LIML Methods for a Dynamic Structural Model in Long Panels

Kentaro Akashi *

First version August, 2017

Translated version July, 2021

Contents

*Faculty of Economics, Gakushuin University. I am grateful for the valuable comments and joint research with Naoto Kunitomo, and I appreciate the translation support by Etsuko Yokoyama. This study was supported by research grants (2016) from Gakushuin University and published in Japanese by the Journal of Faculty of Economics, Gakushuin University, Vol. 57-58.

1 Introduction

First, this study aims to briefly summarize the recent results of theoretical analysis in long panel data. Hsiao (2014) wrote one of the representative textbooks of panel analysis and the book was translated into Japanese by Naoto Kunitomo. The third edition complements a dynamic panel structural model. Moreover, Arellano (2003a) wrote a textbook regarding dynamic panels, which incorporates many empirical examples. However, in these textbooks, discussion on long panel data is limited, which has been increasing in analysis with the accumulation of data in recent years. Hence, we thought of focusing on long panel data. Long panel data possibly become a problem because recent research showed that existing estimation methods built into packaged software and programs (EViews, Stata, and Ox) may not always work well. Distinguishing between short and long panel data in empirical analyses may be difficult. Therefore, in this study, we will focus on fixed-effects estimation because it can be a consistent estimation regardless of whether panel data are short or long and is robust to the assumption of the individual effect.

Second, this work focuses on endogeneity, which is one of the most important issues in econometric empirical analyses. We apply limited information maximum likelihood (LIML) estimation for structural panel analysis based on the simultaneous equation model, which is often useful in testing economic models. We introduce the backward filter and the long difference, which are relatively new data transformations to exclude individual effects. Then, we present that the two LIML estimators related to the transformations have the best properties in long panel data. Of these, the doubly filtered LIML (D-LIML) estimator can be easily calculated using package software after data transformation.

Third, we present some useful results for the procedures of structural panel analysis. Although theoretical analyses of long panel data have discussed the estimation problem, few studies were conducted for hypothesis testing. In particular, a model selection based on the information criterion, the exogeneity test of instrumental variables, and the rank test for identification are constructed based on the Anderson-Rubin test statistic. These procedures will give a deeper panel analysis.

This paper is further organized as follows. Part I summarizes the results of existing long panel data analyses for regression analysis using a simple panel AR(1) model. Then, Part II considers a general model, discusses the estimation methods of several LIML estimators, and shows the simulation results under a finite sample. Part III proposes the test statistics based on the D-LIML estimator for structural analysis and shows the simulation results. The proofs of our theorems are summarized in the Appendix.

2 Part I: Regression Analysis

Part I provides an overview of the results for the regression model and methods of long panel data, which are also used in Parts II and III. A dynamic panel model is given as follows:

$$y_{it} = \pi y_{it-1} + \eta_i + v_{it} \ , \quad |\pi| < 1 \ . \tag{2.1}$$

The individual effect η_i $(i = 1, \cdots, N)$ is just added to the AR(1) model, but this effect makes the estimation problem difficult. The individual effect is a unique formulation of panel analysis and represents the individual attributes that do not change with time. Regarding the error term, we assume the homoscedastic variance $\mathcal{V}ar[v_{it}] = \omega$, which can be extended to the AR(p) model. However, for simplicity, only the results for AR(1) are summarized in Part I.

Example 1.1 : One of the simplest applications of the reduced form (2.1) is the verification of growth rate convergence in the macro growth theory of Barro and Sala-i-Martin (1995). The value of convergence is $\mathcal{E}[y_{it}] = \eta_i/(1-\pi)$, but it varies from country to country.

Acemoglu et al. (2008) also conducted another influential empirical analysis. They analyzed the relationship between a democratization indicator and logarithmic GDP per $y^{(2)}_{it-1}$ capita using data from more than 100 countries from 1960 to 2000 ($T = 40$),

$$y^{(1)}_{it} = \pi_1 y^{(1)}_{it-1} + \pi_2 y^{(2)}_{it-1} + \eta_i + v_{it} \ .$$

The data were corrected to five-year data. Hence, the number of periods was decreased. Arellano (2003a) provided empirical examples of the reduced form.

Let us compare the dynamic panel and the static panel model,

$$y_{it} = \pi z_{it} + \eta_i + v_{it} \ .$$

In the static model, for an exogenous variable z_{it} that is uncorrelated with the error term, the following hypothesis can be considered:

$$\mathrm{H}_0 : \ \mathcal{E}[z_{it}\eta_i] = 0 \ ,$$

and the test of Hausman (1978) is conducted. However, in the dynamic panel, this hypothesis testing would not hold as shown below.

The first problem with the dynamic model is that although the individual effects are random, the reduced form has the endogeneity problem. Substituting y_{it-1} repeatedly, we obtain the following:

$$\begin{aligned} \mathcal{E}[y_{it-1}\eta_i] &= \mathcal{E}\left[\left(v_{it-1} + \pi v_{it-2} + \cdots + \pi^{t-2} v_{i1} + \pi^{t-1} y_{i0} + \frac{1-\pi^{t-1}}{1-\pi}\eta_i\right)\eta_i\right] \\ &= \mathcal{E}\left[\left(\pi^{t-1} y_{i0} + \frac{1-\pi^{t-1}}{1-\pi}\eta_i\right)\eta_i\right] \\ &\neq 0 \ . \end{aligned}$$

From the above equation, y_{it-1} is a function of η_i. Therefore, this function correlates with the individual effect, and the initial value of the first term is also highly possible to correlate with η_i.

Hence, let us consider a fixed-effects estimation that excludes individual effects. The covariance (CV) estimator is the standard method in static models. This estimator is also called the least square dummy variable estimator or the within groups (WG) estimator. Let the mean of within group as $\bar{y}_i = \sum_t y_{it}$, and we have the following:

$$y_{it} - \bar{y}_i = \pi(y_{it-1} - \bar{y}_{i,-1}) + (v_{it} - \bar{v}_i) \ ,$$

which does not depend on individual effects.

