymbol{p}_t$. Then, $\hat{\boldsymbol{p}}_t$ is the eigenvector of $\boldsymbol{D}_N(\boldsymbol{\theta})$ and associated with the first rm largest eigenvalue of $\boldsymbol{D}_N(\boldsymbol{\theta})$. We denote the eigenvalues by $\lambda_t(\boldsymbol{\theta})$ and $\lambda_t(\boldsymbol{\theta})$ is

$$\lambda_t(\boldsymbol{\theta}) = \frac{1}{1 - \hat{\boldsymbol{p}}_t^{'}\hat{\boldsymbol{p}}_t}. \tag{4.30}$$

Thus, the concentrated quasi-log-likelihood function can be expressed as

$$\ell_N(\boldsymbol{\theta}) \propto -\frac{N}{2} log|\boldsymbol{\Sigma}_e| - \frac{N}{2}\sum_{t=1}^{rm} log\left(\lambda_t(\boldsymbol{\theta})\right) + \frac{N}{2}\sum_{t=1}^{rm}\left(\lambda_t(\boldsymbol{\theta}) - 1\right) - \frac{N}{2}\sum_{t=1}^{Tm}\left(\lambda_t(\boldsymbol{\theta})\right). \tag{4.31}$$

In maximising the likelihood above, we consider a number of random initial values.

4.4 Identification

In this section, we establish the order condition on m and T for identification of the number of interactive effects.

4.4.1 Order Condition

In deriving the order condition on m and T, from the first difference model (4.8) we can see that $\boldsymbol{\theta}$ can only be identified from the distinct elements of $Var(\Delta \boldsymbol{z}_{i,t}) = \boldsymbol{\Sigma}_{R_\chi}$. Since $\boldsymbol{Q}$ enters $\boldsymbol{\Sigma}_{R_\chi}$ as $\boldsymbol{A}^* = \boldsymbol{Q}\boldsymbol{Q}^{'}$, we need to consider the unknown elements of $\boldsymbol{A}^*$

under different rank conditions. To identify $\boldsymbol{\theta}$, we need the $rank\left(\boldsymbol{A}^{*}\right)=rank\left(\boldsymbol{Q}\right)=rm<Tm$. Recall also from Section 4.3 that the number of non-redundant elements of $\boldsymbol{Q}$ is given by $Trm^2-\frac{rm(rm-1)}{2}$.

Then the order condition for identification of $\boldsymbol{\theta}$ is given by

$$\frac{Tm(Tm+1)}{2}\geq Trm^2-\frac{rm(rm-1)}{2}+m^2+m(m+1). \tag{4.32}$$

This order condition is satisfied if $T>3$ and the number of endogenous variables $m\geq 1$, for $r=0,1,\ldots,T-2$. The largest number of factors that satisfies the above condition is $T-2$.

4.4.2 Global Identification

Consider the average quasi log- likelihood function defined by (4.14) expressed as

$$\begin{aligned}\bar{\ell}_N\left(\boldsymbol{\theta}\right)=N^{-1}\ell_N(\boldsymbol{\varphi},\boldsymbol{\phi})=&\\ -\frac{T}{2}log(2\pi)-\frac{1}{2}log|\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})|-\frac{1}{2N}\sum_{i=1}^{N}\boldsymbol{\chi}_i^{'}(\boldsymbol{\varphi})\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})^{-1}\boldsymbol{\chi}_i(\boldsymbol{\varphi}).&\end{aligned} \tag{4.33}$$

Assumption 7 (i) $\boldsymbol{\theta}\in\boldsymbol{\Theta}=\boldsymbol{\Theta}_{\varphi}\times\boldsymbol{\Theta}_{\phi}$, where $\boldsymbol{\Theta}_{\varphi}$ is a compact subset of $\mathbb{R}^{n_{\varphi}}$ with $n_{\varphi}=m^2$ and $\boldsymbol{\Theta}_{\phi}=\boldsymbol{\Theta}_{\Psi}\times\boldsymbol{\Theta}_{\Sigma}\times\boldsymbol{\Theta}_{q}$ where $q=vec(\boldsymbol{Q})$, Θ_q is a compact subset of $\mathbb{R}^{n_q}$, $\boldsymbol{\Theta}_{\Psi}$ and $\boldsymbol{\Theta}_{\Sigma}$ are compact subsets of $\mathbb{R}^{n_{\Psi}}$ and $\mathbb{R}^{n_{\Sigma}}$ with $n_q=Trm^2-rm(rm-1)/2$, $n_{\Psi}=m(m+1)/2$ and $n_{\Sigma}=m(m+1)/2$; $\boldsymbol{\theta}_0=(\boldsymbol{\varphi}_0^{'},\boldsymbol{\phi}_0^{'})^{'}=(vec(\boldsymbol{\Phi}_0)^{'},\boldsymbol{\phi}_0^{'})^{'}$ lies in the interior of $\boldsymbol{\Theta}$, (ii) for some $c_{max}>c_{min}>0$, $c_{min}\leq inf_{\phi\in\Theta_{\phi}}\lambda_{min}\left(\boldsymbol{\Sigma}_{\chi}\left(\boldsymbol{\phi}\right)\right)\leq c_{max}$, (iii) $\boldsymbol{A}\left(\boldsymbol{\phi}\right)=\lim_{N\to\infty}N^{-1}\Sigma_{i=1}^{N}E_0\left(\Delta\boldsymbol{Z}_{i,-1}^{'}\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})^{-1}\Delta\boldsymbol{Z}_{i,-1}\right)$ is positive definite almost surely uniformly on $\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\phi}$, where the expectation is taken with respect to the true probability measure.

The global identification condition requires $\lim_{N\to\infty}E_0\left[\bar{\ell}_N(\boldsymbol{\varphi},\boldsymbol{\phi})\right]$ to attain a unique maximum at $\boldsymbol{\theta}_0=(\boldsymbol{\varphi}_0,\boldsymbol{\phi}_0)\in\boldsymbol{\Theta}$. Under Assumptions 1-7 and using results in Hayakawa et al. (2021) that readily extend to the multivariate case, we have that

$$\bar{\ell}_N(\boldsymbol{\varphi}_0,\boldsymbol{\phi}_0)-\bar{\ell}_N(\boldsymbol{\varphi},\boldsymbol{\phi})\overset{a.s.}{\to}\lim_{N\to\infty}E_0\left[\bar{\ell}_N(\boldsymbol{\varphi}_0,\boldsymbol{\phi}_0)-\bar{\ell}_N(\boldsymbol{\varphi},\boldsymbol{\phi})\right]\geq 0, \tag{4.34}$$

where

$$\begin{aligned}2\lim_{N\to\infty}E_0\left[\bar{\ell}_N(\boldsymbol{\varphi}_0,\boldsymbol{\phi}_0)-\bar{\ell}_N(\boldsymbol{\varphi},\boldsymbol{\phi})\right]=&\kappa(\boldsymbol{\phi},\boldsymbol{\phi}_0)+(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)^{'}\boldsymbol{A}(\boldsymbol{\phi})\left(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0\right)+\\ &2\left(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0\right)\varrho_0(\boldsymbol{\phi},\boldsymbol{\phi}_0),\end{aligned} \tag{4.35}$$

with

$$\kappa\left(\boldsymbol{\phi},\boldsymbol{\phi}_0\right)=Tr\left[\boldsymbol{\Sigma}_{\chi}^{-1}(\boldsymbol{\phi})\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi}_0)\right]-\log\left(|\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi}_0)|/|\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})|\right)-T\geq 0, \tag{4.36}$$

and

$$\varrho(\boldsymbol{\phi},\boldsymbol{\phi}_0)=Tr\left\{\left[\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi})-\boldsymbol{\Sigma}_{\chi}(\boldsymbol{\phi}_0)\right]\boldsymbol{\Sigma}_{\chi}^{-1}(\boldsymbol{\phi})\boldsymbol{L}\boldsymbol{B}(\boldsymbol{\Phi}_0)^{-1}\right\}\geq 0. \tag{4.37}$$

Under Assumption 7, $\boldsymbol{A}(\boldsymbol{\phi})$ is a positive definite matrix and we have

$$(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)^{'}\boldsymbol{A}(\boldsymbol{\phi})(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0) \geq \lambda_{min}\left[\boldsymbol{A}(\boldsymbol{\phi})\right](\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)^{'}(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0) > 0, \tag{4.38}$$

with $(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)^{'}\boldsymbol{A}(\boldsymbol{\phi})(\boldsymbol{\varphi}-\boldsymbol{\varphi}_0)=0$ if and only if $\boldsymbol{\varphi}=\boldsymbol{\varphi}_0$. However, as explained in Hayakawa et al. (2021), $w(\boldsymbol{\theta},\boldsymbol{\theta}_0)=\kappa(\boldsymbol{\phi},\boldsymbol{\phi}_0)+2(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\varrho_0(\boldsymbol{\phi},\boldsymbol{\phi}_0)$ is not ensured to be non-negative, and hence global identification of the parameters of interest cannot be guaranteed.

4.4.3 Local Identification

As global identification of the parameters of interest on the parameter space $\boldsymbol{\Theta}$ cannot be guaranteed, we proceed by considering a restriction of $\boldsymbol{\Theta}$ on which identification and consistency can be shown.

Consider the following definition:

Definition 1 Let $N_\epsilon(\boldsymbol{\theta}_0)$ be a set in the closed neighbourhood of $\boldsymbol{\theta}_0$ defined by $N_\epsilon(\boldsymbol{\theta}_0)=\boldsymbol{\theta}\in\Theta_\varphi\times\Theta_\phi:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\epsilon$ for some $\epsilon>0$, such that $w(\boldsymbol{\theta},\boldsymbol{\theta}_0)=\kappa(\boldsymbol{\phi},\boldsymbol{\phi}_0)+2(\boldsymbol{\Phi}-\boldsymbol{\Phi}_0)\varrho_0(\boldsymbol{\phi},\boldsymbol{\phi}_0)\geq 0$ for all values of $\boldsymbol{\varphi}\in\boldsymbol{\Theta}_\varphi$ and $\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi$ where Θ_φ is a compact subset of $\mathbb{R}^{n_\varphi}$ with $n_\varphi=m^2$ and $\Theta_\phi=\Theta_\phi\times\Theta_\Sigma\times\Theta_q$ where $q=vec(\boldsymbol{Q})$, Θ_q is a compact subset of $\mathbb{R}^{n_q}$; Θ_ϕ and Θ_Σ are compact subsets of $\mathbb{R}^{n_\phi}$ and $\mathbb{R}^{n_\Sigma}$ with $n_q=Trm^2-rm(rm-1)/2$, $n_\phi=m(m+1)/2$ and $n_\Sigma=m(m+1)/2$.

Given the local nature of the analysis, henceforth we consider the more restricted parameter space as set out in the following assumption:

Assumption 8 $\boldsymbol{\theta}\in\Theta_\epsilon=N_\epsilon(\boldsymbol{\theta}_0)$, where $N_\epsilon(\boldsymbol{\theta}_0)$ is given in Definition 1; Θ_ϵ is a compact subset of $\mathbb{R}^{n_\theta}$ with $n_\theta=(Trm^2-\frac{rm(rm-1)}{2})+\frac{m(m+1)}{2}+m^2$; $\boldsymbol{\theta}_0=(\boldsymbol{\varphi}_0^{'},\boldsymbol{\phi}_0^{'})^{'}=(vec(\boldsymbol{\Phi}_0)^{'},\boldsymbol{\phi}_0^{'})^{'}$ lies in the interior of Θ_ϵ.

We make the following conjecture.

Conjecture 1 The vector of true parameters $\boldsymbol{\theta}_0=(\boldsymbol{\varphi}_0^{'},\boldsymbol{\phi}_0^{'})^{'}$ is identified on Θ_ϵ.

Remark Conjecture 1 is a natural extension to the multivariate case of the identification analysis of Hayakawa et al. (2021) within the univariate context.

4.5 Asymptotic properties of the estimator

Consider the average log-likelihood function given by equation (4.33). To show the consistency and asymptotic normality of the QML estimator the following conditions need to be met:

(1) $\mathbf{\Theta}_\epsilon$ is a compact subset of $\mathbf{\Theta}$.
(2) $\bar{C}_N(\boldsymbol{\theta}) \overset{a.s.}{\to} C_N(\boldsymbol{\theta})(a\ non-stochastic\ function\ of\ \boldsymbol{\theta})$ uniformly on $\mathbf{\Theta}_\epsilon$, where $\bar{C}_N(\boldsymbol{\theta}) = -2\bar{\ell}_N(\boldsymbol{\theta})$ and $\bar{C}(\boldsymbol{\theta}) = E_0\left(\bar{C}_N(\boldsymbol{\theta})\right)$.
(3)$\boldsymbol{\theta}_0 \in int(\mathbf{\Theta})$ is the unique minimum of $\bar{C}(\boldsymbol{\theta})$.

If the above three conditions are satisfied then it follows that $\hat{\boldsymbol{\theta}}_N \overset{a.s.}{\to} \boldsymbol{\theta}_0$ on $\mathbf{\Theta}_\epsilon$ (See Hayakawa et al. (2021)).

Taking a Taylor expansion of $\frac{\partial \bar{\ell}_N(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ at $\boldsymbol{\theta}_0$, we can derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}$. Then, we need to check the behaviour of the score function $\bar{\boldsymbol{s}}_N(\boldsymbol{\theta}) = \frac{\partial \bar{\ell}_N(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$, and Hessian matrix, $\boldsymbol{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2 \bar{\ell}_N(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$. If $E_0\left[\frac{\bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right] = 0$, the Hessian matrix, $H_N\left(\check{\boldsymbol{\theta}}\right) \overset{a.s.}{\to} H_N(\boldsymbol{\theta}_0)$ and following the mean value theorem, the asymptotic normality of the QML estimator:

$$\sqrt{N}\bar{\boldsymbol{s}}_N\left(\hat{\boldsymbol{\theta}}\right) = \sqrt{N}\bar{\boldsymbol{s}}_N(\boldsymbol{\theta}_0) - \boldsymbol{H}_N(\check{\boldsymbol{\theta}})\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{0}, \tag{4.39}$$

where $\check{\boldsymbol{\theta}}$ lie between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

We summarise the above discussion and the resultant asymptotic distribution in the following theorem:

Theorem 1 Suppose that Assumptions 1- 7(ii),(iii) and 8, as well as the order condition (4.32) and Conjecture 1, hold. Denote the QML estimator of $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}} = arg\max_{\boldsymbol{\theta} \in \mathbf{\Theta}_\epsilon} \bar{\ell}_N(\boldsymbol{\theta})$, where $\bar{\ell}_N(\boldsymbol{\theta})$ is given by (4.33). Then, the QML estimator $\hat{\boldsymbol{\theta}}$ is almost surely locally consistent for $\boldsymbol{\theta}_0$ on $\mathbf{\Theta}_\epsilon$ for values of the VAR coefficient $\mathbf{\Phi}$ sufficiently close to $\mathbf{\Phi}_0$ as formalised by **Definition** 1, and

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \overset{d}{\to} N\left(\mathbf{0}, \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\right), \tag{4.40}$$

where $\boldsymbol{B}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} E_0(N\frac{\partial \bar{\ell}_N(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\frac{\partial \bar{\ell}_N(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'})$ and $\boldsymbol{H}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} E_0(\frac{\partial^2 \bar{\ell}_N(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'})$.

4.6 Monte Carlo simulation

In this section, we provide simulation evidence to show that the proposed estimator for the short T panel VAR with individual and interactive time effects have good finite sample properties.

We generate $\boldsymbol{z}_{i,t}$ as

$$\boldsymbol{z}_{i,t} = \mathbf{\Phi}\boldsymbol{z}_{i,t-1} + \boldsymbol{a}_i + \boldsymbol{\xi}_{i,t},\ t = -S+1, \dots, 0, 1, \dots, T;\ i = 1, 2, \dots, N, \tag{4.41}$$

and

$$\boldsymbol{\xi}_{i,t} = \mathbf{\Gamma}_i'\boldsymbol{f}_t + \boldsymbol{\varepsilon}_{i,t}. \tag{4.42}$$

We assume the number of factors, r, is known and equal to the true number, r_0, which we set equal to two, namely, $r = r_0 = 2$. We start the process with

$$\boldsymbol{z}_{i,-s+1} = \boldsymbol{a}_i + \boldsymbol{\xi}_{i,t},\ \ \boldsymbol{\xi}_{i,t} \overset{iid}{\sim} \left(\mathbf{0}_m, \sum_{j=0}^{s-1} \boldsymbol{\Phi}_0^j \Sigma_\chi (\boldsymbol{\Phi}_0^j)'\right), \tag{4.43}$$

and set $s = 60$. By discarding the first sixty observations, the impact of the initial value can be reduced. The elements in $\boldsymbol{\Phi}$ are denoted by ϕ_{ll}.

For the generating process of the factor loadings, we consider

$$\Gamma_{\tau\ell i} \sim N\left(0, \sigma^2_{\Gamma\tau\ell}\right),\ \ for\ all\ \tau = 1, \ldots r;\ \ \ell = 1, \ldots, m.$$

The factors $\boldsymbol{f}_t$ are generated as

$$f_{\tau t} = \rho_{f\tau} f_{\tau,t-1} + \left(1 - \rho^2_{f\ell\tau}\right)^{\frac{1}{2}} u_{f\tau},\ \ u_{f\tau} \sim N\left(0, \sigma^2_f\right)\ \ for\ \tau = 1, \ldots, r, \tag{4.44}$$

where $\rho_{f\ell\tau} = 0.5$ and $\sigma^2_f = \{1, 5\}$. The individual effects, $\boldsymbol{\alpha}_i$ are generated as

$$\alpha_{\ell i} = a_0 \bar{\boldsymbol{\varepsilon}}_i + a_1 \boldsymbol{\eta}_i,\ \ \ell = 1, \ldots, m, \tag{4.45}$$

where $\bar{\boldsymbol{\varepsilon}}_i = T^{-1}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{i,t}$, $\boldsymbol{\varepsilon}_{i,t} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\eta}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$, with $\boldsymbol{\Sigma}_\eta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. a_0 and a_1 are constants that are used to control the correlation between the individual effects and the idiosyncratic error. Setting $a_0 = 1$ and $a_1 = 1$ allows for the individual effects to be correlated with the idiosyncratic error.

We set the number of endogenous variables to $m = 2$ and consider the following values for the number of individuals N and time units T, $N = (100, 300, 500)$ and $T = (5, 6, 7)$, respectively. All simulations were carried out using 1000 replications.

For the matrix of the autoregressive parameters we consider the following values:

- Case 1. Stationary panel VAR with maximum eigenvalue of Φ, $\lambda_{max}(\boldsymbol{\Phi}) = 0.3$ $\boldsymbol{\Phi} = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$.
- Case 2. Stationary panel VAR with maximum eigenvalue of Φ, $\lambda_{max}(\boldsymbol{\Phi}) = 0.6$ $\boldsymbol{\Phi} = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}$.
- Case 3. Stationary panel VAR with maximum eigenvalue of Φ, $\lambda_{max}(\boldsymbol{\Phi}) = 0.8$ $\boldsymbol{\Phi} = \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$.

4.6.1 Monte Carlo results

We report results for the bias, MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) of the QML estimator for our panel VAR model. We focus on the results for the parameters of interest namely, $\mathbf{\Phi}_{11}$ and $\mathbf{\Phi}_{21}$ as the results for $\mathbf{\Phi}_{12}$ and $\mathbf{\Phi}_{22}$ are very similar. Throughout, the considered sample sizes are $N = \{100, 300, 500\}$ and $T = \{5, 6, 7\}$, the number of factors is 2, and for the variance of the factors we consider the values $\sigma_f^2 = \{1, 5\}$.

Tables 4.1-4.3 report the bias, MAE, and RMSE for cases 1-3 of the autoregressive parameters and $\sigma_f^2 = 1$. In these simulation results we can observe that the bias and RMSE decrease as N increases. Also, when the maximum eigenvalue of $\mathbf{\Phi}$ increases, the bias slightly increases across most of the cases. Furthermore, the bias of $\mathbf{\Phi}_{21}$ is smaller than the bias of $\mathbf{\Phi}_{11}$. For the MAE of the QML estimator, we see that the MAE of $\mathbf{\Phi}_{11}$ is between 0.14 and 0.18 when $N = 100$ and $T = 5$. From Tables 4.1- 4.3 we also see that the MAE of $\mathbf{\Phi}$ decreases as N increases, and when T increases, the MAE of $\mathbf{\Phi}$ decreases. As the maximum eigenvalue of $\mathbf{\Phi}$ increases, the MAE of $\mathbf{\Phi}$ slightly increases. Similarly, we find that the MAE of $\mathbf{\Phi}_{21}$ is smaller than the MAE of $\mathbf{\Phi}_{11}$. With regard to the RMSE of $\mathbf{\Phi}$, this decreases as N increases. Also, the RMSE of $\mathbf{\Phi}_{21}$ is smaller than the RMSE of $\mathbf{\Phi}_{11}$.

In Tables 4.4-4.6 we report the bias, MAE, RMSE for cases 1-3 with $\sigma_f^2 = 5$. In this case the variance of the factors σ_f^2 increases from 1 to 5. These tables show that as σ_f^2 increases, the bias of $\mathbf{\Phi}$ slightly decreases. Also, we see that the bias of $\mathbf{\Phi}_{21}$ is smaller than that of $\mathbf{\Phi}_{11}$ across most cases. Similarly, the MAE of $\mathbf{\Phi}$ decreases when the variance of the factors σ_f^2 increases. The MAE of $\mathbf{\Phi}$ decreases as N increases from 100 to 500. The RMSE of $\mathbf{\Phi}$ also decreases when variance of the factors σ_f^2 increases. The RMSE of $\mathbf{\Phi}$ decreases when N increases.

Based on the above simulation findings, we find that (1) the variance of σ_f^2 influences the performance of the QML estimator of the short T panel VAR model with interactive effects. (2) The performance of the QML estimator is influenced by the maximum eigenvalue of $\mathbf{\Phi}$.

Table 4.1 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 1$, and $\lambda_{max}(\mathbf{\Phi}) = 0.3$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.0234	0.1414	0.2137	-0.0166	0.1140	0.1809	-0.0180	0.0981	0.1686
300	-0.0213	0.0899	0.1462	-0.0198	0.0665	0.1163	-0.0241	0.0626	0.1370
500	-0.0210	0.0642	0.1090	-0.0203	0.0577	0.1213	-0.0158	0.0471	0.1090
Φ_{21}									
100	0.0063	0.1193	0.1832	0.0196	0.0955	0.1541	0.0027	0.0858	0.1454
300	0.0027	0.0690	0.1110	-0.0013	0.0591	0.1106	0.0035	0.0495	0.0979
500	-0.0026	0.0529	0.0892	0.0025	0.0414	0.0751	-0.0010	0.0399	0.0839

$\mathbf{\Phi}_{j,l}$, for $j = 1, \dots m$, $l = 1, \dots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

Table 4.2 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 1$, and $\lambda_{max}(\mathbf{\Phi}) = 0.6$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.0586	0.1694	0.2451	-0.0511	0.1466	0.2229	-0.0351	0.1127	0.1893
300	-0.0401	0.1173	0.1810	-0.0275	0.0834	0.1346	-0.0146	0.0647	0.1123
500	-0.0349	0.0954	0.1568	-0.0298	0.0740	0.1262	-0.0199	0.0514	0.1048
Φ_{21}									
100	-0.0025	0.1439	0.2098	0.0122	0.1156	0.1724	-0.0009	0.0979	0.1541
300	0.0004	0.0943	0.1422	0.0021	0.0695	0.1127	0.0003	0.0555	0.1044
500	-0.0024	0.0770	0.1152	-0.0009	0.0561	0.0927	0.0022	0.0412	0.0776

$\mathbf{\Phi}_{j,l}$, for $j = 1, \dots m$, $l = 1, \dots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

Table 4.3 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 1$, and $\lambda_{max}(\mathbf{\Phi}) = 0.8$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.1135	0.1878	0.2726	-0.0981	0.1677	0.2589	-0.0720	0.1331	0.2025
300	-0.0840	0.1555	0.2158	-0.0491	0.1113	0.1706	-0.0361	0.0924	0.1507
500	-0.0668	0.1303	0.2014	-0.0413	0.0997	0.1670	-0.0344	0.0786	0.1458
Φ_{21}									
100	-0.0470	0.1689	0.2405	-0.0244	0.1313	0.1863	-0.0217	0.1126	0.1682
300	-0.0332	0.1191	0.1709	-0.0093	0.0913	0.1367	-0.0027	0.0755	0.1239
500	-0.0215	0.1067	0.1643	-0.0094	0.0772	0.1255	-0.0059	0.0638	0.1137

$\mathbf{\Phi}_{j,l}$, for $j = 1, \ldots m$, $l = 1, \ldots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

Table 4.4 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 5$, and $\lambda_{max}(\mathbf{\Phi}) = 0.3$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.0206	0.1090	0.1851	-0.0142	0.0805	0.1312	-0.0082	0.0646	0.1160
300	-0.0106	0.0662	0.1180	-0.0092	0.0482	0.0920	-0.0087	0.0429	0.0893
500	-0.0094	0.0492	0.0910	-0.0076	0.0399	0.0766	-0.0072	0.0272	0.0572
Φ_{21}									
100	0.0028	0.0856	0.1460	0.0025	0.0657	0.1094	-0.0007	0.0552	0.1046
300	-0.0012	0.0519	0.0943	0.0016	0.0414	0.0819	-0.0017	0.0355	0.0812
500	-0.0035	0.0416	0.0796	-0.0024	0.0336	0.0705	-0.0026	0.0237	0.0480

$\mathbf{\Phi}_{j,l}$, for $j = 1, \ldots m$, $l = 1, \ldots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

Table 4.5 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 5$, and $\lambda_{max}(\mathbf{\Phi}) = 0.6$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.0348	0.1462	0.2388	-0.0211	0.1094	0.1752	-0.0180	0.0823	0.1508
300	-0.0052	0.0922	0.1452	-0.0111	0.0646	0.1246	-0.0101	0.0480	0.0879
500	-0.0105	0.0745	0.1211	-0.0066	0.0477	0.0781	-0.0110	0.0372	0.0775
Φ_{21}									
100	0.0054	0.1160	0.1809	0.0046	0.0915	0.1469	0.0034	0.0700	0.1207
300	0.0116	0.0745	0.1128	0.0084	0.0543	0.0974	0.0030	0.0382	0.0680
500	0.0056	0.0628	0.1013	0.0057	0.0388	0.0623	0.0018	0.0318	0.0632

$\mathbf{\Phi}_{j,l}$, for $j = 1, \dots m$, $l = 1, \dots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

Table 4.6 Bias, MAE and RMSE of Φ_{11}, Φ_{21},$\sigma_f^2 = 5$, and $\lambda_{max}(\mathbf{\Phi}) = 0.8$

N	T=5			T=6			T=7		
(r, r_0)	(2,2)			(2,2)			(2,2)		
	Bias	MAE	RMSE	Bias	MAE	RMSE	Bias	MAE	RMSE
Φ_{11}									
100	-0.0790	0.1598	0.2490	-0.0683	0.1414	0.2273	-0.0552	0.1132	0.1995
300	-0.0474	0.1213	0.1913	-0.0248	0.0836	0.1309	-0.0282	0.0766	0.1358
500	-0.0384	0.1000	0.1593	-0.0227	0.0753	0.1284	-0.0194	0.0583	0.0992
Φ_{21}									
100	-0.0400	0.1329	0.2039	-0.0218	0.1169	0.1914	-0.0206	0.0936	0.1515
300	-0.0103	0.0950	0.1502	0.0090	0.0726	0.1146	0.0022	0.0658	0.1158
500	-0.0089	0.0859	0.1381	0.0053	0.0641	0.1132	0.0092	0.0483	0.0811

$\mathbf{\Phi}_{j,l}$, for $j = 1, \dots m$, $l = 1, \dots$ is the coefficient in row j and column l of $\mathbf{\Phi}$.
r_0 is true value of factor, we assume the number of factor is 2.

4.7 Concluding remarks

In this study, we extend the work by Hayakawa et al. (2021) to the multivariate case by considering the panel VAR model for short T with individual and interactive effects. The Monte Carlo results show that the proposed QML estimator performs reasonably well. But, in line with the theoretical results in Hayakawa et al. (2021), we conjecture that global identification of the QML estimator is not possible. Further research into the identification conditions for the QML estimator in the multivariate case would be of future interest.

In this study, by extending the univariate model to the multivariate context we have not assumed that the coefficients are sparse. It would be of interest in future research to consider quasi maximum likelihood estimation of large scale panel VAR models with sparse coefficients in the presence of cross-sectional dependence. Furthermore, it would also be of interest to extend the panel VAR model to panel error correction setting and investigate nonstationarity under cross-sectional dependence.

Chapter 5

Constructing Optimal Instruments on Dynamic Heterogeneous Panels with Defactored Regressors and Multifactor Error Sructure

5.1 Introduction

In recent years, there has been a dramatic proliferation of research concerned with studies of the panel data model with large cross-section and time-series dimensions, N and T, respectively. The richer the content of the data, the more complicated and general the panel data models can become. In this paper, we consider dynamic panel data model with cross-sectionally heterogeneous slopes and a multi-factor error structure.

By extending Pesaran (2006), Chudik and Pesaran (2015) propose mean group CCE (CCEMG) estimation of panel autoregressive distributed lag models. Chudik and Pesaran (2015) employ a mean group type estimator to deal with the slope heterogeneity, and propose to augment the regression with the cross-sectional averages of dependent variables and covariates and their lags, in order to control the interactive effects.

Recently, Norkutė et al. (2021) has proposed a novel instrumental variable (IV) estimator for the dynamic panel data models. Their approach initially projects out the common factors from the exogenous covariates of the model, and constructs instruments based on defactored covariates in order to build a consistent first step IV estimator. They found that the estimator performs satisfactorily. This estimator has some advantages over the CCEMG estimator of Chudik and Pesaran (2015). Firstly, the IV estimator employs the principal component estimator for defactoring the exogenous covariates, therefore it do not need to seek external variables to approximate the factors when the number of unobserved factors is larger than the number of covariates plus one. By contrast, in this situation the CCE estimation requires additional sets of variables, which are not in the original model of interest but expected to from a part of the dynamic system. Secondly, the CCE estimator is subject to the small T bias of least squares estimators, whilst the IV estimator is not. Chudik and Pesaran (2015) propose to adjust the bias using the jackknife method, which might not be very effective for small or moderate T.

One important issue of the IV approach by Norkutė et al. (2021) is that they are silent about the optimal choice of the instruments. For this approach, the number of valid instruments for each cross-section unit can increase proportionally to T^2. Therefore, how many of which instruments to choose is an important issue.

To date, a number of studies have proposed different types of methods for instruments selection. Donald and Newey (2001) developed the mean square error criteria to choose among valid instruments. By choosing instruments to minimize approximate mean square error (MSE), Donald and Newey (2001) show that the finite sample properties of 2SLS estimators can be improved. Hansen (2007b) takes a different approach by focusing on selecting the weights for averaging across least squares estimators. This method is particularly useful in reducing estimation variance. Other instrument selection methods are shrinkage and parameter penalization. Knight and Fu (2000) propose the asymptotic property of the Lasso type estimator and show that the limiting distributions can have positive probability mass at 0 when the true value of the parameter is 0. Fan and Li (2001) and Fan and Peng (2004) propose penalized likelihood estimators. Hansen (2007a) propose the least square model averaging approach based on the restricted weights. Kuersteiner and Okui (2010) propose the model averaging method that weigh predicted value of endogenous variables in the estimation stage for the two stage least squares (2SLS) estimator. By choosing the weight to minimize the approximate MSE, the optimal IV estimator can be obtained [1]. Chen et al. (2016) provide an weighting scheme based on weighting individual estimators. This method can also be applied in a heterogeneous dynamic panel data model. Windmeijer et al. (2019) consider the case that some of the available instruments can be invalid, e.g. some instruments have a direct effect on the outcome, and some of instruments are associated with unobserved confounders.

In this chapter, we apply the model averaging method of Kuersteiner and Okui (2010) to the IV estimator of Norkutė et al. (2021) for the dynamic heterogeneous panel data model with regressors and a multifacctor error structure. The first step in this process is to compute several 2SLS estimators with different numbers (and sets) of instruments. Next, we estimate the weights for averaging the obtained 2SLS estimators (Kuersteiner and Okui (2010)).

This chapter is organized as follows: Section 2 sets out the model, assumptions, estimation method and its asymptotic property. Section 3 introduces the model average IV estimators. Section 4 studies the Mean Group estimator. The Monte Carlo experiments and its results are reported in Section 5. Section 6 concludes.

[1] Kuersteiner and Okui (2010) call this estimator the model average two stage least square estimator.

Proofs of theorems, lemmas and propositions are contained in the Appendix.

The following notation is used in the remainder of this chapter. $\boldsymbol{P}_i = \boldsymbol{X}_i \left(\boldsymbol{X}_i^{'}\boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i^{'}$ is the projection onto the column space of $\boldsymbol{X}_i$, where $\boldsymbol{X}_i$ is a full column rank matrix for each individual i. A k vector of ones is denoted as $\boldsymbol{\iota}_k$. The l_p norm is denoted by $\|.\|_p$ and Δ is finite positive constant. $(N,T) \xrightarrow{j} \infty$ denote that N and T tend to infinity jointly.

5.2 Model, asymptotic property of IV estimator

5.2.1 The models and Assumptions

Consider the following dynamic heterogeneous panel data model with a multifactor error structure:

$$y_{i,t} = \phi_i y_{i,t-1} + \boldsymbol{x}_{i,t}^{'}\boldsymbol{\beta}_i + u_{i,t}; \quad i = 1, \ldots N;\ t = 1, \ldots, T\,, \tag{5.1}$$

with

$$u_{i,t} = \boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t} + \varepsilon_{i,t}, \tag{5.2}$$

where $\boldsymbol{x}_{i,t}$ is an $k \times 1$ vector of explanatory variables, ϕ_i is a scalar cross-sectionally heterogeneous coefficient for $y_{i,t-1}$ with $sup_{1\leq i\leq N}|\phi_i| < 1$ and $\boldsymbol{\beta}_i$ is an $k \times 1$ vector of cross-sectionally heterogeneous coefficients. $\boldsymbol{f}_{y,t} = \left(f_{y,1t}, f_{y,2t}, \ldots, f_{y,m_y t}\right)^{'}$ is an $m_y \times 1$ vector of unobservable factors, $\boldsymbol{\gamma}_{y,i}$ is an $m_y \times 1$ vector of associated factor loadings, and $\varepsilon_{i,t}$ is the idiosyncratic error term.

Regressors, $\boldsymbol{x}_{i,t}$ is the following process:

$$\boldsymbol{x}_{i,t} = \boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t} + \boldsymbol{v}_{i,t}, \tag{5.3}$$

where $\boldsymbol{f}_{x,t} = (f_{x,1t}, f_{x,2t}, \ldots, f_{x,m_x t})^{'}$ is an $m_x \times 1$ vector of unobservable factors and $\boldsymbol{\Gamma}_{x,i} = \left(\boldsymbol{\gamma}_{1,i}, \ldots, \boldsymbol{\gamma}_{k,i}\right)$ is the $m_x \times k$ associated factor loadings matrix, and $\boldsymbol{v}_{i,t} = (v_{1,it}, v_{2,it}, \ldots, v_{k,it})^{'}$ is the idiosyncratic error term which is independent of $\varepsilon_{i,t}$.

Follow above description, the model can be expressed as

$$y_{i,t} = \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta}_i + u_{i,t}; \qquad i = 1, \ldots N;\ t = 1, \ldots, T\,, \tag{5.4}$$

where $\boldsymbol{w}_{i,t} = \left(y_{i,t-1}, \boldsymbol{x}_{i,t}^{'}\right)^{'}$ and $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i$ with $\boldsymbol{\theta}_i = \left(\phi_i, \boldsymbol{\beta}_i^{'}\right)^{'}$, $\boldsymbol{\lambda}_i \overset{i.i.d.}{\sim} (\boldsymbol{0},\, \boldsymbol{\Sigma}_\lambda)$, $\boldsymbol{\Sigma}_\lambda$ is a fixed positive definite matrix and $\boldsymbol{\theta} = E\left(\boldsymbol{\theta}_i\right)$.

Stacking the T observations for each i, we have

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\theta}_i + \boldsymbol{u}_i \quad ; \quad i = 1, \ldots N, \tag{5.5}$$

where $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\boldsymbol{W}_i = \left(\boldsymbol{y}_{i,-1}, \boldsymbol{X}_i\right) = (\boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,T})'$, $\boldsymbol{y}_{i,-1} = (y_{i,0}, \ldots, y_{i,T-1})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,T})'$ and

$$\boldsymbol{u}_i = \boldsymbol{F}_y \boldsymbol{\gamma}_{y,i} + \boldsymbol{\varepsilon}_i \tag{5.6}$$

with $\boldsymbol{F}_y = \left(\boldsymbol{f}_{y,1}, \boldsymbol{f}_{y,2}, \ldots, \boldsymbol{f}_{y,T}\right)'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \ldots, \varepsilon_{i,T})'$.

Similarly,

$$\boldsymbol{X}_i = \boldsymbol{F}_x \boldsymbol{\Gamma}_{x,i} + \boldsymbol{V}_i, \tag{5.7}$$

where $\boldsymbol{F}_x = \left(\boldsymbol{f}_{x,1}, \boldsymbol{f}_{x,2}, \ldots, \boldsymbol{f}_{x,T}\right)'$ and $\boldsymbol{V}_i = (\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,T})'$. For the observed lags of $\boldsymbol{X}_i$ we have

$$\begin{aligned} \boldsymbol{X}_{i,-1} &= \boldsymbol{F}_{x,-1} \boldsymbol{\Gamma}_{x,i} + \boldsymbol{V}_{i,-1} \\ &\vdots \\ \boldsymbol{X}_{i,-J} &= \boldsymbol{F}_{x,-J} \boldsymbol{\Gamma}_{x,i} + \boldsymbol{V}_{i,-J}, \end{aligned} \tag{5.8}$$

where J is the maximum number of lags that we can observe.

In the model, the factors $\boldsymbol{F}_x, \boldsymbol{F}_{x,-1}, \ldots, \boldsymbol{F}_{x,-J}$ are unobserved. Therefore, we can apply principal components approach to estimate $\boldsymbol{F}_x, \boldsymbol{F}_{x,-1}, \ldots, \boldsymbol{F}_{x,-J}$ by Bai (2003) and Bai (2009). In addition, the number of factors, m_x and m_y can be estimated by Bai and Ng (2002). In this chapter, we focus on constructing the optimal instruments, so we treat the number of factors as given. Next, we consider the following projection matrices:

$$\begin{aligned} \boldsymbol{M}_{F_x} &= \boldsymbol{I}_T - \boldsymbol{F}_x \left(\boldsymbol{F}_x' \boldsymbol{F}_x\right)^{-1} \boldsymbol{F}_x'; \\ \boldsymbol{M}_{F_{x,-1}} &= \boldsymbol{I}_T - \boldsymbol{F}_{x,-1} \left(\boldsymbol{F}_{x,-1}' \boldsymbol{F}_{x,-1}\right)^{-1} \boldsymbol{F}_{x,-1}'; \\ &\vdots \\ \boldsymbol{M}_{F_{x,-J}} &= \boldsymbol{I}_T - \boldsymbol{F}_{x,-J} \left(\boldsymbol{F}_{x,-J}' \boldsymbol{F}_{x,-J}\right)^{-1} \boldsymbol{F}_{x,-J}'. \end{aligned} \tag{5.9}$$

Assume $\boldsymbol{V}_i$ is independent of $\boldsymbol{\varepsilon}_i$, $\boldsymbol{F}_x$, $\boldsymbol{F}_y$ and $\boldsymbol{\gamma}_{y,i}$. Premultiplying $\boldsymbol{X}_i$ by $\boldsymbol{M}_{F_x}$, we can show $\boldsymbol{M}_{F_x} \boldsymbol{X}_i = \boldsymbol{M}_{F_x} \boldsymbol{V}_i$. Similarly, premultiplying $\boldsymbol{X}_{i,-1}$ by $\boldsymbol{M}_{F_{x,-1}}$, we can show $\boldsymbol{M}_{F_{x,-1}} \boldsymbol{X}_{i,-1} = \boldsymbol{M}_{F_{x,-1}} \boldsymbol{V}_{i,-1}$. Now, it is easily seen that

$$E\left(\boldsymbol{X}_i' \boldsymbol{M}_{F_x} \boldsymbol{u}_i\right) = E\left(\boldsymbol{X}_{i,-1}' \boldsymbol{M}_{F_{x,-1}} \boldsymbol{u}_i\right) = \cdots = E\left(\boldsymbol{X}_{i,-J}' \boldsymbol{M}_{F_{x,-J}} \boldsymbol{u}_i\right) = 0 \tag{5.10}$$

Therefore, premultiplying $\boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{X}_{i,-J}$ by $\boldsymbol{M}_{F_{x,-1}}, \ldots, \boldsymbol{M}_{F_{x,-J}}$, respectively, we get the set of IVs:

$$\boldsymbol{Z}_i = \left(\boldsymbol{M}_{F_x} \boldsymbol{X}_i, \boldsymbol{M}_{F_{x,-1}} \boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{M}_{F_{x,-j}} \boldsymbol{X}_{i,-j}, \ldots, \boldsymbol{M}_{F_{x,-J}} \boldsymbol{X}_{i,-J}\right), \tag{5.11}$$

where $\boldsymbol{Z}_i$ is $T \times (J+1)k$ matrix and J is the maximum number of lags of $\boldsymbol{X}_i$.

5.2.2 IV estimation method and asymptotic property

In this section, we introduce IV estimator for dynamic heterogeneous panel data models by Norkutė et al. (2021). Before we introduce estimation, we consider the following assumptions:

Assumption 1 *:* $\varepsilon_{i,t}$ *is independently distributed across* i *and* t*, with* $E(\varepsilon_{i,t}) = 0$, $E\left(\varepsilon_{i,t}^2\right) = \sigma_{\varepsilon,it}^2$*, and* $E|\varepsilon_{i,t}|^{8+c} \leq \triangle < \infty$ *for a small positive constant* c*.*

Assumption 2 *: (i)* $x_{\ell i,t}$ *and* $\varepsilon_{i,t}$ *are independently distributed for all* ℓ*,* t *and* i *and* $\ell = 1, \ldots, k$*; (ii)* $E(v_{\ell i,t}) = 0$ *and* $E|v_{\ell i,t}|^{8+c} \leq \triangle < \infty$*; (iii)* $T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E|v_{\ell i,t}v_{\ell i,s}|^{1+c} \leq \triangle < \infty$*; (iv)* $E|N^{-1/2}\sum_{i=1}^{N}\left(v_{\ell i,t}v_{\ell i,s} - E\left(v_{\ell i,t}v_{\ell i,s}\right)\right)|^4 \leq \triangle < \infty$ *for all* ℓ*,* t *and* s*; (v) the largest eigenvalue of* $E(\boldsymbol{v}_{\ell i}\boldsymbol{v}_{\ell i}^{'})$ *is bounded uniformly for every* ℓ*,* i *and* T*; (vi)* $N^{-1}T^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{T}\sum_{r=1}^{T}|cov(v_{\ell is}v_{\ell it}, v_{\ell iq}v_{\ell ir})| \leq \Delta < \infty$.

Assumption 3 *:* $\boldsymbol{f}_{y,t} = \boldsymbol{\psi}_y(L)\boldsymbol{e}_{f_y,t}$ *and* $\boldsymbol{f}_{x,t} = \boldsymbol{\psi}_x(L)\boldsymbol{e}_{f_x,t}$*, where* $\boldsymbol{\psi}_y(L)$ *and* $\boldsymbol{\psi}_x(L)$ *are absolutely summable,* $\boldsymbol{e}_{f_y,t} \overset{i.i.d.}{\sim} \left(\boldsymbol{0}, \boldsymbol{\Sigma}_{f_y}\right)$ *and* $\boldsymbol{e}_{f_x,t} \overset{i.i.d.}{\sim} (\boldsymbol{0}, \boldsymbol{\Sigma}_{f_x})$*, where* $\boldsymbol{\Sigma}_{f_y}$ *and* $\boldsymbol{\Sigma}_{f_x}$ *are positive define matrices. All elements of* $\boldsymbol{e}_{f_y,t}$ *and* $\boldsymbol{e}_{f_x,t}$ *has finite fourth order moments and are group wise independent from* $\boldsymbol{v}_{i,t}$ *and* $\varepsilon_{i,t}$*.*

Assumption 4 *:* $\boldsymbol{\gamma}_{y,i} \overset{i.i.d.}{\sim} \left(\boldsymbol{0}, \boldsymbol{\Sigma}_{\gamma_y}\right)$ *and* $\boldsymbol{\Gamma}_{x,i} \overset{i.i.d.}{\sim} (\boldsymbol{0}, \boldsymbol{\Sigma}_{\Gamma_x})$*, where* $\boldsymbol{\Sigma}_{\gamma_y}$ *and* $\boldsymbol{\Sigma}_{\Gamma_x}$ *are positive definite matrices. All elements of* $\boldsymbol{\gamma}_{y,i}$ *and* $\boldsymbol{\Gamma}_{x,i}$ *has finite fourth order moments and independent from* $\boldsymbol{v}_{i,t}$*,* $\varepsilon_{i,t}$*,* $\boldsymbol{e}_{f_x,t}$ *and* $\boldsymbol{e}_{f_y,t}$*.*

Assumption 5 *: (i)* $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{\lambda}_i$*,* $\boldsymbol{\lambda}_i \overset{i.i.d.}{\sim} (\boldsymbol{0}, \boldsymbol{\Sigma}_\lambda)$*, where* $\boldsymbol{\Sigma}_\lambda$ *is a fixed positive definite matrix; (ii)* $\boldsymbol{\lambda}_i$ *is independent from* $\boldsymbol{\Gamma}_{x,i}, \boldsymbol{\gamma}_{y,i}, \boldsymbol{v}_{i,t}, \varepsilon_{i,t}, \boldsymbol{e}_{f_y,t}$ *and* $\boldsymbol{e}_{f_x,t}$*; (iii)* $Prob(|\lambda_{r,i}| > z) \leq 2exp\left(-\frac{z^2}{2(a+bz)}\right)$*, for all* z*,* i *and fixed* $a, b > 0$*, where* $\lambda_{r,i}$ *is the* r*-th element of* $\boldsymbol{\lambda}_i$ *for* $2 \leq r \leq 1 + k$*.*

Assumption 6 *: (i)* $\tilde{\boldsymbol{A}}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\boldsymbol{W}_i, \tilde{\boldsymbol{B}}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\boldsymbol{Z}_i$ *have full column rank for all* i *for large* T*; (ii)* $E\left\|\tilde{\boldsymbol{A}}_{i,T}\right\|^{2+2c} \leq \Delta < \infty, E\left\|\tilde{\boldsymbol{B}}_{i,T}\right\|^{2+2c} \leq \Delta < \infty$ *for all* i *for large* T*.*

Assumption 7 *: (i)* $\boldsymbol{A}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{W}_i$ *and* $\boldsymbol{B}_{i,T} = \frac{1}{T}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{Z}_i$ *have full column rank for all* i *for large* T*; (ii)* $E\|\boldsymbol{A}_{i,T}\|^{2+2c} \leq \triangle < \infty$*,* $E\|\boldsymbol{B}_{i,T}\|^{2+2c} \leq \triangle < \infty$ *for all* i *for large* T*.*

Assumption 8 *: (i)* $E\|\boldsymbol{\lambda}_i\|^4 \leq \Delta$*; (ii)* $E\left\|T^{-1/2}\boldsymbol{V}_i^{'}\boldsymbol{F}_x\right\|^4 \leq \Delta$*;*
(iii) $E\left\|N^{-1/2}T^{-1/2}\sum_{\ell=1}^{k}\sum_{l=1}^{N}\left(\boldsymbol{V}_i^{'}\boldsymbol{v}_{\ell l} - E\left(\left(\boldsymbol{V}_i^{'}\boldsymbol{v}_{\ell l}\right)\boldsymbol{\gamma}_{\ell l}^{'}\right)\right)\right\|^4 \leq \Delta$*;*
(iv) $E\left(T^{-1/2}\sum_{\ell=1}^{k}\sum_{t=1}^{T}\left(v_{\ell i,t}^2 - E(v_{\ell i,t}^2)\right)\right)^2 \leq \Delta$*.*

Assumption 9 *:* $\boldsymbol{A}_i = \mathrm{plim}_{T\to\infty}\tilde{\boldsymbol{A}}_{i,T}$ *has full column rank,* $\boldsymbol{B}_i = \mathrm{plim}_{T\to\infty}\tilde{\boldsymbol{B}}_{i,T}$ *and* $\boldsymbol{\Sigma}_i = \mathrm{plim}_{T\to\infty} T^{-1}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{Z}_i$ *are positive definite, uniformly.*

In practice, we can apply principle components approach to estimate factor structure by Bai (2003) and Bai (2009). For the unknown number of factors, Bai and Ng (2002) provide the information criteria that can be applied. Consider the empirical counterpart of the projection matrices defined in (5.9), we have

$$
\begin{aligned}
\boldsymbol{M}_{\hat{F}_x} &= \boldsymbol{I}_T - \hat{\boldsymbol{F}}_x \left(\hat{\boldsymbol{F}}_x^{'}\hat{\boldsymbol{F}}_x\right)^{-1} \hat{\boldsymbol{F}}_x^{'};\\
\boldsymbol{M}_{\hat{F}_{x,-1}} &= \boldsymbol{I}_T - \hat{\boldsymbol{F}}_{x,-1} \left(\hat{\boldsymbol{F}}_{x,-1}^{'}\hat{\boldsymbol{F}}_{x,-1}\right)^{-1} \hat{\boldsymbol{F}}_{x,-1}^{'};\\
&\vdots\\
\boldsymbol{M}_{\hat{F}_{x,-j}} &= \boldsymbol{I}_T - \hat{\boldsymbol{F}}_{x,-j} \left(\hat{\boldsymbol{F}}_{x,-j}^{'}\hat{\boldsymbol{F}}_{x,-j}\right)^{-1} \hat{\boldsymbol{F}}_{x,-j}^{'};\\
&\vdots\\
\boldsymbol{M}_{\hat{F}_{x,-J}} &= \boldsymbol{I}_T - \hat{\boldsymbol{F}}_{x,-J} \left(\hat{\boldsymbol{F}}_{x,-J}^{'}\hat{\boldsymbol{F}}_{x,-J}\right)^{-1} \hat{\boldsymbol{F}}_{x,-J}^{'}.
\end{aligned}
\tag{5.12}
$$

We have the following associated instrument matrix:

$$
\hat{\boldsymbol{Z}}_i = \left(\boldsymbol{M}_{\hat{F}_x}\boldsymbol{X}_i, \boldsymbol{M}_{\hat{F}_{x,-1}}\boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{M}_{\hat{F}_{x,-j}}\boldsymbol{X}_{i,-j}, \ldots, \boldsymbol{M}_{\hat{F}_{x,-j}}\boldsymbol{X}_{i,-J}\right). \tag{5.13}
$$

The two stage least squares (2SLS) estimator of $\boldsymbol{\theta}_i$ is defined as

$$
\hat{\boldsymbol{\theta}}_i = \left(\hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\hat{\hat{\boldsymbol{A}}}_{i,T}\right)^{-1} \hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\hat{\hat{\boldsymbol{g}}}_{i,T}, \tag{5.14}
$$

where[2]

$$
\hat{\hat{\boldsymbol{A}}}_{i,T} = \frac{1}{T}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{W}_i,\ \ \hat{\hat{\boldsymbol{B}}}_{i,T} = \frac{1}{T}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i \textit{ and } \hat{\hat{\boldsymbol{g}}}_{i,T} = \frac{1}{T}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{y}_i. \tag{5.15}
$$

From the model (5.5) and the 2SLS estimator (5.14), we have

$$
\begin{aligned}
&\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i + \left(\hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\hat{\hat{\boldsymbol{A}}}_{i,T}\right)^{-1} \hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\left(T^{-1}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i\right)\\
\Rightarrow&\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) = \left(\hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\hat{\hat{\boldsymbol{A}}}_{i,T}\right)^{-1} \hat{\hat{\boldsymbol{A}}}_{i,T}^{'}\hat{\hat{\boldsymbol{B}}}_{i,T}^{-1}\left(T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i\right)
\end{aligned}
\tag{5.16}
$$

Following proposition is the limiting property of the $T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i$.

Proposition 1 *Consider the model (5.5). Under Assumptions 1-9, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*, we have*

$$
T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i = T^{-1/2}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{u}_i + \sqrt{T}O_p\left(c_{NT}^{-2}\right), \tag{5.17}
$$

where $c_{NT} = min\{\sqrt{N}, \sqrt{T}\}$.

[2] By Norkutė et al. (2021), we can see that using $\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i$ as instruments is expected to be more efficient than using $\hat{\boldsymbol{Z}}_i$ as instruments for first step IV estimator of $\boldsymbol{\theta}_i$.

Under the above proposition , we can see that $T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i$ is $O_p(1)$ as $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$. We summarise the asymptotic property in the following theorem:

Theorem 1 *: Consider the model (5.5). Under Assumptions 1-9, as* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$*. for each* i*,*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) \xrightarrow{d} N\left(\boldsymbol{0},\ \left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)^{-1}\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)^{-1}\right). \tag{5.18}$$

Thus, above result shows that the individual 2SLS estimator $\hat{\boldsymbol{\theta}}_i$ is $\sqrt{T}$ consistent to $\boldsymbol{\theta}_i$.

5.3 Constructing optimal instruments

In Norkutė et al. (2021), the number of valid instruments for $\boldsymbol{\theta}_i$ can increase proportionally to T^2. Thus, the more instruments variables can be used when T is large. The selection of number of instruments is an important issue because there is a well known trade off between efficiency and bias (e.g., Bekker (1994), Han and Phillips (2006) and Ng and Bai (2009)). We survey the literature that there is no studies clearly to discuss how to choose the instruments in this approach.

Kuersteiner and Okui (2010) provide method of constructing optimal instruments by weighting the predicted value of endogenous variable in the estimation stage. This weights can be found by minimizing the asymptotic MSE of average 2SLS estimators. In this section, we consider a method to choose an optimal set of model averaging approach.

To begin with, we rewrite the model (5.1) as

$$\begin{aligned} y_{i,t} &= \phi_i y_{i,t-1} + \boldsymbol{x}_{i,t}^{'}\boldsymbol{\beta}_i + u_{i,t} = \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta}_i + u_{i,t}, \\ \boldsymbol{w}_{i,t} &= \left(y_{i,t-1}, \boldsymbol{x}_{i,t}^{'}\right)^{'} = \boldsymbol{g}_{i,t}(\boldsymbol{z}_{i,t}) + \boldsymbol{e}_{i,t}, \end{aligned} \tag{5.19}$$

where $\boldsymbol{g}_{i,t}(\boldsymbol{z}_{i,t})$ is a $(1+k)$ vector function of $\boldsymbol{z}_{i,t}$ and $\boldsymbol{z}_{i,t}$ is a vector of exogenous variables [3]. We use a short hand notation $\boldsymbol{g}_{i,t} = \boldsymbol{g}_{i,t}(\boldsymbol{z}_{i,t})$ in the remainder of this chapter. Note $u_{i,t}$ and $\boldsymbol{e}_{i,t}$ are unobserved random variable with finite second moments which do not depend on $\boldsymbol{z}_{i,t}$. Stacking the T observations for model (5.19) for each i, we have

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{W}_i\boldsymbol{\theta}_i + \boldsymbol{u}_i, \\ \boldsymbol{W}_i &= \boldsymbol{G}_i + \boldsymbol{E}_i, \end{aligned} \tag{5.20}$$

where $\boldsymbol{G}_i = \left(\boldsymbol{g}_{i,1}, \ldots, \boldsymbol{g}_{i,T}\right)^{'}$ is the $T \times (1+k)$ matrix and $\boldsymbol{E}_i = \left(\boldsymbol{e}_{i,1}, \ldots, \boldsymbol{e}_{i,T}\right)^{'}$ is the $T \times (1+k)$ matrix. The set of instruments has the form $\boldsymbol{Z}_i^j = (\boldsymbol{\psi}_1(\boldsymbol{Z}_i), \ldots, \boldsymbol{\psi}_j(\boldsymbol{Z}_i))$,

[3]From Norkutė et al. (2021) and previous section, we know that these exogenous variables include defactored of $\boldsymbol{x}_{i,t}$ and its lags.

where $\boldsymbol{\psi}_j$ are function of $\boldsymbol{Z}_i$, for $j = 1, \ldots, J$ with J is the maximum lags of variables that we can observe. Therefore, the number of instruments increase when j increase, that is,

$$\begin{aligned}
\boldsymbol{Z}_i^1 &= (\boldsymbol{\psi}_1(\boldsymbol{Z}_i)) = \left(\boldsymbol{M}_{F_x}\boldsymbol{X}_i, \boldsymbol{M}_{F_{x,-1}}\boldsymbol{X}_{i,-1}\right), \\
\boldsymbol{Z}_i^2 &= (\boldsymbol{\psi}_1(\boldsymbol{Z}_i), \boldsymbol{\psi}_2(\boldsymbol{Z}_i)) = \left(\boldsymbol{M}_{F_x}\boldsymbol{X}_i, \boldsymbol{M}_{F_{x,-1}}\boldsymbol{X}_{i,-1}, \boldsymbol{M}_{F_{x,-2}}\boldsymbol{X}_{i,-2}\right), \\
&\vdots \\
\boldsymbol{Z}_i^J &= (\boldsymbol{\psi}_1(\boldsymbol{Z}_i), \ldots, \boldsymbol{\psi}_J(\boldsymbol{Z}_i)) = \left(\boldsymbol{M}_{F_x}\boldsymbol{X}_i, \ldots, \boldsymbol{M}_{F_{x,-j}}\boldsymbol{X}_{i,-j}, \ldots, \boldsymbol{M}_{F_{x,-J}}\boldsymbol{X}_{i,-J}\right).
\end{aligned}$$

The 2SLS estimator of Norkutė et al. (2021) can also be expressed as

$$\hat{\boldsymbol{\theta}}_i^j = \left(\boldsymbol{W}_i'\boldsymbol{P}_i^j\boldsymbol{W}_i\right)^{-1}\boldsymbol{W}_i'\boldsymbol{P}_i^j\boldsymbol{y}_i, \tag{5.21}$$

where the projection matrix, $\boldsymbol{P}_i^j$, is defined as

$$\boldsymbol{P}_i^j = \boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j\left(\hat{\boldsymbol{Z}}_i^{j'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j\right)^{-1}\hat{\boldsymbol{Z}}_i^{j'}\boldsymbol{M}_{\hat{F}_x} \tag{5.22}$$

with

$$\hat{\boldsymbol{Z}}_i^j = \left(\boldsymbol{\psi}_1(\hat{\boldsymbol{Z}}_i), \ldots, \boldsymbol{\psi}_j(\hat{\boldsymbol{Z}}_i)\right) = \left(\boldsymbol{M}_{\hat{F}_x}\boldsymbol{X}_i, \ldots, \boldsymbol{M}_{\hat{F}_{x,-j}}\boldsymbol{X}_{i,-j}\right),$$

which is an $T \times (j+1)k$ matrix where $1 \leq j \leq J$.

A weighting vector for the projection matrix is defined as

$$\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{iJ})', \tag{5.23}$$

where $\boldsymbol{\omega}_i$ is a $J \times 1$ vector and $\sum_{j=1}^J \omega_{i,j} = 1$. Then, we can weight $\boldsymbol{P}_i^j$ as

$$\boldsymbol{P}_i = \sum_{j=1}^J \omega_{i,j}\boldsymbol{P}_i^j, \tag{5.24}$$

where $\boldsymbol{P}_i$ is symmetric but not idempotent.

We estimate a weight vector $\boldsymbol{\omega}_i$ that minimizes a linear combination of approximate mean square error of $\boldsymbol{\eta}_i'\hat{\boldsymbol{\theta}}_i$, denoted as $s_{\eta_i(\boldsymbol{\omega_i})}$[4]. This is defined as[5] $s_{\eta_i(\boldsymbol{\omega_i})} = \boldsymbol{\eta}_i'\boldsymbol{S}_i(\boldsymbol{\omega_i})\boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i \in \mathbb{R}^{1+k}$.

Before we introduce the criterion of estimation of weighting matrix, we consider the following assumptions:

---

[4]If model include more than one endogenous variable, we choose $\boldsymbol{\omega}_i$ to minimizes a linear combination of the approximate mean square error $s_{\eta_i(\boldsymbol{\omega_i})}$, where it is an estimator of $\boldsymbol{\eta}_i'\boldsymbol{S}_i(\boldsymbol{\omega_i})\boldsymbol{\eta}_i$, where $\hat{\boldsymbol{\eta}}_i \xrightarrow{p} \boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_i$ is user specified.

[5]See Theorem 2 for the definition of $\boldsymbol{S}_i(\boldsymbol{\omega_i})$.

Assumption 10 *Define* $\boldsymbol{\omega}_i^+ = (|\omega_{i,1}|, \ldots, |\omega_{i,J}|)^{'}$. *We have following conditions: (i)* $\sum_{j=1}^{J} \omega_{i,j} = 1$ *; (ii)* $\boldsymbol{\omega}_i \in l_1$ *for all* T*, where* $l_1 = (x = (x_1, \ldots) | \sum_{t=1}^{\infty} |x_t| \leq C_{l_1} < \infty)$ *for some constant* C_{l_1}, $J \leq T$*; (iii) As* $T \to \infty$ *and* $J \to \infty$, $\tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i^+ = \sum_{j=1}^{J} |\omega_{ij}| j \to \infty$*, where* $\tilde{\boldsymbol{j}} = (1, \ldots, J)^{'}$.

Assumption 11 *(i)* $\tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i^+ / \sqrt{T} = \sum_{j=1}^{J} |\omega_{ij}| j / \sqrt{T} \to 0$ *or (ii)* $\tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i^+ / T = \sum_{j=1}^{J} |\omega_{ij}| j / T \to 0$ *and* $J/T \to 0$.

Assumption 12 *(i)* $\bar{\boldsymbol{H}}_i \equiv E(\boldsymbol{g}_{i,t} \boldsymbol{g}_{i,t}^{'})$ *exists and is nonsingular. (ii) for some* $\alpha > 1/2$,

$$sup_{j \leq J} j^{2\alpha} \left(sup_{\boldsymbol{\eta}_i^{'} \boldsymbol{\eta}_i = 1} \boldsymbol{\eta}_i^{'} \boldsymbol{G}_i^{'} \left(\boldsymbol{I}_T - \boldsymbol{P}_i^j \right) \boldsymbol{G}_i \boldsymbol{\eta}_i / T \right) = O_p(1).$$

(iii) Let T_+ *be the set of positive integers. There exists a subset* $\bar{J} \subset T_+$ *with a finite number of elements such that*

$$sup_{j \in \bar{J}} sup_{\boldsymbol{\eta}_i^{'} \boldsymbol{\eta}_i} \boldsymbol{\eta}_i^{'} \boldsymbol{G}_i^{'} \left(\boldsymbol{P}_i^j - \boldsymbol{P}_i^{j+1} \right) \boldsymbol{G}_i \boldsymbol{\eta}_i / T = 0,$$

with probability approaching 1, and for all $j \notin \bar{J}$*, it follow that*

$$inf_{j \notin \bar{J}, j \leq J} j^{2\alpha+1} \left(sup_{\boldsymbol{\eta}_i^{'} \boldsymbol{\eta}_i = 1} \boldsymbol{\eta}_i^{'} \boldsymbol{G}_i^{'} \left(\boldsymbol{P}_i^j - \boldsymbol{P}_i^{j+1} \right) \boldsymbol{G}_i \boldsymbol{\eta}_i / T \right) > 0,$$

with probability approaching 1.

Assumption 13 *(i) Let* $e_{ai,t}$ *be the ath element of* $\boldsymbol{e}_{i,t}$*. Then* $E(u_{i,t}^r e_{ai,t}^s | \boldsymbol{z}_{i,t})$ *are constant and bounded for all* a *and* $r, s \leq 5$*. Denote* $\sigma_{u,i}^2 = E(u_{i,t}^2 | \boldsymbol{z}_{i,t})$, $\boldsymbol{\sigma}_{eu,i} = E(\boldsymbol{e}_{i,t} u_{i,t} | \boldsymbol{z}_{i,t})$ *and* $\boldsymbol{\Sigma}_{e,i} = E(\boldsymbol{e}_{i,t} \boldsymbol{e}_{i,t}^{'} | \boldsymbol{z}_{i,t})$ *(ii)* $max_{t \leq T} \boldsymbol{P}_{i,tt}^J \to 0$*, where* $\boldsymbol{P}_{i,tt}^J$ *denote the* (t,t)*th element of* $\boldsymbol{P}_i^J$ *(iii)* $\boldsymbol{g}_{i,t}$ *is bounded. (iv)* $\boldsymbol{Z}_i^{J'} \boldsymbol{Z}_i^J$ *are nonsingular with probability approaching 1. (v)* $\boldsymbol{P}_{i,tt}$ *denote the* (t,t)*th element of* $\boldsymbol{P}_i$.

The optimal weight $\boldsymbol{\omega}_i^*$ is the solution of $min_{\boldsymbol{\omega}_i \in \Omega} s_{\eta_i(\boldsymbol{\omega}_i)}$, where Ω is some set. In this chapter, we consider two versions of Ω:

$$\begin{aligned} &(i) \quad \Omega_p = \left\{ \boldsymbol{\omega}_i \in q | \boldsymbol{\omega}_i^{'} \iota_J = 1; \omega_{i,j} \in [0,1], \forall j \leq J \right\}, \\ &(ii) \quad \Omega_u = \left\{ \boldsymbol{\omega}_i \in q | \boldsymbol{\omega}_i^{'} \iota_J = 1; \forall j \leq J \right\}, \end{aligned}$$

where q is a space of absolutely summable sequences and ι_J is a $J \times 1$ vector of one. In the beginning, we consider positive weights, such that $\Omega = \Omega_p$. Let $\boldsymbol{g}_{i,t} = \boldsymbol{\Pi}_i \boldsymbol{z}_{i,t}$,[6] where $\boldsymbol{\Pi}_i$ is the projection coefficient matrix and the estimator $\hat{\boldsymbol{H}}_i = \frac{\hat{\boldsymbol{G}}_i^{'} \hat{\boldsymbol{G}}_i}{T} = \frac{\left(\hat{\boldsymbol{Z}}_i^j \hat{\boldsymbol{\Pi}}_i \right)^{'} \left(\hat{\boldsymbol{Z}}_i^j \hat{\boldsymbol{\Pi}}_i \right)}{T}$[7]. Denote $\tilde{\boldsymbol{E}}_i$[8] be some preliminary residual from first stage regression

[6]In first stage regression, Donald and Newey (2001) and Kuersteiner and Okui (2010) use a nonparametric reduced form. In this chapter, we use linear form in first stage regression.

[7]We use simple LS estimation for first stage regression, such that $\hat{\boldsymbol{\Pi}}_i = \left(\hat{\boldsymbol{Z}}_i^{j'} \hat{\boldsymbol{Z}}_i^j \right)^{-1} \hat{\boldsymbol{Z}}_i^{j'} \boldsymbol{W}_i$.

[8]$\tilde{\boldsymbol{E}}_i = \boldsymbol{W}_i - \boldsymbol{Z}_i^j \hat{\boldsymbol{\Pi}}_i$.

which is $T \times (1+k)$ matrices. Similarly, define the residuals $\tilde{\boldsymbol{u}}_i = \boldsymbol{y}_i - \boldsymbol{W}_i \hat{\boldsymbol{\theta}}_{pre,i}$[9]. Let $\tilde{\boldsymbol{e}}_{\eta,i} = \tilde{\boldsymbol{E}}_i \hat{\boldsymbol{H}}_i^{-1} \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i$ is a some fixed user-specified vector. Define

$$\hat{\sigma}_{u,i}^2 = \frac{\tilde{\boldsymbol{u}}_i' \tilde{\boldsymbol{u}}_i}{T}, \ \hat{\sigma}_{\eta,i}^2 = \frac{\tilde{\boldsymbol{e}}_{\eta,i}' \tilde{\boldsymbol{e}}_{\eta,i}}{T}, \ \hat{\sigma}_{\eta u,i} = \frac{\tilde{\boldsymbol{e}}_{\eta,i}' \tilde{\boldsymbol{u}}_i}{T}. \tag{5.25}$$

Let $\hat{\boldsymbol{U}}_i = \left(\hat{\mathbf{e}}_{\eta,i}^1, \ldots, \hat{\mathbf{e}}_{\eta,i}^J\right)' \left(\hat{\mathbf{e}}_{\eta,i}^1, \ldots, \hat{\mathbf{e}}_{\eta,i}^J\right)$ is the $J \times J$ matrix, where $\hat{\mathbf{e}}_{\eta,i}^j = \left(\boldsymbol{P}_i^J - \boldsymbol{P}_i^j\right) \boldsymbol{W}_i \hat{\boldsymbol{H}}_i^{-1} \boldsymbol{\eta}_i$ is a $T \times 1$ vector. Define $\boldsymbol{\Gamma}_i$ be the $J \times J$ matrix whose (j', j) element is $min(j', j)$ and let $\tilde{\boldsymbol{j}} = (1, \ldots, J)'$. When we consider $\Omega = \Omega_p$, the criterion $\hat{s}_{\eta_i(\boldsymbol{\omega_i})}$ is

$$\hat{s}_{\eta_i(\boldsymbol{\omega_i})} = \hat{\sigma}_{\eta u,i}^2 \frac{(\tilde{\boldsymbol{j}}' \boldsymbol{\omega}_i)^2}{T} + \hat{\sigma}_{u,i}^2 \frac{\boldsymbol{\omega}_i' \hat{\boldsymbol{U}}_i \boldsymbol{\omega}_i - \hat{\sigma}_{\eta,i}^2 (J - 2\tilde{\boldsymbol{j}}' \boldsymbol{\omega}_i + \boldsymbol{\omega}_i' \boldsymbol{\Gamma}_i \boldsymbol{\omega}_i)}{T}. \tag{5.26}$$

Then, the optimal weight is given by

$$\boldsymbol{\omega}_i^* = \underset{\omega_i \in \Omega_p}{\arg\min}\, \hat{s}_{\eta_i}(\boldsymbol{\omega}_i). \tag{5.27}$$

When we consider $\Omega = \Omega_u$, the criterion $\hat{s}_{\eta_i(\boldsymbol{\omega}_i)}$ is

$$\begin{aligned} \hat{s}_{\eta_i(\boldsymbol{\omega_i})} = \hat{\sigma}_{\eta u,i}^2 \frac{(\tilde{\boldsymbol{j}}' \boldsymbol{\omega}_i)^2}{T} + \left(\hat{\sigma}_{\eta,i}^2 \hat{\sigma}_{u,i}^2 + \hat{\sigma}_{\eta u,i}^2\right) \frac{\boldsymbol{\omega}_i' \Gamma_i \boldsymbol{\omega}_i}{T} - \frac{\tilde{\boldsymbol{j}}' \boldsymbol{\omega}_i}{T} \hat{b}_T + \\ \hat{\sigma}_{u,i}^2 \frac{\boldsymbol{\omega}_i' \hat{\boldsymbol{U}}_i \boldsymbol{\omega}_i - \hat{\sigma}_{\eta,i}^2 (J - 2\tilde{\boldsymbol{j}}' \boldsymbol{\omega}_i + \boldsymbol{\omega}_i' \Gamma_i \boldsymbol{\omega}_i)}{T}, \end{aligned} \tag{5.28}$$

where

$$\begin{aligned} \hat{b}_T = \boldsymbol{\eta}_i' \hat{\boldsymbol{H}}_i^{-1} [2(\hat{\sigma}_{u,i}^2 \hat{\boldsymbol{\Sigma}}_{e_i} + d \hat{\boldsymbol{\sigma}}_{eu,i} \hat{\boldsymbol{\sigma}}_{eu,i}' + \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{G}}_i \hat{\boldsymbol{\sigma}}_{eu,i}' \hat{\boldsymbol{H}}_i^{-1} \hat{\boldsymbol{\sigma}}_{eu,i} \hat{\boldsymbol{G}}_i' + \\ \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\boldsymbol{G}}_i \hat{\boldsymbol{\sigma}}_{eu,i}' \hat{\boldsymbol{H}}_i^{-1} \hat{\boldsymbol{G}}_i \hat{\boldsymbol{\sigma}}_{eu,i}' + \hat{\boldsymbol{\sigma}}_{eu,i} \hat{\boldsymbol{G}}_i' \hat{\boldsymbol{H}}_i^{-1} \hat{\boldsymbol{\sigma}}_{eu,i} \hat{\boldsymbol{G}}_i' \right))] \hat{\boldsymbol{H}}_i^{-1} \boldsymbol{\eta}_i, \end{aligned} \tag{5.29}$$

with $d = dim(\boldsymbol{\theta}_i)$.

The optimal weight is given by

$$\boldsymbol{\omega}_i^* = \underset{\omega_i \in \Omega_u}{\arg\min}\, \hat{s}_{\eta_i}(\boldsymbol{\omega}_i). \tag{5.30}$$

Thus, we can weight $\boldsymbol{P}_i^j$ by the optimal weight $\boldsymbol{\omega}_i^* = \left(\omega_{i,1}^*, \ldots, \omega_{i,J}^*\right)'$ as

$$\boldsymbol{P}_i^* = \sum_{j=1}^{J} \omega_{i,j}^* \boldsymbol{P}_i^j, \tag{5.31}$$

[9] We use 2SLS estimator for $\hat{\boldsymbol{\theta}}_{pre,i}$, such that $\hat{\boldsymbol{\theta}}_{pre,i} = \left(\boldsymbol{W}_i' \boldsymbol{P}_i^j \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i' \boldsymbol{P}_i^j \boldsymbol{y}_i$, and denoted that $\hat{\boldsymbol{\theta}}_{pre,i}$ does not depend on the weighting vector. Also, the number of lags j can be selected by the first-stage Mallows criterion.

The model average 2SLS estimator, $\hat{\boldsymbol{\theta}}_i^*$, now is defined as

$$\hat{\boldsymbol{\theta}}_i^* = \left(\boldsymbol{W}_i^{'}\boldsymbol{P}_i^*\boldsymbol{W}_i\right)^{-1}\boldsymbol{W}_i^{'}\boldsymbol{P}_i^*\boldsymbol{y}_i. \tag{5.32}$$

By minimizing the approximation to the higher order MSE of $\boldsymbol{\theta}_i$, we can select the optimal weights. Following Kuersteiner and Okui (2010), the theorem provides the approximate MSE of $\boldsymbol{\theta}_i$ conditional on the exogenous variable $\boldsymbol{z}_{i,t}$, $E\left[\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)^{'}|\boldsymbol{z}_{i,t}\right]$ by $\sigma_{u,i}^2\boldsymbol{H}_i^{-1}+\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)$, where $\sigma_{u,i}^2\boldsymbol{H}_i^{-1}$ is the first-order asymptotic variance and $\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)$ is the dominant term in the approximate MSE of $\boldsymbol{\theta}_i$.

Theorem 2 *: Consider the model (5.19) and suppose Assumption 1-13 are satisfied. Define* $\mu_{i,t}(\boldsymbol{\omega}_i)=E(u_{i,t}^2\boldsymbol{e}_{i,t})\boldsymbol{P}_{i,tt}$ *and* $\boldsymbol{\mu}_i(\boldsymbol{\omega}_i)=\left(\mu_{i,1}(\boldsymbol{\omega}_i),\ldots,\mu_{i,T}(\boldsymbol{\omega}_i)\right)^{'}$*, where* $\boldsymbol{P}_{i,tt}$ *is* (t,t) *element of* $T\times T$ *matrix of* $\boldsymbol{P}_i$*. As* $\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^{+}\rightarrow\infty$, $T\rightarrow\infty$, *we have*

$$\begin{aligned}
&T\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)^{'}=\hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right)+\hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right)\\
&E\left[\hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right)|\boldsymbol{z}_{i,t}\right]=\sigma_{u,i}^2\boldsymbol{H}_i^{-1}+\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)+\boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right)\\
&\left(\hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right)+\boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right)\right)/tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right)=o_p(1),
\end{aligned} \tag{5.33}$$

with

$$\begin{aligned}
&\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)=\boldsymbol{H}^{-1}[Cum[u_{i,t},u_{i,t},\boldsymbol{e}_{i,t},\boldsymbol{e}_{i,t}^{'}]\frac{\sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})}{T}+\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+\\
&E\left(u_{i,1}^2\boldsymbol{e}_{i,1}\right)\sum_{t=1}^{T}\boldsymbol{g}_{i,t}^{'}\boldsymbol{P}_{i,tt}/T+\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{P}_{i,tt}E\left(u_{i,1}^2\boldsymbol{e}_{i,1}\right)/T+\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{\mu}_i\left(\boldsymbol{\omega}_i\right)}{T}+\\
&\frac{\boldsymbol{\mu}_i^{'}\left(\boldsymbol{\omega}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\frac{\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i\right)^2}{T}+\left(\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\right)\frac{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right)}{T}-\\
&2\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\left[\sigma_u^2\boldsymbol{\Sigma}_{e,i}+d\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}+\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\right]-\\
&2\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T^2}\left(\sum_{t=1}^{T}\left(\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\right)\right)]\boldsymbol{H}^{-1}
\end{aligned} \tag{5.34}$$

and[10]

$$
\begin{aligned}
&d = dim\left(\boldsymbol{\theta}_i\right), \\
&Cum\left[u_{i,t}, u_{i,t}, \boldsymbol{e}_{i,t}, \boldsymbol{e}_{i,t}^{'}\right] = E\left(u_{i,t}^2 \boldsymbol{e}_{i,t}\boldsymbol{e}_{i,t}^{'}\right) - \sigma_{u,i}^2 \boldsymbol{\Sigma}_{e,i} - 2\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}, \\
&\hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right) = \boldsymbol{H}_i^{-1}\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right)\boldsymbol{H}_i^{-1}, \\
&\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) = \left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'} \\
&- \boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'} - \left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^{'} \\
&- \boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}\boldsymbol{H}_i^{-1}\left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right) - \left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'} \\
&- \boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)^{'} - \left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}.
\end{aligned}
$$

In above theorem, $\sigma_{u,i}^2\boldsymbol{H}_i^{-1}$ is the first-order asymptotic variance. And $\hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right)$ and $\boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right)$ go to zero faster than $\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)$. Therefore, $\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)$ is the dominant term in the MSE of estimator. The bias could be eliminate by setting $\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i = 0$, so we need to consider an expansion of model average 2SLS estimator that contains additional higher order term. By using the average predicted value of the endogenous variable in the estimation stage, we can construct optimal instruments for 2SLS estimator on dynamic heterogeneous panel data model with defactored regressors and a multifactor error.

5.4 Mean group 2SLS estimator

Our interest parameters are over group of the coefficients, $\boldsymbol{\theta}$. From Pesaran and Smith (1995), we know cross-section estimator, $\hat{\boldsymbol{\theta}}$, is inconsistent as $N, T \to \infty$. In practice, we can estimate individual consistent estimators of $\boldsymbol{\theta}_i$, and then calculate the coefficient means to get Mean Group estimators. Now, we consider the Mean Group estimator of $\boldsymbol{\theta}$:

$$
\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_i^{*}. \tag{5.35}
$$

From Assumption 5, we can show that the asymptotic property of $\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*}$, as

$$
\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*} - \boldsymbol{\theta} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^{*} - \boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^{*} - \boldsymbol{\theta}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i, \tag{5.36}
$$

[10] The model average 2SLS estimator has the form as $\sqrt{T}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}) = \hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i$. Let the decomposition of $\hat{\boldsymbol{H}}_i$ and $\hat{\boldsymbol{h}}_i$ are $\hat{\boldsymbol{H}}_i = \boldsymbol{H}_i + \boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H + \boldsymbol{c}_{3i}^H + \boldsymbol{r}_i^H$ and $\hat{\boldsymbol{h}}_i = \boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h$, The detail of these decomposition and the definition of $\hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right)$ and $\boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right)$ are in the proof of Theorem 2 in Appendix c.

where the first of right hand side

$$
\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right) = \\
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{j=1}^{J}\omega_{i,j}^*\left(\left(\frac{\boldsymbol{W}_i'\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j}{T}\right)\left(\frac{\hat{\boldsymbol{Z}}_i^{j'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j}{T}\right)^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j}{T}\right)\right)^{-1} \qquad (5.37)\\
&\times\left(\frac{\boldsymbol{W}_i'\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j}{T}\right)\left(\frac{\hat{\boldsymbol{Z}}_i^{j'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i^j}{T}\right)^{-1}\left(\frac{\hat{\boldsymbol{Z}}_i^{j'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i}{T}\right)\\
&= O_p(1),
\end{aligned}
$$

which implies $\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right) = O_p\left(\frac{1}{\sqrt{NT}}\right)$.