With the application of ordinary least squares (OLS) estimation after data transformation, the CV estimator is given as follows:

$$\tilde{\pi}_{\mathrm{CV}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)(y_{it-1}-\bar{y}_{i,-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it-1}-\bar{y}_{i,-1})^2} .$$

However, in the dynamic panel model,

$$\begin{aligned}\mathcal{E}[(y_{it-1}-\bar{y}_{i,-1})(v_{it}-\bar{v}_i)] &= \mathcal{E}[(y_{it-1}-\bar{y}_{i,-1})(-\bar{v}_i)] \\ &\neq 0 ,\end{aligned}$$

endogeneity will occur because of data transformation, and then, the CV estimator is biased. According to Nickel (1981) and Anderson and Hsiao (1982), the bias is as follows:

$$\begin{aligned}\tilde{\pi}_{\mathrm{CV}} - \pi &\xrightarrow[N\to\infty]{p} \frac{-\frac{1+\pi}{T-1}\left[1-\frac{1}{T}\frac{1-\pi^T}{1-\pi}\right]}{1-\frac{2\pi}{(T-1)(1-\pi)}\left[1-\frac{1-\pi^T}{T(1-\pi)}\right]} \\ &\xrightarrow[N,\,T\to\infty]{p} 0 ,\end{aligned}$$

and the inconsistency in the short panel data ($T < \infty$) is well known. In the long panel data ($T \to \infty$), the bias of the CV estimator becomes weaker and is the consistent estimation. However, constructing the t-test using the CV estimator is difficult, as will be further discussed later. Hence, the instrumental variable method or the maximum likelihood method in the dynamic panel should be considered.

2.1 Long Panel Data

Considering the properties of long panel data, although the total number of panel data is $n = NT$, either N or T must be considered large enough for the asymptotic inference.[2] A long panel is defined by $T \to \infty$. In actual, T is finite such as $T = 100$, and such data are not observed even nowadays. What matters is the relative ratio of T to N,

$$\frac{T}{N} \to c > 0 .$$

A problem arises when the ratio cannot be ignored as 0. We formally represent an estimator $\hat{\pi}$ as follows:

$$\hat{\pi} = (\pi + b_0) + \frac{z}{\sqrt{NT}} + \cdots ,$$

[2]Depending on the estimators n may become $N(T-1)$ or $N(T-2)$, but it does not affect the asymptotic theory because $(T-1)/T \to 1$.

where $z \sim \mathcal{N}(b_1, v)$ due to the asymptotic normality. As will be further discussed later, the bias term b_0 and the noncentral parameter b_1 often depend on the ratio of the sequences such as T/N or its reciprocal N/T. Hence, the ratio is the important value for a long panel data. For example, if $(N, T) = (100,\ 10)$ or $(200,\ 20)$, then the ratio becomes $c = 0.1$. However, the small value might have a non-negligible effect on estimators from the numerical experiments of the previous studies. Such a situation may occur because of the accumulation of data in recent years. As for hypothesis testing, notably, the t-test or χ^2-test cannot be conducted when $b_1 \neq 0$, and the CV estimator under long panel data is an example.

In the usual short panel data $(N \to \infty, T < \infty)$ of empirical analyses, c is regarded as 0. Thus, the estimation is based on the asymptotic theory in the cross sectional data $(N \to \infty,\ T = 1)$. In the long panel data, we are interested in the behavior of the estimator when T also increases, and thus, we consider it based on the double asymptotics $(N \to \infty, T \to \infty)$. Alternatively, the repeated measurements of the time series data $(N < \infty,\ T \to \infty)$ can also be included in the long panel data (cf. Anderson, 1978a). Therefore, we allow the situation $N < T$, such as $(N, T) = (5, 30)$. In terms of application, this study focuses on the estimation method such that the estimators does not depend on the sample size of N. That is, we consider the consistent estimators regardless of whether N is fixed or tends to infinity.

2.2 Incidental Parameters Problem

The second problem is the initial values. (2.1) is equivalent to the following state-space representation:

$$\begin{aligned} y_{it} &= w_{it} + \mu_i\ , \\ w_{it} &= \pi w_{it-1} + v_{it}\ , \end{aligned} \tag{2.2}$$

where $\mu_i = \eta_i/(1-\pi)$. Assuming that μ_i follows a certain distribution, the individual effect becomes a random effect, which means $\mathcal{V}ar[\mu_i] = \omega_\mu$. For instance, if y_{it} is considered a household income, then μ_i may follow the Pareto distribution. Meanwhile, if y_{it} is the growth rate of each country, then assuming what type of distribution μ_i follows may be difficult. However, if the heteroscedastic variance $\mathcal{V}ar[\mu_i] = \omega_{\mu i}$ $(i = 1, \cdots, N)$, then it depends on many parameters. The random-effects maximum likelihood estimation (MLE) is valid when the model has a finite number of unknown parameters.

Anderson and Hsiao (1981) verified for the first time that the initial value is another issue in the dynamic panel model, even if the random effect is assumed for the individual effect. We do not have much information on the distribution of the initial value $y_{i0} = w_{i0} + \mu_i$ and make almost no assumption if w_{i0} is considered as fixed.

When $(\mu_i,\ v_{it})$ is normally distributed the likelihood function is given as follows:

$$\begin{aligned}
& f_i(y_{i0},\ y_{i1}, \cdots, y_{iT}) \\
= \ & \left(\frac{1}{\sqrt{2\pi\omega}}\right)^T \exp\left\{-\frac{1}{2\omega}\sum_{t=1}^{T}\left[(y_{it} - y_{i0} + w_{i0}) - \pi(y_{it-1} - y_{i0} + w_{i0})\right]^2\right\} \\
& \times \frac{1}{\sqrt{2\pi\omega_\mu}} \exp\left\{-\frac{1}{2\omega_\mu}(y_{i0} - w_{i0})^2\right\} .
\end{aligned}$$

The random-effects MLE $\tilde{\pi}_{\mathrm{RM}}$ is obtained by maximizing this log-likelihood function with respect to $(\pi,\ \omega\ , \omega_\mu)$ and $(w_{10}, \cdots, w_{N0})$. They pointed out the incidental parameters problem such that the random-effects MLE would be inconsistent in the short panel data, which is caused by many equations:

$$\frac{\partial \ell_i}{\partial w_{i0}} = 0\ , \quad (i = 1,\ \cdots,\ N)\ , \tag{2.3}$$

where $\ell_i = \log f_i$ is the log-likelihood function of each individual. By estimating w_{i0}, $\tilde{\pi}_{\mathrm{RM}}$ becomes inconsistent if T is not large. The reason is that $\tilde{\pi}_{\mathrm{RM}}$ cannot be solved independently from the normal equations of (2.3). This has been known as the incidental parameters problem, as noted by Neyman and Scott (1948).

As $T \to \infty$,

$$\tilde{\pi}_{\mathrm{RM}} = \tilde{\pi}_{\mathrm{CV}} + o_p(1)\ ,$$

they also showed that the random-effects MLE is asymptotically equivalent to the CV estimator. The CV estimator is consistent in the long panel data but has the noncentrality parameter as we will show later.