Then, we can see

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{MA2SLSMG}^* - \boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i + o_p(1). \qquad (5.38)$$

As $N \to \infty$, we can see

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\lambda}_i \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_\lambda^*\right). \qquad (5.39)$$

Therefore, we know that $\hat{\boldsymbol{\theta}}_{MA2SLSMG}^*$ is $\sqrt{N}$ consistent.

And the variance estimator of $\hat{\boldsymbol{\theta}}_{MA2SLSMG}^*$ is given by

$$\hat{\boldsymbol{\Sigma}}_\lambda^* = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right)\left(\hat{\boldsymbol{\theta}}_i^* - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right)'. \qquad (5.40)$$

Follow Norkutė et al. (2021), we can show that $\hat{\boldsymbol{\Sigma}}_\lambda^*$ is consistent and it does not have small T bias. Firstly, we decompose (5.40) as

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right)\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta} + \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right)' = \\
&\sum_{i=1}^{N}\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i' + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right)' + \sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i + \sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{\theta}_i\right) - \\
&N\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right)'\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MA2SLSMG}^*\right).
\end{aligned}
\qquad (5.41)
$$

Then we can show consistent of $\hat{\boldsymbol{\Sigma}}_{\lambda}^{*}$ as

$$\hat{\boldsymbol{\Sigma}}_{\lambda}^{*}-\boldsymbol{\Sigma}_{\lambda}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\boldsymbol{\lambda}_i\boldsymbol{\lambda}_i^{'}-\boldsymbol{\Sigma}_{\lambda}^{*}\right)+\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^{*}-\boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i^{*}-\boldsymbol{\theta}_i\right)^{'}$$
$$+\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\theta}}_i^{*}-\boldsymbol{\theta}_i\right)\boldsymbol{\lambda}_i+\frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\left(\hat{\boldsymbol{\theta}}_i^{*}-\boldsymbol{\theta}_i\right)- \tag{5.42}$$
$$\frac{N}{N-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*}\right)^{'}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*}\right)$$
$$=O_p\left(\frac{1}{\sqrt{N}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left((\frac{1}{N}\right).$$

Then, we can see that the asymptotic property of $\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*}$ as

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{MA2SLSMG}^{*}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}_{\lambda}^{*}\right). \tag{5.43}$$

5.5 Monte Carlo simulation design

We compare the performance of four estimators. First, we consider MA2SLSMG estimator with $\Omega=\Omega_u$. The criterion for finding weights of MA2SLSMG estimator is (5.28). Second, we consider P-MA2SLSMG estimator with $\Omega=\Omega_p$ and the criterion is (5.28). Third, we consider Ps-MA2SLSMG estimator with $\Omega=\Omega_p$ and the criterion is (5.26). Fourth, we consider 2SLSMG estimator with all available instruments.

In this experiments, we consider two instruments set with true factors projection matrix:

$$a:\boldsymbol{Z}_i^2=\left(\boldsymbol{M}_{F_x}\boldsymbol{X}_i,\boldsymbol{M}_{F_{x,-1}}\boldsymbol{X}_{i,-1},\boldsymbol{M}_{F_{x,-2}}\boldsymbol{X}_{i,-2}\right);$$
$$b:\boldsymbol{Z}_i^5=\left(\boldsymbol{M}_{F_x}\boldsymbol{X}_i,\boldsymbol{M}_{F_{x,-1}}\boldsymbol{X}_{i,-1},\ldots,\boldsymbol{M}_{F_{x,-5}}\boldsymbol{X}_{i,-5}\right).$$

Denote MA2SLSMG^a, P-MA2SLSMG^a and Ps-MA2SLSMG^a use 6 instruments. MA2SLSMG^b,P-MA2SLSMG^a and Ps-MA2SLSMG^b use 12 instruments. We use the similar Monte Carlo design as Norkutė et al. (2021) to investigate the performance of MA2SLSMG, P-MA2SLSMG and Ps-MA2SLSMG estimators.

5.5.1 Dynamic heterogeneous panels data model with multifactor error structure

We consider the data generating process as following:

$$y_{i,t}=\phi_i y_{i,t-1}+\beta_{1,i}x_{1i,t}+\beta_{2,i}x_{2i,t}+u_{i,t},\quad i=1,\ldots N;\ t=-49,\ldots,T, \tag{5.44}$$
$$\tag{5.45}$$

where

$$u_{i,t}=\sum_{s=1}^{m_y}\gamma_{si}f_{s,t}+\varepsilon_{i,t}, \tag{5.46}$$

with

$$f_{s,t} = \rho_f f_{s,t-1} + (1-\rho_f^2)^{1/2}\zeta_{s,t}, \tag{5.47}$$

with $\zeta_{s,t} \overset{i.i.d.}{\sim} N(0,1)$ for $s = 1, \dots m_y$. We set $m_y = 3$ and $\rho_f \in \{0.1, 0.9\}$. We allow the idiosyncratic error term, $\varepsilon_{i,t}$, is drawn as

$$\varepsilon_{i,t} = \varsigma_\varepsilon \sigma_{i,t} \left(\epsilon_{i,t} - 1\right)/\sqrt{2}, \tag{5.48}$$

where

$$\begin{aligned} &\varsigma_\varepsilon = \frac{\pi_\mu}{1-\pi_\mu} m_y, \quad \sigma_{i,t}^2 = \eta_i \varphi_t, \quad \eta_i \overset{i.i.d.}{\sim} \chi_2^2/2, \quad \varphi_t = t/T, \\ &and \; \epsilon_{i,t} \overset{i.i.d.}{\sim} \chi_1^2. \end{aligned} \tag{5.49}$$

We set $\pi_\mu \in \{1/4\}$.

The process for the regressors are drawn as

$$x_{\ell it} = \sum_{s=1}^{m_x} \gamma_{\ell si} f_{s,t} + v_{\ell it}, \quad i = 1, \dots N;\, t = -49, \dots, T\,;\, \ell = 1, 2, \tag{5.50}$$

where $k = 2$, $m_x = 2$, $\boldsymbol{f}_{y,t} = (f_{1t}, f_{2t}, f_{3t})'$ and $\boldsymbol{f}_{x,t} = (f_{1t}, f_{2t})'$.

We set

$$v_{\ell i,t} = \rho_{v,\ell} v_{\ell i,t-1} + \left(1-\rho_{v,\ell}^2\right)^{\frac{1}{2}} \varpi_{\ell i,t}, for\, \ell = 1, 2, \tag{5.51}$$

where $\rho_{v,\ell} \in \{0.1, 0.9\}$ for all ℓ and $\varpi_{\ell i,t} \overset{i.i.d.}{\sim} N\left(0, \sigma_{\varpi_{it}}^2\right)$, $\sigma_{\varpi_{it}}^2 \overset{i.i.d.}{\sim} U\left(0.5, 1.5\right)$. Factor loadings in $u_{i,t}$ are generated as $\gamma_{si} \overset{i.i.d.}{\sim} N\left(0,1\right)$, for $s = 1, \dots, m_y = 3$, and the factor loadings in x_{1it} and x_{2it} are generated as

$$\begin{aligned} \gamma_{1si} &= \rho_{\gamma,1s}\gamma_{3i} + \left(1-\rho_{\gamma,1s}^2\right)^{1/2} \xi_{1si}; \quad \xi_{1si} \overset{i.i.d.}{\sim} N\left(0,1\right); \\ \gamma_{2si} &= \rho_{\gamma,2s}\gamma_{si} + \left(1-\rho_{\gamma,2s}^2\right)^{1/2} \xi_{2si}; \quad \xi_{2si} \overset{i.i.d.}{\sim} N\left(0,1\right); \end{aligned} \tag{5.52}$$

for $s = 1, \dots, m_x = 2$. We set $\rho_{\gamma,11} = \rho_{\gamma,12} \in \{0, 0.3\}$ and $\rho_{\gamma,21} = \rho_{\gamma,22} = \{0.5\}$. The factor loading matrix is defined as

$$\boldsymbol{\Gamma}_i = \begin{bmatrix} \gamma_{1i} & \gamma_{11i} & \gamma_{21i} \\ \gamma_{2i} & \gamma_{12i} & \gamma_{22i} \\ \gamma_{3i} & 0 & 0 \end{bmatrix}. \tag{5.53}$$

The slope coefficients are generated as

$$\phi_i = \phi + \eta_{\phi i}, \;\; \beta_{1,i} = \beta_1 + \eta_{\beta_1 i}\, and\, \beta_{2,i} = \beta_2 + \eta_{\beta_2 i}. \tag{5.54}$$

Here we consider $\phi \in \{0.5, 0.8\}$, $\beta_1 = 3$ and $\beta_2 = 1$. For the design of heterogenous slopes, $\eta_{\phi i} \overset{i.i.d.}{\sim} U(-0.2, 0.2)$, and

$$\eta_{\beta_\ell i} = \left[(0.4)^2/12\right]^{1/2} \rho_\beta \xi_{\beta \ell i} + \left(1 - \rho_\beta^2\right)^{1/2} \eta_{\phi i}, \tag{5.55}$$

where

$$\xi_{\beta \ell i} = \frac{\bar{v}^2_{\ell i} - \bar{v}^2_{\ell}}{\left[N^{-1} \sum_{i=1}^{N} \left(\bar{v}^2_{\ell i} - \bar{v}^2_{\ell}\right)^2\right]^{1/2}}, \tag{5.56}$$

with $\bar{v}^2_{\ell i} = T^{-1} \sum_{t=1}^{T} v^2_{\ell it}$, $\bar{v}^2_{\ell} = N^{-1} \sum_{i=1}^{N} \bar{v}^2_{\ell i}$, for $\ell = 1, 2$. To investigate size and power, we consider null hypothesis, $H_0 : \phi = \phi$, $H_0 : \beta_1 = \beta_1$ and $H_0 : \beta_2 = \beta_2$. For the size of t-test, we consider null hypothesis, $H_0 : \phi = \phi + 0.1$, $H_0 : \beta_1 = \beta_1 + 0.1$ and $H_0 : \beta_2 = \beta_2 + 0.1$. All tests are carried out at 5% significance level. We consider $T \in \{25, 50, 100\}$, $N \in \{25, 50, 100\}$, and all experiments are replicated 1000 times. Depending of endogeneity increases when factor more persistent or correlation between the factor loadings increases because the strength of IVs increase when $|\phi_i|$ increase, $|\rho_{v,\ell}|$ increasing or $|\boldsymbol{\beta}_i|$ increases[11]. In this simulation setting, we fixed correlation between the factor loadings and $|\boldsymbol{\beta}_{\ell i}|$. By controling $\rho_{v,\ell}$ and ρ_f, we can investigate the effect of strength of IVs and degree of endogeneity. The following table summarise Monte Carlo designs:

Endogeneity \ IVs	Weak	Strong
Low	Case A: ($\rho_{v,\ell} = 0.1$, $\rho_f = 0.1$)	Case C: ($\rho_{v,\ell} = 0.9$, $\rho_f = 0.1$)
High	Case B ($\rho_{v,\ell} = 0.1$, $\rho_f = 0.9$)	Case D ($\rho_{v,\ell} = 0.9$, $\rho_f = 0.9$)

5.5.2 Monte Carlo simulation results

Table 5.1 to Table 5.8 report the bias and MSE of MA2SLSMG^{a,b}, P-MA2SLSMG^{a,b}, Ps-MA2SLSMG^{a,b}, 2SLSMG^{a,b} estimates, and size (%) and power (%) of associated t-tests for dynamic heterogeneous panel data model with $\{\phi\} = \{0.5, 0.8\}$ and $\{\beta_1, \beta_2\} = \{3, 1\}$.

Table 5.1 to Table 5.2 consider the case of weak instruments and low degree of endogeneity with 6 and 12 instruments, respectively. In Table 5.1, the results show that the bias of ϕ of MA2SLSMG^a is smaller than that of 2SLSMG^a in most scenarios. But the MSE of MA2SLSMG^a is larger than that of 2SLSMG^a. The size of the t-test associated with the MA2SLSMG^a is similar with 2SLSMG^a in most scenarios and when N and T are large they are close to the nominal value. The bias, MSE, size and power perform similarly in terms of 2SLSMG^a estimators .

Table 5.2 reports results for the case in which the 12 instruments are employed. The results show that the bias of ϕ of MA2SLSMG^b is smaller than 2SLSMG^b in most scenarios. The MSE of MA2SLSMG^b is slightly larger than 2SLSMG^b when both N and T are relatively small but they become similar as N and T get larger. As shown

[11]See Appendix c for the discussion of IVs strength and degree of endogeneity

in Table 5.2, the MSE of P-MA2SLSMG^b is smaller than that of MA2SLSMG^b. In the case of weak instruments and low degree of endogeneity, P-MA2SLSMG estimators show similar performance with 2SLSMG^b estimators in terms of bias, MSE, size and power.

In Table 5.3 and Table 5.4, we consider the case of weak instruments and high degree of endogeneity with 6 and 12 instruments, respectively. Looking at Table 5.3 and Table 5.4, it is apparent that the performance of MA2SLSMG^a estimator is worse than that of 2SLSMG^a in most scenarios. In Table 5.3, the performance of P-MA2SLSMG estimator is similar to that of 2SLSMG estimator in all scenarios. When the number of instruments increases to 12, the bias and MSE of P-MA2SLSMG estimators of ϕ is smaller than that of 2SLSMG estimator in Table 5.4.

In Table 5.5 and Table 5.6, we consider the case of strong instruments and low degree of endogeneity with 6 and 12 instruments, respectively. MA2SLSMG estimator still performs worse than P-MA2SLSMG estimator and 2SLSMG estimator in Table 5.5 and Table 5.6. In Table 5.5, the performance of P-MA2SLSMG estimator is similar to that of 2SLSMG estimator. When instruments increase, the bias of P-MA2SLSMG estimator is smaller than 2SLSMG estimator in most scenarios.

In Table 5.7 and Table 5.8, we consider the case of strong instruments and high degree of endogeneity with 6 and 12 instruments, respectively. P-MA2SLS estimator of ϕ performs better than 2SLSMG estimator when the number of instruments increases. The bias of of P-MA2SLS^b estimator of β is smaller than that of 2SLSMG^b estimator but the MSE is slightly larger than 2SLSMG estimator of β.

5.6 Conclusions

This chapter develops model average 2SLS mean group estimators for dynamic heterogeneous panel data models with defactored regressors and a multifactor errors structure to reduce the uncertainty of instruments selection. First, we use lag defector regressor as instruments to obtain consistent IV estimator of cross-sectionally heterogeneous slope vector $\boldsymbol{\theta}_i$. Then, we apply modal average method to construct optimal instruments by weighting the first stage regression. Finally, we construct mean group estimator by weighting cross-sectionally heterogeneous model average IV estimator. When the degree of endogeneity is high, instruments weak and instruments contain many uninformative instruments, it is recommended to use P-MA2SLS estimators. To compare this method with other instruments selection methods (Fan et al. (2020) ,Lee and Shin (2020), Abadie et al. (2019), Belloni et al. (2012), Fan and Li (2001) etc.) would be an interesting area in further research. Norkutė et al. (2021) allow idiosyncratic errors are heteroskedasticity in equation for $y_{i,t}$ but Kuersteiner and Okui (2010) assume idiosyncratic errors are homoscedastic. It would be interesting to extend this assumption to robust weighting method.

Table 5.1 (Case A: Low degree of endogeneity, Weak IVs) Bias, MSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias($\times$100)			MSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.060	-0.131	-0.258	0.123	0.107	0.030	6.2	4.7	5.2	83.9	95.5	99.6
50	-0.012	-0.108	-0.036	0.077	0.038	0.018	6.7	5.6	3.6	94.5	99.8	100.0
100	-0.033	0.028	-0.054	0.063	0.030	0.016	6.9	5.2	4.8	97.4	100.0	100.0
P-MA2SLSMG^a												
25	-0.264	-0.193	-0.241	0.083	0.043	0.022	6.3	5.8	5.8	91.5	99.2	100.0
50	-0.251	-0.186	-0.236	0.068	0.031	0.016	7.4	5.1	4.4	96.2	100.0	100.0
100	-0.054	-0.112	-0.012	0.056	0.031	0.015	5.6	6.4	5.0	98.8	100.0	100.0
Ps-MA2SLSMG^a												
25	-0.073	0.047	-0.110	0.113	0.064	0.031	6.8	6.7	5.7	86.2	97.2	99.6
50	-0.047	-0.069	0.021	0.066	0.032	0.017	5.7	5.0	5.0	96.6	99.8	100.0
100	0.118	0.015	-0.033	0.059	0.030	0.014	5.6	6.1	4.4	98.7	100.0	100.0
2SLSMG^a												
25	-0.264	-0.193	-0.241	0.083	0.043	0.022	6.3	5.8	5.8	91.5	99.2	100.0
50	-0.251	-0.186	-0.236	0.068	0.031	0.016	7.4	5.1	4.4	96.2	100.0	100.0
100	-0.054	-0.112	-0.012	0.056	0.031	0.015	5.6	6.4	5.0	98.8	100.0	100.0

Results for β_1.

	Bias($\times$100)			MSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.477	0.102	-0.130	0.587	0.275	0.123	7.4	4.1	3.7	32.1	55.2	78.5
50	-0.002	-0.096	0.032	0.199	0.116	0.053	5.1	4.5	3.5	61.3	81.1	96.9
100	0.021	0.130	0.050	0.119	0.057	0.029	4.6	5.3	3.8	82.4	97.9	99.9
P-MA2SLSMG^a												
25	0.023	0.099	-0.052	0.494	0.244	0.116	6.3	5.2	4.3	37.1	58.5	82.7
50	-0.051	-0.006	-0.061	0.228	0.097	0.057	7.5	3.8	5.4	59.0	85.6	97.3
100	0.126	-0.124	0.018	0.119	0.058	0.029	5.8	4.6	4.1	82.4	97.4	100.0
Ps-MA2SLSMG^a												
25	0.000	-0.376	0.220	0.584	0.348	0.170	3.7	5.2	6.4	33.1	50.7	78.4
50	-0.044	-0.021	0.102	0.199	0.105	0.057	5.2	3.5	5.9	61.5	84.8	98.4
100	0.168	0.059	-0.094	0.125	0.061	0.028	5.5	5.3	4.2	81.2	98.3	100.0
2SLSMG^a												
25	0.023	0.099	-0.052	0.494	0.244	0.116	6.3	5.2	4.3	37.1	58.5	82.7
50	-0.051	-0.006	-0.061	0.228	0.097	0.057	7.5	3.8	5.4	59.0	85.6	97.3
100	0.126	-0.124	0.018	0.119	0.058	0.029	5.8	4.6	4.1	82.4	97.4	100.0

$y_{i,t}$ is generated as $y_{i,t} = \phi_i y_{i,t-1} + \beta_{1,i}x_{1i,t} + \beta_{2,i}x_{2i,t} + u_{i,t}$, $u_{i,t} = \sum_{s=1}^{m_y} \gamma_{si} f_{s,t} + \varepsilon_{i,t}$, $x_{\ell it} = \sum_{\ell=1}^{k} \phi_\ell x_{\ell i,t-1} + \sum_{s=1}^{m_x} \gamma_{\ell si} f_{s,t} + v_{\ell it}$, $\phi_\ell = 0$, $\ell = 1,2$; $i = 1, \dots, N$; $t = -50, \dots, T$ and the first 50 observations are discarded; $v_{\ell i,t} = \rho_{v,\ell} v_{\ell i,t-1} + \left(1 - \rho_{v,\ell}^2\right)^{\frac{1}{2}} \varpi_{\ell i,t}$, $\rho_{v,\ell} = 0.5$ for all ℓ and $\varpi_{\ell i,t} \overset{i.i.d.}{\sim} U(0.5, 1.5)$; $f_{s,t} = \rho_f f_{s,t-1} + (1-\rho_f^2)^{1/2}\zeta_{s,t}$, $\zeta_{s,t} \overset{i.i.d.}{\sim} N(0,1)$ for $s = 1, \dots m_y = 3$; $\varepsilon_{i,t} = \varsigma_\varepsilon \sigma_{it} \left(\epsilon_{it} - 1\right)/\sqrt{2}$, $\epsilon_{it} \overset{i.i.d.}{\sim} \chi_1^2$ with $\sigma_{it}^2 = \eta_i \varphi_t$, $\eta_i \overset{i.i.d.}{\sim} \chi_2^2/2$, and $\varphi_t = t/T \, for \; t = 0, \dots, T$; $\varsigma_\varepsilon = \frac{\pi_\mu}{1-\pi_\mu} m_y$; $\gamma_{si} \overset{i.i.d.}{\sim} N(0,1)$, for $s = 1, \dots, m_y$, $\gamma_{1si} = \rho_{\gamma,1s}\gamma_{3i} + \left(1 - \rho_{\gamma,1s}^2\right)^{1/2} \xi_{1si}$, $\gamma_{2si} = \rho_{\gamma,2s}\gamma_{si} + \left(1 - \rho_{\gamma,2s}^2\right)^{1/2} \xi_{2si}$, $\xi_{\ell si} \overset{i.i.d.}{\sim} N(0,1) \; for \; \ell = 1,2, \; s = 1, \dots, m_x = 2$. $\phi_i = \phi + \eta_{\phi i}$, $\beta_{1,i} = \beta_1 + \eta_{\beta_1 i}$; $\beta_{1,i} = \beta_1 + \eta_{\beta_1 i}$ and $\beta_{2,i} = \beta_2 + \eta_{\beta_2 i}$, $\eta_{\phi i} \overset{i.i.d.}{\sim} U(-0.2, 0.2)$, and $\eta_{\beta_\ell i} = \left[(0.4)^2/12\right]^{1/2} \rho_\beta \xi_{\beta \ell i} + \left(1 - \rho_\beta^2\right)^{1/2} \eta_{\phi i}$, where $\xi_{\beta \ell i} = \frac{\bar{v}_{\ell i}^2 - \bar{\bar{v}}_\ell^2}{\left[N^{-1}\sum_{i=1}^N \left(\bar{v}_{\ell i}^2 - \bar{\bar{v}}_\ell^2\right)^2\right]^{1/2}}$ with $\bar{v}_{\ell i}^2 = T^{-1}\sum_{t=1}^T v_{\ell it}^2$, $\bar{\bar{v}}_\ell^2 = N^{-1}\sum_{i=1}^N \bar{v}_{\ell i}^2$. We set $\rho_\beta = 0.4$ for $\ell = 1,2$.. We set $\rho_{\gamma,11} = \rho_{\gamma,12} = \rho_{\gamma,21} = \rho_{\gamma,22} \in \{0.8\}$ for correlated factor loading in $x_{1i,t}$ and $u_{i,t}$.

Table 5.2 (Case A: Low degree of endogeneity, Weak IVs) Bias, MSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.												
	Bias($\times$100)			MSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.099	-0.106	-0.212	0.090	0.045	0.024	7.1	5.5	6.3	90.6	99.6	100.0
50	0.040	-0.119	-0.052	0.070	0.036	0.018	6.4	5.7	5.2	94.9	99.9	100.0
100	-0.033	0.036	-0.058	0.061	0.031	0.016	6.5	5.6	5.6	97.4	99.9	100.0
P-MA2SLSMG^b												
25	-0.181	-0.145	-0.186	0.071	0.036	0.019	6.6	5.9	6.8	94.7	99.7	100.0
50	-0.210	-0.176	-0.210	0.064	0.030	0.015	6.9	5.9	5.2	97.1	100.0	100.0
100	-0.042	-0.118	-0.022	0.055	0.030	0.014	5.5	6.5	5.1	99.1	100.0	100.0
Ps-MA2SLSMG^b												
25	-0.096	0.010	-0.163	0.109	0.061	0.029	6.8	6.4	5.7	86.7	97.2	99.7
50	-0.062	-0.085	-0.001	0.066	0.032	0.017	6.0	4.9	4.8	96.7	99.8	100.0
100	0.106	0.005	-0.046	0.059	0.030	0.014	5.5	6.1	4.5	98.4	100.0	100.0
2SLSMG^b												
25	-0.166	-0.132	-0.169	0.070	0.036	0.019	6.8	5.9	6.6	94.5	99.7	100.0
50	-0.193	-0.163	-0.196	0.063	0.030	0.015	7.0	5.9	5.0	97.2	100.0	100.0
100	-0.037	-0.114	-0.017	0.055	0.030	0.014	5.4	6.6	5.2	99.0	100.0	100.0

Results for β_1.												
	Bias($\times$100)			MSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.460	0.071	-0.169	0.519	0.233	0.110	8.4	5.2	3.7	34.9	58.5	81.5
50	-0.018	-0.080	0.038	0.199	0.115	0.051	5.1	5.0	3.2	61.1	84.0	98.2
100	0.030	0.114	0.051	0.118	0.057	0.029	4.6	5.4	4.2	81.6	98.0	100.0
P-MA2SLSMG^b												
25	0.032	-0.003	-0.034	0.452	0.226	0.105	5.6	4.8	4.5	38.5	59.6	84.5
50	-0.059	-0.020	-0.061	0.222	0.094	0.056	7.8	3.5	5.8	59.9	86.1	97.4
100	0.118	-0.117	0.016	0.118	0.057	0.029	5.8	4.4	4.2	82.8	97.4	100
Ps-MA2SLSMG^b												
25	0.070	-0.347	0.188	0.581	0.340	0.152	3.5	4.5	5.7	33.2	51.4	78.5
50	-0.031	-0.035	0.102	0.199	0.106	0.058	5.3	3.6	5.4	60.9	84.4	98.2
100	0.176	0.061	-0.089	0.125	0.061	0.028	5.6	5.3	4.1	81.2	98.3	100.0
2SLSMG^b												
25	0.036	-0.005	-0.024	0.451	0.224	0.105	5.8	4.6	4.3	38.4	60.4	84.6
50	-0.051	-0.020	-0.054	0.222	0.094	0.055	7.7	3.4	5.8	60.1	85.9	97.5
100	0.121	-0.114	0.021	0.117	0.057	0.029	5.8	4.4	4.1	82.7	97.2	100.0

The DGP is the same as that for Table 5.1 except the number of instruments are 12.

Table 5.3 (Case B: High degree of endogeneity, Weak IVs) Bias, MSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	2.706	2.880	2.593	0.360	0.247	0.200	16.2	19.7	27.3	82.6	96.1	99.1
50	2.005	1.966	1.977	0.197	0.138	0.096	10.6	16.9	23.3	90.3	98.1	99.9
100	1.027	1.126	1.025	0.113	0.066	0.039	9.2	9.0	12.8	94.4	99.6	100.0
P-MA2SLSMG^a												
25	1.739	2.003	2.012	0.170	0.132	0.113	12.7	17.9	25.5	93	99.6	100.0
50	1.166	1.126	1.180	0.120	0.064	0.051	10.1	9.8	17.0	94.8	99.9	100.0
100	0.536	0.461	0.615	0.073	0.040	0.024	6.0	6.1	8.1	96.4	100.0	100.0
Ps-MA2SLSMG^a												
25	1.140	1.402	1.127	0.242	0.151	0.111	8.4	9.8	10.8	79.8	92.3	97.1
50	0.701	0.369	0.599	0.243	0.080	0.052	6.6	5.5	6.1	88.3	96.3	98.8
100	0.338	0.224	0.208	0.089	0.046	0.022	5.9	5.7	5.0	94.1	98.8	99.4
2SLSMG^b												
25	1.739	2.003	2.012	0.170	0.132	0.113	12.7	17.9	25.5	93	99.6	100.0
50	1.166	1.126	1.180	0.120	0.064	0.051	10.1	9.8	17.0	94.8	99.9	100.0
100	0.536	0.461	0.615	0.073	0.040	0.024	6.0	6.1	8.1	100	96.4	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-1.189	0.216	0.138	3.062	0.580	0.279	6.7	5.8	5.0	22.3	36.6	55.6
50	-0.052	-0.105	-0.147	0.507	0.335	0.127	5.5	5.5	5.3	35.8	54.7	79.6
100	-0.253	-0.354	-0.344	0.274	0.140	0.075	5.6	5.4	5.7	53.6	75.9	93.0
P-MA2SLSMG^a												
25	0.027	0.189	0.200	0.837	0.460	0.222	5.4	5.9	4.4	26	41.9	63.3
50	-0.449	0.035	-0.306	0.506	0.223	0.122	7.1	5.2	5.6	35.2	61.5	81.6
100	-0.203	-0.524	-0.362	0.255	0.122	0.064	5.3	4.6	6.1	54.2	76.4	93.6
Ps-MA2SLSMG^a												
25	0.695	0.105	0.296	1.903	0.796	0.500	4.7	5.2	4.8	23.2	35.3	56.6
50	-0.444	-0.520	-0.202	0.561	0.280	0.277	5.5	4.1	6.1	37.2	54.7	77.4
100	-0.093	-0.278	-0.273	0.307	0.144	0.071	6.2	5.2	5.3	54.1	76.6	93.8
2SLSMG^b												
25	0.027	0.189	0.200	0.837	0.460	0.222	5.4	5.9	4.4	26	41.9	63.3
50	-0.449	0.035	-0.306	0.506	0.223	0.122	7.1	5.2	5.6	35.2	61.5	81.6
100	-0.203	-0.524	-0.362	0.255	0.122	0.064	5.3	4.6	6.1	54.2	76.4	93.6

The DGP is the same as that for Table 5.1 and the number of instruments are 6.

Table 5.4 (Case B: High degree of endogeneity, Weak IVs) Bias, MSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	3.828	3.933	3.618	0.355	0.311	0.261	28.3	42.3	50.5	97	99.8	100.0
50	3.938	3.735	3.795	0.304	0.243	0.218	29.0	42.3	61.0	98.6	100.0	99.9
100	2.726	2.952	2.867	0.174	0.150	0.128	20.7	32.6	48.8	99.5	100.0	100.0
P-MA2SLSMG^b												
25	2.912	3.181	3.136	0.229	0.220	0.198	23.0	35.3	47.5	99	99.9	100.0
50	2.564	2.513	2.617	0.177	0.124	0.126	20.3	27.5	44.6	99.4	100.0	100.0
100	1.711	1.568	1.741	0.101	0.066	0.056	12.0	17.2	28.7	99.7	100.0	100.0
Ps-MA2SLSMG^b												
25	1.066	1.312	1.081	0.253	0.157	0.110	8.0	9.6	10.5	78.7	91.8	96.7
50	0.683	0.390	0.563	0.244	0.088	0.053	6.5	6.5	5.8	87.8	95.6	98.6
100	0.339	0.229	0.202	0.089	0.046	0.023	6.1	5.4	4.6	94.0	98.8	99.5
2SLSMG^b												
25	3.052	3.339	3.295	0.240	0.233	0.213	24.2	37.6	49.2	99.1	99.9	100.0
50	2.794	2.732	2.854	0.191	0.138	0.142	22.1	30.3	48.4	99.6	100.0	100.0
100	1.942	1.789	1.967	0.111	0.074	0.065	13.4	19.2	35.0	99.9	100.0	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.354	0.212	-0.049	1.000	0.442	0.193	7.0	5.8	4.8	25.6	42.8	64.7
50	-0.010	-0.114	-0.171	0.390	0.232	0.097	4.5	5.9	4.2	39.1	61.3	85.7
100	-0.234	-0.279	-0.369	0.254	0.129	0.063	5.2	4.9	4.7	55.6	78.7	96.0
P-MA2SLSMG^b												
25	0.251	0.277	0.211	0.703	0.398	0.193	4.4	5.8	4.7	29.1	45.5	68.8
50	-0.250	0.129	-0.139	0.454	0.207	0.107	7.2	5.2	6.1	37.5	65.4	86.0
100	-0.153	-0.450	-0.293	0.236	0.111	0.061	5.6	4.7	6.0	56.6	79.7	95.7
Ps-MA2SLSMG^b												
25	0.309	-0.077	0.161	1.282	0.839	0.517	4.9	5.1	4.6	22.5	34.1	55.5
50	-0.483	-0.521	-0.240	0.560	0.314	0.282	5.7	4.3	5.8	36.7	53.3	76.9
100	-0.074	-0.278	-0.290	0.309	0.144	0.073	6.2	5.3	5.1	54.1	77.1	93.1
2SLSMG^b												
25	0.301	0.291	0.206	0.703	0.399	0.192	4.3	5.6	4.8	29.1	46.2	68.9
50	-0.223	0.166	-0.103	0.451	0.208	0.107	6.9	5.2	5.5	37.3	65.6	86.1
100	-0.127	-0.415	-0.259	0.234	0.109	0.060	5.8	4.3	5.9	56.6	80.0	95.8

The DGP is the same as that for Table 5.1 except the number of instruments are 12.

Table 5.5 (Case C: Low degree of endogeneity, Strong IVs) Bias, MSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.151	-0.242	-0.347	0.183	0.094	0.048	6.3	5.7	6.0	72.1	91.6	97.7
50	-0.131	-0.169	-0.060	0.100	0.051	0.025	5.1	6.5	5.4	88.5	97.9	99.7
100	-0.149	-0.023	-0.092	0.067	0.035	0.017	5.5	5.4	4.8	96.6	99.6	100.0
P-MA2SLSMG^a												
25	-0.348	-0.352	-0.303	0.103	0.052	0.028	6.6	5.8	6.3	83.9	97.5	99.9
50	-0.154	-0.144	-0.250	0.074	0.036	0.020	6.6	5.8	5.3	93.9	99.9	100.0
100	-0.025	-0.125	-0.026	0.064	0.032	0.015	5.3	5.1	5.1	96.8	100.0	100.0
Ps-MA2SLSMG^a												
25	0.124	-0.060	-0.275	0.453	0.137	0.081	6.4	4.7	5.0	67.3	86.1	94.1
50	-0.004	0.015	0.008	0.126	0.058	0.029	6.7	5.5	4.9	85.2	97.2	99.3
100	0.044	-0.006	0.000	0.074	0.036	0.000	5.3	5.1	5.7	95.3	99.3	99.8
2SLSMG^a												
25	-0.348	-0.352	-0.303	0.103	0.052	0.028	6.6	5.8	6.3	83.9	97.5	99.9
50	-0.154	-0.144	-0.250	0.074	0.036	0.020	6.6	5.8	5.3	93.9	99.9	100.0
100	-0.025	-0.125	-0.026	0.064	0.032	0.015	5.3	5.1	5.1	96.8	100.0	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.116	0.321	0.304	1.894	0.836	0.452	5.5	5.1	5.2	18.2	26.5	43.5
50	0.490	0.335	0.358	0.505	0.229	0.128	7.0	4.5	5.8	40.5	59.5	84.2
100	0.361	0.209	0.238	0.183	0.096	0.050	4.4	4.3	5.8	69.8	89.1	99.0
P-MA2SLSMG^a												
25	0.895	0.286	0.410	1.365	0.672	0.358	6.0	5.3	6.2	21.5	31.9	47.8
50	0.231	0.292	0.269	0.410	0.198	0.110	5.9	4.5	5.1	41.1	63.2	88.2
100	0.222	0.088	0.145	0.186	0.084	0.042	5.6	4.7	4.9	70.1	92.9	99.5
Ps-MA2SLSMG^a												
25	0.272	-0.250	0.581	3.907	1.982	1.234	5.5	5.4	6.6	13.9	19.4	34.1
50	-0.089	0.037	0.167	0.793	0.350	0.215	5.6	4.4	6.4	33.7	52.8	75.0
100	-0.008	0.085	0.000	0.243	0.110	0.000	6.2	5.0	5.0	63.0	85.1	96.7
2SLSMG^a												
25	0.895	0.286	0.410	1.365	0.672	0.358	6.0	5.3	6.2	21.5	31.9	47.8
50	0.231	0.292	0.269	0.410	0.198	0.110	5.9	4.5	5.1	41.1	63.2	88.2
100	0.222	0.088	0.145	0.186	0.084	0.042	5.6	4.7	4.9	70.1	92.9	99.5

The DGP is the same as that for Table 5.1 and the number of instruments are 6.

Table 5.6 (Case C: Low degree of endogeneity, Strong IVs) Bias, MSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.365	-0.434	-0.563	0.105	0.055	0.034	7.1	6.1	8.6	85.2	97.1	100.0
50	0.024	-0.065	-0.033	0.082	0.041	0.022	5.8	6.7	5.9	93.1	99.4	100.0
100	-0.051	0.013	-0.002	0.069	0.034	0.017	6.1	6.2	4.8	96.1	99.5	100.0
P-MA2SLSMG^b												
25	-0.358	-0.436	-0.393	0.082	0.043	0.023	6.2	7.4	7.1	90.3	99.2	100.0
50	-0.170	-0.150	-0.224	0.066	0.032	0.018	6.5	5.3	5.8	96.3	100.0	100.0
100	-0.009	-0.116	-0.026	0.059	0.031	0.014	5.6	5.8	5.5	97.2	100.0	100.0
Ps-MA2SLSMG^b												
25	-0.153	-0.289	-0.452	0.171	0.115	0.068	5.8	4.5	5.6	71.4	90.6	95.6
50	-0.144	-0.203	-0.185	0.104	0.049	0.025	6.1	6.0	5.5	89.5	98.4	99.5
100	-0.116	-0.134	0.000	0.071	0.034	0.000	5.7	5.2	5.6	95.9	99.5	99.7
2SLSMG^b												
25	-0.366	-0.449	-0.407	0.081	0.043	0.023	6.3	7.4	7.3	90.8	99.3	100.0
50	-0.183	-0.174	-0.238	0.066	0.032	0.018	6.2	4.9	5.8	96.3	100.0	100.0
100	-0.027	-0.129	-0.043	0.059	0.031	0.014	5.8	6.0	5.4	97.2	100.0	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	0.114	0.607	0.662	1.236	0.546	0.283	7.9	4.9	5.1	22.6	34.5	54.9
50	0.327	0.228	0.327	0.412	0.194	0.101	6.5	4.3	6.1	43.1	64.3	88.2
100	0.253	0.155	0.111	0.176	0.087	0.046	4.6	4.0	4.7	70.5	90	99.3
P-MA2SLSMG^b												
25	0.674	0.350	0.621	1.079	0.541	0.272	6	5.2	6.1	24.7	37.8	58.2
50	0.247	0.309	0.206	0.351	0.174	0.097	6.5	5.5	5.3	42.2	68.4	90.8
100	0.206	0.092	0.148	0.167	0.079	0.039	5.2	5.3	4.8	72.3	92.9	99.5
Ps-MA2SLSMG^b												
25	0.442	0.517	1.057	2.664	1.281	0.910	5.1	4.6	5.8	16.4	23.5	39.3
50	0.125	0.459	0.543	0.721	0.304	0.180	5.2	4.5	5.4	38.7	60.2	82.8
100	0.311	0.268	0.000	0.218	0.099	0.000	5.9	5.0	4.0	68.1	88.9	98.5
2SLSMG^b												
25	0.688	0.395	0.650	1.074	0.532	0.268	6.3	4.7	5.8	24.9	38.5	59.0
50	0.264	0.354	0.226	0.349	0.173	0.097	6.6	5.5	5.2	43.0	68.9	91.2
100	0.240	0.114	0.176	0.166	0.079	0.039	5.0	5.1	4.5	73.0	93.2	99.6

The DGP is the same as that for Table 5.1 except the number of instruments are 12.

Table 5.7 (Case D: High degree of endogeneity, Strong IVs) Bias, MSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	3.099	3.258	3.021	0.424	0.338	0.235	14.6	22.2	27.8	76.8	91.4	97.5
50	3.347	3.202	3.058	0.386	0.243	0.179	18.7	23.4	33.1	88.0	97.2	99.2
100	1.936	2.001	2.107	0.178	0.121	0.094	11.4	15.6	26.0	92.4	98.8	99.8
P-MA2SLSMG^a												
25	2.385	2.325	2.608	0.232	0.161	0.148	14.5	21.9	32.0	92.1	99.1	100.0
50	2.134	1.955	2.130	0.180	0.105	0.100	14.5	15.5	27.7	93.9	99.5	100.0
100	1.325	1.275	1.472	0.114	0.064	0.050	10.7	9.8	19.3	95.5	99.8	100.0
Ps-MA2SLSMG^a												
25	1.910	2.035	1.851	0.397	0.266	0.155	10.0	12.5	13.7	74.8	87.9	95.1
50	1.429	1.438	1.454	0.231	0.180	0.119	8.7	9.5	13.1	83.3	92.7	97.7
100	0.916	0.824	0.807	0.155	0.100	0.050	7.6	6.7	9.0	88.8	96.0	99.1
2SLSMG^a												
25	2.385	2.325	2.608	0.232	0.161	0.148	14.5	21.9	32.0	92.1	99.1	100.0
50	2.134	1.955	2.130	0.180	0.105	0.100	14.5	15.5	27.7	93.9	99.5	100.0
100	1.325	1.275	1.472	0.114	0.064	0.050	10.7	9.8	19.3	95.5	99.8	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-5.349	-3.727	-3.043	9.332	4.398	2.076	7.3	6.1	5.7	8.1	7.9	11.0
50	-4.440	-4.395	-4.569	4.018	2.221	1.253	7.6	7.1	7.3	8.7	12.6	14.6
100	-3.065	-3.457	-3.653	2.119	1.137	0.645	6.8	7.5	9.5	12.1	19.5	25.4
P-MA2SLSMG^a												
25	-2.163	-3.503	-3.766	5.497	3.135	1.704	6.3	6.4	6.0	9.5	11.5	14.8
50	-4.064	-2.688	-3.734	3.313	1.607	0.991	5.3	6.1	7.0	9.5	14.2	18.4
100	-2.361	-2.734	-3.056	1.664	0.981	0.487	5.9	7.5	6.5	13.3	20.3	27.3
Ps-MA2SLSMG^a												
25	-1.346	-2.933	-1.562	12.377	6.287	3.214	4.8	4.6	5.1	8.0	8.3	12.7
50	-3.407	-2.662	-2.803	5.372	3.307	1.506	6.3	5.7	6.9	9.7	12.4	17.6
100	-1.740	-2.363	-1.591	2.480	1.399	0.618	5.9	6.3	4.9	13.7	19.0	30.6
2SLSMG^a												
25	-2.163	-3.503	-3.766	5.497	3.135	1.704	6.3	6.4	6.0	9.5	11.5	14.8
50	-4.064	-2.688	-3.734	3.313	1.607	0.991	5.3	6.1	7.0	9.5	14.2	18.4
100	-2.361	-2.734	-3.056	1.664	0.981	0.487	5.9	7.5	6.5	13.3	20.3	27.3

The DGP is the same as that for Table 5.1 and the number of instruments are 6.

Table 5.8 (Case D: High degree of endogeneity, Strong IVs) Bias, MSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.5, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	3.405	3.467	3.136	0.293	0.248	0.195	23.9	35.1	45.3	95.5	99.9	100
50	4.645	4.382	4.347	0.384	0.306	0.266	34.3	49.6	66.9	98.8	100	100
100	3.702	3.766	3.809	0.261	0.216	0.205	26.8	44.5	63.4	99.6	100	100
P-MA2SLSMG^b												
25	2.880	2.976	3.091	0.218	0.190	0.172	21.0	33.8	47.6	98.2	100	100
50	3.425	3.188	3.398	0.239	0.171	0.183	26.9	36.3	56.9	99.8	100	100
100	2.625	2.551	2.720	0.158	0.114	0.108	20.0	29.0	50.3	99.4	100	100
Ps-MA2SLSMG^b												
25	1.934	2.042	1.801	0.388	0.251	0.151	11.2	12.5	14.2	75	89.2	95.1
50	1.341	1.406	1.429	0.216	0.161	0.098	8.3	9.8	12.1	83.7	93.7	98.3
100	0.948	0.823	0.822	0.151	0.096	0.046	8.0	7.5	9.6	90.4	96.4	99.3
2SLSMG^b												
25	2.952	3.069	3.170	0.223	0.197	0.179	21.6	35.4	49.0	98.7	100.0	100.0
50	3.593	3.357	3.572	0.252	0.184	0.198	27.9	38.8	59.7	99.9	100.0	100.0
100	2.840	2.760	2.927	0.171	0.127	0.122	22.2	32.8	54.8	99.6	100.0	100.0

Results for β_1.

	Bias(×100)			MSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-4.859	-3.458	-3.125	5.585	2.521	1.250	7.9	5.5	6.1	9.7	11.2	15.2
50	-5.657	-5.736	-5.563	2.930	1.835	1.068	8.2	9.6	11.5	7.9	13.5	14.7
100	-5.207	-5.618	-5.709	1.789	1.139	0.745	7.8	10.5	18.0	11	15.3	20.4
P-MA2SLSMG^b												
25	-2.521	-3.647	-4.109	4.094	2.379	1.362	7.0	7.2	6.8	9.8	12	16.3
50	-5.646	-4.120	-4.909	2.785	1.407	0.982	7.4	7.8	10.4	9.2	14.3	19
100	-3.805	-4.285	-4.518	1.498	0.920	0.566	6.7	9.7	10.9	11.9	17.5	24.1
Ps-MA2SLSMG^b												
25	-1.975	-3.021	-1.484	12.084	6.242	3.233	4.6	5.1	5.0	7.3	8.0	13.0
50	-3.317	-2.504	-2.626	5.274	3.102	1.409	6.1	5.3	6.4	10.1	12.1	18.4
100	-1.879	-2.400	-1.575	2.462	1.358	0.613	6.2	6.2	5.2	13.7	19.2	30.7
2SLSMG^b												
25	-2.592	-3.719	-4.173	4.068	2.361	1.364	6.9	7.0	7.3	9.7	11.6	16.3
50	-5.814	-4.347	-5.082	2.767	1.406	0.996	7.4	8.0	10.3	9.4	13.9	19.0
100	-4.085	-4.556	-4.765	1.478	0.932	0.583	6.7	9.8	12.3	11.4	17.2	24.0

The DGP is the same as that for Table 5.1 except the number of instruments are 12.

Chapter 6

Conclusion

In this thesis, we have developed estimation and inferential methods for linear dynamic panel data models with unobserved additive and interactive effects for short panels and large panels.

In Chapter 2, after a review of Nickell (1981) bias, the conventional instrumental variable (IV) of Anderson and Hsiao (1981, 1982) and generalised method of moments (GMM) estimation methods of Arellano and Bond (1991) and Blundell and Bond (1998) for short dynamic panel data models, the transformed maximum likelihood (TML) estimator of Hayakawa and Pesaran (2015a), the bias-corrected method of moments (BMM) estimator of Chudik and Pesaran (2017), and the double filter instrumental variable (DFIV) estimator of Hayakawa et al. (2019) are discussed. To assess the finite sample behaviour of these estimators, a Monte Carlo experiment is conduced. We particularly investigate the effects of different initial conditions. We find that in terms of the bias and the RMSE, the TML estimator and the BMM estimator perform better for all the designs. The size of the t-tests based on the TML estimator and the BMM estimator is close to the nominal level. The DFIV estimator is the least efficient among these three estimators in our design.

In Chapter 3, we have extended the quasi maximum likelihood (QML) estimator for short dynamic panel data models with interactive effects proposed by Hayakawa et al. (2021) to the case where the errors are cross-sectionally heteroskedastic, using a similar approach discussed in Hayakawa and Pesaran (2015a). Extending the error to cross-sectionally heteroskedastic is important in empirical studies, because the heteroskedasticity is generic and standard in real world (See Hansen (2020)). We apply our method to an annual panel data set which consists of 16,502 US firms over the period from 1960 to 2017 to empirically assess the trade off theory (Graham and Harvey (2001)). In particular We find that the speed of adjustment (SOA) of the US firms is between 74% and 86% from 1960 to 1999, whilst it decreases around 32% from 2000 to 2007, followed by around 60% after 2008. The value of SOA we have found is higher than that in other existing research (Ozkan (2001), Fama and French (2002), Kayhan and Titman (2007), Flannery and Rangan (2006), Lemmon et al. (2008) and Dang et al. (2014)), which suggests the importance of controlling unobserved interactive effects. As a future research agenda, it would be intriguing to extend the current framework to accommodate an asymmetric partial adjustment model. Similarly, it would be illuminating to extend to permit heterogeneous dynamic partial adjustment capital structure.

In Chapter 4, we extended the QML estimator of Hayakawa et al. (2021) for the estimation of short panel vector autoregressive (VAR) models with interactive effects. Holtz-Eakin et al. (1988) and Binder et al. (2005), among others, consider estimation and inference on cross-sectionally independent short panel VAR models. The finite sample evidence provided by Juodis (2018) shows that the TML estimator outperforms the GMM based estimators. A few studies focus on the estimation of panel VAR models with cross section dependence (Mutl (2009) and Huang (2008)), however, these methods are not asymptotically justified for the models with interactive effects. In the Monte Carlo results, we show that the proposed QML estimator performs reasonably well. But, in line with the theoretical results in Hayakawa et al. (2021), we conjecture that global identification of the QML estimator is not possible. Further research into the identification conditions for the QML estimator in the multivariate case would be of future interest. Also it would be of interest to consider quasi maximum likelihood estimation of large scale panel VAR models with sparse coefficients in the presence of cross-sectional dependence. Furthermore, it would also be of interest to extend the panel VAR model to panel error correction setting and investigate nonstationarity under cross-sectional dependence.

Based on Kuersteiner and Okui (2010), in Chapter 5 we have proposed a method to choose a set of weights to average instrumental variable (IV) estimators proposed by Norkutė et al. (2021) for large dynamic panel data models with interactive effects. Norkutė et al. (2021) provides the IV estimators that use lagged defactored covariates in the model as instrumental variables. Therefore, this IV estimator does not need to search for instruments outside the model. However, when T increases, the number of valid instruments rises. Norkutė et al. (2021) is silent about how to choose a set of instruments to avoid the problems due to having too many instruments or weak instruments. To tackle this problem, we take the approach proposed by Kuersteiner and Okui (2010). In particular, they propose a procedure to choose a set of weights to average the IV estimators, which are computed using different subset of available instrumental variables. The weights are chosen so that the mean squared error of the average of the IV estimators is theoretically minimised. The estimator is called model average two stage least square (2SLS) estimator, and we apply this approach to the IV estimator of Norkutė et al. (2021). The empirical evidence shows that the proposed model average 2SLS estimator can reduce the bias but not the mean squared errors (MSE). As a future research agenda, we may apply another instruments selection method, such as Windmeijer et al. (2019), to see whether it can improve MSE.

Appendix

Appendix A

Appendix to Chapter 2

A.1 Bias of LS estimator

Given t, the LS estimator as

$$\operatorname*{plim}_{N\to\infty}\left(\hat{\phi}_t-\phi\right)=\frac{\operatorname{plim}_{N\to\infty}1/N\sum_{i=1}^{N}(y_{i,t-1}-\bar{y}_{i,-1})(u_{i,t}-\bar{u}_i)}{\operatorname{plim}_{N\to\infty}1/N\sum_{i=1}^{N}(y_{i,t-1}-\bar{y}_{i,-1})^2}. \tag{A.1}$$

Suppose

$$A_t=\operatorname*{plim}_{N\to\infty}1/N\sum_{i=1}^{N}(y_{i,t-1}-\bar{y}_{i,-1})(u_{i,t}-\bar{u}_i)$$

and by taking expectation across i, we have

$$\begin{aligned}A_t=&E_i(y_{i,t-1}-\bar{y}_{i,-1})(u_{i,t}-\bar{u}_i)\\=&E_i(y_{i,t-1}u_{i,t})-E_i(y_{i,t-1}\bar{u}_i)-E_i(\bar{y}_{i,-1}u_{i,t})+E(\bar{y}_{i,-1}\bar{u}_i),\end{aligned} \tag{A.2}$$

where $E_i(y_{i,t-1}u_{i,t})=0$. By setting stationarity assumption in the AR(1) model, we have

$$y_{i,t}=\frac{\alpha_i}{(1-\phi)}+\sum_{j=0}^{\infty}\phi^j u_{i,t-j}. \tag{A.3}$$

Then substitute (A.3) to (A.2), we have

$$\begin{aligned}A_t=&-E_i\left\{(\sum_{j=0}^{\infty}u_{i,t-j-1}\phi^j)(\frac{1}{T}\sum_{s=1}^{T}u_{i,t})\right\}-E_i\left\{\frac{u_{i,t}}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}u_{is-j-1}\phi^j\right\}+\\&E_i\left\{(\frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}u_{i,t-j-1}\phi^j)(\frac{1}{T}\sum_{s=1}^{T}u_{i,t})\right\}.\end{aligned} \tag{A.4}$$

And we know $E_i u_{i,t} = E_i u_{i,t}\alpha_i = 0$ and $E_i u_{i,t}^2 = \sigma_u^2$, then

$$
\begin{aligned}
A_t =& -\frac{1}{T}E_i\left\{\left(u_{i,t-1} + u_{i,t-2}\phi^1 + u_{i,t-3}\phi^2 + \ldots\right)\left(u_{i1} + \ldots + u_{i,t-1} + u_{i,t} + \ldots + u_{i,T}\right)\right\} - \\
&\frac{1}{T}E_i\left\{u_{i,t}\sum_{s=1}^{T}\left(u_{is-1}\phi^0 + u_{is-2}\phi^1 + \ldots + u_{is-t-1}\phi^t + \ldots + u_{is-T-1}\phi^T + \ldots\right)\right\} + \\
&\frac{1}{T}E_i\left\{\left(\sum_{s=1}^{T}u_{is-1}\phi^0 + \sum_{s=1}^{T}u_{is-2}\phi^1 + \ldots + \sum_{s=1}^{T}u_{is-t-1}\phi^t + \ldots + \sum_{s=1}^{T}u_{is-T-1}\phi^T + \ldots\right)\left(\frac{1}{T}\sum_{s=1}^{T}u_{is}\right.\right. \\
=& -\frac{\sigma_u^2}{T}\frac{(1-\phi^{t-1})}{1-\phi} - \frac{\sigma_u^2}{T}\frac{(1-\phi^{T-t})}{(1-\phi)} + \frac{\sigma_u^2}{T}\left[\frac{1}{1-\phi} - \frac{1}{T}\frac{(1-\phi^T)}{(1-\phi)^2}\right] \\
=& -\frac{\sigma_u^2}{T(1-\phi)}\left\{1 - \phi^{t-1} - \phi^{T-t} + \frac{1}{T}\frac{(1-\phi^T)}{(1-\phi)}\right\}.
\end{aligned}
\tag{A.5}
$$

For B_t, we have

$$
\begin{aligned}
B_t =& E_i\left(y_{i,t-1} - y_{i.-1}\right)^2 \\
=& E_i\left(\sum_{j=0}^{\infty}\phi^j u_{i,t-j-1} - \frac{1}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}\phi^j u_{is-j-1}\right)^2 \\
=& E_i\left(\sum_{j=0}^{\infty}\phi^j u_{i,t-j-1}\right)^2 - \frac{2}{T}E_i\left(\sum_{j=0}^{\infty}\phi^j u_{i,t-j-1}\right)\left(\sum_{s=1}^{\infty}\sum_{j=0}^{\infty}\phi^j u_{is-j-1}\right) + \frac{1}{T^2}E_i\left(\sum_{s=1}^{T}\sum_{j=0}^{\infty}u_{is-j-1}\right)^2 \\
=& E_i\left(\phi^o u_{i,t-1} + \phi^1 u_{i,t-2} + \ldots\right)^2 - \\
&\frac{2}{T}E_i\left(\left(\phi^o u_{i,t-1} + \phi^1 u_{i,t-2} + \ldots\right)\left(\phi^0\sum_{s=1}^{T}u_{is-1} + \phi^1\sum_{s=1}^{T}u_{is-2} + \ldots\right)\right) \\
=& \frac{\sigma_u^2}{1-\phi^2} - \frac{2\sigma_u^2}{T(1-\phi^2)}\left\{\frac{1-\phi^t}{1-\phi} + \phi\frac{(1-\phi^{T-t})}{1-\phi}\right\} + \frac{\sigma_u^2}{T(1-\phi)^2}\left\{1 - \frac{2\phi(1-\phi^T)}{T\left(1-\phi^2\right)}\right\} \\
=& \frac{\sigma_u^2}{1-\phi^2}\left(1 - \frac{1}{T}\right) + \frac{\sigma_u^2}{T(1-\phi^2)}\left[1 - 2\left(\frac{1-\phi^t+\phi-\phi^{T-t+1}}{1-\phi}\right)\right] + \\
&\frac{2\phi}{1-\phi^2}\left[\frac{\sigma_u^2\left(1-\phi^2\right)}{2T\left(1-\phi^2\right)} - \frac{\sigma_u^2\left(1-\phi^T\right)}{T^2\left(1-\phi\right)^2}\right] \\
=& \frac{\sigma_u^2}{1-\phi^2}\left(1 - \frac{1}{T}\right) + \\
&\frac{2\phi}{1-\phi^2}\left\{-\frac{\sigma_u^2}{T(1-\phi)}\left(\frac{-(1-\phi)}{2\phi} + \frac{1-\phi t+\phi-\phi^{T-t+1}}{\phi} - \frac{(1-\phi^2)}{2\phi(1-\phi)} + \frac{(1-\phi^T)}{T(1-\phi)}\right)\right\} \\
=& \frac{\sigma_u^2}{1-\phi^2}\left(1 - \frac{1}{T}\right) + \frac{2\phi}{1-\phi^2}A_t.
\end{aligned}
\tag{A.6}
$$

Then we can know the bias of within group estimator is

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}(\hat{\phi}_t-\phi) &= \frac{A_t}{B_t} = \left\{\frac{\frac{\sigma_u^2}{1-\phi^2}\left(1-\frac{1}{T}\right)+\frac{2\phi}{1-\phi^2}A_t}{A_t}\right\}^{-1} \\
&= \left\{\frac{\frac{\sigma_u^2}{1-\phi}\left(\frac{T-1}{T}\right)+\frac{2\phi}{1-\phi^2}-\frac{\sigma_u^2}{T(1-\phi)}\left[1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{\left(1-\phi^T\right)}{(1-\phi)}\right]}{-\frac{\sigma_u^2}{1-\phi^2}\left[1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{\left(1-\phi^T\right)}{(1-\phi)}\right]}\right\}^{-1} \\
&= \left\{\frac{2\phi}{1-\phi^2}-\left[\frac{1+\phi}{T-1}(1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{1-\phi^T}{1-\phi})\right]^{-1}\right\}^{-1} \\
&= \left\{\frac{2\phi}{(1-\phi)(1+\phi)}-\frac{T-1}{1+\phi}\left[1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{\left(1-\phi^T\right)}{(1-\phi)}\right]^{-1}\right\}^{-1} \\
&= -\frac{1+\phi}{T-1}\left(1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{\left(1-\phi^T\right)}{(1-\phi)}\right)\times \\
&\left\{1-\frac{2\phi}{(1-\phi)(T-1)}\left[1-\phi^{t-1}-\phi^{T-t}+\frac{1}{T}\frac{\left(1-\phi^T\right)}{(1-\phi)}\right]\right\}^{-1}
\end{aligned}
\tag{A.7}
$$

We can observe that $y_{i,t-1}$ is correlated with $u_{i\cdot}$. Therefore, LS estimator is inconsistent and the bias is of order T^{-1}. Apart from that, within group estimator exhibits a downward bias when $\phi > 0$.

A.2 Hsiao et al. (2002) Transformed likelihood estimation in standard dynamic panel data models with fixed effects

Consider the model include regressors, then the model can be written as

$$y_{i,t} = \phi y_{i,t-1} + \beta x_{i,t} + \alpha_i + \mu_{i,t}, t = 1, \ldots, T; i = 1, \ldots, N. \tag{A.8}$$

We assume y_{i0} and x_{i0} are available. By taking first difference, the model can be expressed as

$$\Delta y_{i,t} = \phi \Delta y_{i,t-1} + \beta \Delta x_{i,t} + \Delta \mu_{i,t}, \;\; t = 2, 3, \ldots, T; i = 1, \ldots, N \tag{A.9}$$

where $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$, $\Delta x_{i,t} = x_{i,t} - x_{i,t-1}$ and $\Delta u_{i,t} = u_{i,t} - u_{i,t-1}$. However, we cannot find $\Delta y_{i,1}$ because $y_{i,-1}$ is not observed. Therefore, starting from $\Delta y_{i,-m+1}$

and by continuous substitution, $\Delta y_{i,1}$ can be expressed as

$$\Delta y_{i,1} = \phi^m \Delta y_{i,-m+1} + \beta \sum_{j=0}^{m-1} \phi^j \Delta x_{i,1-j} + \sum_{j=0}^{m-1} \phi^j \Delta u_{i,1-j},\ i = 1...N. \tag{A.10}$$

Because $\Delta x_{i,1-j}$, j=1,2,... are unobserved, the mean of $\Delta y_{i,1}$ conditional on $\Delta y_{i,-m+1}$ and $\Delta u_{i,1-j}$ for j=0,1,2,... as

$$\begin{aligned}\eta_{i,1} =& E(\Delta y_{i,1} | \Delta y_{i,-m+1}, \Delta x_{i,1}, \Delta x_{i,0}, ...) \\ =& \phi^m \Delta y_{i,-m+1} + \beta \sum_{j=0}^{m-1} \phi^j \Delta x_{i,1-j}, i = 1, 2, ..., N.\end{aligned} \tag{A.11}$$

Because $\eta_{i,1}$ is a free parameter, we will encounter incidental parameter problem. To deal with this problem, $\eta_{i,1}$ should be a function of a finite number parameters. Meanwhile, $x_{i,t}$ should be generated by

$$\begin{aligned}x_{i,t} =& u_i + gt + \sum_{j=0}^{\infty} a_j \epsilon_{i,t-j},\ \sum_{j=0}^{\infty} |a_j| < \infty, \\ & or \\ \Delta x_{i,t} =& g + \sum_{j=0}^{\infty} d_j \epsilon_{i,t-j},\ \sum_{j=0}^{\infty} |d_j| < \infty,\end{aligned} \tag{A.12}$$

where $\epsilon_{i,t} \sim i.i.d.(0, \sigma_\epsilon^2)$. If the generating processes $x_{i,t}$ follow random walks with drift and the drift parameters are different in different individual or $x_{i,t}$ have different trend in different individual. Thus, the transformed likelihood method will confront the incidental parameters problem. we continue assume that $|\phi| < 1$ and $m \to \infty$, or m is finite and $E(\Delta y_{i,t-1} | \Delta x_{i,1}, \Delta x_{i,2}, ..., \Delta x_{i,T})$ is the same for all i.

Hsiao et al. (2002) distinguish two case: strictly exogenous and weakly exogenous. When the disturbance term $u_{i,t}$ is independent with current, lagged and future value of $x_{i,t}$, regressors are strictly exogenous. When the disturbance term $u_{i,t}$ is independent with current and lagged value of $x_{i,t}$, regressors are weakly exogenous.

For the model with strict exogenous regressors, the joint probability density function of $\Delta \boldsymbol{y}_i$ condition on $\Delta \boldsymbol{x}_i$ can be written as

$$\begin{aligned}f(\Delta \boldsymbol{y_i} | \Delta \boldsymbol{x_i}) =& f(\Delta y_{i,T} | \Delta y_{i,T-1}, ..., \Delta y_{i1}, \Delta \boldsymbol{x_i}) \\ & f(\Delta y_{i,T-1} | \Delta y_{i,T-2}, ..., \Delta y_{i,1}, \Delta \boldsymbol{x_i}) \\ & f(\Delta y_{i,T-2} | \Delta y_{i,T-3}, ..., \Delta y_{i,1}, \Delta x_i)... \\ & f(\Delta y_{i,2} | \Delta y_{i1}, \Delta \boldsymbol{x}_i) f(\Delta y_{i,1} | \Delta \boldsymbol{x_i}).\end{aligned} \tag{A.13}$$

And, $\eta_{i,1} = E(\Delta y_{i,1} | \Delta y_{i,-m+1}, \Delta x_{i,1}, \Delta x_{i,0}, ...)$ can be expressed as

$$\eta_{i1} = b^* + \beta \Delta x_{i1} + \beta \Sigma_{j=1}^{m-1} \phi^j E(\Delta x_{i,1-j} | \Delta x_i) + q_{i1}. \tag{A.14}$$

Under assumption, $b^* = b = 0$ and $q_{i,1} = \eta_{i,1} - E(\eta_{i,1}|\Delta x_i)$. From x generating process, we know

$$\Delta x_{i,t} = g + \sum_{j=0}^{\infty} d_j^* \varepsilon_{i,t-j}, \ \sum_{j=0}^{\infty} |d_j^*| < \infty, \tag{A.15}$$

If m is finite, we have

$$E(\Delta x_{i,1-j}|\Delta \boldsymbol{x_i}) = b_j + \pi_j^{'} \Delta \boldsymbol{x_i}. \tag{A.16}$$

Therefore, $f(\Delta y_{i,1}|\Delta \boldsymbol{x_i})$ can be written as

$$\Delta y_{i1} = b^* + \boldsymbol{\pi}^{'} \Delta \boldsymbol{x}_i + \nu_{i1}, \tag{A.17}$$

where π is a unknown $T \times 1$ vector and $\nu_{i1} = q_{i,1} + \sum_{j=0}^{m-1} \phi^j \Delta u_{i,1-j}$.

For model include weak exogenous regressors, $\Delta \boldsymbol{w}_{i,t} = (\Delta y_{i,t}, \Delta x_{i,t})$. The joint pdf of $(\Delta \boldsymbol{w}_{i,1}, \ldots, \Delta \boldsymbol{w}_{i,T})$ is

$$f(\Delta \boldsymbol{w}_{i,T}|\Im_{i,T-1}) f(\Delta \boldsymbol{w}_{i,T-1}|\Im_{i,T-2}), ..., f(\Delta \boldsymbol{w}_{i,2}|\Im_{i,1}) f(\Delta \boldsymbol{w}_{i,1}|\Im_{i,0}), \tag{A.18}$$

where $\Im_{i,t} = (\Delta \boldsymbol{w}_{i,t}, \Delta \boldsymbol{w}_{i,t-1}, ..., \Delta \boldsymbol{w}_{i,1}), t = 1, 2, ..., T-1$, and $\Im_{i,0}$ normalized at unity. Thus, the likelihood function can be expressed as

$$\prod_{i=1}^{N} \prod_{t=1}^{T} f(\Delta y_{i,t}, |\ \Im_{i,t-1}, \Delta x_{i,t}). \tag{A.19}$$

As we know

$$E(\Delta y_{i,1}|\Delta x_{i,1}) = b^* + \beta \sum_{j=0}^{m-1} \phi^j E(\Delta x_{i,1-j}|\Delta x_{i1}) + \sum_{j=0}^{m-1} \phi^j E(\Delta u_{i,1-j}|\Delta x_{i1}). \tag{A.20}$$

By using projection, $E(\Delta x_{i,1-j}) = \psi_{j0} + \psi_{j1} \Delta x_{i,1}$, and $E(\Delta u_{i,1-j}) = \varphi_{j0} + \varphi_{j1} \Delta x_{i,1}$. The $\psi_{j0}, \psi_{j1}, \varphi_{j0}, \varphi_{j1}$ can derived from the parameters of the joint probability density function. Under above result, $E(\Delta y_{i,1}|\Delta x_{i,1}) = \delta_{j0} + \delta \Delta x_{i,1}$, where δ_{j0}, δ_{j1} can be treat as free parameters. Thus,

$$\Delta y_{i,1} = \delta_0 + \delta_1 \Delta x_{i,1} + \xi_{i,1}, \tag{A.21}$$

where

$$\xi_{i,1} = \beta \sum_{j=1}^{m-1} \phi^j (\Delta x_{i,1-j} - \psi_{j0} - \psi_{j1} \Delta x_{i1}) + \sum_{j=0}^{m-1} \phi^j (\Delta u_{i,1-j} - \varphi_{j,0} - \varphi_{j,1} \Delta x_{i,1}). \tag{A.22}$$

Therefore, the likelihood function can be expressed as

$$(2\pi)^{\frac{-NT}{2}} |\boldsymbol{\Omega}|^{\frac{-N}{2}} \exp\{-\frac{1}{2} \sum_{i=1}^{N} \Delta \boldsymbol{u}_i^{*'} \boldsymbol{\Omega}^{-1} \Delta \boldsymbol{u}_i^*\}, \tag{A.23}$$

where

$$\Delta\boldsymbol{u}_i^* = [\Delta y_{i,1} - b^* - \boldsymbol{\pi}^{*'}\Delta\boldsymbol{x}_i, \Delta y_{i,2} - \phi\Delta y_{i,1} - \beta\Delta x_{i,2}, \ldots, \Delta y_{i,T} - \phi\Delta y_{i,T-1} - \beta\Delta x_{i,T}]' \tag{A.24}$$

By minimizing $\sum_{i=1}^{N} \Delta\boldsymbol{u}_i^{*'}\boldsymbol{\Omega}^{-1}\Delta\boldsymbol{u}_i^*$, the MLE of (ϕ, β) is given by

$$\begin{aligned}\begin{pmatrix}\hat{\phi}\\ \hat{\beta}\end{pmatrix} =& [(\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\tilde{\boldsymbol{G}}_i) - (\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\Delta\tilde{\boldsymbol{X}}_i^*)(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}}_i^{*'}\Delta\tilde{\boldsymbol{X}}_i^*)^{-1}(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}}_i^{*'}\tilde{\boldsymbol{G}}_i)]^{-1}\times\\ & [(\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\Delta\tilde{\boldsymbol{y}}_i) - (\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\Delta\tilde{\boldsymbol{X}}_i^*)(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}}_i^{*'}\Delta\tilde{\boldsymbol{X}}_i^*)^{-1}(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}}_i^{*'}\Delta\tilde{\boldsymbol{y}}_i)],\end{aligned} \tag{A.25}$$

where $\Delta\tilde{\boldsymbol{X}}_i^* = (\boldsymbol{p}_1, \boldsymbol{p}_1\Delta\boldsymbol{x}_i')$, $\tilde{\boldsymbol{G}}_i = \left(\Delta\tilde{\boldsymbol{y}}_{i,-1}, \Delta\tilde{\boldsymbol{x}}_{\boldsymbol{i}}\right)$ and $\Delta\tilde{\boldsymbol{y}}_i = \boldsymbol{P}\Delta\boldsymbol{y}_i$ with $\Delta\tilde{\boldsymbol{y}}_{i,-1} = \boldsymbol{P}\left(0, \Delta y_{i,1}, \ldots, \Delta y_{i,T-1}\right)'$ and $\Delta\tilde{\boldsymbol{x}}_i = \boldsymbol{P}\left(0, \Delta x_{i,2}, \ldots, \Delta x_{i,T}\right)'$, $\boldsymbol{p}_1$ is the first column of $\boldsymbol{P}$ and $\boldsymbol{P} = \boldsymbol{D}^{-1/2}\boldsymbol{C}$. $\boldsymbol{C}$ and $\boldsymbol{D}$ is defined in Eqs. (3.5) and Eqs (3.6) of Hsiao et al. (2002), respectively.

A.3 Transformed likelihood estimators

$\boldsymbol{\delta} = \binom{b^*}{\boldsymbol{\pi}}$, $\boldsymbol{\theta} = \binom{\phi}{\beta}$, $\boldsymbol{v_i} = \boldsymbol{P}\Delta\boldsymbol{u_i}^* = \Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}} - \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*\boldsymbol{\delta} - \tilde{\boldsymbol{G}}_{\boldsymbol{i}}\boldsymbol{\theta}$ *and* $\boldsymbol{\psi}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \sum_{i=1}^{N}\boldsymbol{v_i}'\boldsymbol{v_i}$. Differentiate $\boldsymbol{\psi}(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and equal to zero

$$\frac{\partial\boldsymbol{\psi}(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -2\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}'(\Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}} - \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*\boldsymbol{\delta} - \tilde{\boldsymbol{G}}_{\boldsymbol{i}}\boldsymbol{\theta}) = 0. \tag{A.26}$$

$$(\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}'\tilde{\boldsymbol{G}}_{\boldsymbol{i}})\hat{\boldsymbol{\theta}} = \sum_{i=1}^{N}\tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}'(\Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}} - \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*\hat{\boldsymbol{\delta}}), \tag{A.27}$$

then we get

$$\hat{\boldsymbol{\theta}} = (\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}'\tilde{\boldsymbol{G}}_i)^{-1}(\sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\Delta\tilde{\boldsymbol{y}}_i - \sum_{i=1}^{N}\tilde{\boldsymbol{G}}_i^{'}\Delta\tilde{\boldsymbol{X}}_i^*\hat{\boldsymbol{\delta}}). \tag{A.28}$$

Differentiate $\boldsymbol{\psi}(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\delta}$ and equal to zero

$$\frac{\partial\boldsymbol{\psi}(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial\boldsymbol{\delta}} = -2\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}_{\boldsymbol{i}}}^{*'}(\Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}} - \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}^*\boldsymbol{\delta} - \tilde{\boldsymbol{G}}_{\boldsymbol{i}}\boldsymbol{\theta}) = 0. \tag{A.29}$$

$$(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}_{\boldsymbol{i}}}^{*'}\Delta\tilde{\boldsymbol{X}_{\boldsymbol{i}}}^*)\hat{\boldsymbol{\delta}} = \sum_{i=1}^{N}\Delta\tilde{\boldsymbol{X}_{\boldsymbol{i}}}^{*'}(\Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}} - \tilde{\boldsymbol{G}}_{\boldsymbol{i}}\hat{\boldsymbol{\theta}}), \tag{A.30}$$

then we get

$$\hat{\boldsymbol{\delta}} = (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*})^{-1} (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}_i} - \sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}_i}\hat{\boldsymbol{\theta}}). \tag{A.31}$$

Substitute (A.28) to (A.30), then

$$\begin{aligned}&(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*})\hat{\boldsymbol{\delta}} = \\ &[(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}_i}) - (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{y}_i} - \sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*}\hat{\boldsymbol{\delta}})]\end{aligned} \tag{A.32}$$

$$\begin{aligned}\Rightarrow &[(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*}) + (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*})]\hat{\boldsymbol{\delta}} = \\ &[(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}_i}) - (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{y}_i})].\end{aligned} \tag{A.33}$$

$$\begin{aligned}\Rightarrow \hat{\boldsymbol{\delta}} = &[(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*}) + (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*})]^{-1} \times \\ &[(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}_i}) - (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{y}_i})].\end{aligned} \tag{A.34}$$

Substitute (A.31) to (A.28), then

$$\begin{aligned}&\hat{\boldsymbol{\theta}} = (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i)^{-1} \times \\ &[(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{y}}_i) - (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*})(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*})^{-1} (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}}_i - \sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}_i}\hat{\boldsymbol{\theta}})]\end{aligned} \tag{A.35}$$

$$\begin{aligned}\Rightarrow &[(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \tilde{\boldsymbol{G}}_i) - (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*})(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*})^{-1} (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \tilde{\boldsymbol{G}}_i)]\hat{\boldsymbol{\theta}} = \\ &[(\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{y}}_i) - (\sum_{i=1}^{N} \tilde{\boldsymbol{G}_i}^{'} \Delta\tilde{\boldsymbol{X}_i}^{*})(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{X}_i}^{*})^{-1} (\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}_i}^{*'} \Delta\tilde{\boldsymbol{y}_i})]\end{aligned} \tag{A.36}$$

$$\Rightarrow \hat{\boldsymbol{\theta}} = [(\sum_{i=1}^{N} \tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}' \tilde{\boldsymbol{G}}_{\boldsymbol{i}}) - (\sum_{i=1}^{N} \tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}' \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*})(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*\prime} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*})^{-1}(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*\prime} \tilde{\boldsymbol{G}}_{\boldsymbol{i}})]^{-1} \times$$
$$[(\sum_{i=1}^{N} \tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}' \Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}}) - (\sum_{i=1}^{N} \tilde{\boldsymbol{G}}_{\boldsymbol{i}}{}' \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*})(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*\prime} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*})^{-1}(\sum_{i=1}^{N} \Delta\tilde{\boldsymbol{X}}_{\boldsymbol{i}}{}^{*\prime} \Delta\tilde{\boldsymbol{y}}_{\boldsymbol{i}})]. \tag{A.37}$$

A.4 Derivatives first- and second-order derivatives of log-likelihood function

Likelihood function: $(2\pi)^{-\frac{NT}{2}}|\boldsymbol{\Omega}|^{-\frac{N}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{N}\Delta\boldsymbol{u_i}^{*}(\boldsymbol{\Omega})^{-1}\Delta\boldsymbol{u_i}^{*}\}$, where $\Delta\boldsymbol{u}_i^{*} = [\Delta y_{i,1} - b^{*} - \boldsymbol{\pi}^{*\prime}\Delta\boldsymbol{X_i}, \Delta y_{i2} - \phi\Delta y_{i1} - \beta\Delta x_{i2}, ..., \Delta y_{i,T} - \phi\Delta y_{i,T-1} - \beta\Delta x_{i,T}]'$ and

$$\boldsymbol{\Omega} = \sigma_u^2 \begin{bmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix} = \sigma_u^2\boldsymbol{\Omega}^{*}.$$

Let $\boldsymbol{\varphi} = (b^{*}, \boldsymbol{\pi'}, \phi, \beta)'$, and

$$\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}} = \begin{bmatrix} 1 & \Delta\boldsymbol{x}_i' & 0 & 0 \\ 0 & 0 & \Delta y_{i,1} & \Delta x_{i,2} \\ 0 & 0 & \Delta y_{i,2} & \Delta x_{i,2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \Delta y_{i,T-1} & \Delta x_{i,T}, \end{bmatrix}.$$

$(\boldsymbol{\Omega^{*}})^{-1} = \frac{adj(\boldsymbol{\Omega^{*}})}{|\boldsymbol{\Omega^{*}}|}$ and we know $|\boldsymbol{\Omega^{*}}| = 1 + T(\omega - 1)$,

$$adj(\boldsymbol{\Omega^{*}}) = \begin{bmatrix} (1) & (2) & \cdots & (3) & (4) \\ (5) & (6) & \cdots & (7) & (8) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (9) & (10) & \cdots & (11) & (12) \\ (13) & (14) & \cdots & (15) & (16) \end{bmatrix},$$

For (1):

$$\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix}_{2\times 2} = 3,$$

$$\begin{vmatrix} -2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{vmatrix}_{3\times 3} = 4, \cdots,$$

$$\begin{vmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix}_{(T-1)\times(T-1)} = T.$$

For (2) and (5):

$$(-1)\begin{vmatrix} -1 & -1 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-1)\times(T-1)} = \begin{vmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)(T-2)} = T-1.$$

For (3) and (9)

$$(-1)^{(T-1)+1}\begin{vmatrix} -1 & 0 & 0 & \cdots & 0 \\ 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-1)\times(T-1)} = (-1)^{(T+1)}\begin{vmatrix} -1 & 0 & 0 & \cdots & 0 \\ 2 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$(-1)^{(T-(T-1))}\begin{vmatrix} -1 & 0 \\ -1 & 2 \end{vmatrix} = 2.$$

For (4) and (13)

$$\begin{vmatrix} -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 \end{vmatrix} = ... = \begin{vmatrix} -1 & 2 \\ 0 & -1 \end{vmatrix} = 1.$$

For (6)

$$\begin{vmatrix} \omega & 0 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 \end{vmatrix}_{(T-1)\times(T-1)} = (-1)^{(1+1)}\omega \begin{vmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$\omega(T-1)$.