In empirical analyses, deciding whether it is short or long panel data is difficult. The use of different estimation methods depending on whether T is fixed or not would be inconvenient. Therefore, in this study, we also focus on the estimators that are consistent even if $T \to \infty$ or $T < \infty$. Moreover, Section 3.7.1 shows that some maximum likelihood estimator can avoid the incidental parameters problem.

2.3 Fixed-Effects Estimation

If the individual effects are fixed rather than random, then a method that can consistently estimate is called a fixed-effects estimation. Anderson and Hsiao (1981) considered a fixed-effects method that does not depend on the assumptions of individual effects or initial values. A simple instrumental variable estimator is presented and is one of the main estimation methods in the early stages of panel analysis. The instrumental variable (IV) estimator is based on the following orthogonal condition,

$$\mathcal{E}[y_{it-2}\Delta v_{it}] = 0 \ .$$

Taking the first-difference in (2.1),

$$\begin{aligned} \Delta y_{it} &= y_{it} - y_{it-1} \\ &= \pi \Delta y_{it-1} + \Delta v_{it} \ , \end{aligned}$$

where the difference $\Delta u_{it} = u_{it} - u_{it-1}$ does not include the individual effects and is uncorrelated with the level y_{it-2}, which becomes an instrumental variable. Similarly, the difference Δy_{it-2} is also an instrumental variable, and thus, the orthogonal condition $\mathcal{E}[\Delta y_{it-2}\Delta u_{it}] = 0$ is satisfied. The first-difference of panel data is obtained from $\mathbf{D}_T\mathbf{y}_i$, where $\mathbf{y}_i = (y_{i1}, \cdots, \ y_{iT})'$ and

$$\underset{(T-1)\times T}{\mathbf{D}_T} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} . \tag{2.4}$$

We refer to the estimators of Anderson and Hsiao (1981) as AH estimators, and they are as follows:

$$\begin{aligned} \tilde{\pi}_{\mathrm{IV}} &= \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it}\Delta y_{it-2}}{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}\Delta y_{it-2}} \ , \\ \bar{\pi}_{\mathrm{IV}} &= \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it} y_{it-2}}{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1} y_{it-2}} \ . \end{aligned}$$

The assumptions for the following theorems are given by the following:

(a1) $\{v_{it}\}$ $(t = 1, \cdots, T; \ i = 1, \cdots, N)$ are i.i.d across time and individuals. v_{it} is independent of y_{i0} with $\mathcal{E}[v_{it}] = 0$, $\mathcal{V}ar[v_{it}] = \omega$, and has a finite moment up to the eighth order.

(a2) The initial observations satisfy the following:

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

where $w_{i0} = \sum_{h=0}^{\infty}\pi^s v_{i,-h}$ is independent of η_i.

(a3) η_i are i.i.d. across individuals. η_i is independent from $\{v_{it}\}$ with $\mathcal{E}[\eta_i] =$

0, $\mathcal{V}ar[\eta_i] = \omega_\eta$, and has a finite moment up to the fourth order.

(a4) $(1/N)\sum_{i=1}^{N} \eta_i^2 = O(1)$ and $v_{it} \sim \mathcal{N}(0, \omega)$.

These assumptions are the same as those of Alvarez and Arellano (2003) and simplify the derivation and expression of theorems. Assumptions (a2) and (a3) are for setting a random-effects model, but the fixed-effects estimators may not need the assumptions of the individual effects or initial conditions. If the random-effects model is correct, then the issue will be whether the fixed-effects estimator can achieve the same efficiency as the random-effects MLE. Assumption (a1) means the homoscedasticity of the error terms, and assumption (a4) is used in deriving the lower bound of efficiency. Assumption (a2) is regarding the initial conditions. If the data is generated from a sufficient past before the initial value, then the next relation may be natural,

$$\lim_{t\to\infty} \frac{1-\pi^{t-1}}{1-\pi}\eta_i = \frac{1}{1-\pi}\eta_i \ .$$

From the expression of (2.2),

$$y_{it} - y_{it-1} = w_{it} - w_{it-1} \ .$$

Notably, the individual effect disappears by the first-difference Δy_{it} at all t. Hayakawa (2008) examined the IV estimator when the initial condition is different from $\eta_i/(1-\pi)$, but the initial effect diminishes in the stationary process and does not affect the consistency.

The following results hold.[3]

Theorem 1.1 (Anderson and Hsiao, 1981) : *Supposing assumptions (a1)-(a3) hold, then as $N \to \infty$ or $T \to \infty$ or both, $\tilde{\pi}_{IV} \xrightarrow{p} \pi$ and $\bar{\pi}_{IV} \xrightarrow{p} \pi$.*

Provided $T \to \infty$,

$$\sqrt{NT}\,(\tilde{\pi}_{\mathrm{IV}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{2(1+\pi)(3-\pi)}{(1-\pi)^2}\right) ,$$

$$\sqrt{NT}\,(\bar{\pi}_{\mathrm{IV}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(\,0,\ 2(1+\pi)\,\right) .$$

[3]The derivation of the asymptotic variance is based on the studies by Hsiao and Zhang (2015) and Phillips and Han (2014).

The AH estimators are consistent estimators for short and long panel data, and assumptions (a2) and (a3) are not necessary for consistency. Regarding the two asymptotic variances, $\bar{\pi}_{\text{IV}}$ is smaller because $(3-\pi)/(1-\pi)^2 > 1$. However, Anderson and Hsiao (1981) pointed out that AH estimators are the simplest IV estimators, so that efficiency can be further improved.

The next issue is the improvement of efficiency, but White (1999, Ch.4) explained that increasing the instrumental variables can improve efficiency in the usual situation. Therefore, the variability of an estimator is suppressed by the orthogonal conditions as the correct constraint increases. As a lagged endogenous variable, the instrumental variable is not only y_{it-2} but also y_{it-3}. Therefore, in each period t ,

$$\mathcal{E}[y_{it-s}\Delta v_{it}] = 0 \ , \quad s = 2, 3, \cdots, t \ ,$$

that is, $(t-1)$ orthogonal conditions exist. As a whole,

$$\sum_{t=2}^{T}(t-1) = \frac{T(T-1)}{2} \ .$$

Let the $(T-1)\times 1$ vector of the first-order difference be

$$\Delta\mathbf{y}_i = \pi\Delta\mathbf{y}_{i,-1} + \Delta\mathbf{v}_i \ .$$

If the $(T-1)\times T(T-1)/2$ matrix of the instrumental variables is

$$\mathbf{Z}_i = \begin{pmatrix} y_{i0} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & y_{i0} & \cdots & y_{iT-1} \end{pmatrix} , \tag{2.5}$$

then the orthogonal conditions are collectively $\mathcal{E}[\mathbf{Z}_i'\mathbf{v}_i] = \mathbf{0}$.