For (7) and (10)

$$(-1)^T \begin{vmatrix} \omega & 0 & 0 & \cdots & 0 \\ -1 & -1 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-1)\times(T-1)} = \omega \begin{vmatrix} -1 & 0 & \cdots & 0 \\ 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$(-1)^{T-(T+1)}\omega \begin{vmatrix} -1 & 0 \\ -1 & 2 \end{vmatrix} = 2\omega.$$

For (8) and (14)

$$\begin{vmatrix} \omega & 0 & 0 & \cdots & 0 \\ -1 & -1 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{vmatrix}_{(T-1)\times(T-1)} = \omega \begin{vmatrix} -1 & 0 & \cdots & 0 \\ 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$\omega \begin{vmatrix} -1 & 2 \\ 0 & -1 \end{vmatrix}_{2\times 2} = \omega.$$

For (11)

$$\begin{vmatrix} \omega & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 2 \end{vmatrix}_{(T-1)\times(T-1)} = 2^{2(T-1)} \begin{vmatrix} \omega & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$2[\omega \begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(t-3)\times(T-3)} + (-1)^3 \begin{vmatrix} -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-3)\times(T-3)}] =$$

$$2[(T-2)\omega - \begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-4)\times(T-4)}] = 2((T-2)\omega - (T-3)).$$

For (12) and (15)

$$\begin{vmatrix} \omega & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -1 & -1 \end{vmatrix}_{(T-1)\times(T-1)} = (-1)^2 \begin{vmatrix} \omega & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$\omega \begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(t-3)\times(T-3)} + (-1)^3 \begin{vmatrix} -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-3)\times(T-3)} =$$

$$(T-2)\omega - \begin{vmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-4)\times(T-4)} = (T-2)\omega - (T-3).$$

For (16)

$$\begin{vmatrix} \omega & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 2 \end{vmatrix}_{(T-1)\times(T-1)} =$$

$$\omega \begin{vmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} + (-1)^{(2+1)}(-1) \begin{vmatrix} -1 & -1 & 0 & \cdots & 0 \\ 0 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-2)\times(T-2)} =$$

$$(T-1)\omega - \begin{vmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix}_{(T-3)\times(T-3)} = (T-1)\omega - (T-2).$$

Then we can see that

$$(\Omega^*)^{-1} = (1 + T(\omega - 1))^{-1} \times$$
$$\begin{bmatrix} T & (T-1) & \cdots & 2 & 1 \\ (T-1) & (T-1)\omega & \cdots & 2\omega & \omega \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2 & 2\omega & \cdots & 2[(T-2)\omega - (T-3)] & (T-2)\omega - (T-3) \\ 1 & \omega & \cdots & (T-2)\omega - (T-3) & (T-1)\omega - (T-2) \end{bmatrix}$$

log-likelihood function:

$$\ln L = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{i=1}^{N}[(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\boldsymbol{\varphi})^{'}\boldsymbol{\Omega}^{-1}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\boldsymbol{\varphi})]$$
$$= -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln(\sigma_u^2) - \frac{N}{2}\ln[1+T(\omega-1)] -$$
$$\frac{1}{2}\sum_{i=1}^{N}[(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\boldsymbol{\varphi})^{'}\boldsymbol{\Omega}^{-1}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\boldsymbol{\varphi})].$$

Differentiate log-likelihood function with respect to $\boldsymbol{\varphi}$ and equate to zero

$$\frac{\partial \ln L}{\partial \boldsymbol{\varphi}} = -\frac{1}{2}\times(-2)\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}^{'}(\hat{\boldsymbol{\Omega}})^{-1}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})] = 0$$
$$\Rightarrow \sum_{i=1}^{N}(\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}(\hat{\boldsymbol{\Omega}})^{-1}\Delta \boldsymbol{y_i}) = \sum_{i=1}^{N}(\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}^{\prime}(\hat{\boldsymbol{\Omega}})^{-1}\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})$$
$$\Rightarrow \hat{\boldsymbol{\varphi}} = (\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}^{\prime}(\hat{\boldsymbol{\Omega}}^{*})^{-1}\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}})^{-1}(\sum_{i=1}^{N}\Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}(\hat{\boldsymbol{\Omega}}^{*})^{-1}\Delta \boldsymbol{y_i}),$$

where, $(\hat{\boldsymbol{\Omega}})^{-1} = \frac{1}{\sigma_u^2}(\hat{\boldsymbol{\Omega}}^{*})^{-1}$.

Differentiate log-likelihood function with respect to σ_u^2 and equate to zero

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{NT}{2\hat{\sigma}_u^2} - (-1)\frac{1}{2(\hat{\sigma}_u^2)^2}\sum_{i=1}^{N}[\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})^{'}(\hat{\boldsymbol{\Omega}}^{*})^{-1}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})] = 0$$
$$\Rightarrow \hat{\sigma}_u^2 = \frac{1}{NT}\sum_{i=1}^{N}[\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\boldsymbol{\varphi})^{'}(\hat{\boldsymbol{\Omega}}^{*})^{-1}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})].$$

Differentiate log-likelihood function with respect to ω and equate to zero

By $\frac{\partial \frac{A(x)}{B(x)}}{\partial x} = \frac{B(x)\frac{\partial A(x)}{\partial x} - \frac{B(x)}{\partial x}A(x)}{(B(x))^2}$,

$$\frac{\partial \ln L}{\partial \omega} = -\frac{NT}{2[1+T(\hat{\omega}-1)]} - \frac{1}{2\sigma_u^2[1+T(\hat{\omega}-1)]^2}\frac{1}{NT}\sum_{i=1}^{N}[\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})^{'}\boldsymbol{\Phi}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})] = 0$$
$$\Rightarrow 1+T(\hat{\omega}-1) = \frac{1}{NT}\sum_{i=1}^{N}[\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})^{'}\boldsymbol{\Phi}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})]$$
$$\Rightarrow \hat{\omega} = \frac{T-1}{T} + \frac{1}{NT^2}\sum_{i=1}^{N}[\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})^{'}\boldsymbol{\Phi}(\Delta \boldsymbol{y_i} - \Delta\tilde{\boldsymbol{W}}_{\boldsymbol{i}}\hat{\boldsymbol{\varphi}})],$$

where

$$
\begin{aligned}
\mathbf{\Phi} = [1+T(\omega-1)] & \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & (T-1) & \dots & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 2 & \cdots & 2(T-2) & (T-2) \\ 0 & 1 & \cdots & (T-2) & (T-1) \end{bmatrix} \\
- T & \begin{bmatrix} T & (T-1) & & 2 & 1 \\ (T-1) & (T-1)\omega & & 2\omega & \omega \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2 & 2\omega & \cdots & 2[(T-2)\omega-(T-3)] & (T-2)\omega-(T-3) \\ 1 & \omega & \cdots & (T-2)\omega-(T-3) & (T-1)\omega-(T-2) \end{bmatrix} \\
= & \begin{bmatrix} -T^2 & -T(T-1) & \cdots & -2T & -T) \\ -T(T-1) & -(T-1)^2 & \cdots & -2(T-1) & -(T-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -2T & -2(T-1) & \dots & -4 & -2 \\ -T & -(T-1) & \dots & -2 & -1 \end{bmatrix} \\
= - & \begin{bmatrix} T^2 & T(T-1) & \cdots & 2T & T) \\ T(T-1) & (T-1)^2 & \cdots & 2(T-1) & (T-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2T & 2(T-1) & \dots & 4 & 2 \\ T & (T-1) & \dots & 2 & 1 \end{bmatrix}.
\end{aligned}
$$

Second order differential:

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} = -\sum_{i=1}^{N} \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}}{}' \boldsymbol{\Omega}^{-1} \tilde{\boldsymbol{W}}_{\boldsymbol{i}} = -\frac{1}{\sigma_u^2} \sum_{i=1}^{N} \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}}{}' (\boldsymbol{\Omega}^*)^{-1} \tilde{\boldsymbol{W}}_{\boldsymbol{i}}.$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\varphi} \partial \sigma_u^2} = -\frac{1}{\sigma_u^4} \sum_{i=1}^{N} [\Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} (\boldsymbol{\Omega}^*)^{-1} (\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})].$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\varphi} \partial \omega} = -\frac{1}{\sigma_u^2 [1 + T(\omega - 1)]^2} \sum_{i=1}^{N} [\Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}}{}' \boldsymbol{\Phi} (\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})].$$

$$\frac{\partial^2 \ln L}{\partial \omega^2} = -(-1) \frac{NT}{2[1 + T(\omega - 1)]^2} +$$

$$\frac{1}{2}(-2) \frac{T}{\sigma_u^2 [1 + T(\omega - 1)]^3} \sum_{i=1}^{N} [\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})' (\boldsymbol{\Phi}) (\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})].$$

$$\frac{\partial^2 \ln L}{\partial (\sigma^2)^2} = -(-1) \frac{NT}{2\sigma_u^4} + \frac{-2}{2\sigma^6} \sum_{i=1}^{N} [\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})' (\boldsymbol{\Omega}^*)^{-1} (\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})].$$

$$\frac{\partial^2 \ln L}{\partial \omega \partial \sigma_u^2} = (-1) \frac{1}{2\sigma_u^4 [1 + T(\omega - 1)]^2} \sum_{i=1}^{N} [(\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})' (\boldsymbol{\Phi}) (\Delta \boldsymbol{y}_{\boldsymbol{i}} - \Delta \tilde{\boldsymbol{W}}_{\boldsymbol{i}} \boldsymbol{\varphi})].$$

Appendix B

Appendix to Chapter 3

B.1 Computing for the QML estimator

To compute the pseudo QML estimators, we use the eigenvalue approach by Hayakawa et al. (2021). Form pseudo log-likelihood function (3.16), we have

$$\begin{gathered}\left|\boldsymbol{\Omega}\left(\omega\right)+\boldsymbol{Q}\boldsymbol{Q}^{'}\right|=\left|\boldsymbol{\Omega}\left(\omega\right)\right|\left|\boldsymbol{I}_{m}+\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}\right| \\ \textit{and} \\ \left(\boldsymbol{\Omega}\left(\omega\right)+\boldsymbol{Q}\boldsymbol{Q}^{'}\right)^{-1}=\boldsymbol{\Omega}^{-1}\left(\omega\right)-\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}\boldsymbol{A}^{-1}\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right),\end{gathered} \tag{B.1}$$

where $\boldsymbol{A}=\boldsymbol{I}_{m}+\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}$. Then, the pseudo log likelihood function (3.16) can be written as

$$\begin{aligned}N^{-1}\ell_{p,N}\left(\boldsymbol{\theta}\right)\propto&-\frac{T}{2}\ln\left(\sigma^{2}\right)-\frac{1}{2}\ln\left|\boldsymbol{\Omega}\left(\omega\right)\right|-\frac{1}{2}\left|\boldsymbol{A}\right|- \\ &\frac{1}{2\sigma^{2}}\left(tr\left(\boldsymbol{B}_{N}\boldsymbol{\Omega}^{-1}\left(\omega\right)\right)-tr\left(\boldsymbol{B}_{N}\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right)\right)\right),\end{aligned} \tag{B.2}$$

where

$$\left|\boldsymbol{\Omega}\left(\omega\right)\right|=1+T\left(\omega-1\right),\ \textit{and}\ \boldsymbol{B}_{N}=\sum_{i=1}^{N}\boldsymbol{\xi}_{i}\left(\boldsymbol{\varphi}\right)\boldsymbol{\xi}_{i}^{'}\left(\boldsymbol{\varphi}\right)/N \tag{B.3}$$

Let $\boldsymbol{P}=\boldsymbol{\Omega}^{-1/2}\left(\omega\right)\boldsymbol{Q}\boldsymbol{A}^{-1/2}$ and $rank\left(\boldsymbol{P}\right)=m$, we have

$$\boldsymbol{I}_{m}-\boldsymbol{P}^{'}\boldsymbol{P}=\boldsymbol{I}_{m}-\boldsymbol{A}^{-1/2}\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}\boldsymbol{A}^{-1/2}, \tag{B.4}$$

and we know

$$\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}\left(\omega\right)\boldsymbol{Q}=\boldsymbol{A}-\boldsymbol{I}_{m}, \tag{B.5}$$

and

$$\begin{aligned}\boldsymbol{I}_m - \boldsymbol{P}^{'}\boldsymbol{P} &= \boldsymbol{I}_m - \boldsymbol{A}^{-1/2}\left(\boldsymbol{A} - \boldsymbol{I}_m\right)\boldsymbol{A}^{-1/2} \\ &= \boldsymbol{I}_m - \boldsymbol{A}^{-1}\end{aligned} \tag{B.6}$$

Therefore, we have

$$\begin{aligned}&\boldsymbol{A}^{-1} = \boldsymbol{I}_m - \boldsymbol{P}^{'}\boldsymbol{P} \\ &and \\ &tr\left(\boldsymbol{B}_N\boldsymbol{\Omega}^{-1}(\omega)\,\boldsymbol{Q}\boldsymbol{A}^{-1}\boldsymbol{Q}^{'}\boldsymbol{\Omega}^{-1}(\omega)\right) = \sigma^2 tr\left(\boldsymbol{P}^{'}\boldsymbol{C}_N(\boldsymbol{\phi})\,\boldsymbol{P}\right),\end{aligned} \tag{B.7}$$

where

$$\boldsymbol{C}_N(\boldsymbol{\phi}) = \sigma^{-2}\boldsymbol{\Omega}^{-1/2}(\omega)\,\boldsymbol{B}_N(\boldsymbol{\varphi}), \tag{B.8}$$

and $\boldsymbol{\phi} = \left(\boldsymbol{\varphi}^{'}, \omega, \sigma^2\right)^{'}$. Then, the pseudo log likelihood function (B.2) can be expressed as

$$N^{-1}\ell_N(\boldsymbol{\phi}, \boldsymbol{P}) \propto \\ -\frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln\left[1 + T(\omega - 1)\right] + \frac{1}{2}\ln\left|\boldsymbol{I}_m - \boldsymbol{P}^{'}\boldsymbol{P}\right| - \frac{1}{2}\left\{tr\left(\boldsymbol{C}_N(\boldsymbol{\phi})\right) - tr\left(\boldsymbol{P}^{'}\boldsymbol{C}_N(\boldsymbol{\phi})\,\boldsymbol{P}\right)\right\} \tag{B.9}$$

As we know that $\boldsymbol{P}^{'}\boldsymbol{P}$ can be diagonalised by an orthonormal transformation. To fix the rotation problem, we impose the $m(m-1)/2$ orthogonality conditions

$$\boldsymbol{p}_s^{'}\boldsymbol{p}_t = 0 \ for\ all\ s \neq t = 1, 2, \ldots, m, \tag{B.10}$$

where $\boldsymbol{p}_t$ and $\boldsymbol{p}_s$ are the t^{th} and s^{th} column of $\boldsymbol{P}$, respectively. Based on the restriction (B.10), the pseudo log likelihood function (B.9) can be written as

$$N^{-1}\ell_N(\boldsymbol{\phi}, \boldsymbol{P}) \propto \\ -\frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln\left[1 + T(\omega - 1)\right] + \frac{1}{2}\sum_{t=1}^{m}\ln\left(1 - \boldsymbol{p}_t^{'}\boldsymbol{p}_t\right) + \frac{1}{2}\sum_{t=1}^{m}\boldsymbol{p}_t^{'}\boldsymbol{C}_N(\boldsymbol{\phi})\,\boldsymbol{p}_t - \frac{1}{2}tr\left(\boldsymbol{C}_N(\boldsymbol{\phi})\right). \tag{B.11}$$

Then, taking the first derivatives with respect to $\boldsymbol{p}_t$ and setting this derivative to zero, as

$$\boldsymbol{C}_N(\boldsymbol{\phi})\,\hat{\boldsymbol{p}}_t - \left(\frac{1}{1 - \hat{\boldsymbol{p}}_t^{'}\hat{\boldsymbol{p}}_t}\right)\hat{\boldsymbol{p}}_t = 0,\ \ for\ t = 1, \ldots, m. \tag{B.12}$$

The concentrated pseudo log likelihood function in term of $\boldsymbol{\phi}$ can be written as

$$N^{-1}\ell_N(\boldsymbol{\phi}, m) \propto \\ -\frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2}\ln\left[1 + T(\omega - 1)\right] - \frac{1}{2}\sum_{t=1}^{m}\ln\left(\lambda_t(\boldsymbol{\phi}_i)\right) + \frac{1}{2}\sum_{t=1}^{m}\left(\lambda_t(\boldsymbol{\phi}) - 1\right) - \frac{1}{2}\sum_{t=1}^{T}\lambda_t(\boldsymbol{\phi}), \tag{B.13}$$

where $\lambda_t(\boldsymbol{\phi})$ is the t^{th} eigenvalue of $\boldsymbol{C}_N(\boldsymbol{\phi})$. By maximised this pseudo log likelihood function respect to $\boldsymbol{\phi} = \left(\boldsymbol{\varphi}', \omega, \sigma^2\right)$, we can obtain the pseudo estimator $\hat{\boldsymbol{\phi}} = \left(\hat{\boldsymbol{\phi}}, \hat{\omega}, \hat{\sigma}^2\right)$.

B.2 Monte Carlo simulation

B.2.1 Monte Carlo design

We consider the following dynamic panel data model with one regressor and two unobserved factors:

$$y_{it} = \rho y_{i,t-1} + \beta x_{i,t} + u_{it},\ for\ t = -49, \ldots, 1, \ldots, T;\ i = 1, \ldots, N, \tag{B.14}$$

with

$$u_{it} = \alpha_i + \sum_{s=1}^{m_y} \gamma_{ys,i} f_{s,t} + e_{it}. \tag{B.15}$$

The fixed effect α_i are generated as $\alpha_i \overset{iid}{\sim} N(0,1)$. The factor loadings, $\gamma_{ys,i}$ are generated as

$$\gamma_{ys,i} \overset{iid}{\sim} (0, m_0/\sqrt{m_0}). \tag{B.16}$$

The idiosyncratic errors, $e_{i,t}$ are generated as $e_{i,t} \sim N(0, \sigma_i^2)$ $\sigma_i^2 \sim U[0.5, 1.5]$.

The regressors, $x_{i,t}$ are generate as

$$x_{i,t} = \alpha_{xi} + \sum_{s=1}^{m_x} \gamma_{xs,i} f_{s,t} + v_{i,t},\ v_{i,t} = \rho_x v_{i,t-1} + \sqrt{(1-\rho_x^2)}\varepsilon_{i,t},\ for\ t = 1, \ldots, T, \tag{B.17}$$

with $\rho_x = 0.95$, and $\varepsilon_{i,t} \overset{iid}{\sim} N(0,1)$. The factor loadings of regressor, $\gamma_{xs,i}$ are generated as $\gamma_{xs,i} \overset{iid}{\sim} N\left(0, m_x/\sqrt{m_x}\right)$. We set $\alpha_{xi} = \alpha_i + v_i$, where $v_i \overset{iid}{\sim} N(0,1)$, for all i.

We generate unobserved common factors, $f_{s,t}$

$$f_{s,t} = \rho_{f,s} f_{s,t-1} + \sqrt{\left(1-\rho_{f,s}^2\right)}\varsigma_{fs,t},\ \varsigma_{fs,t} \overset{iid}{\sim} N(0,1),\ for\ s = 1, 2, \ldots, m_0,\ and\ t = 1, \ldots, T, \tag{B.18}$$

with $\rho_{f,s} = 0.5$.

In this Monte Carlo experiments, we consider $T = (5, 10)$ and $N = (100, 300, 500, 1000)$. The parameter $\rho = \{0.4, 0.8\}$ and $\beta = 1$. The number of factor, $m_0 = 2$.

B.2.2 Monte Carlo results

Table B.1 Bias, MSE of Robust QML estimates and Size (%) of the associated t-tests for the dynamic panel data model with $\{\rho, \beta\} = \{0.4, 1\}$, using the estimated number of factors, $\hat{m}$.

ρ	Bias(×100)				RMSE(×100)				Size(×100)			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
5	0.42	0.13	-0.01	-0.09	6.76	3.57	2.55	1.77	8.90	5.75	5.35	4.00
10	0.01	0.00	0.00	-0.02	2.75	1.57	1.22	0.85	5.90	5.25	4.85	5.50
β												
5	-0.48	-0.25	-0.09	0.02	10.97	6.31	4.73	3.34	5.55	4.65	4.25	4.45
10	0.33	-0.15	-0.01	0.02	7.17	4.15	3.14	2.28	4.90	4.20	4.55	4.40

$y_{i,t}$ is generated as $y_{it} = \rho y_{i,t-1} + \beta x_{i,t} + u_{it}$, $u_{it} = \alpha_i + \sum_{s=1}^{m_y} \gamma_{ys,i} f_{s,t} + e_{it}$, for $i = 1, \ldots, N$; $t = -49, \ldots, 1, \ldots, T$. The fixed effect α_i are generated as $\alpha_i \overset{iid}{\sim} N(0,1)$. The factor loadings, $\gamma_{ys,i} \overset{iid}{\sim} \left(0, m_0/\sqrt{m_0}\right)$. The idiosyncratic errors are generated as $e_{i,t} \sim U[0.5, 1.5]$. The regressors, $x_{i,t}$ are generate as $x_{i,t} = \alpha_{xi} + \sum_{s=1}^{m_x} \gamma_{xs,i} f_{s,t} + v_{i,t}$, where $v_{i,t} = \rho_x v_{i,t-1} + \sqrt{(1-\rho_x^2)}\varepsilon_{i,t}$, $\rho_x = 0.95$, $\varepsilon_{i,t} \overset{iid}{\sim} N(0,1)$ and $m_x = 2$. We set $\alpha_{xi} = \alpha_i + v_i$, where $v_i \overset{iid}{\sim} N(0,1)$, for all i. The common factors are generated as $f_{s,t} = \rho_{f,s} f_{s,t-1} + \sqrt{\left(1-\rho_{f,s}^2\right)}\varsigma_{fs,t}$, where $\varsigma_{fs,t} \overset{iid}{\sim} N(0,1)$, for $s = 1, 2, \ldots, m_0$, and $\rho_{f,s} = 0.5$. The number of factor, $m_0 = 1$.

Table B.2 Bias, MSE of Robust QML estimates and Size (%) of the associated t-tests for the dynamic panel data model with $\{\rho, \beta\} = \{0.8, 1\}$, using the estimated number of factors, $\hat{m}$.

ρ	Bias(×100)				RMSE(×100)				Size(×100)			
T/N	100	300	500	1000	100	300	500	1000	100	300	500	1000
5	0.03	-0.03	-0.02	0.06	6.25	3.26	2.60	1.80	7.65	5.75	4.80	4.95
10	-0.01	0.06	-0.02	-0.02	2.12	1.17	0.89	0.66	5.40	4.80	4.05	5.05
β												
5	-0.72	-0.25	0.12	-0.06	11.18	5.97	4.82	3.46	6.50	3.60	5.20	5.00
10	0.11	-0.03	-0.02	-0.06	7.35	4.05	3.16	2.35	5.15	3.90	3.50	4.85

B.3 Summary statistics

Table B.3 Summary statistics

Variables	Number of observations	*Mean*	*Median*	*Std.Dev.*	*Min.*	*Max.*	*Skew*	*Kurt*
Leverage	315621	0.2741	0.2335	0.2613	0.0000	2.5661	2.1415	12.2456
Profitability	315621	−0.0080	0.0747	0.3506	−4.6060	0.4128	−5.2004	42.1898
Growth opportunities	315621	1.5451	0.9933	1.9953	0.0000	27.9605	5.3999	45.4098
Tangibility	315621	0.0455	0.0383	0.0339	0.0000	0.2519	1.8726	8.3645
Size	315621	0.3216	0.2677	0.2409	0.0000	0.9298	0.6869	2.4917
Non-debt tax shields	315621	4.3679	4.2578	2.8161	−2.2026	10.9785	0.1233	2.3068

The data set is a panel of US firms collected from the CRSP/Compustat database over the period 1960-2017. Leverage ratio greater than one means the firm owns more liabilities than it does assets. It indicates that the firm is extremely leveraged.