Notably, the error term becomes the moving average process MA(1) by taking the difference with the serial correlation $\mathcal{E}[\Delta v_{it-1}, \Delta v_{it}] = -\omega$,

$$\begin{aligned} \underset{(T-1)\times(T-1)}{\mathcal{E}\left[\Delta\mathbf{v}_i\Delta\mathbf{v}_i'\right]} &= \omega\mathbf{W} \\ &= \omega\begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} . \end{aligned} \tag{2.6}$$

Arellano and Bond (1991) proposed the generalized method of moment (GMM) estimator which efficiently estimates under serial correlation. This method belongs to the efficient GMM estimator in the framework of the moment method. Their AB estimator becomes

$$\tilde{\pi}_{\text{GM}} = \frac{\sum_{i=1}^{N} \Delta \mathbf{y}_{i,-1}' \mathbf{Z}_i \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{W} \mathbf{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \Delta \mathbf{y}_i}{\sum_{i=1}^{N} \Delta \mathbf{y}_{i,-1}' \mathbf{Z}_i \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{W} \mathbf{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' \Delta \mathbf{y}_{i,-1}} . \tag{2.7}$$

Theorem 1.2 (Arellano and Bond, 1991) : *Supposing assumptions (a1)-(a3) hold, then as $N \to \infty$ and T is fixed,*

$$\sqrt{N}\,(\tilde{\pi}_{\text{GM}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(0,\ \frac{\omega}{\phi_T} \right) ,$$

where

$$\phi_T = \mathcal{E}\left[\Delta \mathbf{y}_{i,-1}' \mathbf{Z}_i\right] \mathcal{E}\left[\mathbf{Z}_i' \mathbf{W} \mathbf{Z}_i\right]^{-1} \mathcal{E}\left[\mathbf{Z}_i' \Delta \mathbf{y}_{i,-1}\right] .$$

The above equation is the result under the short panel data. Therefore, the asymptotic variance depends on T. Arellano and Bond (1991) estimated the UK wage equation in short panel data ($N = 611,\ T = 6$) and stated that efficiency was significantly improved by the AB estimator. Moreover, they found that the estimation results were stable compared with those using the AH estimator. After that, the AB estimator is known as a representative in the estimation of the dynamic panel model and is installed in the package software.

However, in the AB estimator, the number of instrumental variables increases rapidly on the order of $O(T^2)$ as T increases. Wooldrige (2002, Ch. 11) showed that the finite sample properties of GMM were not so good when the instrumental variables were increased. The use of many instrumental variables is not recommended. Although these are empirical discussions, the next section will clarify the point of the problems.

2.4 Forward Orthogonal Deviation

This section considers the properties of the GMM estimator in the long panel data. In preparation for that, we look at a simpler expression of the GMM estimator. Although the serial correlations by $\mathcal{E}[\Delta \mathbf{v}_i \Delta \mathbf{v}_i'] = \omega \mathbf{D}_T \mathbf{D}_T'$ exist, the data transformation in which the serial correlation does not occur from the beginning is given as follows:

$$\underset{(T-1)\times T}{\mathbf{D}_f} = (\mathbf{D}_T\mathbf{D}_T^{'})^{-\frac{1}{2}}\mathbf{D}_T \ . \qquad (2.8)$$

Then,

$$\mathbf{D}_f\mathbf{D}_f^{'} = \mathbf{I}_{T-1} \ ,$$

by the definition. In particular, Arellano and Bover (1995) noted that

$$\mathbf{D}_f = \text{diag}\left[\sqrt{\frac{T-1}{T}}, ..., \sqrt{\frac{1}{2}}\right]\begin{pmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix} .$$

Considering that the sum of each column is zero, the individual effects disappear from the regression equation when this transformation is applied.

$$\mathbf{y}_i^{(f)} = \pi\mathbf{y}_{i,-1}^{(f)} + \mathbf{v}_i^{(f)} \ , \qquad (2.9)$$

where $(\mathbf{y}_i^{(f)}, \mathbf{y}_{i,-1}^{(f)}) = \mathbf{D}_f(\mathbf{y}_i, \mathbf{y}_{i,-1})$ and $\mathbf{v}_i^{(f)} = \mathbf{D}_f\mathbf{v}_i$. As this transformation is orthogonal, the homoscedasticity is maintained, and no serial correlation exists:

$$\begin{aligned} \mathcal{E}\left[v_{it}^{(f)2}\right] &= \omega \ , \\ \mathcal{E}\left[v_{it\pm s}^{(f)}v_{it}^{(f)}\right] &= 0 \ , \ (s \neq 0) \ . \end{aligned}$$

The transformed error becomes

$$v_{it}^{(f)} = f_t\left[v_{it} - \frac{1}{T-t}(v_{it+1} + \cdots + v_{iT})\right] \ ,$$

where

$$f_t^2 = \frac{T-t}{T-t+1} \ ,$$

with respect to $t = 1, \cdots, T-1$. As the average after period t is subtracted, the orthogonal condition of each t can be

$$\mathcal{E}\left[y_{is}v_{it}^{(f)}\right] = 0 \ , \ \ (s = 0, 1, \cdots, t-1) \ .$$

However, notably, the $y_{it-1}^{(f)}$ on the right-hand side of (2.9) correlates with the transformed error term,

$$\mathcal{E}\left[y_{it-1}^{(f)}v_{it}^{(f)}\right] \neq 0 \ .$$

From the above properties, we may call the forward orthogonal deviation the forward filter in this work.

Next, the AB estimator is equivalent to the GMM estimator with the forward filter. Given the $n \times 1$ vector $\mathbf{y}^{(f)}$,

$$\mathbf{y}^{(f)} = \left(\mathbf{y}_1^{(f)\prime}, \cdots, \mathbf{y}_N^{(f)\prime}\right)^{\prime} .$$

Similarly, $\mathbf{y}_{-1}^{(f)}$ is defined. Then, the simple expression of the AB estimator becomes

$$\tilde{\pi}_{\mathrm{GM}} = \frac{\mathbf{y}_{-1}^{(f)\prime}\mathbf{P}\mathbf{y}^{(f)}}{\mathbf{y}_{-1}^{(f)\prime}\mathbf{P}\mathbf{y}_{-1}^{(f)}} , \tag{2.10}$$

where the projection matrix is $\mathbf{P} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ constructed by

$$\mathbf{Z} = \left(\mathbf{Z}_1', \cdots, \mathbf{Z}_N'\right)' .$$

As the serial correlation disappears, the weighted matrix becomes $\mathbf{W} = \mathbf{I}$, so that it is expressed by the two-stage least squares estimator. As for another expression,

$$\tilde{\pi}_{\mathrm{GM}} = \frac{\sum_{t=1}^{T-1}\mathbf{y}_{t-1}^{(f)\prime}\mathbf{P}_t\mathbf{y}_t^{(f)}}{\sum_{t=1}^{T-1}\mathbf{y}_{t-1}^{(f)\prime}\mathbf{P}_t\mathbf{y}_{t-1}^{(f)}} , \tag{2.11}$$

where the $N \times 1$ vector $\mathbf{y}_t^{(f)}$ is

$$\mathbf{y}_t^{(f)} = \left(y_{1t}, \cdots y_{Nt}\right)' , \quad (t = 1, \cdots, T-1) .$$

Similarly, $\mathbf{y}_{t,-1}^{(f)}$ is defined. The projection matrix of each t becomes $\mathbf{P}_t = \mathbf{Z}_t\left(\mathbf{Z}_t'\mathbf{Z}_t\right)^{-1}\mathbf{Z}_t'$, and then, the i-th row of the $N \times t$ matrix $\mathbf{Z}_t$ is $(y_{i0}, \cdots, y_{it-1})$. Although the forward filter is complicated at first glance, calculating the asymptotic property in the long panel data based on (2.7) is difficult. That is, the calculation of (2.11) is still easier.

Let us check why they are numerically equivalent. Using the relation of (2.8),

$$\begin{aligned}
\tilde{\pi}_{\mathrm{GM}} &= \frac{\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{D}_T'\mathbf{Z}_i\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{D}_T\mathbf{y}_i}{\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{D}_T'\mathbf{Z}_i\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{D}_T\mathbf{D}_T'\mathbf{Z}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{D}_T\mathbf{y}_{i,-1}} \\
&= \frac{\mathbf{y}_{-1}'\left(\mathbf{I}_N \otimes \mathbf{D}_f'\right)\mathbf{Z}\left[\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{D}_f\mathbf{D}_f'\right)\mathbf{Z}\right]^{-1}\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{D}_f\right)\mathbf{y}}{\mathbf{y}_{-1}'\left(\mathbf{I}_N \otimes \mathbf{D}_f'\right)\mathbf{Z}\left[\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{D}_f\mathbf{D}_f'\right)\mathbf{Z}\right]^{-1}\mathbf{Z}'\left(\mathbf{I}_N \otimes \mathbf{D}_f\right)\mathbf{y}_{-1}} .
\end{aligned}$$

Therefore, (2.7) is equal to (2.10). Next, rearranging the rows with $\vec{\mathbf{J}}_n'\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_t)$ gives

$$\tilde{\pi}_{\mathrm{GM}} = \frac{\mathbf{y}_{-1}^{(f)\prime}\vec{\mathbf{J}}_n\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)\left[\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)'\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)\right]^{-1}\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)'\vec{\mathbf{J}}_n'\mathbf{y}^{(f)}}{\mathbf{y}_{-1}^{(f)\prime}\vec{\mathbf{J}}_n\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)\left[\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)'\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)\right]^{-1}\left(\vec{\mathbf{J}}_n'\mathbf{Z}\right)'\vec{\mathbf{J}}_n'\mathbf{y}_{-1}^{(f)}} .$$

Hence, (2.10) is equal to (2.11) because $\vec{\mathbf{J}}_n \vec{\mathbf{J}}_n' = \mathbf{I}_n$.

The CV estimator can also be expressed in the form of OLS, because

$$\underset{T\times T}{\mathbf{D}_f' \mathbf{D}_f} = \mathbf{Q}_T = \mathbf{I}_T - \frac{1}{T}\boldsymbol{\iota}\boldsymbol{\iota}' .$$

Using this relation, we have

$$\tilde{\pi}_{\mathrm{CV}} = \frac{\mathbf{y}_{-1}^{(f)'} \mathbf{y}^{(f)}}{\mathbf{y}_{-1}^{(f)'} \mathbf{y}_{-1}^{(f)}} .$$

Regarding the LIML estimator, Alonso-Borrego and Arellano (1999) examined the estimator using the same instrumental variables as that in the AB estimator. The detail will be discussed in Part II, given as follows:

$$\tilde{\pi}_{\mathrm{LI}} = \frac{\mathbf{y}_{-1}^{(f)'} \mathbf{P} \mathbf{y}^{(f)} - \tilde{\lambda} \mathbf{y}_{-1}^{(f)'} \mathbf{y}^{(f)}}{\mathbf{y}_{-1}^{(f)'} \mathbf{P} \mathbf{y}_{-1}^{(f)} - \tilde{\lambda} \mathbf{y}_{-1}^{(f)'} \mathbf{y}_{-1}^{(f)}} ,$$

where $\tilde{\lambda}$ is the minimum eigenvalue of some eigenvalue problem.

Using the representations by the forward filters, Alvarez and Arellano (2003) derived the properties of these estimators under the double asymptotics (N, $T \to \infty$) in a long panel data.

Theorem 1.3 (Alvarez and Arellano, 2003) : *Supposing assumptions (a1)-(a3) hold, then*

[i] *as $T \to \infty$, regardless of N is fixed or tends to infinity, provided $N/T^3 \to 0$,*

$$\sqrt{NT}\,(\tilde{\pi}_{\mathrm{CV}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(-\sqrt{d}(1+\pi), 1-\pi^2\right) ,$$

where $d = \lim N/T$ such that $0 \leq d < \infty$.

[ii] *As N and T tend to infinity, provided $(\log T)^2/N \to 0$,*

$$\sqrt{NT}\,(\tilde{\pi}_{\mathrm{GM}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(-\sqrt{c}(1+\pi), 1-\pi^2\right) ,$$

where $c = \lim T/N$ such that $0 \leq c < \infty$.

[iii] *As N and T tend to infinity,*

$$\sqrt{NT}\,(\tilde{\pi}_{\mathrm{LI}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}\left(-\frac{\sqrt{c}}{2-c}(1+\pi), 1-\pi^2\right) ,$$

where $c = \lim T/N$ such that $0 \leq c \leq 2$.