Figure B.1 Number of firms from 1960 to 1967

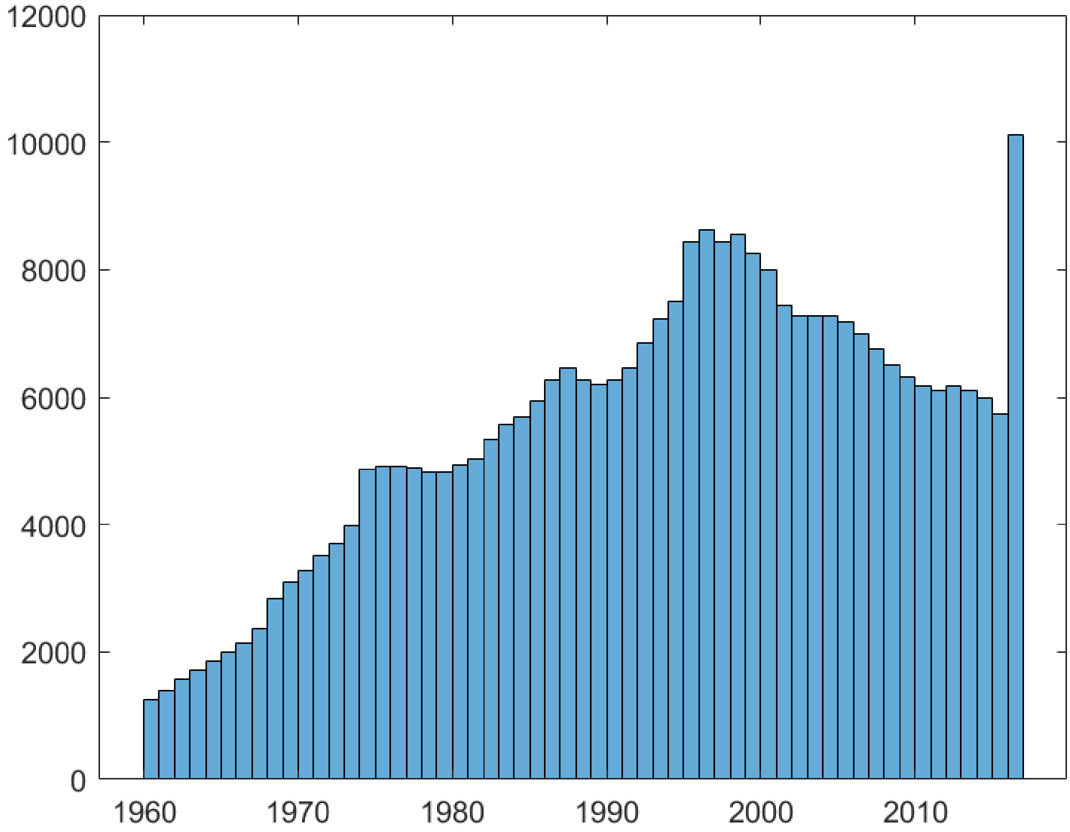


Figure B.2 Mean of leverage ratio from 1960 to 1967

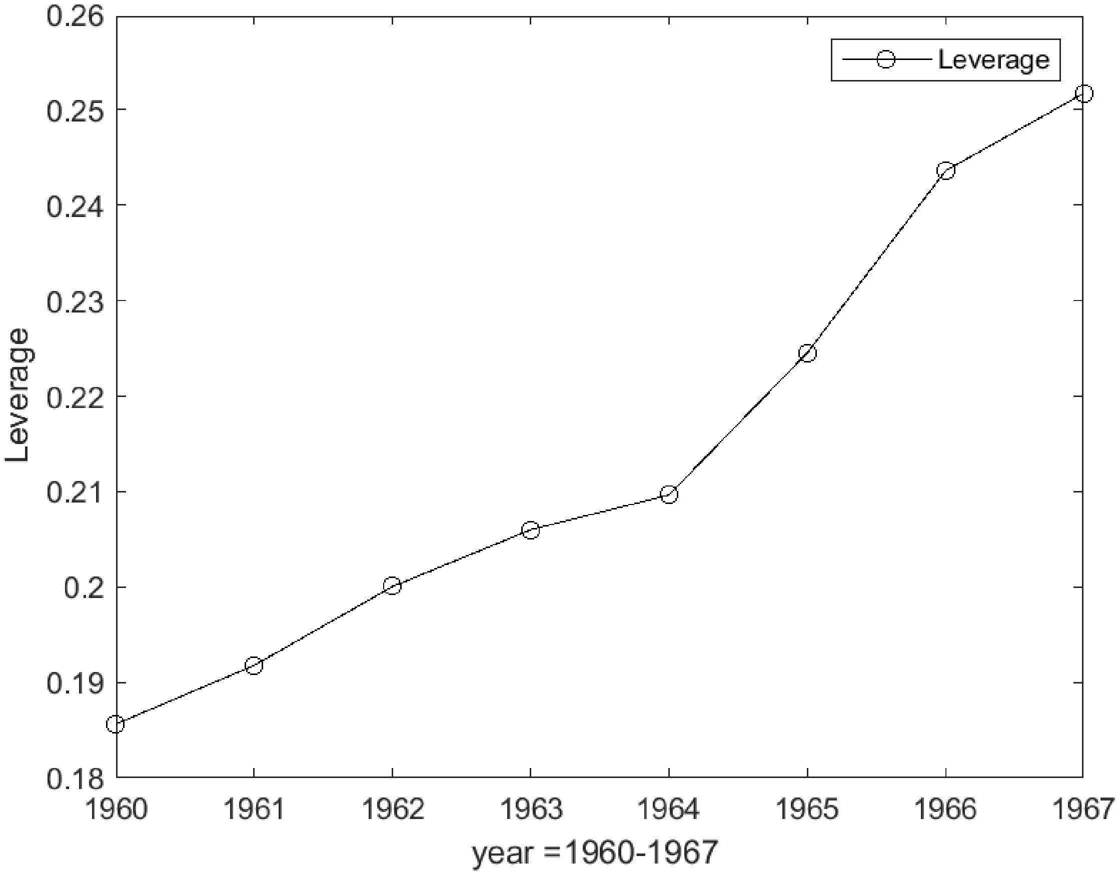


Figure B.3 Mean of leverage ratio from 1968 to 1975

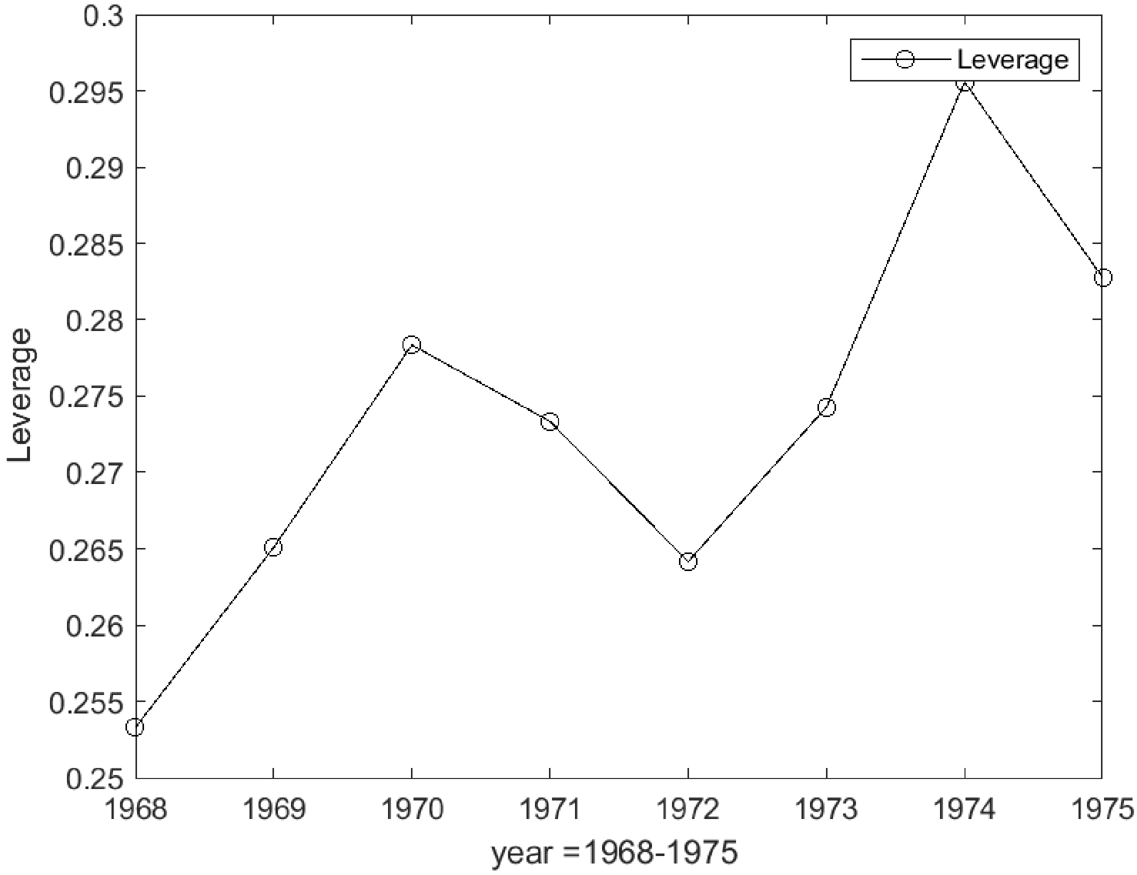


Figure B.4 Mean of leverage ratio from 1976 to 1983

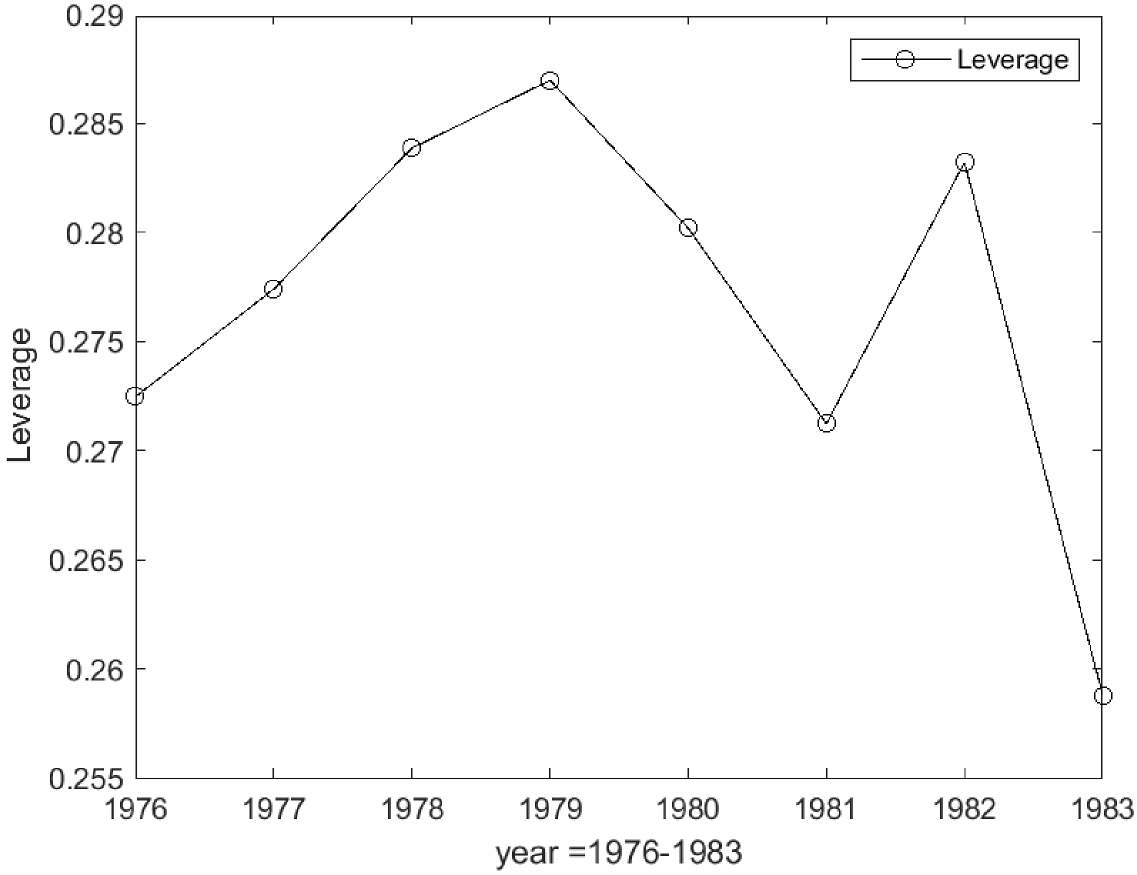


Figure B.5 Mean of leverage ratio from 1984 to 1991

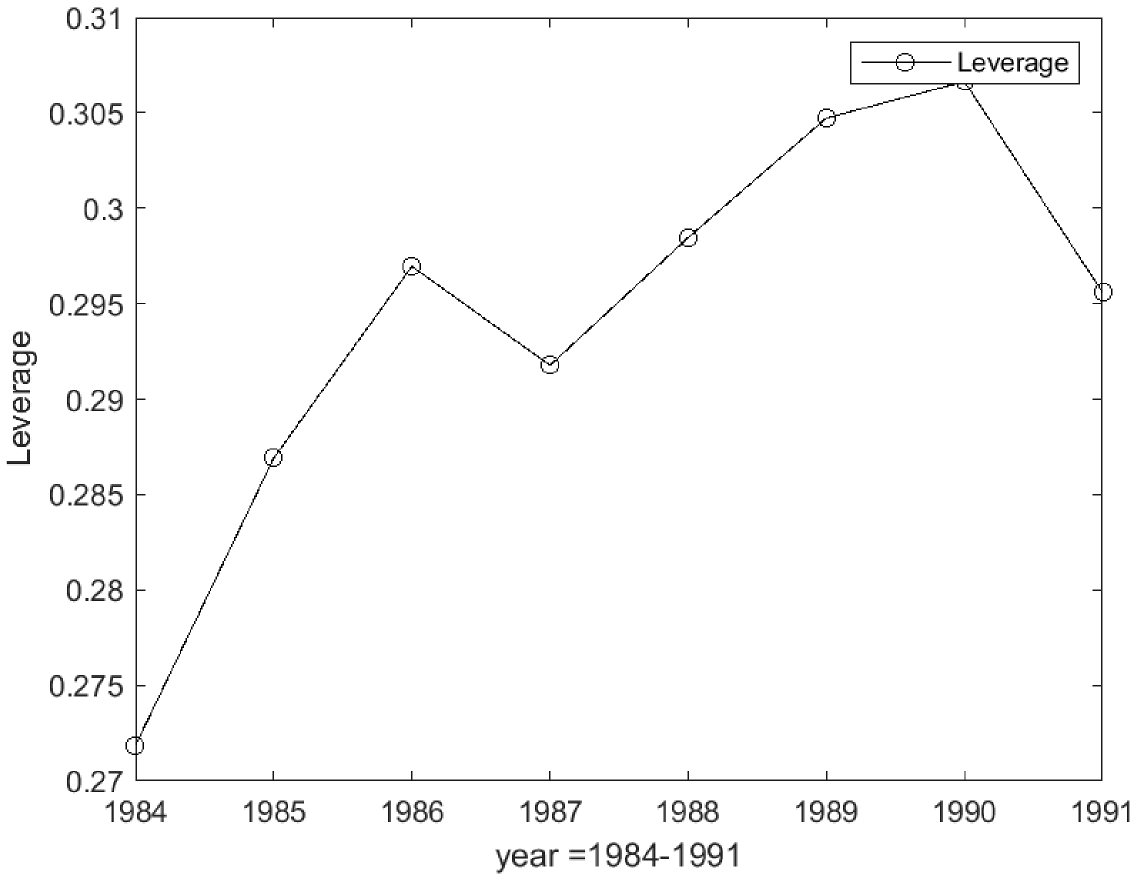


Figure B.6 Mean of leverage ratio from 1992 to 1999

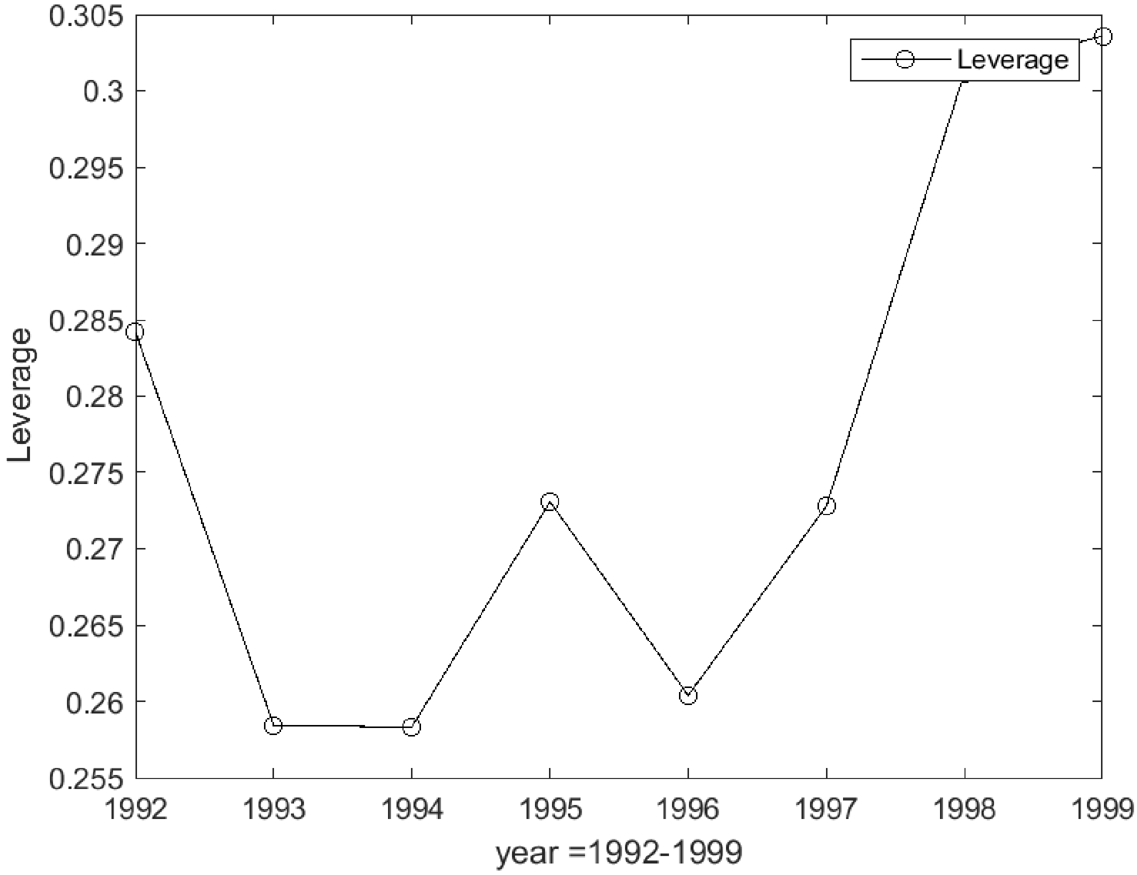


Figure B.7 Mean of leverage ratio from 2000 to 2007

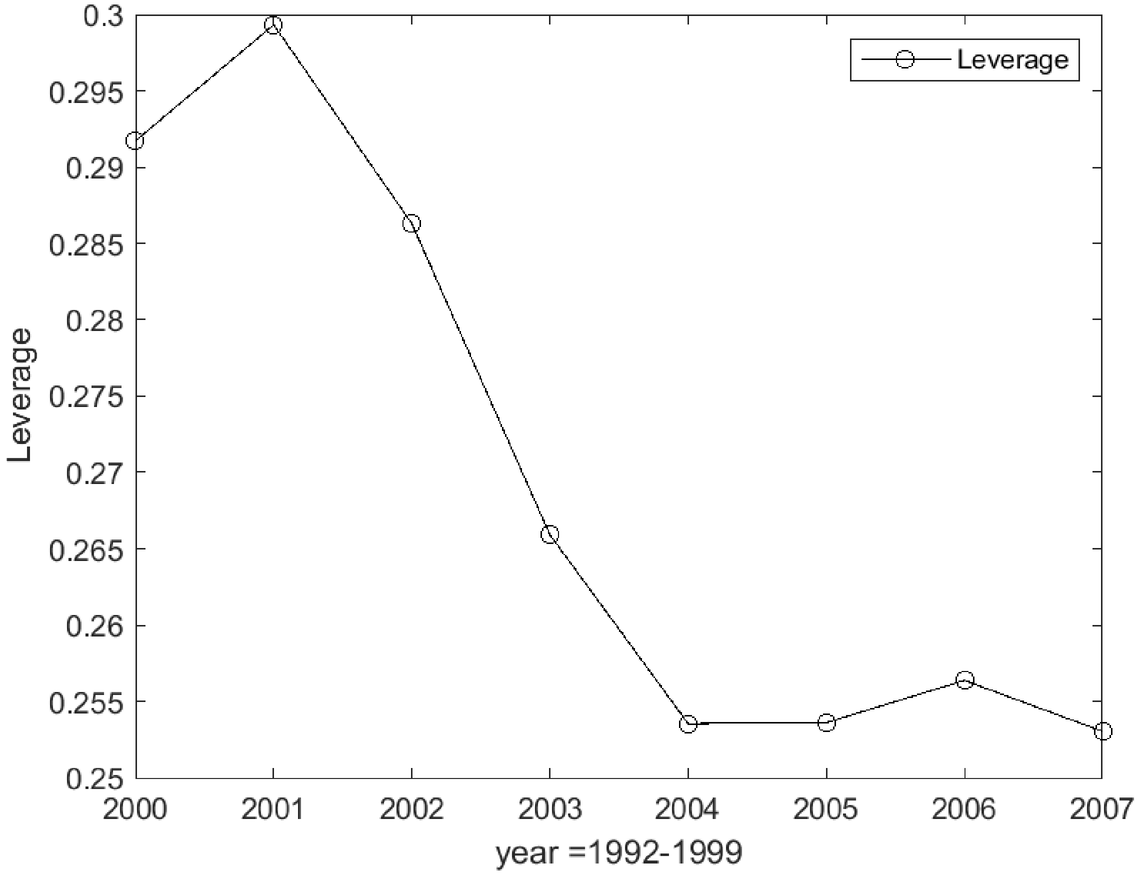
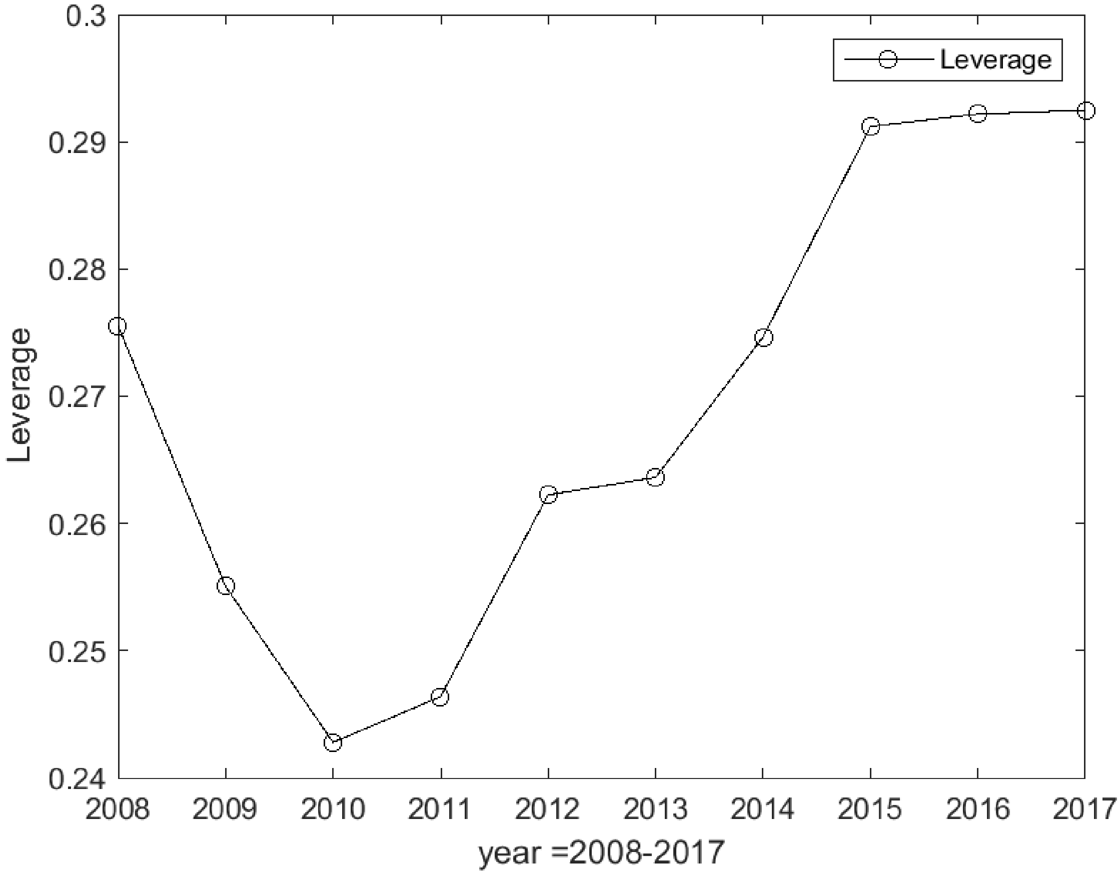


Figure B.8 Mean of leverage ratio from 2008 to 2017



Appendix C

Appendix to Chapter 5

C.1 Proofs of Lemmas and Theorem for IV estimation

Lemma 1 *: Suppose $\{X_{i,T}\}$ are independent across i for all T with $E\left(X_{i,T}\right) = \mu_{i,T}$ and $E|X_{i,T}|^{1+c} < \Delta < \infty$ for some $c > 0$ and all i, T. Then $N^{-1}\sum_{i=1}^{N}\left(X_{i,T} - \mu_{i,T}\right) \xrightarrow{p} 0$ as $(N,T) \xrightarrow{j} \infty$.*

Proof :
See proof Lemma 1 in Hansen (2007b).

Lemma 2 *: Suppose a $k \times 1$ vector, $\{\boldsymbol{x}_{i,T}\}$, are independent across i for all T with $E\left(\boldsymbol{x}_{i,T}\right) = 0$, $E\left(\boldsymbol{x}_{i,T}\boldsymbol{x}_{i,T}^{'}\right) = \boldsymbol{\Sigma}_{i,T}$, and $E\left\|\boldsymbol{x}_{i,T}\right\|^{2+c} < \Delta < \infty$ for some $c > 0$. Assume $\boldsymbol{\Sigma} = \lim_{T,N\to\infty} N^{-1}\sum_{i=1}^{N}\boldsymbol{\Sigma}_{i,T}$ is positive definite and the minimum eigenvalue of $\boldsymbol{\Sigma}$ is strictly positive. Then, $N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_{i,T} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma})$ as $(N,T) \xrightarrow{j} \infty$.*

Proof :
See proof Lemma 2 in Hansen (2007b).

Lemma 3 *: As $(N,T) \xrightarrow{j} \infty$ such that $N/T \to c$ with $0 < c < \infty$, for $i = 1, \ldots, N$ and $\ell = 1, 2, \ldots, k$,*

$$T^{-1}\left\|\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right\|^2 = T^{-1}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{f}}_{x,t} - \boldsymbol{G}_x^{'}\boldsymbol{f}_{x,t}\right\|^2 = O_p\left(c_{NT}^{-2}\right), \tag{C.1}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)^{'}\hat{\boldsymbol{F}}_x}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.2}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)^{'}\boldsymbol{F}_x}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.3}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x \boldsymbol{G}_x\right)' \boldsymbol{F}_y}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.4}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x \boldsymbol{G}_x\right)' \boldsymbol{\varepsilon}_i}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.5}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x \boldsymbol{G}_x\right)' \boldsymbol{v}_{\ell,i}}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.6}$$

$$\frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x \boldsymbol{G}_x\right)' \boldsymbol{W}_i}{T} = O_p\left(c_{NT}^{-2}\right), \tag{C.7}$$

$$N^{-1/2} \sum_{i=1}^{N} \frac{\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x \boldsymbol{G}_x\right)' \boldsymbol{v}_{\ell,i}}{T} \boldsymbol{\gamma}_{\ell i}' = O\left(N^{-1/2}\right) + O_p\left(c_{NT}^{-2}\right), \tag{C.8}$$

$$\boldsymbol{G}_x \boldsymbol{G}_x' - \left(\frac{\boldsymbol{F}_x' \boldsymbol{F}_x}{T}\right)^{-1} = O_p\left(c_{NT}^{-2}\right), \tag{C.9}$$

$$\frac{\boldsymbol{F}_x' \hat{\boldsymbol{F}}_x}{T} \xrightarrow{p} \boldsymbol{\Lambda}_x \ as \ (N,T) \xrightarrow{j} \infty, \tag{C.10}$$

where $\boldsymbol{F}_x = \boldsymbol{F}_x \boldsymbol{G}_x$, $\boldsymbol{\Gamma}_{xi} = \boldsymbol{G}_x^{-1} \boldsymbol{\Gamma}_{xi}$, *and* $\boldsymbol{G}_x$ *and* $\boldsymbol{\Gamma}_x$ *are invertible* $m_x \times m_x$ *matrices.*

Proof :
See proof Lemma B.4 in Norkutė et al. (2021).
Lemma 4 *: As* $(N,T) \xrightarrow{j} \infty$ *such that* $N/T \to c$ *with* $0 < c < \infty$.

(i) $T^{-1/2} \boldsymbol{X}_i' \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{\hat{F}_x}\right) \boldsymbol{u}_i = \sqrt{T} O_p\left(c_{NT}^{-2}\right)$.

(ii) $T^{-1/2} \boldsymbol{X}_{i,-1}' \boldsymbol{M}_{\hat{F}_{x,-1}} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x}\right) \boldsymbol{u}_i = \sqrt{T} O_p\left(c_{NT}^{-2}\right)$.

(iii) $T^{-1/2} \boldsymbol{X}_{i,-1}' \left(\boldsymbol{M}_{\hat{F}_{x,-1}} - \boldsymbol{M}_{F_{x,-1}}\right) \boldsymbol{M}_{F_x} \boldsymbol{u}_i = \sqrt{T} O_p\left(c_{NT}^{-2}\right)$.

(iv) $T^{-1/2} \boldsymbol{X}_{i,-J}' \boldsymbol{M}_{\hat{F}_{x,-J}} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x}\right) \boldsymbol{u}_i = \sqrt{T} O_p\left(c_{NT}^{-2}\right)$.

(v) $T^{-1/2} \boldsymbol{X}_{i,-J}' \left(\boldsymbol{M}_{\hat{F}_{x,-J}} - \boldsymbol{M}_{F_{x,-J}}\right) \boldsymbol{M}_{F_x} \boldsymbol{u}_i = \sqrt{T} O_p\left(c_{NT}^{-2}\right)$.

Proof :
By using $\frac{\hat{\boldsymbol{F}}_x'\hat{\boldsymbol{F}}_x}{T} = \boldsymbol{I}_{m_x}$, we have $\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x} = \boldsymbol{P}_{F_x} - \boldsymbol{P}_{\hat{F}_x} = -\left(\frac{\hat{\boldsymbol{F}}_x'\hat{\boldsymbol{F}}_x}{T} - \boldsymbol{P}_{F_x}\right)$. We can decompose Lemma 4 (i) as

$$
\begin{aligned}
&T^{-1/2}\boldsymbol{X}_i'\left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x}\right)\boldsymbol{u}_i = -T^{-1/2}\boldsymbol{X}_i'\left(\frac{\hat{\boldsymbol{F}}_x'\hat{\boldsymbol{F}}_x}{T} - \boldsymbol{P}_{F_x}\right)\boldsymbol{u}_i = \\
&-T^{-1/2}\frac{\boldsymbol{X}_i'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\boldsymbol{G}_x'\boldsymbol{F}_x'\boldsymbol{u}_i - T^{-1/2}\frac{\boldsymbol{X}_i'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i \\
&-T^{-1/2}\frac{\boldsymbol{X}_i'\boldsymbol{F}_x}{T}\boldsymbol{G}_x\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i - T^{-1/2}\frac{\boldsymbol{X}_i'\boldsymbol{F}_x}{T}\left(\boldsymbol{G}_x\boldsymbol{G}_x' - \left(\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right)^{-1}\right)\boldsymbol{F}_x'\boldsymbol{u}_i \\
&= -\left(b_1 + b_2 + b_3 + b_4\right).
\end{aligned}
\tag{C.11}
$$

By using Cauchy–Schwarz inequality, we have

$$
\begin{aligned}
&|b_1| \leq T^{1/2}\left\|\frac{\boldsymbol{X}_i'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|\|\boldsymbol{G}_x\|\left\|\frac{\boldsymbol{F}_x'\boldsymbol{u}_i}{T}\right\| \leq \\
&T^{1/2}\|\boldsymbol{\Gamma}_{x,i}\|\left\|\frac{\boldsymbol{F}_x'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|\|\boldsymbol{G}_x\|\left\|\frac{\boldsymbol{F}_x}{\sqrt{T}}\right\|\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\| \\
&+T^{1/2}\left\|\frac{\boldsymbol{V}_i'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|\|\boldsymbol{G}_x\|\left\|\frac{\boldsymbol{F}_x}{\sqrt{T}}\right\|\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\| = \sqrt{T}O_p\left(c_{NT}^{-2}\right),
\end{aligned}
\tag{C.12}
$$

from $\left\|\frac{\boldsymbol{F}_x'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\| = O_p\left(c_{NT}^{-2}\right)$, $\left\|\frac{\boldsymbol{V}_i'\left(\hat{\boldsymbol{F}}_x - \boldsymbol{F}_x\boldsymbol{G}\right)}{T}\right\| = O_p\left(c_{NT}^{-2}\right)$ by Lemma 3, and from Assumption 4 we know $\|\boldsymbol{G}_x\| = O_p(1)$, $\|\boldsymbol{\Gamma}_{x,i}\| = O_p(1)$, $\frac{\|\boldsymbol{F}_x\|}{\sqrt{T}} = O_p(1)$, $\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\| \leq \|\boldsymbol{\gamma_i}\|\frac{\|\boldsymbol{F}_y\|}{\sqrt{T}} + \|\boldsymbol{\lambda}_i\|\frac{\|\boldsymbol{F}_y\|}{\sqrt{T}} + \frac{\|\varepsilon_i\|}{\sqrt{T}} = O_p(1)$, $c_{NT} = min\left(\sqrt{N}, \sqrt{T}\right)$.

$$
\begin{aligned}
|b_2| &\leq \sqrt{T}\left\|\frac{\boldsymbol{X}_i'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x^0\boldsymbol{G}_x\right)}{T}\right\|\left\|\frac{\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i}{T}\right\| \\
&\leq \sqrt{T}\left\|\boldsymbol{\Gamma}_{x,i}\right\|\left\|\frac{\boldsymbol{F}_x'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|\left\|\frac{\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i}{T}\right\|+ \qquad (C.13)\\
&\sqrt{T}\left\|\frac{\boldsymbol{V}_i'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|\left\|\frac{\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i}{T}\right\| = \sqrt{T}O_p\left(c_{NT}^{-4}\right),
\end{aligned}
$$

and by Lemma 3 we know $\left\|\frac{\boldsymbol{u}_i'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'}{T}\right\| \leq \|\boldsymbol{\gamma}_i\|\left\|\frac{\boldsymbol{F}_x'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|+\|\boldsymbol{\lambda}_i\|\left\|\frac{\boldsymbol{F}_y'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\|+\left\|\frac{\boldsymbol{\varepsilon}_i'\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)}{T}\right\| = O_p\left(c_{NT}^{-2}\right)$.

$$
\begin{aligned}
|b_3| &\leq \sqrt{T}\left\|\frac{\boldsymbol{X}_i'\boldsymbol{F}_x}{T}\right\|\|\boldsymbol{G}_x\|\left\|\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i/T\right\| \\
&\leq \sqrt{T}\left\|\boldsymbol{\Gamma}_{x,i}\right\|\left\|\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right\|\|\boldsymbol{G}_x\|\left\|\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i/T\right\|+ \qquad (C.14)\\
&\sqrt{T}\left\|\frac{\boldsymbol{V}_i}{\sqrt{T}}\right\|\left\|\frac{\boldsymbol{F}_x}{\sqrt{T}}\right\|\|\boldsymbol{G}_x\|\left\|\left(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x\right)'\boldsymbol{u}_i/T\right\| = \sqrt{T}O_p\left(c_{NT}^{-2}\right).
\end{aligned}
$$

from Lemma 3, we have $\left\|\boldsymbol{u}_i'(\hat{\boldsymbol{F}}_x-\boldsymbol{F}_x\boldsymbol{G}_x)/T\right\| = O_p(c_{NT}^{-2})$. And by Assumption 2, 3, 4, we have $\|\boldsymbol{G}_x\| = O_p(1)$, $\|\boldsymbol{\Gamma}_{xi}\| = O_p(1)$, $\left\|\frac{\boldsymbol{V}_i}{\sqrt{T}}\right\| = o_P(1)$ and $\frac{\|\boldsymbol{F}_x\|}{\sqrt{T}} = O_p(1)$.

$$
\begin{aligned}
|b_4| &\leq \sqrt{T}\left\|\frac{\boldsymbol{X}_i'\boldsymbol{F}_x}{T}\right\|\left\|\boldsymbol{G}_x\boldsymbol{G}_x'-\left(\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right)^{-1}\right\|\left\|\frac{\boldsymbol{F}_x'\boldsymbol{u}_i}{T}\right\| \\
&\leq \sqrt{T}\left\|\boldsymbol{\Gamma}_x\right\|\left\|\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{\sqrt{T}}\right\|\left\|\boldsymbol{G}_x\boldsymbol{G}_x'-\left(\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right)^{-1}\right\|\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\|+ \qquad (C.15)\\
&\sqrt{T}\left\|\frac{\boldsymbol{V}_i}{\sqrt{T}}\right\|\left\|\frac{\boldsymbol{F}_x}{\sqrt{T}}\right\|\left\|\boldsymbol{G}_x\boldsymbol{G}_x'-\left(\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right)^{-1}\right\|\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\| = \sqrt{T}O_p\left(c_{NT}^{-2}\right),
\end{aligned}
$$

from Lemma 3, we have $\boldsymbol{G}_x\boldsymbol{G}_x'-\left(\frac{\boldsymbol{F}_x'\boldsymbol{F}_x}{T}\right)^{-1} = O_p(c_{NT}^{-2})$. By Assumption 2, 3 and 4, we have $\left\|\frac{\boldsymbol{V}_i}{\sqrt{T}}\right\| = O_p(1)$, $\|\boldsymbol{\Gamma}_x\| = O_p(1)$ and $\left\|\frac{\boldsymbol{u}_i}{\sqrt{T}}\right\| = O_p(1)$.

Then, we can see

$$\left\| T^{-1/2} \hat{\boldsymbol{X}}_i^{'} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{\hat{F}_x} \right) \boldsymbol{u}_i \right\| = \sqrt{T} O_p \left(c_{NT}^{-2} \right). \tag{C.16}$$

Thus,

$$T^{-1/2} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i = T^{-1/2} \boldsymbol{X}_i' \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p \left(c_{NT}^{-2} \right) \tag{C.17}$$

Lemma 4 (iv) (v) can be shown in a similarly way. Also can see the proof of Lemma 10 in Appendix A, Norkutė et al. (2021).

Proof of Proposition 1 :
Consider $T^{-1/2} \hat{\boldsymbol{Z}}_i^{'} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i$, where the instruments variable set as

$$\hat{\boldsymbol{Z}}_i = \left(\boldsymbol{M}_{\hat{F}_x} \boldsymbol{X}_i, \boldsymbol{M}_{\hat{F}_{x,-1}} \boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{M}_{\hat{F}_{x,-j}} \boldsymbol{X}_{i,-j}, \ldots, \boldsymbol{M}_{\hat{F}_{x,-j}} \boldsymbol{X}_{i,-J} \right). \tag{C.18}$$

Then, the first component of $\boldsymbol{M}_{\hat{F}_x} \hat{\boldsymbol{Z}}_i$ in $T^{-1/2} \hat{\boldsymbol{Z}}_i^{'} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i$ as

$$\begin{aligned} & T^{-1/2} \boldsymbol{X}_i' \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_i' \boldsymbol{M}_{F_x} \boldsymbol{u}_i + T^{-1/2} \boldsymbol{X}_i^{'} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x} \right) \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_i' \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}), \end{aligned} \tag{C.19}$$

by the Lemma 4 (i). Next we consider the second component of $\boldsymbol{M}_{\hat{F}_x} \hat{\boldsymbol{Z}}_i$ in $T^{-1/2} \hat{\boldsymbol{Z}}_i^{'} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i$ as

$$\begin{aligned} & T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{\hat{F}_{x,-1}} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{\hat{F}_{x,-1}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{\hat{F}_{x,-1}} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x} \right) \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{\hat{F}_{x,-1}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}) \\ & = T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{F_{x,-1}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \left(\boldsymbol{M}_{\hat{F}_{x,-1}} - \boldsymbol{M}_{F_{x,-1}} \right) \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}) \\ & = T^{-1/2} \boldsymbol{X}_{i,-1}^{'} \boldsymbol{M}_{F_{x,-1}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}). \end{aligned}$$

Finally we consider the last component of $\boldsymbol{M}_{\hat{F}_x} \hat{\boldsymbol{Z}}_i$ in $T^{-1/2} \hat{\boldsymbol{Z}}_i^{'} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i$ as

$$\begin{aligned} & T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{\hat{F}_{x,-J}} \boldsymbol{M}_{\hat{F}_x} \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{\hat{F}_{x,-J}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{\hat{F}_{x,-J}} \left(\boldsymbol{M}_{\hat{F}_x} - \boldsymbol{M}_{F_x} \right) \boldsymbol{u}_i \\ & = T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{\hat{F}_{x,-J}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}) \\ & = T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{F_{x,-J}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \left(\boldsymbol{M}_{\hat{F}_{x,-J}} - \boldsymbol{M}_{F_{x,-J}} \right) \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}) \\ & = T^{-1/2} \boldsymbol{X}_{i,-J}^{'} \boldsymbol{M}_{F_{x,-J}} \boldsymbol{M}_{F_x} \boldsymbol{u}_i + \sqrt{T} O_p(c_{NT}^{-2}), \end{aligned}$$

by Lemma 4 (iv) and (v). Therefore, we can show that

$$T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i = T^{-1/2}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{u}_i + \sqrt{T}O_p(c_{NT}^{-2}). \tag{C.20}$$

Proof of Theorem 1 :

First we consider 2SLS estimator without model average, as

$$\begin{aligned}
\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) &= \left(\boldsymbol{W}_i^{'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i\left(\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i\right)^{-1}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{W}_i\right)^{-1} \times \\
&\boldsymbol{W}_i^{'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i\left(\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\hat{\boldsymbol{Z}}_i\right)^{-1}\left(T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i\right) \\
&= \left(\hat{\tilde{\boldsymbol{A}}}_{i,T}^{'}\hat{\tilde{\boldsymbol{B}}}_{i,T}^{-1}\hat{\tilde{\boldsymbol{A}}}_{i,T}\right)^{-1}\hat{\tilde{\boldsymbol{A}}}_{i,T}^{'}\hat{\tilde{\boldsymbol{B}}}_{i,T}^{-1}\left(T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i\right),
\end{aligned} \tag{C.21}$$

where $\hat{\boldsymbol{Z}}_i = \left(\boldsymbol{M}_{\hat{F}_x}\boldsymbol{X}_i, \boldsymbol{M}_{\hat{F}_{x,-1}}\boldsymbol{X}_{i,-1}, \ldots, \boldsymbol{M}_{\hat{F}_{x,-J}}\boldsymbol{X}_{i,-J}\right)$.

As $(N,T) \xrightarrow{j} \infty$ as $N/T \to c$ for $0 < c < \infty$ and by Proposition 1, we have

$$T^{-1/2}\hat{\boldsymbol{Z}}_i^{'}\boldsymbol{M}_{\hat{F}_x}\boldsymbol{u}_i = T^{-1/2}\boldsymbol{Z}_i^{'}\boldsymbol{M}_{F_x}\boldsymbol{u}_i + o_p(1). \tag{C.22}$$

Under Assumptions $1-9$ and for each i, $T^{-1/2}\boldsymbol{Z}_i^{'}\boldsymbol{u}_i \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_i\right)$. And we can see that $\hat{\tilde{\boldsymbol{A}}}_{i,T} \xrightarrow{p} \tilde{\boldsymbol{A}}_{i,T}$ and $\hat{\tilde{\boldsymbol{B}}}_{i,T} \xrightarrow{p} \tilde{\boldsymbol{B}}_{i,T}$ as $T \to \infty$.

From Assumption 9, we see that $\operatorname{plim}_{T\to\infty}\hat{\tilde{\boldsymbol{A}}}_{i,T} = \boldsymbol{A}_i$ and $\operatorname{plim}_{T\to\infty}\hat{\tilde{\boldsymbol{B}}}_{i,T} = \boldsymbol{B}_i$. Thus, we have following asymptotic property :

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) \xrightarrow{d} N\left(\boldsymbol{0}, \left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)^{-1}\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{\Sigma}_i\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\left(\boldsymbol{A}_i^{'}\boldsymbol{B}_i^{-1}\boldsymbol{A}_i\right)^{-1}\right). \tag{C.23}$$

C.2 Proofs of Lemmas and Theorem for Model average 2SLS estimator

For ease of reading, we repeat the model (5.19) as

$$\begin{aligned}
y_{i,t} &= \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta}_i + u_{i,t}, \\
\boldsymbol{w}_{i,t} &= \left(y_{i,t-1}, \boldsymbol{x}_{i,t}^{'}\right)^{'} = g(\boldsymbol{z}_{i,t}) + \boldsymbol{e}_{i,t}.
\end{aligned} \tag{C.24}$$

Consider model average $2SLS$ estimator on above model, we have

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right) = \hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i. \tag{C.25}$$

We define $\boldsymbol{H}_i = \frac{\boldsymbol{G}_i^{'}\boldsymbol{G}_i}{T}$ and $\boldsymbol{h}_i = \frac{\boldsymbol{G}_i^{'}\boldsymbol{u}_i}{\sqrt{T}}$, where $\boldsymbol{G}_i = \left(\boldsymbol{g}_{i,1}, \ldots, \boldsymbol{g}_{i,T}\right)^{'}$ and $\boldsymbol{g}_{i,t} = g(\boldsymbol{z}_{i,t})$.

Following Alvarez and Arellano (2003), Donald and Newey (2001), Okui (2009) and Kuersteiner and Okui (2010), we have following Lemmas:

Lemma 5 *If there is a decomposition,* $\hat{\boldsymbol{h}}_i = \boldsymbol{h}_i + \boldsymbol{c}_i^h + \boldsymbol{r}_i^h$, $\tilde{\boldsymbol{h}}_i = \boldsymbol{h}_i + \boldsymbol{c}_i^h$, $\hat{\boldsymbol{H}}_i = \boldsymbol{H}_i + \boldsymbol{c}_i^H + \boldsymbol{r}_i^H$, *and*

$$\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'} - \tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_i^{H'} - \boldsymbol{c}_i^H\boldsymbol{H}_i^{-1}\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'} = \hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) + \boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right), \tag{C.26}$$

such that $\boldsymbol{c}_i^h = o_p(1)$, $\boldsymbol{h}_i = O_p(1)$, *and* $\boldsymbol{H}_i = O_p(1)$, *the determinant of* $\boldsymbol{H}_i$ *is bounded away from zero with probability* 1, $\rho_{i\omega,T} = tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right)$, *and* $\rho_{i\omega,T} = o_p(1)$,

$$\begin{aligned}
&\left\|\boldsymbol{c}_i^H\right\|^2 = o_p\left(\rho_{i\omega,T}\right), \ \left\|\boldsymbol{r}_i^h\right\| = o_p\left(\rho_{i\omega,T}\right), \ \left\|\boldsymbol{r}_i^H\right\| = o_p\left(\rho_{i\omega,T}\right), \\
&\boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right) = o_p\left(\rho_{i\omega,T}\right), \ E\left[\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right)|\boldsymbol{z}_{i,t}\right] = \sigma_i^2\boldsymbol{H}_i + \boldsymbol{H}_i\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\boldsymbol{H}_i + o_p\left(\rho_{i\omega,T}\right),
\end{aligned}$$

Then, we have

$$\begin{aligned}
&T\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)^{'} = \hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right) + \hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right) \\
&E\left[\hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right)|\boldsymbol{z}_{i,t}\right] = \sigma_{u,i}^2\boldsymbol{H}_i^{-1} + \boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right) + \boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right) \\
&\left(\hat{\boldsymbol{q}}_i\left(\boldsymbol{\omega}_i\right) + \boldsymbol{T}_i\left(\boldsymbol{\omega}_i\right)\right)/tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right) = o_p(1), \ \textit{as } T \to \infty, \ \tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^+ \to \infty.
\end{aligned} \tag{C.27}$$

Proof:
From model average 2SLS estimator, we have $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right) = \hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i$. And we decompose $\hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i$ into terms that are linear and quadratic in the difference of estimates and true values, as

$$\hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i = \boldsymbol{H}_i^{-1}\hat{\boldsymbol{h}}_i - \boldsymbol{H}_i^{-1}\left(\hat{\boldsymbol{H}}_i - \boldsymbol{H}_i\right)\boldsymbol{H}_i^{-1}\hat{\boldsymbol{h}}_i + \boldsymbol{H}_i^{-1}\left(\hat{\boldsymbol{H}}_i - \boldsymbol{H}_i\right)\boldsymbol{H}_i^{-1}\left(\hat{\boldsymbol{H}}_i - \boldsymbol{H}_i\right)\hat{\boldsymbol{H}}_i\hat{\boldsymbol{h}}_i.$$

And we know[1] $\left\|\boldsymbol{c}_i^H\right\|^2 = o_p\left(\rho_{i\omega,T}\right)$, $\left\|\boldsymbol{r}_i^H\right\| = o_p\left(\rho_{i\omega,T}\right)$, $\left\|\boldsymbol{r}_i^h\right\| = o_p\left(\rho_{i\omega,T}\right)$. Then, we have

$$\begin{aligned}
&\hat{\boldsymbol{h}}_i = \tilde{\boldsymbol{h}}_i + o_p\left(\rho_{i\omega,T}\right) \\
&\hat{\boldsymbol{H}}_i - \boldsymbol{H}_i = \boldsymbol{c}_i^H + o_p\left(\rho_{i\omega,T}\right).
\end{aligned} \tag{C.28}$$

Therefore,

$$\begin{aligned}
\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right) &= \hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i \\
&= \boldsymbol{H}_i^{-1}\tilde{\boldsymbol{h}}_i - \boldsymbol{H}_i^{-1}\boldsymbol{c}_i^H\boldsymbol{H}_i^{-1}\tilde{\boldsymbol{h}}_i + o_p\left(\rho_{i\omega,T}\right) \\
&= \hat{\boldsymbol{H}}_i^{-1}\left(\tilde{\boldsymbol{h}}_i - \boldsymbol{c}_i^H\boldsymbol{H}_i^{-1}\tilde{\boldsymbol{h}}_i\right) + o_p\left(\rho_{i\omega,T}\right).
\end{aligned}$$

[1]See the proof of Theorem 2 for $\left\|\boldsymbol{c}_i^H\right\|^2$, $\left\|\boldsymbol{r}_i^H\right\|$ and $\left\|\boldsymbol{r}_i^h\right\|$.

Next, let $\tilde{\boldsymbol{\tau}}_i = \tilde{\boldsymbol{h}}_i - \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i$ and

$$\begin{aligned}
\tilde{\boldsymbol{\tau}}_i \tilde{\boldsymbol{\tau}}_i^{'} &= \left(\tilde{\boldsymbol{h}}_i - \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i\right)\left(\tilde{\boldsymbol{h}}_i - \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i\right)^{'} \\
&= \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{'} - \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{'} \boldsymbol{H}_i^{-1} \boldsymbol{c}_i^{H'} - \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{'} + \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{'} \boldsymbol{H}_i^{-1} \boldsymbol{c}_i^H \\
&= \hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) + \boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right) + \boldsymbol{c}_i^H \boldsymbol{H}_i^{-1} \tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{'} \boldsymbol{H}_i^{-1} \boldsymbol{c}_i^H \\
&= \hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) + o_p\left(\rho_{i\omega,T}\right),
\end{aligned}$$

where[2] $\boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right) = o_p\left(\rho_{i\omega,T}\right)$, $\left\|\boldsymbol{c}_i^H\right\| = o_p\left(\rho_{i\omega,T}\right)$.

Then, we have

$$\begin{aligned}
T\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)^{'} &= \boldsymbol{H}_i^{-1}\left(\tilde{\boldsymbol{\tau}}_i \tilde{\boldsymbol{\tau}}_i^{'}\right)\boldsymbol{H}_i^{-1} + o_p\left(\rho_{i\omega,T}\right) \\
&= \boldsymbol{H}_i^{-1}\left(\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) + o_p\left(\rho_{i\omega,T}\right)\right)\boldsymbol{H}_i^{-1} + o_p\left(\rho_{i\omega,T}\right) \\
&= \boldsymbol{H}_i^{-1}\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right)\boldsymbol{H}_i^{-1} + o_p\left(\rho_{i\omega,T}\right) \\
&= \hat{\boldsymbol{Q}}_i\left(\boldsymbol{\omega}_i\right) + o_p\left(\rho_{i\omega,T}\right).
\end{aligned}$$

Lemma 6

(i) $tr\left(\boldsymbol{P}_i\right) = \sum_{j=1}^{J} \omega_{i,j} \cdot j = \tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i$ *(Hansen (2007a), Lemma 1.1).*

(ii) $\sum_{t=1}^{T} \left(\boldsymbol{P}_{i,tt}\right)^2 = o_p(\tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i^{+})$.

(iii) $\boldsymbol{H}_i = \frac{\boldsymbol{G}_i^{'} \boldsymbol{G}_i}{T} = O_p(1)$ *and* $\boldsymbol{h}_i = \frac{\boldsymbol{G}_i^{'} \boldsymbol{u}_i}{\sqrt{T}} = O_p(1)$.

Proof : (i)

$$tr\left(\boldsymbol{P}_i\right) = tr\left(\sum_{j=1}^{J} \omega_{i,j} \boldsymbol{P}_i^j\right) = \sum_{j=1}^{J} \omega_{i,j} \cdot j = \tilde{\boldsymbol{j}}^{'} \boldsymbol{\omega}_i.$$

[2] See the proof of Theorem 2 for $\boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right)$ and $\left\|\boldsymbol{c}_i^H\right\|$.

(ii) By Assumption (13) and Lemma 6 (i), it imply

$$\sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})^2 = \sum_{t=1}^{T}\sum_{j,j'=1}^{J}\omega_{ij}\omega_{ij'}\boldsymbol{P}_{i,tt}^{j}\boldsymbol{P}_{i,tt}^{j'} \leq$$

$$\sum_{t=1}^{T}\sum_{j,j'=1}^{J}|\omega_{ij}||\omega_{ij'}|\boldsymbol{P}_{i,tt}^{j}\boldsymbol{P}_{i,tt}^{j'} \leq$$

$$max_t\left(\boldsymbol{P}_{i,tt}^{J}\right)\left(\sum_{j=1}^{J}|\omega_{ij'}|\right)\sum_{t=1}^{T}\sum_{j=1}^{J}|\omega_{ij}|\boldsymbol{P}_{i,tt}^{j} \leq C\cdot max_t(\boldsymbol{P}_{i,tt}^{J})tr\left(\boldsymbol{P}_i^{+}\right) =$$

$$o_p(1)\left(\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i^{+}\right) = o_p\left(\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i^{+}\right),$$

$$\textit{where } \boldsymbol{\omega}_i^{+} = (|\omega_{i,1}|,\ldots,|\omega_{i,J}|) \textit{ and } \boldsymbol{P}_i^{+} = \sum_{j=1}^{J}|\omega_{i,j}|\boldsymbol{P}_i^{j}.$$

(iii) $\{\boldsymbol{G}_i'\boldsymbol{G}_i\}$ and $\{\boldsymbol{G}_i'\boldsymbol{u}_i\}$ obeys a LLN. And by assumption, $\boldsymbol{G}_i$ and $\boldsymbol{u}_i$ are orthogonal so the 2SLS estimator is consistent. Then, by CLT, we can show $\boldsymbol{H}_i$ and $\boldsymbol{h}_i$ are $O_p(1)$ [3].