The model is quite simple, but the forward filter and many instruments make the derivation complicated. Hence, their derivation of the noncentrality parameter is a pioneering result in the theoretical analysis of the long panel data. In the long panel data the number of instrumental variables can be $O(T^2) \to \infty$. We will discuss the many instruments problem in Prat II. Even in this situation, the three estimators are consistent and have the same asymptotic variance. In a short panel data, the CV estimator is inconsistent, because d becomes infinite. Meanwhile, the noncentrality parameter of the AB estimator disappears because $c = 0$. In the case of a long panel data, the noncentrality parameter appears because of the effects of a large number of instrumental variables and the data transformation. Thus, the property of the GMM estimator differs between the short and long panel data.

Bias correction may be applied to the noncentrality parameter. Hahn and Kuersteiner (2002) considered the bias-corrected CV estimator and relating t test. The GMM and LIML estimators also have the noncentrality parameter so that the t test cannot be used as it is. T may be reduced to T_0 ($T_0 < T$) and estimate it as short panel data, but this method would not be a fundamental solution because the speed of convergence drops from $\sqrt{NT}$ to $\sqrt{NT_0}$. In our simple model, π is the only unknown parameter. Thus, the bias-corrected t test statistic is given as follows:

$$\tilde{t} = \frac{1}{\sqrt{1 - (\tilde{\pi}_{\mathrm{GM}})^2}} \left(\sqrt{NT}(\tilde{\pi}_{\mathrm{GM}} - \pi) + \sqrt{\frac{T}{N}}(1 + \pi) \right) ,$$

where π is assigned to a hypothetical value.

Next, we consider the property of the GMM estimator proposed by Blundell and Bond (1998) in the long panel data, which is called the system GMM estimator. This estimator widely used as often as the AB estimator. They pointed out that the AB estimator is less efficient depending on the variance ratio $\psi = \omega_\eta / \omega$ when π is close to the unit root or the variance of the individual effect is large. For an instrumental variable to be valid, in addition to no correlation with the error term, the condition of correlation with the endogenous variable is required. However, when the variable is close to the unit root, the instrumental variable y_{it-1} as the level becomes a weak instrumental variable, which can hardly explain Δy_{it}. Conversely, if the regression equation does not take the difference and the level y_{it} is used, then the difference Δy_{it-1} can be used as the instrumental variable. Therefore, the influence of the near unit root will be mild. For the following equations,

$$\begin{bmatrix} \Delta y_{it} \\ y_{it} \end{bmatrix} = \pi \begin{bmatrix} \Delta y_{it-1} \\ y_{it-1} \end{bmatrix} + \begin{bmatrix} \Delta v_{it} \\ \eta_i + v_{it} \end{bmatrix} ,$$

the system GMM estimator is given as follows:

$$\tilde{\pi}_{\text{SG}} = \frac{\sum_{t=3}^{T} \Delta\mathbf{y}_{t-1}'\mathbf{P}_t\Delta\mathbf{y}_t + \sum_{t=3}^{T} \mathbf{y}_{t-1}'\mathbf{P}_t^{(\Delta)}\mathbf{y}_t}{\sum_{t=3}^{T} \Delta\mathbf{y}_{t-1}'\mathbf{P}_t\Delta\mathbf{y}_{t-1} + \sum_{t=3}^{T} \mathbf{y}_{t-1}'\mathbf{P}_t^{(\Delta)}\mathbf{y}_{t-1}} \ ,$$

where $\mathbf{P}_t^{(\Delta)} = \Delta\mathbf{y}_{t-1}(\Delta\mathbf{y}_{t-1}'\Delta\mathbf{y}_{t-1})^{-1}\Delta\mathbf{y}_{t-1}'$.

The result of the short panel data shows that efficiency is significantly improved over the AB estimator, particularly in $\pi = 0.9$ and so on. Moreover, the result of the long panel data is as follows.

Theorem 1.4 (Hayakawa, 2006a) : *Supposing assumptions, then as N and T tend to infinity, provied that $T/N \to c$ $(0 \leq c \leq 1)$,*

$$\tilde{\pi}_{\text{SG}} - \pi \quad \xrightarrow{p} \quad -\frac{c}{c + \frac{3-2\pi}{1+\pi}} \ .$$

That is, the system GMM estimator is inconsistent in the long panel data.[4] Although the AB estimator uses the instrumental variables of the same order $O(T^2)$, the AB estimator is consistent. Alvarez and Arellano (2003) showed that a GMM estimator that does not use the optimal weighted matrix $\mathbf{W}$ results in inconsistency in the long panel data. Thus, the inefficiency of the system GMM estimator may cause inconsistency.

These GMM estimators were derived in short panel data, and they are still useful methods for short panel data. From the above discussions, the properties of estimators significantly change between short and long panel data, which is one of the motivations of research on long panel data in recent years.

2.5 Optimal Instrumental Variable

The IV and GMM estimators are robust to the assumptions of individual effects and initial conditions in the sense that they use only orthogonal conditions. However, the GMM estimator may show poor properties under many orthogonal conditions. Meanwhile, Wooldridge (2002, Ch.8) noted that an argument called the optimal instrumental variable exists, which searches for an efficient estimator with the minimum necessary orthogonal conditions. Arellano (2003b) considered the optimal instrumental variable $z_{it-1}^{(*)}$ in the dynamic panel model. In the case

[4] Bun and Windmeijer (2010) also discussed the bias of the system GMM estimator.

of the AR(1) model,

$$\breve{\pi}_{\mathrm{IV}} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-1} z_{it-1}^{(*)} y_{it}^{(f)}}{\sum_{i=1}^{N}\sum_{t=1}^{T-1} z_{it-1}^{(*)} y_{it-1}^{(f)}} ,$$

where the condition of the optimal instrumental variable is given as follows:

$$\begin{aligned} z_{it-1}^{(*)} &= \mathcal{E}\left[y_{it-1}^{(f)} | y_{it-1}\right] \\ &= f_t \left[1 - \frac{\pi(1-\pi^{T-t})}{(T-t)(1-\pi)}\right]\left[w_{it-1} + O_p\left(\frac{1}{\sqrt{t}}\right)\right] . \end{aligned}$$

Similar to AH estimators, the estimator uses one orthogonal condition for each t, but $\breve{\pi}_{\mathrm{IV}}$ is infeasible because it depends on an unknown parameter. For the optimum instrumental variable, the following conditions should be noted. If t is large, then the instrumental variable does not depend on the individual effect, and if T is large, then it becomes almost w_{it-1}.

Hayakawa (2006b) considered the transformation $\mathbf{y}_{i,-1}^{(b)} = \mathbf{D}_b \mathbf{y}_{i,-1}$, which is almost equivalent to the following,

$$\underset{(T-1)\times T}{\mathbf{D}_b} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & 1 & 0 \\ -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & 1 \end{pmatrix} ,$$

where the sum of each column is zero and the individual effect disappears.

$$\begin{aligned} y_{it-1}^{(b)} &= b_t \left[y_{it-1} - \frac{1}{t-1}(y_{i0} + \cdots + y_{it-2})\right] \\ &= w_{it-1} - \frac{1}{t-1}(w_{i0} + \cdots + w_{it-2}) , \end{aligned}$$

$$b_t = 1 , \tag{2.12}$$

for $t = 2, \cdots, T$.[5] Such a transformation is called the backward orthogonal deviation, or the recursive mean adjustment in So and Shin (1999) in the time series analysis. In this work, we call this transformation the backward filter. In contrast to the forward filter, the historical average is subtracted to have the orthogonal conditions:

$$\mathcal{E}\left[y_{it-1}^{(b)} u_{it}^{(f)}\right] = 0 , \quad (t = 2, \cdots, T-1) .$$

[5]It may be replaced with $b_t^2 = (t-1)/t$ or f_t. As orthogonalizing the instrumental variables is not necessary, it is simply set to 1. However, $b_t = f_t$ is used only in Section 3.7.2.

This instrumental variable does not depend on individual effects, w_{it-1} is the main term, and thus, the condition of the optimal IV is satisfied. Considering that the IV estimator is replaced with the optimal instrumental variable,

$$\hat{\pi}_{\rm IV} \;=\; \frac{\sum_{i=1}^{N}\sum_{t=2}^{T-1} y_{it-1}^{(b)} y_{it}^{(f)}}{\sum_{i=1}^{N}\sum_{t=2}^{T-1} y_{it-1}^{(b)} y_{it-1}^{(f)}} \;.$$

This is equivalent to the estimator proposed by Hayakawa (2009), and the next result can be considered in the AR(p) model.

Theorem 1.5 **(Hayakawa, 2009)** : *Supposing Assumptions (a1)-(a3) hold, then, as N and T tend to infinity,*

$$\sqrt{NT}\,(\hat{\pi}_{\rm IV} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}(0, 1-\pi^2) \;.$$

Similar to the AH estimator $\tilde{\pi}_{\rm IV}$, this estimator does not require assumption (a3) because the individual effects disappear from the regression equation and the instrumental variables even under a finite sample. Therefore, this estimator is not affected by the variance ratio ψ under a finite sample. In addition, the result holds with $T \to \infty$ alone. Above all, the noncentrality parameter disappears and efficiency does not decrease compared with that of the GMM estimator, which is an important result. In general, a trade-off exists such that when the number of instrumental variables is large, the noncentrality parameter becomes large while the efficiency increases. The IV estimator decreases the number of instrumental variables, so that the noncentrality parameter becomes small. However, as this estimator uses asymptotically optimal instruments, efficiency can be maintained.

Finally, we consider. the lower bound of the asymptotic efficiency. When assuming the individual effects η_i $(i = 1, \cdots, N)$ as incidental parameters, the lower bound for π is not obvious because the parameters involved are infinite. The following result also holds for the panel VAR model in Holtz-Eakin et al. (1988).

Theorem 1.6 **(Hahn and Kuersteiner, 2002)** : *Supposing assumptions (a1), (a2), and (a4) hold, then as N and T tend to infinity, the asymptotic distribution of any regular estimator of* π *cannot be more concentrated than* $\mathcal{N}(0, 1-\pi^2)$.

The lower bound is $1 - \pi^2$ and some of the estimators that we have seen above attain the bound. As suggested, the slope π can be estimated efficiently without depending on the information of the individual effects. On the contrary, if an asymptotic variance depends on the variance ratio ψ, then an estimator is inefficient in long panel data.

We consider the optimal instrumental variables more intuitively. In the case of the AR(1) model of the time series ($N = 1$), the individual effect can be $\eta_i = \eta = 0$. If $y_{t-1} = w_{t-1}$, then we should use the OLS estimator $\hat{\pi}_{\text{LS}}$ without using the instrumental variable $(y_{t-2}, y_{t-3}, \cdots)$,

$$\begin{aligned} y_t &= \pi y_{t-1} + v_t \ , \\ \hat{\pi}_{\text{LS}} &= \frac{\sum_{t=1}^{T} y_{t-1} y_t}{\sum_{t=1}^{T} y_{t-1}^2} \ . \end{aligned}$$

If the error term is normally distributed, as is well known, then the OLS is the same as the MLE. The Cramer-Rao lower bound becomes $1-\pi^2$, and the coefficient is estimated independently from the intercept. That is, the optimal instrumental variable is $\mathcal{E}[y_{t-1}|y_{t-1}] = y_{t-1}$, which is the explanatory variable itself. As for $\hat{\pi}_{\text{IV}}$, for a sufficiently large t,

$$y_{t-1}^{(b)} = y_{t-1} + o_p(1) \ , \quad y_{t-1}^{(f)} = y_{t-1} + o_p(1) \ .$$

Although $\hat{\pi}_{\text{IV}}$ is the IV estimator, it can be interpreted as it is fairly close to the OLS estimator.

From the above discussions, regression analysis of long panel data has several efficient estimators. From the viewpoint that no condition for the data sequence exists, $\hat{\pi}_{\text{IV}}$ has the most desirable result. As Anderson and Hsiao (1981) pointed out at the beginning, in terms of the results, efficient estimation can be constructed by the IV method if the appropriate data transformation is used.

In the next part, we will examine the structural analysis in the long panel data, but the result may be different from that of the regression analysis because the estimation problem becomes more difficult.

3 Part II: Structural Analysis

The simultaneous equation model was developed to verify the economic theory. This model raises the issue of endogeneity and identification in the field of statistics and is still one of the central issues in econometrics today. The structural form considers an estimation problem in which endogenous variables are included on the right-hand side. In the dynamic panel model, the instrumental variables have been already used for the reduced form so that the estimation method does not change. However, in structural estimation, whether variables are predetermined is important. Therefore, there are test procedures that are not used in the regression analysis, such as the overidentification test and the identification test for structural parameters. Then, the economic model can be verified deeply.