Lemma 7

(i) *Let* $e_{g,i}(\boldsymbol{\omega}_i) = \boldsymbol{G}_i'(\boldsymbol{I}_T - \boldsymbol{P}_i)(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{G}_i/T$ *and* $\Delta(\boldsymbol{\omega}_i) = tr\left(e_{g,i}(\boldsymbol{\omega}_i)\right)$. *Then,* $\Delta(\boldsymbol{\omega}_i) = o_p(1)$.

(ii) $\frac{\boldsymbol{G}_i'(\boldsymbol{I}_T-\boldsymbol{P}_i)\boldsymbol{u}_i}{\sqrt{T}} = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\right)$.

(iii) $E\left(\boldsymbol{E}_i'\boldsymbol{P}_i\boldsymbol{u}_i|\boldsymbol{z}_{i,t}\right) = \boldsymbol{\sigma}_{eu,i}\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i$.

(iv) *Let* $f(\boldsymbol{\omega}_i):\boldsymbol{\omega}_i \to \mathbb{R}$ *with* $f(\boldsymbol{\omega}_i) > 0$ *be a function of* $\boldsymbol{\omega}_i$ *such that* $f(\boldsymbol{\omega}_i)\to\infty$ *as* $T\to\infty$. *Then* $\sqrt{\frac{f(\boldsymbol{\omega}_i)\Delta(\boldsymbol{\omega}_i)}{T}} = O_p\left(f(\boldsymbol{\omega}_i)/T + \Delta(\boldsymbol{\omega}_i)\right)$.

(v) $\tilde{\boldsymbol{j}}'\boldsymbol{\Gamma}_i\tilde{\boldsymbol{j}} \leq C\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i^{+}$.

(vi) $E\left(\boldsymbol{E}_i'\boldsymbol{P}_i\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{P}_i\boldsymbol{E}_i\right) = \boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}'\left(\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i\right)^2 + \left(\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i} + \boldsymbol{\sigma}_{ue,i}\boldsymbol{\sigma}_{eu,i}'\right)\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) +$ $Cum\left(u_{i,t},u_{i,t},\boldsymbol{e}_{i,t},\boldsymbol{e}_{i,t}'\right)\sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})^2$.

(vii) $E\left(\boldsymbol{G}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{P}_i\boldsymbol{E}_i|\boldsymbol{z}_{i,t}\right) = \sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{P}_{i,tt}E\left(u_{i,t}^2\boldsymbol{e}_{i,t}'\right) = O_p(\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i^{+})$.

(viii) $E\left(\boldsymbol{G}_i'(\boldsymbol{I}_T-\boldsymbol{P}_i)\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{P}_i\boldsymbol{E}_i/T|\boldsymbol{z}_{i,t}\right) = \boldsymbol{G}_i'(\boldsymbol{I}_T-\boldsymbol{P}_i)\boldsymbol{\mu}_i(\boldsymbol{\omega}_i)/T = o_p\left(\frac{\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i^{+}}{T} + \Delta(\boldsymbol{\omega}_i)\right)$, *where* $\mu_{i,t}(\boldsymbol{\omega}_i) = E(u_{i,t}^2\boldsymbol{e}_{i,t})\boldsymbol{P}_{i,tt}$ *and* $\boldsymbol{\mu}_i(\boldsymbol{\omega}_i) = (\mu_{i,1}(\boldsymbol{\omega}_i),\ldots,\mu_{i,T}(\boldsymbol{\omega}_i))'$.

(ix) $E\left(\boldsymbol{G}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{G}_i\boldsymbol{H}_i^{-1}\boldsymbol{E}_i'\boldsymbol{P}_i\boldsymbol{E}_i|\boldsymbol{z}_{i,t}\right)/T^2 = O_p(T^{-1}) + \left(\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i/T\right)$.

---

[3] See Chapter 7 and Chapter 12 of Hansen (2020).

(x) $E\left(\boldsymbol{G}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{P}_i\boldsymbol{E}_i\boldsymbol{H}_i^{-1}\left(\boldsymbol{E}_i'\boldsymbol{G}_i+\boldsymbol{G}_i'\boldsymbol{E}_i|\boldsymbol{z}_{i,t}\right)\right)/T^2=$
$O_p(T^{-1})+\left(\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i/T\right)\left(\sum_{t=1}^T\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}'\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}/T+\sum_{t=1}^T\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}'\boldsymbol{H}_i^{-1}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}'/T\right).$

(xi) $E\left(\boldsymbol{G}_i\boldsymbol{G}_i'\boldsymbol{H}_i^{-1}\boldsymbol{E}_i'\boldsymbol{G}_i|\boldsymbol{z}\right)=\sum_{t=1}^T\boldsymbol{g}_{i,t}\boldsymbol{g}_{i,t}'\boldsymbol{H}_i^{-1}E(u_{i,t}^2\boldsymbol{e}_{i,t})\boldsymbol{G}_i'/T^2=O_p(T^{-1}).$

Proof :
See proof of Lemma A.6 in Kuersteiner and Okui (2010).

Lemma 8 *Let*

$$\Xi\left(\boldsymbol{\omega}_i\right)=tr\left(\frac{\boldsymbol{G}_i'\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}\right). \tag{C.29}$$

From previous, we know $\rho_{i\omega,T}=tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right)$*. Then, we have*

$$\left(\Xi\left(\boldsymbol{\omega}_i\right)\right)^2=o_p\left(\rho_{i\omega,T}\right). \tag{C.30}$$

Proof :
See proof of Lemma A.7 in Kuersteiner and Okui (2010).

Lemma 9 *If for some sequence* $L\leq J$*,* $L\to\infty$*,* $L\notin\bar{J}$[4]*,* $sup_{m\notin\bar{J},m\leq L}|\sum_{j=1}^m\omega_{i,j}|=O_p(1/\sqrt{T})$ *as* $J\to\infty$*, and* $\sum_{j=1}^J\omega_{i,j}=1$ *for any* J*, then it follows that* $\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\to\infty$ *as* $J\to\infty$*.*

Proof :
See proof of Lemma A.3 in Kuersteiner and Okui (2010).

Proof of Theorem 2:

The model average 2SLS estimator can be expressed as

$$\begin{aligned}\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)&=\left(\frac{\boldsymbol{W}_i'\boldsymbol{P}_i\boldsymbol{W}_i}{T}\right)^{-1}\frac{\boldsymbol{W}_i'\boldsymbol{P}_i\boldsymbol{u}_i}{\sqrt{T}}\\&=\hat{\boldsymbol{H}}_i^{-1}\hat{\boldsymbol{h}}_i.\end{aligned} \tag{C.31}$$

Expand this forms as

$$\begin{aligned}\hat{\boldsymbol{h}}_i=\frac{\boldsymbol{W}_i'\boldsymbol{P}_i\boldsymbol{u}_i}{\sqrt{T}}&=\frac{\left(\boldsymbol{G}_i+\boldsymbol{E}_i\right)'\boldsymbol{P}_i\boldsymbol{u}_i}{\sqrt{T}}\\&=\frac{\boldsymbol{G}_i'\boldsymbol{u}_i}{\sqrt{T}}-\frac{\boldsymbol{G}_i\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{u}_i}{\sqrt{T}}+\frac{\boldsymbol{E}_i'\boldsymbol{P}_i\boldsymbol{u}_i}{\sqrt{T}}\\&=\boldsymbol{h}_i+\boldsymbol{c}_{1i}^h+\boldsymbol{c}_{2i}^h\\&=\boldsymbol{h}_i+\boldsymbol{c}_i^h,\end{aligned} \tag{C.32}$$

where $\boldsymbol{h}_i=\frac{\boldsymbol{G}_i'\boldsymbol{u}_i}{\sqrt{T}}$, $\boldsymbol{c}_{1i}^h=-\frac{\boldsymbol{G}_i(\boldsymbol{I}_T-\boldsymbol{P}_i)\boldsymbol{u}_i}{\sqrt{T}}$, $\boldsymbol{c}_{2i}^h=\frac{\boldsymbol{E}_i'\boldsymbol{P}_i\boldsymbol{u}_i}{\sqrt{T}}$, and $\boldsymbol{c}_i^h=\boldsymbol{c}_{1i}^h+\boldsymbol{c}_{2i}^h$.

[4] $\bar{J}$ is define in Assumption 12 (3).

By Lemma 6 (iii), we know $\boldsymbol{h}_i = O_p(1)$. And by Lemma 7 (i), (ii), we have

$$\boldsymbol{c}_{1i}^h = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\right) = o_p(1), \tag{C.33}$$

where $\Delta(\boldsymbol{\omega}_i) = o_p(1)$. By Lemma 7 (iii) we have

$$\boldsymbol{c}_{2i}^h = O_p\left(max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})^2}\right)/\sqrt{T}\right) = o_p(1). \tag{C.34}$$

For $\boldsymbol{c}_{2i}^h$, we know $\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i/\sqrt{T} = o(1)$ because $|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|/\sqrt{T} \leq \tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^+/\sqrt{T} = o(1)$. And, by Lemma 6 (ii) we have $\sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})^2 = o_p\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^+\right)$. By Lemma 7 (v) we have $\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i = O\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^+\right)$. Therefore,

$$\boldsymbol{c}_i^h = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\right) + O_p\left(max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}(\boldsymbol{P}_{i,tt})^2}\right)/\sqrt{T}\right) = o_p(1). \tag{C.35}$$

Next, we consider the decomposition of $\hat{\boldsymbol{H}}_i$ as

$$\begin{aligned}
\hat{\boldsymbol{H}}_i = \frac{\boldsymbol{W}_i^{'}\boldsymbol{P}_i\boldsymbol{W}_i}{T} &= \frac{(\boldsymbol{G}_i + \boldsymbol{E}_i)^{'}\boldsymbol{P}_i(\boldsymbol{G}_i + \boldsymbol{E}_i)}{T} \\
&= \frac{\boldsymbol{G}_i^{'}\boldsymbol{P}_i\boldsymbol{G}_i + \boldsymbol{G}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i + \boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{G}_i + \boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i}{T} \\
&= \frac{\boldsymbol{G}_i^{'}\boldsymbol{G}_i}{T} - \frac{\boldsymbol{G}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{G}_i}{T} + \frac{\boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i}{T} \\
&+ \frac{\boldsymbol{E}_i^{'}\boldsymbol{G}_i + \boldsymbol{G}_i\boldsymbol{E}_i}{T} + \frac{\boldsymbol{E}_i\boldsymbol{P}_i\boldsymbol{E}_i}{T} \\
&+ \frac{\boldsymbol{E}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{G}_i + \boldsymbol{G}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{E}_i}{T} \\
&= \boldsymbol{H}_i + \boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H + \boldsymbol{c}_{3i}^H + \boldsymbol{r}_i^H,
\end{aligned} \tag{C.36}$$

where

$$\begin{aligned}
\boldsymbol{H}_i &= \frac{\boldsymbol{G}_i^{'}\boldsymbol{G}_i}{T}, \\
\boldsymbol{c}_{1i}^H &= -\boldsymbol{G}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{G}_i/T, \\
\boldsymbol{c}_{2i}^H &= \left(\boldsymbol{E}_i^{'}\boldsymbol{G}_i + \boldsymbol{G}_i^{'}\boldsymbol{E}_i\right)/T, \\
\boldsymbol{c}_{3i}^H &= \boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i/T, \\
\boldsymbol{r}^H &= \left(\boldsymbol{E}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{G}_i + \boldsymbol{G}_i^{'}(\boldsymbol{I}_T - \boldsymbol{P}_i)\boldsymbol{E}_i\right)/T.
\end{aligned} \tag{C.37}$$

By Lemma 6 (iii), we know $\boldsymbol{H}_i = O_p(1)$. And by the definition and Lemma 8 and Lemma 7 (i), we have $\boldsymbol{c}_{1i}^H = O(\Xi(\boldsymbol{\omega}_i)) = o_p(1)$. By CLT, we have $\boldsymbol{c}_{2i}^H = O_p\left(1/\sqrt{T}\right)$. And by similar arguments as before, we have

$$\boldsymbol{c}_{3i}^H = O_p\left(max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right). \tag{C.38}$$

Next, we analyze $\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{1i}^H\right\|$, $\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{2i}^H\right\|$ and $\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{3i}^H\right\|$.

①

By Lemma 6 and Lemma 7 (i), we have

$$\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{1i}^H\right\| = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\,\Xi(\boldsymbol{\omega}_i)\right) = O_p\left((\Xi(\boldsymbol{\omega}_i))^2\right) = o_p(\rho_{i\omega,T}). \tag{C.39}$$

②

Let $f(\boldsymbol{\omega}_i) = T\left(tr\left(\boldsymbol{S}_i(\boldsymbol{\omega}_i)\right) - \Delta(\boldsymbol{\omega}_i)\right)$. By Lemma 9, we know that $\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i \to \infty$ as $T \to \infty$. This implies that $f(\boldsymbol{\omega}_i) \to \infty$. Then, by Lemma 7 (iv), it is hold that

$$\Delta(\boldsymbol{\omega}_i)^{1/2}/\sqrt{T} = o_p\left(\frac{f(\boldsymbol{\omega}_i)}{T} + \Delta(\boldsymbol{\omega}_i)\right) = o_p(tr\left(\boldsymbol{S}_i(\boldsymbol{\omega}_i)\right)) = o_p(\rho_{i\omega,T}). \tag{C.40}$$

Therefore, we have

$$\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{2i}^H\right\| = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\right)O_p\left(1/\sqrt{T}\right) = O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}/\sqrt{T}\right) = o_p(\rho_{i\omega,T}). \tag{C.41}$$

③

By Lemma 7(i), $\Delta(\boldsymbol{\omega}_i) = o_p(1)$ and the fact that $\boldsymbol{c}_{3i}^H = O_p(tr\left(\boldsymbol{S}_i(\boldsymbol{\omega}_i)\right)) = O_p(\rho_{i\omega,T})$., we can show

$$\begin{aligned}\left\|\boldsymbol{c}_{1i}^h\right\| \cdot \left\|\boldsymbol{c}_{3i}^H\right\| &= O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\right)O_p\left(max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right)\\ &= O_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\,max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right)\\ &= o_p\left(max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right)\\ &= o_p(\rho_{i\omega,T}).\end{aligned} \tag{C.42}$$

Next, we analyze $\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{1i}^{H}\right\|$, $\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{2i}^{H}\right\|$ and $\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{3i}^{H}\right\|$.

4

By Lemma 7 (i) and Lemma 8, we have

$$\begin{aligned}
\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{1i}^{H}\right\| &= O_p\left(max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/\sqrt{T}\right) O\left(\Xi\left(\boldsymbol{\omega}_i\right)\right) \\
&= O_p\left(\Xi\left(\boldsymbol{\omega}_i\right) max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/\sqrt{T}\right) \\
&= o_p\left(\Delta\left(\boldsymbol{\omega}_i\right)^{1/2} max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/\sqrt{T}\right)
\end{aligned} \tag{C.43}$$

By Lemma 7 (iv), it is implied that

$$\Delta\left(\boldsymbol{\omega}_i\right)^{1/2}|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|/\sqrt{T} \leq (\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i)^2/T + \Delta(\boldsymbol{\omega}_i) = O\left(\rho_{i\omega,T}\right). \tag{C.44}$$

And $\Delta\left(\boldsymbol{\omega}_i\right)^{1/2} = o_p(1)$ such that $o_p\left(\Delta(\boldsymbol{\omega}_i)^{1/2}\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i/\sqrt{T}\right) = o_p(\rho_{i\omega,T})$. By Lemma 7 (iv) and let $f(\boldsymbol{\omega}_i) = \left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2$, we have

$$\begin{aligned}
&\sqrt{\frac{\Delta\left(\boldsymbol{\omega}_i\right)\left(\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2\right)}{T}} = O_p\left(\frac{\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}{T} + \Delta\left(\boldsymbol{\omega}_i\right)\right) \\
&= O_p(\rho_{i\omega,T}).
\end{aligned} \tag{C.45}$$

Thus, we have $\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{1i}^{H}\right\| = o_p(\rho_{i\omega,T})$.

5

By similar arguments as before, we have

$$\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{2i}^{H}\right\| = O_p\left(max\left(|\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right), \tag{C.46}$$

where $\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i/T = O\left(tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right)\right)$ and $\sqrt{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}/T = o_p\left(tr\left(\boldsymbol{S}_i\left(\boldsymbol{\omega}_i\right)\right)\right)$.

6

Also, we have

$$\left\|\boldsymbol{c}_{2i}^{h}\right\| \cdot \left\|\boldsymbol{c}_{3i}^{H}\right\| = O_p\left(max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|^2, \left(\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2\right)\right)/T^{3/2}\right) \qquad \text{(C.47)}$$
$$= o_p\left(\rho_{i\omega,T}\right),$$

where $\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|/T\right)^{3/2} = o\left(\rho_{i\omega,T}\right)$ and $\left(\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2\right)/T = O_p\left(\rho_{i\omega,T}\right)$.

We also have

7

$$\left\|\boldsymbol{c}_{2i}^{h}\right\|^2 \cdot \left\|\boldsymbol{c}_{1i}^{H}\right\| = o_p\left(\rho_{i\omega,T}\right)$$

8

$$\left\|\boldsymbol{c}_{2i}^{h}\right\|^2 \cdot \left\|\boldsymbol{c}_{2i}^{H}\right\| = o_p\left(\rho_{i\omega,T}\right)$$

9

$$\left\|\boldsymbol{c}_{2i}^{h}\right\|^2 \cdot \left\|\boldsymbol{c}_{3i}^{H}\right\| = o_p\left(\rho_{i\omega,T}\right)$$

10

$$\left\|\boldsymbol{c}_{1i}^{H}\right\|^2 = O_p\left(\Xi\left(\boldsymbol{\omega}_i\right)^2\right) = o_p\left(\rho_{i\omega,T}\right)$$

11

$$\left\|\boldsymbol{c}_{2i}^{H}\right\|^2 = O_p\left(1/T\right) = o_p\left(\rho_{i\omega,T}\right)$$

12

$$\left\|\boldsymbol{c}_{3i}^{H}\right\|^2 = O_p\left(\left(max\left(|\tilde{\boldsymbol{j}}'\boldsymbol{\omega}_i|, \sqrt{\left(\boldsymbol{\omega}_i'\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right) + \sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2}\right)/T\right)^2\right) = o_p\left(\rho_{i\omega,T}\right)$$

Therefore, by Cauchy-Schwarz inequality, $\left\|\boldsymbol{c}_i^H\right\|^2 = o_p\left(\rho_{i\omega,T}\right)$. By Lemma 7 *(iv)*, we can show that each term of $\boldsymbol{c}_i^H$ is $O_p(\Delta(\boldsymbol{\omega}_i)^{1/2}/\sqrt{T}) = o_p(g(\boldsymbol{\omega}_i)/T + \Delta(\boldsymbol{\omega}_i)) = o_p(\rho_{i\omega,T})$ for $g(\omega_i) = T(tr(\boldsymbol{S}(\boldsymbol{\omega}_i)) - \Delta(\boldsymbol{\omega}_i))$. Therefore, we have $\left\|\boldsymbol{r}_i^H\right\| = o_p(\rho_{i\omega,T})$. From above discussion, we know that $\hat{\boldsymbol{H}}_i = \boldsymbol{H}_i + o_p(1)$ and $\hat{\boldsymbol{h}}_i = \boldsymbol{h}_i + o_p(1)$.

Now, we can discuss Lemma 5 by using above results.

From equation $(C.26)$, we have

$$\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'} - \tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_i^{H'} - \boldsymbol{c}_i^{H}\boldsymbol{H}_i^{-1}\tilde{\boldsymbol{h}}_i\tilde{\boldsymbol{h}}_i^{'} =$$

$$\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'} - \left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'}$$

$$-\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'} = \hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) + \boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right), \tag{C.48}$$

where

$$\begin{aligned}
\boldsymbol{r}_i^D\left(\boldsymbol{\omega}_i\right) = & -\boldsymbol{h}_i\boldsymbol{c}_{1i}^{h'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'} - \left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{c}_{1i}^h\boldsymbol{h}_i^{'} \\
& - \boldsymbol{c}_{1i}^h\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'} - \left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{c}_{1i}^{h'} \\
& - \boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{3i}^{H'} - \boldsymbol{c}_{3i}^H\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'} - \boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{3i}^{H'} - \boldsymbol{c}_{3i}^H\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'} \\
& - \left(\boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'} \\
& - \left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\left(\boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'} = o_p\left(\rho_{i\omega,T}\right).
\end{aligned} \tag{C.49}$$

and

$$\begin{aligned}
\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right) = & \left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)\left(\boldsymbol{h}_i + \boldsymbol{c}_{1i}^h + \boldsymbol{c}_{2i}^h\right)^{'} \\
& - \boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)^{'} - \left(\sum_{q=1}^{3}\boldsymbol{c}_{qi}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^{'} \\
& - \boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}\boldsymbol{H}_i^{-1}\left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right) - \left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'} \\
& - \boldsymbol{c}_{2i}^h\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)^{'} - \left(\boldsymbol{c}_{1i}^H + \boldsymbol{c}_{2i}^H\right)\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}.
\end{aligned} \tag{C.50}$$

Next, calculating the expectation of $\hat{\boldsymbol{D}}_i\left(\omega\right)$.

$$\begin{aligned}
& E\left(\boldsymbol{h}_i\boldsymbol{h}_i^{'}|\boldsymbol{z}_{i,t}\right) = E\left(\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i|\boldsymbol{z}_{i,t}\right) = \sigma_{u,i}^2\boldsymbol{H}_i. \\
& E\left(\boldsymbol{h}_i\boldsymbol{c}_{1i}^{h'}|\boldsymbol{z}_{i,t}\right) = E\left(\left(-\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\left(\boldsymbol{I}_T - \boldsymbol{P}_i\right)\right)\boldsymbol{G}_i/T|\boldsymbol{z}_{i,t}\right) = -\sigma_{u,i}^2\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T - \boldsymbol{P}_i\right)\boldsymbol{G}_i/T. \\
& E\left(\boldsymbol{c}_{1i}^h\boldsymbol{h}_i^{'}|\boldsymbol{z}_{i,t}\right) = E\left(-\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T - \boldsymbol{P}_i\right)\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i/T|\boldsymbol{z}_{i,t}\right) = -\sigma_{u,i}^2\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T - \boldsymbol{P}_i\right)\boldsymbol{G}_i/T.
\end{aligned} \tag{C.51}$$

By Lemma 7 (vii), we have

$$\begin{aligned}
E\left(\boldsymbol{h}_i \boldsymbol{c}_{2i}^{h'}|\boldsymbol{z}_{i,t}\right) &= E\left(\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i/T|\boldsymbol{z}_{i,t}\right) \\
&= E\left(u_{i,1}^2 e_{1i,t}\right)\sum_{t=1}^{T}\boldsymbol{g}_{i,t}^{'}\boldsymbol{P}_{i,tt}/T = O_p\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^{+}/T\right)
\end{aligned} \tag{C.52}$$

$$E\left(\boldsymbol{c}_{2i}^{h}\boldsymbol{h}_i^{'}|\boldsymbol{z}_{i,t}\right) = E\left(\boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i/T|\boldsymbol{z}_{i,t}\right) = O_p\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i/T\right). \tag{C.53}$$

$$\begin{aligned}
E\left(\boldsymbol{c}_{1i}^{h}\boldsymbol{c}_{1i}^{h'}|\boldsymbol{z}_{i,t}\right) &= E\left(\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{u}_i\boldsymbol{u}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i|\boldsymbol{z}_{i,t}\right) \\
&= \sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}.
\end{aligned} \tag{C.54}$$

By Lemma 7 $(viii)$, we have

$$\begin{aligned}
&E\left(\boldsymbol{c}_{1i}^{h}\boldsymbol{c}_{2i}^{h'}|\boldsymbol{z}_{i,t}\right) = -E\left(\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i/T|\boldsymbol{z}_{i,t}\right) \\
&= -\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{\mu}_i(\boldsymbol{\omega}_i)/T.
\end{aligned} \tag{C.55}$$

Again, we have

$$E\left(\boldsymbol{c}_{2i}^{h}\boldsymbol{c}_{1i}^{h'}|\boldsymbol{z}_{i,t}\right) = -\boldsymbol{\mu}_i^{'}(\boldsymbol{\omega}_i)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i/T. \tag{C.56}$$

By Lemma 7 $(viii)$, we have

$$\begin{aligned}
E\left(\boldsymbol{c}_{2i}^{h}\boldsymbol{c}_{2i}^{h'}|\boldsymbol{z}_{i,t}\right) &= E\left(\frac{\boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i}{T}|\boldsymbol{z}_{i,t}\right) \\
&= \boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\frac{\left(\boldsymbol{j}^{'}\boldsymbol{\omega}_i\right)^2}{T} + \left(\sigma_{e,i}^2\boldsymbol{\Sigma}_{e,i}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\right)\frac{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i^{'}\right)}{T} \\
&+ Cum\left(u_{i,t},u_{i,t},\boldsymbol{e}_{i,t},\boldsymbol{e}_{i,t}^{'}\right)\sum_{t=1}^{T}\left(\boldsymbol{P}_{i,tt}\right)^2,
\end{aligned} \tag{C.57}$$

where $Cum\left(u_{i,t},u_{i,t},\boldsymbol{e}_{i,t},\boldsymbol{e}_{i,t}^{'}\right) = E\left(u_{i,t}^2\boldsymbol{e}_{i,t}\boldsymbol{e}_{i,t}^{'}\right) - \sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i} - 2\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}$.

$$\begin{aligned}
E\left(\boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{1i}^{H}|\boldsymbol{z}_{i,t}\right) &= -E\left(\frac{\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i\boldsymbol{H}_i^{-1}\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T^2}|\boldsymbol{z}_{i,t}\right) \\
&= -\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}.
\end{aligned} \tag{C.58}$$

Also,

$$E\left(\boldsymbol{c}_{1i}^{H}\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^{'}|\boldsymbol{z}_{i,t}\right)=-\sigma_{u,i}^{2}\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i/T. \tag{C.59}$$

By Lemma 7 (xii), we have

$$\begin{aligned}E\left(\boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^{H}|\boldsymbol{z}_{i,t}\right)&=E\left(\frac{\boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\left(\boldsymbol{E}_i^{'}\boldsymbol{G}_i+\boldsymbol{G}_i^{'}\boldsymbol{E}_i\right)}{T}|\boldsymbol{z}_{i,t}\right)\\&=O_p\left(T^{-1}\right).\end{aligned} \tag{C.60}$$

and

$$E\left(\boldsymbol{c}_{2i}^{H}\boldsymbol{H}_i^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^{'}|\boldsymbol{z}_{i,t}\right)=O_p\left(T^{-1}\right). \tag{C.61}$$

Also,

$$\begin{aligned}E\left(\boldsymbol{h}_i\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{3i}^{H}|\boldsymbol{z}_{i,t}\right)&=E\left(\frac{\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i\boldsymbol{H}_i^{-1}\boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i}{T^2}|\boldsymbol{z}_{i,t}\right)\\&=\sigma_{u,i}^{2}\boldsymbol{\Sigma}_{e,i}\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}+O_p\left(T^{-1}\right),\end{aligned} \tag{C.62}$$

by Lemma 7 (ix).

Next,

$$\begin{aligned}&E\left(\boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{1i}^{H}|\boldsymbol{z}_{i,t}\right)\\&=-E\left(\frac{\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i\boldsymbol{H}_i^{-1}\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T^2}|\boldsymbol{z}_{i,t}\right)\\&=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{P}_{i,tt}E\left(u_{i,t}^{2}\boldsymbol{e}_{i,t}^{'}\right)\boldsymbol{H}_i^{-1}\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}\\&=O_p\left(\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i^{+}/T\right)\Xi(\boldsymbol{\omega}_i)\right)\\&=o_p\left(\rho_{\omega,T}\right),\end{aligned} \tag{C.63}$$

by Lemma 7 (vii) and

$$\begin{aligned}&E\left(\boldsymbol{h}_i\boldsymbol{c}_{2i}^{h'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^{H}|\boldsymbol{z}_{i,t}\right)\\&=E\left(\frac{\boldsymbol{G}_i^{'}\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{P}_i\boldsymbol{E}_i\boldsymbol{H}_i^{-1}\left(\boldsymbol{E}_i^{'}\boldsymbol{G}_i+\boldsymbol{G}_i^{'}\boldsymbol{u}_i\right)}{T^2}|\boldsymbol{z}_{i,t}\right)\\&=O_p\left(T^{-1}\right)+\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{g}_{i,t}^{'}+\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\right),\end{aligned} \tag{C.64}$$

by Lemma 7 (xi).

Similarly,

$$
\begin{aligned}
&E\left(\boldsymbol{c}_{2i}^{h}\boldsymbol{h}_i^{'}\boldsymbol{H}_i^{-1}\boldsymbol{c}_{2i}^{H}|\boldsymbol{z}_{i,t}\right)\\
&=E\left(\frac{\boldsymbol{E}_i^{'}\boldsymbol{P}_i\boldsymbol{u}_i\boldsymbol{u}_i^{'}\boldsymbol{G}_i\boldsymbol{H}_i^{-1}\left(\boldsymbol{E}_i^{'}\boldsymbol{G}_i+\boldsymbol{G}_i^{'}\boldsymbol{u}_i\right)}{T^2}|\boldsymbol{z}_{i,t}\right)\\
&=O_p\left(T^{-1}\right)+\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\left(d\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}+\boldsymbol{\sigma}_{eu,i}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\right).
\end{aligned}
\tag{C.65}
$$

$$
\begin{aligned}
&E\left(\hat{\boldsymbol{D}}_i\left(\boldsymbol{\omega}_i\right)|\boldsymbol{z}_{i,t}\right)=\\
&\sigma_{u,i}^2\boldsymbol{H}_i-2\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}\\
&+E(u_{i1}^2e_{i1})\sum_{t=1}^{T}\boldsymbol{g}_{i,t}^{'}\boldsymbol{P}_{i,tt}/T+\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{P}_{i,tt}E(u_{i,1}^2e_{i,1})/T+\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{\mu}_i\left(\boldsymbol{\omega}_i\right)}{T}\\
&+\frac{\boldsymbol{\mu}_i^{'}(\boldsymbol{\omega}_i)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\frac{\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i\right)^2}{T}+\left(\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\right)\frac{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right)}{T}\\
&+o_p\left(\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\right)+2\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+O_p\left(T^{-1}\right)-2\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}-\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\times\\
&2\left(d\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}+\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}/T+\sum_{t=1}^{T}\left(\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}/T\right)\right)\\
&+o_p\left(\rho_{i\omega,T}\right)=\\
&\sigma_{u,i}^2\boldsymbol{H}_i+\sigma_{u,i}^2\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}+E\left(u_{i,1}^2e_{i,1}\right)\sum_{t=1}^{T}\boldsymbol{g}_{i,t}^{'}\boldsymbol{P}_{i,tt}/T\\
&+\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{P}_{i,tt}E\left(u_{i,1}^2e_{i,1}\right)/T+\frac{\boldsymbol{G}_i^{'}\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{\mu}_i\left(\boldsymbol{\omega}_i\right)}{T}+\frac{\boldsymbol{\mu}_i^{'}\left(\boldsymbol{\omega}_i\right)\left(\boldsymbol{I}_T-\boldsymbol{P}_i\right)\boldsymbol{G}_i}{T}\\
&+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\frac{\left(\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i\right)^2}{T}+\left(\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}\right)\frac{\left(\boldsymbol{\omega}_i^{'}\boldsymbol{\Gamma}_i\boldsymbol{\omega}_i\right)}{T}-\\
&2\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T}\left[\sigma_{u,i}^2\boldsymbol{\Sigma}_{e,i}+d\boldsymbol{\sigma}_{eu,i}\boldsymbol{\sigma}_{eu,i}^{'}+\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\right]-\\
&2\frac{\tilde{\boldsymbol{j}}^{'}\boldsymbol{\omega}_i}{T^2}\left[\sum_{t=1}^{T}\left(\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{g}_{i,t}\boldsymbol{\sigma}_{eu,i}^{'}+\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\boldsymbol{H}_i^{-1}\boldsymbol{\sigma}_{eu,i}\boldsymbol{g}_{i,t}^{'}\right)\right]\\
&+o_p\left(\rho_{i\omega,T}\right).
\end{aligned}
\tag{C.66}
$$

C.3 Endogeneity and strength of IVs

From the Monte Carlo results of Lee and Shin (2020) and Kuersteiner and Okui (2010), we can find that the mean square error (MSE) and median absolute deviation (MAD) of model averaging 2SLS estimator is slightly larger than 2SLS estimator with all avaialbe instruments when endogeneity is low. It is happen because the weight estimation often give the outliers in model average 2SLS estimator.

Here, we would like to see the endogeneity of the model. For ease of discussion, we repeat the model (5.19)

$$\begin{aligned} y_{i,t} &= \phi_i y_{i,t-1} + \boldsymbol{x}_{i,t}^{'}\boldsymbol{\beta}_i + u_{i,t} = \boldsymbol{w}_{i,t}^{'}\boldsymbol{\theta}_i + u_{i,t}, \\ \boldsymbol{w}_{i,t} &= \left(y_{i,t-1}, \boldsymbol{x}_{i,t}^{'}\right)^{'} = \boldsymbol{g}(\boldsymbol{z}_{i,t}) + \boldsymbol{e}_{i,t}, \end{aligned} \tag{C.67}$$

where $u_{i,t} = \boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t} + \varepsilon_{i,t}$ and $\boldsymbol{x}_{i,t} = \boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t} + \boldsymbol{v}_{i,t}$. From the model, we can see the endogeneity come from $Cov(y_{i,t-1}, u_{i,t})$. By continuous substitution, $y_{i,t-1}$ can be expressed as

$$\begin{aligned} y_{i,t-1} &= \boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j \boldsymbol{x}_{i,t-j-1} + \sum_{j=0}^{\infty}\phi_i^j u_{i,t-j-1} \\ &= \boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j \left(\boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t-j-1} + \boldsymbol{v}_{i,t-j-1}\right) + \sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t-j-1} + \varepsilon_{i,t-j-1}\right) \end{aligned} \tag{C.68}$$

Stacking the T observations for each i, we have

$$\begin{aligned} \boldsymbol{y}_{i,-1} &= \sum_{j=0}^{\infty}\phi_i^j \boldsymbol{X}_{i,-j-1}\boldsymbol{\beta}_i + \sum_{j=0}^{\infty}\phi_i^j \boldsymbol{u}_{i,-j-1} \\ &= \sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{F}_{x,-j-1}\boldsymbol{\Gamma}_{x,i} + \boldsymbol{V}_{i,-j-1}\right)\boldsymbol{\beta}_i + \sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{F}_{y,-j-1}\gamma_{y,i} + \boldsymbol{\varepsilon}_{i,-j-1}\right) \end{aligned} \tag{C.69}$$

From above equation, we can see that the lag defactor regressors are the feasible instrumental variables for $\boldsymbol{y}_{i,-1}$. As well as, we can show that

$$E\left(\boldsymbol{X}_{i,-1}^{'}\boldsymbol{M}_{F_{x,-1}}\boldsymbol{u}_i\right) = \ldots = E\left(\boldsymbol{X}_{i,-\infty}^{'}\boldsymbol{M}_{F_{x,-\infty}}\boldsymbol{u}_i\right) = 0 \tag{C.70}$$

Assume $\boldsymbol{x}_{i,t}$ and $u_{i,t}$ are uncorrelated and $\boldsymbol{f}_{y,t} = \rho_f \boldsymbol{f}_{y,t-1} + \boldsymbol{\zeta}_t$, we can see the degree of endogeneity is controlled by

$$
\begin{aligned}
&E\left(y_{i,t-1}u_{i,t}\right) = E\left[\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t-j-1}+\boldsymbol{v}_{i,t-j-1}\right)+\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t-j-1}+\varepsilon_{i,t-j-1}\right)\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t}+\varepsilon_{i,t}\right)\right]\\
&=E\Bigg[\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t-j-1}+\boldsymbol{v}_{i,t-j-1}\right)+\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t-j-1}+\varepsilon_{i,t-j-1}\right)\right)\times\\
&\left(\boldsymbol{\gamma}_{y,i}^{'}\left(\rho_f\boldsymbol{f}_{y,t-1}+\boldsymbol{\zeta}_{y,t}\right)+\varepsilon_{i,t}\right)\Bigg]\\
&=E\Bigg[\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{x,i}^{'}\boldsymbol{f}_{x,t-j-1}+\boldsymbol{v}_{i,t-j-1}\right)+\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{f}_{y,t-j-1}+\varepsilon_{i,t-j-1}\right)\right)\times\\
&\left(\boldsymbol{\gamma}_{y,i}^{'}\left(\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{t-j-1}\right)+\varepsilon_{i,t}\right)\Bigg]\\
&=E\Bigg[\left(\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{xi}^{'}\sum_{s=0}^{\infty}\rho_f^s\boldsymbol{\zeta}_{x,t-j-s-1}+\boldsymbol{v}_{i,t-j-1}\right)\right)+\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{yi}^{'}\sum_{s=0}^{\infty}\rho_f^s\boldsymbol{\zeta}_{y,t-j-s-1}+\varepsilon_{i,t-j-1}\right)\right)\\
&\left(\boldsymbol{\gamma}_{y,i}^{'}\left(\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{t-j-1}\right)+\varepsilon_{i,t}\right)\Bigg]\\
&=E\left[\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{xi}^{'}\left(\boldsymbol{\zeta}_{x,t-j-1}+\rho_f\boldsymbol{\zeta}_{x,t-j-2}+\ldots\right)+\boldsymbol{v}_{i,t-j-1}\right)\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{y,t-j-1}+\varepsilon_{i,t}\right)\right]\\
&+E\left[\left(\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{yi}^{'}\left(\boldsymbol{\zeta}_{y,t-j-1}+\rho_f\boldsymbol{\zeta}_{y,t-j-2}+\ldots\right)+\varepsilon_{i,t-j-1}\right)\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{y,t-j-1}+\varepsilon_{i,t}\right)\right]\\
&=A+B.
\end{aligned}
\tag{C.71}
$$

$$
\begin{aligned}
A&=E\left[\left(\boldsymbol{\beta}_i^{'}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\Gamma}_{x,i}^{'}\left(\boldsymbol{\zeta}_{x,t-j-1}+\rho_f\boldsymbol{\zeta}_{x,t-j-2}+\ldots\right)+\boldsymbol{v}_{i,t-j-1}\right)\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{y,t-j-1}+\varepsilon_{i,t}\right)\right]\\
&=E\left[\left(\beta_{1i}\sum_{j=0}^{\infty}\phi_i^j\left(\boldsymbol{\gamma}_{x1,i}^{'}\left(\boldsymbol{\zeta}_{t-j-1}+\rho_f\boldsymbol{\zeta}_{t-j-2}+\ldots\right)+v_{i,t-j-1}\right)\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{t-j-1}+\varepsilon_{i,t}\right)\right]\\
&=E\left[\beta_{1i}\sum_{j=0}^{\infty}\phi_i^j\rho_f^j\left(\boldsymbol{\gamma}_{x1,i}^{'}\boldsymbol{\zeta}_{t-j-1}\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{\zeta}_{t-j-1}\right)+\beta_{1i}\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+1}\left(\boldsymbol{\gamma}_{x1,i}^{'}\boldsymbol{\zeta}_{t-j-2}\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{\zeta}_{t-j-2}\right)+\ldots\right]\\
&=E\left[\beta_{1i}\sum_{s=0}^{\infty}\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+s}\left(\boldsymbol{\gamma}_{x1,i}^{'}\boldsymbol{\zeta}_{t-j-s-1}\right)\left(\boldsymbol{\gamma}_{y,i}^{'}\boldsymbol{\zeta}_{t-j-s-1}\right)\right],
\end{aligned}
\tag{C.72}
$$

where $\boldsymbol{\gamma}_{x1,i}$ is a m_x by 1 vector of factor loading in $x_{1i,t}$.

$$
\begin{aligned}
B &= E\left[\left(\sum_{j=0}^{\infty}\phi_i^j\left(\gamma_{y,i}^{'}\left(\boldsymbol{\zeta}_{y,t-j-1}+\rho_f\boldsymbol{\zeta}_{y,t-j-2}+\ldots\right)+\varepsilon_{i,t-j-1}\right)\right)\left(\gamma_{y,i}^{'}\sum_{j=0}^{\infty}\rho_f^j\boldsymbol{\zeta}_{y,t-j-1}+\varepsilon_{i,t}\right)\right] \\
&= E\left[\sum_{j=0}^{\infty}\phi_i^j\rho_f^j\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{y,t-j-1}\right)^2+\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+1}\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{y,t-j-2}\right)^2+\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+2}\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{y,t-j-3}\right)^2+\ldots\right] \\
&= E\left[\sum_{s=0}^{\infty}\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+s}\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{y,t-j-s-1}\right)^2\right].
\end{aligned}
\tag{C.73}
$$

$$
\begin{aligned}
&E\left(y_{i,t-1}u_{i,t}\right) = A+B = \\
&E\left[\beta_{1i}\sum_{s=0}^{\infty}\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+s}\left(\gamma_{x1,i}^{'}\boldsymbol{\zeta}_{t-j-s-1}\right)\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{t-j-s-1}\right)\right]+E\left[\sum_{s=0}^{\infty}\sum_{j=0}^{\infty}\phi_i^j\rho_f^{j+s}\left(\gamma_{y,i}^{'}\boldsymbol{\zeta}_{y,t-j-s-1}\right)^2\right].
\end{aligned}
\tag{C.74}
$$

From above equation, we observe the endogeneity increase when factors more persistent, correlation between the factor loadings increasing or $|\phi_i|$ increasing.

Next, we investigate the strength of IVs. The strength of IVs is controlled by

$$
\begin{aligned}
E\left(y_{i,t-1}v_{\ell i,t-l}\right) &= E\left(\beta_{\ell i}\sum_{j=0}^{\infty}\phi_i^j x_{\ell i,t-j-1}+\sum_{j=0}^{\infty}\phi_i^j u_{i,t-j-1}\right)v_{\ell i,t-l} \\
&= \beta_{\ell i}\sum_{j=0}^{\infty}\phi_i^j v_{\ell i,t-1-j}v_{\ell i,t-l}
\end{aligned}
\tag{C.75}
$$

Thus, the strength of IVs increase when $|\phi_i|$ more persistent, $v_{\ell i,t-l}$ increasing or $|\beta_{\ell i}|$ increasing.

As the degree of endogeneity is low, Lee and Shin (2020) have shown the bias of MA2SLS is smaller than the 2SLS estimator with all available instruments but the MSE of MA2SLS estimator is larger than 2SLS estimator. In our simulation, we will investigate this situation.

C.4 Supplementary material

In this supplement, we provides Monte Carlo results under different scenario. Table 9 to Table 16 consider dynamic heterogeneous panel data model with $\{\phi\}=\{0.8\}$ and $\{\beta_1,\beta_2\}=\{3,1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Table C.1 (Case A: Low degree of endogeneity, Weak IVs) Bias, RMSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.268	-0.200	-0.248	0.094	0.049	0.026	5.4	5.3	5.3	87.4	98.0	99.9
50	-0.079	-0.094	-0.062	0.069	0.034	0.017	5.8	5.6	4.4	96.0	99.9	100.0
100	-0.034	0.017	-0.060	0.060	0.029	0.015	6.7	5.0	5.5	98.4	100.0	100.0
P-MA2SLSMG^a												
25	-0.344	-0.249	-0.285	0.079	0.040	0.020	6.5	6.2	6.7	92.6	99.5	100.0
50	-0.252	-0.202	-0.261	0.066	0.031	0.016	6.7	5.8	5.1	96.3	100.0	100.0
100	-0.062	-0.116	-0.027	0.056	0.030	0.015	5.8	6.4	5.3	99.1	100.0	100.0
Ps-MA2SLSMG^a												
25	-0.154	-0.122	-0.135	0.108	0.060	0.041	6.6	7.0	5.1	86.7	96.5	98.5
50	-0.055	-0.069	-0.004	0.068	0.035	0.021	5.4	4.7	5.4	96.2	99.3	99.9
100	0.090	0.003	-0.045	0.059	0.031	0.015	6.3	6.1	4.9	98.4	99.9	100.0
2SLSMG^a												
25	-0.344	-0.249	-0.285	0.079	0.040	0.020	6.5	6.2	6.7	92.6	99.5	100.0
50	-0.252	-0.202	-0.261	0.066	0.031	0.016	6.7	5.8	5.1	96.3	100.0	100.0
100	-0.062	-0.116	-0.026	0.056	0.030	0.015	5.8	6.4	5.3	99.1	100.0	100.0

Results for β_1.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.708	-0.031	-0.028	0.607	0.287	0.137	6.6	5.2	4.5	30.9	54.0	76.9
50	-0.025	-0.140	0.036	0.214	0.123	0.056	4.7	4.9	4.6	58.9	80.0	97.3
100	0.015	0.143	0.047	0.121	0.058	0.030	4.7	4.4	4.1	80.9	97.8	100.0
P-MA2SLSMG^a												
25	-0.044	0.104	-0.049	0.518	0.259	0.119	6.3	5.3	4.9	34.3	55.6	80.3
50	-0.069	-0.010	-0.055	0.241	0.109	0.059	7.2	4.5	5.5	56.5	83.7	96.9
100	0.113	-0.134	0.013	0.119	0.059	0.030	5.9	4.1	4.4	81.5	97.0	100.0
Ps-MA2SLSMG^a												
25	-0.001	-0.444	0.058	0.563	0.351	0.248	4.6	5.6	5.8	29.8	48.4	73.6
50	-0.150	0.045	0.102	0.217	0.131	0.066	4.9	5.3	4.9	57.1	81.2	96.3
100	0.146	0.017	-0.118	0.130	0.070	0.031	5.2	4.5	4.4	79.6	97.0	99.9
2SLSMG^a												
25	-0.044	0.104	-0.049	0.518	0.259	0.119	6.3	5.3	4.9	34.3	55.6	80.3
50	-0.069	-0.010	-0.055	0.241	0.109	0.059	7.2	4.5	5.5	56.5	83.7	96.9
100	0.113	-0.134	0.013	0.119	0.059	0.030	5.9	4.1	4.4	81.5	97.0	100.0

The DGP is same as that for Table 5.1 except the number of instruments are 6.

Table C.2 (Case A: Low degree of endogeneity, Weak IVs) Bias, RMSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.298	-0.209	-0.332	0.077	0.037	0.019	7.3	5.8	5.9	93.3	99.9	100.0
50	-0.058	-0.183	-0.123	0.061	0.032	0.016	5.9	5.7	5.1	97.0	99.8	100.0
100	-0.099	-0.032	-0.110	0.057	0.028	0.015	6.6	5.9	5.3	98.7	100.0	100.0
P-MA2SLSMG^b												
25	-0.307	-0.288	-0.329	0.064	0.032	0.018	5.8	5.7	7.4	95.8	99.9	100.0
50	-0.280	-0.234	-0.301	0.060	0.029	0.015	7.0	6.1	5.6	97.5	100.0	100.0
100	-0.078	-0.162	-0.072	0.055	0.029	0.014	5.7	6.0	5.3	99.3	100.0	100.0
Ps-MA2SLSMG^b												
25	-0.177	-0.108	-0.164	0.110	0.060	0.042	6.8	6.7	4.8	86.2	96.5	98.3
50	-0.051	-0.094	-0.031	0.070	0.035	0.024	6.2	4.5	5.7	96.3	99.1	99.8
100	0.073	-0.009	-0.054	0.060	0.031	0.015	5.7	5.6	4.4	98.4	99.9	99.9
2SLSMG^b												
25	-0.294	-0.276	-0.317	0.064	0.032	0.017	5.7	5.6	7.6	95.7	99.8	100.0
50	-0.264	-0.220	-0.287	0.060	0.029	0.015	7.0	6.1	5.6	97.6	100.0	100.0
100	-0.069	-0.156	-0.065	0.054	0.029	0.014	5.8	5.9	5.3	99.3	100.0	100.0

Results for β_1.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.553	-0.023	-0.117	0.511	0.236	0.110	7.7	4.9	4.3	34.0	58.2	82.3
50	0.083	-0.118	0.034	0.197	0.118	0.051	4.5	5.2	3.0	61.7	82.2	98.1
100	0.020	0.101	0.046	0.117	0.056	0.029	4.4	4.9	4.0	82.3	98.0	100.0
P-MA2SLSMG^b												
25	0.005	0.013	-0.011	0.449	0.230	0.105	5.6	5.1	4.9	37.9	60.1	85.4
50	-0.087	-0.025	-0.055	0.224	0.099	0.056	7.1	4.2	5.5	58.8	85.5	97.5
100	0.128	-0.122	0.012	0.117	0.057	0.029	5.6	4.3	4.5	82.3	97.4	100.0
Ps-MA2SLSMG^b												
25	0.104	-0.399	0.090	0.569	0.360	0.239	4.2	5.8	5.9	30.3	47.7	73.4
50	-0.164	0.053	0.124	0.219	0.132	0.073	4.8	5.4	5.6	56.6	81.2	96.4
100	0.143	0.025	-0.111	0.132	0.071	0.031	5.3	4.7	4.0	79.1	97.2	99.9
2SLSMG^b												
25	0.013	0.020	0.005	0.447	0.229	0.105	5.6	5.4	4.7	38.0	60.5	85.4
50	-0.081	-0.021	-0.045	0.223	0.099	0.056	7.4	4.2	5.5	59.1	85.9	97.8
100	0.132	-0.118	0.017	0.117	0.057	0.029	5.6	4.3	4.5	82.3	97.3	100.0

The DGP is same as that for Table 5.1 except the number of instruments are 12.

Table C.3 (Case B: High degree of endogeneity, Weak IVs) Bias, RMSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	1.203	1.294	1.184	0.152	0.088	0.067	12.4	13.2	15.3	88.6	98.4	99.8
50	1.114	1.114	1.060	0.102	0.066	0.040	9.6	12.7	14.9	95.0	99.4	99.7
100	0.729	0.789	0.704	0.084	0.044	0.025	8.1	7.2	10.3	96.4	99.7	100.0
P-MA2SLSMG^a												
25	0.950	1.082	1.092	0.095	0.067	0.047	9.7	12.7	18.4	95.6	99.7	100.0
50	0.845	0.834	0.865	0.090	0.044	0.031	8.7	8.3	12.8	96.8	99.7	100.0
100	0.545	0.487	0.624	0.066	0.036	0.022	6.5	7.6	10.1	97.8	100.0	100.0
Ps-MA2SLSMG^a												
25	1.010	1.019	0.860	0.173	0.232	0.059	8.8	9.6	10.8	86.9	94.3	98.1
50	0.936	0.677	0.821	0.193	0.062	0.040	8.5	7.3	9.3	91.6	97.5	99.4
100	0.461	0.518	0.476	0.090	0.110	0.027	6.9	7.1	6.4	94.6	98.6	99.6
2SLSMG^a												
25	0.950	1.082	1.092	0.095	0.067	0.047	9.7	12.7	18.4	95.6	99.7	100.0
50	0.845	0.834	0.865	0.090	0.044	0.031	8.7	8.3	12.8	96.8	99.7	100.0
100	0.545	0.487	0.624	0.066	0.036	0.022	6.5	7.6	10.1	97.8	100.0	100.0

Results for β_1.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	0.034	0.193	0.058	1.315	0.626	0.328	7.0	5.7	5.1	23.8	37.2	55.4
50	0.461	0.229	0.285	0.476	0.288	0.125	4.2	6.4	4.4	40.1	57.6	82.2
100	0.024	0.102	-0.015	0.267	0.145	0.074	6.1	5.1	5.9	53.6	79.5	94.9
P-MA2SLSMG^a												
25	0.369	0.530	0.422	0.834	0.445	0.221	5.2	5.6	3.6	27.2	45.2	66.0
50	0.129	0.425	0.116	0.484	0.204	0.116	6.3	4.8	5.0	40.8	66.0	86.5
100	0.128	-0.180	0.110	0.231	0.112	0.064	4.8	3.2	5.8	56.9	81.5	97.0
Ps-MA2SLSMG^a												
25	0.680	-0.280	0.583	1.369	0.810	0.424	4.7	6.3	5.5	22.5	31.3	54.9
50	-0.049	-0.023	0.178	0.585	0.311	0.205	4.8	5.7	6.5	36.7	56.4	77.3
100	0.158	0.026	-0.030	0.344	0.259	0.081	6.6	5.5	4.8	55.6	76.2	94.2
2SLSMG^a												
25	0.369	0.530	0.422	0.834	0.445	0.221	5.2	5.6	3.6	27.2	45.2	66.0
50	0.129	0.425	0.116	0.484	0.204	0.116	6.3	4.8	5.0	40.8	66.0	86.5
100	0.128	-0.180	0.110	0.231	0.112	0.064	4.8	3.2	5.8	56.9	81.5	97.0

The DGP is same as that for Table 5.1 except the number of instruments are 6.

Table C.4 (Case B: High degree of endogeneity, Weak IVs) Bias, RMSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	1.426	1.527	1.350	0.117	0.081	0.057	16.9	21.7	24.7	97.1	99.8	100.0
50	1.654	1.531	1.592	0.099	0.067	0.051	14.4	20.4	29.9	99.3	100.0	100.0
100	1.194	1.278	1.229	0.079	0.050	0.035	10.3	13.7	22.5	99.2	100.0	100.0
P-MA2SLSMG^b												
25	1.241	1.403	1.380	0.092	0.072	0.055	13.1	18.1	25.9	99.1	100.0	100.0
50	1.295	1.263	1.299	0.087	0.051	0.043	13.2	14.4	25.5	99.3	100.0	100.0
100	1.044	0.927	1.040	0.070	0.040	0.028	9.7	10.5	17.8	99.5	100.0	100.0
Ps-MA2SLSMG^b												
25	1.086	0.988	0.845	0.183	0.235	0.062	8.9	9.3	10.0	86.6	93.8	97.8
50	0.984	0.691	0.824	0.198	0.079	0.042	7.9	7.1	9.2	91.1	97.2	99.3
100	0.499	0.471	0.473	0.094	0.081	0.028	7.2	6.7	5.7	94.5	98.5	99.5
2SLSMG^b												
25	1.268	1.434	1.413	0.093	0.074	0.056	13.2	18.6	26.1	99.2	100.0	100.0
50	1.344	1.309	1.350	0.088	0.052	0.044	13.2	15.0	26.9	99.3	100.0	100.0
100	1.102	0.991	1.101	0.071	0.042	0.030	10.1	11.0	18.7	99.6	100.0	100.0

Results for β_1.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.033	0.223	0.116	0.858	0.433	0.189	7.2	5.8	4.7	24.9	45.5	68.8
50	0.340	0.260	0.176	0.395	0.232	0.098	5.0	5.8	4.8	42.4	64.9	87.7
100	0.014	0.156	0.044	0.235	0.127	0.064	5.0	5.4	5.8	57.3	83.2	97.7
P-MA2SLSMG^b												
25	0.348	0.377	0.303	0.661	0.392	0.189	4.7	6.5	4.2	29.8	48.4	70.5
50	0.088	0.351	0.189	0.438	0.196	0.102	7.0	4.1	4.8	40.8	69.5	89.5
100	0.182	-0.111	0.131	0.225	0.104	0.060	5.3	3.7	5.5	59.7	82.6	97.4
Ps-MA2SLSMG^b												
25	0.725	-0.188	0.603	1.366	0.819	0.486	4.3	6.4	5.9	23.0	31.6	53.5
50	0.064	0.031	0.215	0.597	0.366	0.221	4.7	5.2	6.7	36.9	57.0	76.4
100	0.239	0.037	-0.037	0.367	0.240	0.082	7.0	5.3	4.3	55.8	76.2	93.8
2SLSMG^b												
25	0.381	0.367	0.304	0.665	0.394	0.190	4.7	6.3	4.6	29.9	48.4	70.8
50	0.099	0.363	0.199	0.438	0.198	0.103	7.0	4.0	5.2	41.0	69.5	89.1
100	0.198	-0.096	0.138	0.228	0.105	0.060	5.5	3.9	5.7	59.9	83.0	97.3

The DGP is same as that for Table 5.1 except the number of instruments are 12.

Table C.5 (Case C: Low degree of endogeneity, Strong IVs) Bias, RMSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-0.173	-0.124	-0.283	0.098	0.052	0.027	6.3	5.3	4.8	86.5	97.4	99.6
50	0.000	-0.089	-0.015	0.063	0.035	0.017	5.9	6.9	6.0	97.2	99.8	100.0
100	-0.069	0.007	-0.038	0.058	0.028	0.014	6.1	5.0	4.6	98.2	100.0	100.0
P-MA2SLSMG^a												
25	-0.317	-0.256	-0.266	0.075	0.040	0.019	5.6	5.8	5.0	91.0	99.5	100.0
50	-0.142	-0.115	-0.187	0.062	0.029	0.015	6.6	6.0	5.3	97.5	100.0	100.0
100	0.019	-0.077	0.011	0.055	0.029	0.013	5.4	5.6	5.2	98.8	100.0	100.0
Ps-MA2SLSMG^a												
25	-0.037	0.134	-0.166	0.127	0.083	0.037	6.5	5.8	6.7	82.7	96.5	98.4
50	-0.065	-0.047	0.016	0.078	0.035	0.020	7.0	5.9	5.4	95.0	99.7	99.5
100	0.046	0.021	-0.038	0.061	0.029	0.014	6.3	5.4	4.3	97.2	99.9	100.0
2SLSMG^a												
25	-0.317	-0.256	-0.266	0.075	0.040	0.019	5.6	5.8	5.0	91.0	99.5	100.0
50	-0.142	-0.115	-0.187	0.062	0.029	0.015	6.6	6.0	5.3	97.5	100.0	100.0
100	0.019	-0.077	0.011	0.055	0.029	0.013	5.4	5.6	5.2	98.8	100.0	100.0

Results for β_1.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	0.433	0.475	0.173	2.551	1.146	0.616	5.4	4.3	4.3	15.4	20.9	33.8
50	0.577	0.155	0.337	0.635	0.326	0.150	6.7	5.1	5.5	35.8	50.5	76.3
100	0.163	0.220	0.201	0.205	0.100	0.054	4.3	5.1	4.9	63.5	88.4	98.9
P-MA2SLSMG^a												
25	0.790	0.178	0.834	1.716	0.874	0.443	6.9	5.6	4.8	18.6	24.7	42.5
50	-0.050	0.380	0.301	1.072	0.266	0.137	5.3	5.3	5.4	36.4	57.5	81.5
100	0.126	0.041	0.107	0.205	0.095	0.047	6.1	4.8	4.7	65.2	89.0	99.2
Ps-MA2SLSMG^a												
25	-0.464	-0.507	0.470	4.769	3.533	1.562	5.1	4.8	4.9	11.8	14.8	24.6
50	-0.183	-0.035	0.359	1.042	0.498	0.584	5.2	4.3	4.9	29.0	43.7	67.4
100	0.150	-0.038	0.052	0.265	0.126	0.070	6.1	5.6	5.2	58.5	81.4	96.7
2SLSMG^a												
25	0.790	0.178	0.834	1.716	0.874	0.443	6.9	5.6	4.8	18.6	24.7	42.5
50	-0.050	0.380	0.301	1.072	0.266	0.137	5.3	5.3	5.4	36.4	57.5	81.5
100	0.126	0.041	0.107	0.205	0.095	0.047	6.1	4.8	4.7	65.2	89.0	99.2

The DGP is same as that for Table 5.1 except the number of instruments are 6.

Table C.6 (Case C: Low degree of endogeneity, Strong IVs) Bias, RMSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-0.367	-0.330	-0.450	0.073	0.038	0.021	6.1	5.9	6.4	93.7	99.6	100.0
50	-0.004	-0.107	-0.021	0.061	0.031	0.016	5.7	5.9	5.8	97.2	99.7	100.0
100	-0.063	0.019	-0.036	0.057	0.028	0.014	5.9	5.5	5.0	98.6	100.0	100.0
P-MA2SLSMG^b												
25	-0.390	-0.345	-0.389	0.066	0.035	0.017	5.6	6.0	6.5	94.7	99.8	100.0
50	-0.205	-0.142	-0.222	0.059	0.028	0.015	6.7	5.8	5.6	98.1	100.0	100.0
100	-0.015	-0.105	-0.010	0.054	0.029	0.013	5.0	5.8	5.1	98.9	100.0	100.0
Ps-MA2SLSMG^b												
25	-0.162	-0.071	-0.260	0.117	0.081	0.032	6.9	5.6	6.2	84.5	96.7	99.1
50	-0.118	-0.139	-0.076	0.068	0.032	0.019	6.3	6.1	4.8	95.9	99.7	99.7
100	0.001	-0.041	-0.082	0.059	0.029	0.014	6.2	5.5	4.5	97.9	99.9	100.0
2SLSMG^b												
25	-0.392	-0.354	-0.395	0.065	0.035	0.017	5.8	5.8	6.0	94.9	99.8	100.0
50	-0.211	-0.153	-0.230	0.059	0.028	0.015	6.7	5.7	5.6	98.1	100.0	100.0
100	-0.025	-0.112	-0.017	0.054	0.029	0.013	5.2	5.7	5.0	99.1	100.0	100.0

Results for β_1.

	Bias(×100)			RMSE(×100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	0.390	0.744	0.610	1.391	0.609	0.346	5.3	4.6	5.5	19.3	30.4	50.8
50	0.554	0.323	0.408	0.465	0.216	0.109	6.4	4.5	4.9	42.8	61.7	86.8
100	0.290	0.220	0.184	0.173	0.086	0.047	5.0	4.5	5.4	71.4	91.6	99.3
P-MA2SLSMG^b												
25	0.949	0.408	0.941	1.204	0.614	0.318	6.2	5.3	5.8	21.9	33.1	54.7
50	0.356	0.395	0.320	0.383	0.200	0.106	5.7	5.3	5.9	41.9	65.0	88.9
100	0.222	0.124	0.159	0.179	0.083	0.040	6.0	4.6	4.0	71.2	92.3	99.5
Ps-MA2SLSMG^b												
25	-0.430	0.270	1.092	4.191	3.919	1.349	5.2	4.7	6.0	12.9	17.7	29.9
50	0.055	0.249	0.586	0.807	0.466	0.426	4.9	4.3	5.1	31.3	50.2	72.2
100	0.263	0.100	0.156	0.238	0.116	0.062	5.8	4.9	5.4	63.4	85.2	98.0
2SLSMG^b												
25	0.968	0.414	0.954	1.188	0.599	0.314	6.4	5.3	6.0	22.3	33.0	55.2
50	0.382	0.424	0.329	0.381	0.198	0.103	5.7	5.8	5.7	42.3	65.8	89.6
100	0.250	0.142	0.185	0.178	0.083	0.040	5.9	4.8	3.9	72.1	92.7	99.5

The DGP is same as that for Table 5.1 except the number of instruments are 12.

Table C.7 (Case D: High degree of endogeneity, Strong IVs) Bias, RMSE of MA2SLSMG^a, P-MA2SLSMG^a, Ps-MA2SLSMG^a, 2SLSMG^a estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.												
	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	1.114	1.149	0.992	0.158	0.092	0.062	9.6	13.4	13.1	88.7	96.3	99.2
50	1.217	1.106	1.116	0.128	0.069	0.040	10.8	12.9	14.8	94.5	99.0	100.0
100	0.601	0.686	0.709	0.078	0.043	0.026	7.1	7.9	12.1	95.9	99.5	99.9
P-MA2SLSMG^a												
25	0.872	0.899	1.026	0.101	0.058	0.042	9.3	11.7	14.6	95.0	99.7	100.0
50	0.882	0.783	0.838	0.085	0.043	0.031	10.1	8.5	12.8	96.8	99.8	100.0
100	0.576	0.510	0.636	0.068	0.034	0.022	7.4	6.5	10.7	97.0	100.0	100.0
Ps-MA2SLSMG^a												
25	0.804	1.071	0.874	0.187	0.129	0.057	9.5	9.3	9.0	83.5	93.9	98.6
50	0.622	0.780	0.738	0.156	0.073	0.063	7.0	7.6	11.2	91.6	97.7	98.4
100	0.474	0.421	0.345	0.100	0.046	0.027	6.7	6.8	6.4	93.9	98.5	99.4
2SLSMG^a												
25	0.872	0.899	1.026	0.101	0.058	0.042	9.3	11.7	14.6	95.0	99.7	100.0
50	0.882	0.783	0.838	0.085	0.043	0.031	10.1	8.5	12.8	96.8	99.8	100.0
100	0.576	0.510	0.636	0.068	0.034	0.022	7.4	6.5	10.7	97.0	100.0	100.0

Results for β_1.												
	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^a												
25	-3.259	-0.650	-0.557	12.646	5.252	2.623	6.9	6.3	5.3	8.2	9.7	13.2
50	-1.717	-1.832	-2.341	5.085	2.648	1.338	5.1	4.6	5.2	8.4	11.7	15.9
100	-1.279	-1.988	-1.609	2.830	1.376	0.716	6.0	6.7	6.1	13.1	18.6	28.4
P-MA2SLSMG^a												
25	-0.561	-1.496	-1.266	6.404	3.454	1.796	5.7	5.6	5.6	9.5	11.1	13.5
50	-2.029	-0.224	-1.695	4.139	1.884	1.084	5.8	4.9	5.5	10.0	15.6	20.2
100	-1.239	-1.850	-1.745	2.065	1.132	0.518	4.1	6.5	4.4	13.3	19.0	29.6
Ps-MA2SLSMG^a												
25	-0.545	-1.788	0.026	21.612	9.841	4.872	5.2	4.1	4.7	7.6	7.1	10.9
50	-0.168	-2.386	-1.024	9.862	3.787	4.215	6.1	5.5	4.5	9.2	9.1	17.3
100	-0.787	-1.221	-1.476	5.629	1.775	1.510	6.8	5.0	5.7	13.0	17.2	28.0
2SLSMG^a												
25	-0.561	-1.496	-1.266	6.404	3.454	1.796	5.7	5.6	5.6	9.5	11.1	13.5
50	-2.029	-0.224	-1.695	4.139	1.884	1.084	5.8	4.9	5.5	10.0	15.6	20.2
100	-1.239	-1.850	-1.745	2.065	1.132	0.518	4.1	6.5	4.4	13.3	19.0	29.6

The DGP is same as that for Table 5.1 except the number of instruments are 6.

Table C.8 (Case D: High degree of endogeneity, Strong IVs) Bias, RMSE of MA2SLSMG^b, P-MA2SLSMG^b, Ps-MA2SLSMG^b, 2SLSMG^b estimates and Size (%) and power (%) of the associated t-tests for the dynamic heterogeneous panel data model with $\{\phi, \beta_1, \beta_2\} = \{0.8, 3, 1\}$, correlated factor loadings in $x_{1i,t}$ and $u_{i,t}$.

Results for ϕ.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	1.121	1.165	0.931	0.095	0.062	0.039	12.2	15.8	17.8	97.4	99.8	100.0
50	1.585	1.444	1.476	0.099	0.062	0.046	14.9	18.2	28.0	98.5	100.0	100.0
100	1.122	1.192	1.184	0.076	0.048	0.033	10.9	13.2	20.8	99.3	100.0	100.0
P-MA2SLSMG^b												
25	0.954	1.055	1.078	0.082	0.055	0.039	10.2	14.2	20.5	98.5	99.9	100.0
50	1.221	1.107	1.206	0.083	0.045	0.039	12.1	12.9	21.6	99	100.0	100.0
100	0.950	0.847	0.970	0.068	0.038	0.026	9.6	9.7	15.9	99.6	100.0	100.0
Ps-MA2SLSMG^b												
25	0.779	1.007	0.928	0.182	0.129	0.064	10.4	9.8	9.6	84.7	93.9	98.6
50	0.632	0.781	0.707	0.159	0.074	0.062	7.4	8.0	11.4	91.4	97.8	98.5
100	0.455	0.423	0.338	0.097	0.046	0.027	6.7	6.8	6.2	94.3	98.5	99.5
2SLSMG^b												
25	0.967	1.076	1.093	0.082	0.055	0.039	10.4	14.3	21.1	98.6	99.9	100.0
50	1.248	1.142	1.240	0.083	0.046	0.040	12.4	13.5	22.2	99.0	100.0	100.0
100	0.999	0.888	1.012	0.069	0.038	0.027	9.9	10.2	17.3	99.6	100.0	100.0

Results for β_1.

	Bias($\times$100)			RMSE($\times$100)			Size			Power		
T/N	25	50	100	25	50	100	25	50	100	25	50	100
MA2SLSMG^b												
25	-2.512	-0.842	-1.044	6.574	2.750	1.291	7.2	6.7	5.2	10.3	14.4	18.5
50	-1.889	-2.251	-2.328	3.170	1.697	0.868	5.7	5.1	6.4	11.1	14.7	21.5
100	-1.935	-2.607	-2.462	1.907	1.066	0.540	6.5	6.5	6.8	13.6	17.5	29.7
P-MA2SLSMG^b												
25	-0.414	-1.453	-1.555	4.225	2.544	1.329	6.0	6.0	5.6	11.3	13.6	19.1
50	-2.647	-1.365	-2.137	2.944	1.397	0.869	6.5	4.8	5.9	10.9	16.1	23.0
100	-1.631	-2.158	-2.301	1.659	0.931	0.453	4.8	7.1	5.1	13.5	19.4	30.8
Ps-MA2SLSMG^b												
25	-0.402	-1.640	0.534	21.090	10.174	4.935	4.4	4.4	4.6	7.8	6.9	11.0
50	-0.344	-2.256	-0.762	10.274	3.917	3.750	6.7	5.2	4.2	8.8	9.3	17.5
100	-1.030	-1.088	-1.471	4.764	1.844	1.523	6.9	5.3	5.2	13.0	17.6	28.4
2SLSMG^b												
25	-0.465	-1.547	-1.561	4.164	2.555	1.323	6.0	6.5	5.8	10.9	13.4	18.8
50	-2.617	-1.473	-2.199	2.908	1.382	0.862	6.6	4.8	6.1	11.3	15.9	23.4
100	-1.663	-2.222	-2.344	1.647	0.930	0.453	5.2	7.2	5.3	13.5	19.1	30.8

The DGP is same as that for Table 5.1 except the number of instruments are 12.

Bibliography

Abadie, A., J. Gu, and S. Shen (2019). Instrumental variable estimation with first-stage heterogeneity.

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

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

Anderson, W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. Journal of Econometrics 18, 47–82.

Arellano, M. (1989). A note on the anderson-hsiao estimator for panel data. Economics Letters 31, 337–341.

Arellano, M. and S. Bond (1991). Some test of speci
cation for panel data:monte carlo evidence and an application to employment equation. Review of Economic Studies 58, 277–297.

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

Arioglu, E. and K. Tuan (2014). Speed of adjustment: evidence from borsa istanbul. Borsa Istanbul Review 14(2), 126–131.

Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

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

Bai, J. (2013). Likelihood approach to dynamic panel data models with interactive effects. Working paper.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Barclay, M. J. and C. W. Smith Jr (1995). The maturity structure of corporate debt. the Journal of Finance 50(2), 609–631.

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

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80, 2369–2429.

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregression with unit root and cointegration. Econometric

Theory 21, 795–837.
Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.
Bun, M., M. A. Carree, and A. Juodis (2017). On maximum likelihood estimation of dynamic panel data models. Oxford Bulletin of Economics and Statistics 79, 463–494.
Bun, M. J. G. and F. Windmeijer (2010). The weak instrument problem of the system gmm estimator in dynamic panel data models. Econometrics Journal 13, 95–126.
Byoun, S. (2008). How and when do firms adjust their capital structures toward targets? The Journal of Finance 63(6), 3069–3096.
Chen, X., D. T. Jacho-Chávez, and O. Linton (2016). Averaging of an increasing number of moment condition estimators. Econometric Theory 32, 30–70.
Cheng, H. and Q. Zhou (2018). Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. Journal of Econometrics 207, 114–128.
Chudik, A. and M. H. Pesaran (2013). Large panel data models with cross-sectional dependence: a survey. CAFE Research Paper (13.15).
Chudik, A. and M. H. Pesaran (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188, 393–420.
Chudik, A. and M. H. Pesaran (2017). A bias-corrected method of moments approach to estimation of dynamic short-t panels. USC-INET Research Paper (17-26).
Chudik, A., M. H. Pesaran, et al. (2020). An augmented anderson-hsiao estimator for dynamic short-t panels. Technical report, Federal Reserve Bank of Dallas.
Dang, V. A., M. Kim, and Y. Shin (2012). Asymmetric capital structure adjustments: New evidence from dynamic panel threshold models. Journal of Empirical Finance 19(4), 465–482.
Dang, V. A., M. Kim, and Y. Shin (2014). Asymmetric adjustment toward optimal capital structure: Evidence from a crisis. International Review of Financial Analysis 33, 226–242.
DeAngelo, H. and R. W. Masulis (1980). Leverage and dividend irrelevancy under corporate and personal taxation. The Journal of Finance 35(2), 453–464.
DeAngelo, H. and R. Roll (2015). How stable are corporate capital structures? The Journal of Finance 70(1), 373–418.
Donald, S. G. and W. K. Newey (2001). Choosing the number of instruments. Econometrica 69, 1161–1191.
Drobetz, W. and G. Wanzenried (2006). What determines the speed of adjustment to the target capital structure? Applied Financial Economics 16(13), 941–958.
Fama, E. F. and K. R. French (2002). Testing trade-off and pecking order predictions about dividends and debt. The review of financial studies 15(1), 1–33.
Fan, J., Y. Guo, and Z. Zhu (2020). When is best subset selection the "best"?

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

Fan, J. and H. Peng (2004). Nonconcave penalized likelihood with a diverging number of parameters. The Annals of Statistics 32, 928–961.

Ferri, M. G. and W. H. Jones (1979). Determinants of financial structure: A new methodological approach. The Journal of finance 34(3), 631–644.

Flannery, M. J. and K. P. Rangan (2006). Partial adjustment toward target capital structures. Journal of financial economics 79(3), 469–506.

Graham, J. R. and C. R. Harvey (2001). The theory and practice of corporate finance: Evidence from the field. Journal of financial economics 60(2-3), 187–243.

Hall, T. W. (2012). The collateral channel: Evidence on leverage and asset tangibility. Journal of Corporate Finance 18(3), 570–583.

Halling, M., J. Yu, and J. Zechner (2016). Leverage dynamics over the business cycle. Journal of Financial Economics 122(1), 21–41.

Han, C. and P. C. B. Phillips (2006). Gmm with many moment conditions. Econometrica 74, 147–192.

Hansen, B. E. (2007a). Least squares model averaging. Econometrica 75, 1175–1189.

Hansen, B. E. (2020). Econometrics (draft graduate textbook).

Hansen, C. B. (2007b). Asymptotic properties of a robust variance matrix estimator for panel data when t is large. Journal of Econometrics 141, 597–620.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator in panel ar(p) models when both n and t are large. Econometric Theory 25, 873–890.

Hayakawa, K. and M. H. Pesaran (2015a). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. Journal of Econometrics 188, 111–134.

Hayakawa, K. and M. H. Pesaran (2015b). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. Journal of econometrics 188(1), 111–134.

Hayakawa, K., M. H. Pesaran, and L. V. Smith (2021). Short t dynamic panel data models with individual, time and interactive effects. Working paper.

Hayakawa, K., M. Qi, and J. Breitung (2019). Double filter instrumental variable estimation of panel data models with weakly exogenous variables. Econometric Reviews 38, 1055–1088.

Hayashi, K., P. M. Bentler, and K.-H. Yuan (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. Structural Equation Modeling: A Multidisciplinary Journal 14(3), 505–526.

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

Hsiao, C. (2014). Analysis of panel data. Number 54. Cambridge university press.

Hsiao, C., M. Hashem Pesaran, and A. Kamil Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics 109, 107–150.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of econometrics 109(1), 107–150.

Huang, R. and J. R. Ritter (2009). Testing theories of capital structure and estimating the speed of adjustment. Journal of Financial and Quantitative analysis, 237–271.

Huang, X. (2008). Panel vector autoregression under cross-sectional dependence. Econometrics Journal 11, 219–243.

Iliev, P. and I. Welch (2010). Reconciling estimates of the speed of adjustment of leverage ratios. Available at SSRN 1542691.

Juodis, A. (2018). First difference transformation in panel var models: Robustness, estimation, and inference. Econometric Reviews 37, 650–693.

Kayhan, A. and S. Titman (2007). Firms' histories and their capital structures. Journal of financial Economics 83(1), 1–32.

Kim, E. H. (1978). A mean-variance theory of optimal capital structure and corporate debt capacity. The journal of Finance 33(1), 45–63.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. The Annals of Statistics 28, 1356–1378.

Kraus, A. and R. H. Litzenberger (1973). A state-preference model of optimal financial leverage. The journal of finance 28(4), 911–922.

Krishnaswami, S. and V. Subramaniam (1999). Information asymmetry, valuation, and the corporate spin-off decision. Journal of Financial economics 53(1), 73–112.

Kruiniger, H. (2013). Quasi ml estimation of the panel ar(1) model with arbitrary initial conditions. Journal of Econometrics 173, 175–188.

Kuersteiner, G. and R. Okui (2010). Constructing optimal instruments by first-stage prediction averaging. Econometrica 78, 697–718.

Lee, S. and Y. Shin (2020). Complete subset averaging with many instruments.

Lemmon, M. L., M. R. Roberts, and J. F. Zender (2008). Back to the beginning: persistence and the cross-section of corporate capital structure. The journal of finance 63(4), 1575–1608.

Modigliani, F. and M. H. Miller (1958). The cost of capital, corporation finance and the theory of investment. The American economic review 48(3), 261–297.

Modigliani, F. and M. H. Miller (1963). Corporate income taxes and the cost of capital: a correction. The American economic review 53(3), 433–443.

Moon, H. R. and P. C. B. Phillips (2000). Estimation of autoregressive roots near unity using panel data. Econometric Theory 16, 927–997.

Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33, 158–195.

Mutl, J. (2009). Panel var models with spatial dependence. Working paper.

Myers, S. C. (1977). Determinants of corporate borrowing. Journal of financial economics 5(2), 147–175.

Myers, S. C. (1984). Capital structure puzzle. Technical report, National Bureau of Economic Research.

Myers, S. C. and N. S. Majluf (1984). Corporate financing and investment decisions when firms have informationthat investors do not have. Technical report, National Bureau of Economic Research.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, 1–32.

Ng, S. and J. Bai (2009). Selecting instrumental variables in a data rich environment. Journal of Time Series Econometrics 1.

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

Norkutė, M., V. Sarafidis, T. Yamagata, and G. Cui (2021). Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. Journal of Econometrics 220(2), 416–446.

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

Ozkan, A. (2001). Determinants of capital structure and adjustment to long run target: evidence from uk company panel data. Journal of business finance & accounting 28(1-2), 175–198.

Pesaran, M. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68, 79 – 113.

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

Rajan, R. G. and L. Zingales (1995). What do we know about capital structure? some evidence from international data. The journal of Finance 50(5), 1421–1460.

Ross, S. A., R. Westerfield, and B. D. Jordan (2014). Fundamentals of corporate finance. McGraw Hill LLC.

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. Econometric Reviews 31(5), 483–531.

Scott Jr, J. H. (1976). A theory of optimal capital structure. The Bell Journal of Economics, 33–54.

Westerlund, J., H. Karabiyik, P. K. Narayan, and S. Narayan (2021). Estimating the speed of adjustment of leverage in the presence of interactive effects. Journal of Financial Econometrics.

Windmeijer, F., H. Farbmacher, N. Davies, and G. D. Smith (2019). On the use of the lasso for instrumental variables estimation with some invalid instruments. Journal of the American Statistical Association 114, 1339–1350.