In the dynamic panel structural model, Bhargava and Sargan (1983) investigated the random-effects LIML, and Moral-Benito (2013) considered the partial system method. Moreover, Alonso-Borrego and Arellano (1999), Akashi and Kunitomo (2012, 2015), and Hsiao and Zhou (2015) proposed the fixed-effects LIML estimators. Huang and Quibria (2013) also used the fixed-effects LIML estimator in their empirical analysis. In the next section, we consider the formulation and estimation problem of the dynamic structural panel model.

3.1 Dynamic Panel Structural Equation Model

Let us start with the simplest structural equation,

$$y_{it}^{(1)} = \beta y_{it}^{(2)} + \gamma y_{it-1}^{(1)} + \alpha_i + u_{it}, \quad \mathcal{E}[y_{it}^{(2)}(\alpha_i + u_{it})] \neq 0 \ ,$$

where α_i is an individual effect. The difference from the reduced form is that the variable $y_{it}^{(2)}$ on the right-hand side in period t correlates with the error term u_{it}.

Following the notation of Anderson and Rubin (1949), β and γ stand for the structural parameters, and the reduced form parameter is represented by π. Before considering a general model, we provide the examples of dynamic structural panel models based on the economic models. The following is an example of why simultaneity occurs in profit maximization.

Example 2.1 : Endogeneity can occur in the analysis of production functions, as explained by Hayashi (2000, Ch.3). For simplicity, let $y_{it}^{(1)} = \alpha_i (y_{it}^{(2)})^{\beta} \exp(u_{it})$ be a Cobb-Douglas production function, where $y_{it}^{(2)}$ is the amount of labor, and α_i is the total factor productivity combined with the initial technology. If a firm maximizes the expected profit given α_i, then the first order condition is given as follows:

$$\alpha_i \beta (y_{it}^{(2)})^{\beta-1} = z_t \ ,$$

where $\mathcal{E}[\exp(u_{it})] = 1$ is assumed. For the price taker, the real wage z_t is an exogenous variable. The logarithmic value of the supply function and the factor demand function are as follows:

$$\begin{aligned} \log(y_{it}^{(1)}) &= \beta \log(y_{it}^{(2)}) + \log(\alpha_i) + u_{it} \ , \\ \log(y_{it}^{(2)}) &= \pi \log(z_t) - \pi \log(\beta \alpha_i) \ , \end{aligned}$$

where $\pi = 1/(\beta - 1)$. Therefore, $\log(y_{it}^{(2)})$ on the right-hand side becomes the endogenous variable and correlates with the structural error through the individual effects.

The next one is an example of the dynamics in structural and reduced forms.

Example 2.2 : Let $y_{it}^{(1)} = \beta^* y_{it}^{(2)}$ be a linear production function, where $y_{it}^{(2)}$ is the capital. In empirical analyses, the output often has no data, and thus, the amount of sales $y_{it}^{(2)}$ is used as the proxy variable. However, sales are most likely to include the inventory in the previous term,

$$\begin{aligned} y_{it}^{(3)} &= (1-\alpha^*) y_{it}^{(1)} + \alpha^* y_{it-1}^{(1)} + u_{it} \\ &= (1-\alpha^*)\beta^* y_{it}^{(2)} + \alpha^* \beta^* y_{it-1}^{(2)} + u_{it} \\ &= \beta y_{it}^{(2)} + \gamma y_{it-1}^{(2)} + u_{it} \ . \end{aligned}$$

If the inventory-sales ratio $\alpha^* = (1 + \beta/\gamma)^{-1}$ is obtained, then the original capital coefficient $\beta^* = \beta/\alpha^*$ can be estimated. Using the identity of capital accumulation as the reduced form,

$$y_{it}^{(2)} = (1-\delta) y_{it-1}^{(2)} + v_{it} \ ,$$

where δ is a depletion rate. If the investment v_{it} is determined by the error term as an innovation with $\mathcal{E}[v_{it}] = \eta_i > 0$, then it becomes a panel AR(1) model.

We consider how the limited information method (single-equation method) is useful in the example of utility maximization.

Example 2.3 : Let the Stone-Geary utility function be $\sum_{g=1}^{G^*} \beta_g \log(y_{it}^{(g)*} - u_{it}^{(g)*})$ and the budget constraint be $\sum_{g=1}^{G^*} z_t^{(g)} y_{it}^{(g)*} \leq z_{it}$. $z_t^{(g)}$ is the price of good g, z_{it} is income, and $u_{it}^{(g)*} \geq 0$ is called the minimum required amount, which is different in preferences and changes over time such that it cannot be observed by an econometrician. When maximized under a budget constraint, marginal utilities are equal between two goods. In the first and second goods,

$$\frac{\beta_1}{z_t^{(1)}(y_{it}^{(1)*} - u_{it}^{(1)*})} = \frac{\beta_2}{z_t^{(2)}(y_{it}^{(2)*} - u_{it}^{(2)*})} \ .$$

Then, the following structural equation is obtained,

$$\beta_2 y_{it}^{(1)} = \beta_1 y_{it}^{(2)} + \gamma_1 z_t^{(1)} + \gamma_2 z_t^{(2)} + u_{it}^{(1)} \ , \tag{3.1}$$

where

$$y_{it}^{(g)} = z_t^{(g)} y_{it}^{(g)*}$$

is the expenditure function that is the endogenous variable,

$$\gamma_1 = \beta_2\mu_1 \ , \ \gamma_2 = -\beta_1\mu_2 \ , \ u_{it} = \beta_2(u_{it}^{(1)*} - \mu_1)z_t^{(1)} - \beta_1(u_{it}^{(2)*} - \mu_2)z_t^{(2)} \ ,$$

where $\mu_g = \mathcal{E}[u_{it}^{(g)*}]$. Normalization such as $\sum_{g=1}^{G_*}\beta_g = 1$ is required because the utility function allows a monotonic transformation. The reduced form is known as the linear expenditure system,

$$y_{it}^{(g)} = \sum_{k=1}^{K}\pi_{gk}z_{it}^{(k)} + v_{it}^{(g)} \ , \ \ (g = 1, \cdots, G^*) \ ,$$

where the number of instrumental variables is that of goods and income, that is, $K = G^* + 1$. In empirical analyses, the number of goods or services is reduced by some classifications, but originally, many goods or services exist. When verifying the optimization problem, estimating many structural forms simultaneously would be difficult. The limited information method can estimate the first structural equation of interest or can estimate individually. Instrumental variables become many ($K \to \infty$) even with one structural estimation, but the LIML estimator is known to be robust in this situation.

An empirical analysis may not be strictly derived from an economic theory. However, a reverse causality exists, then, we can start with two structural equations. For instance, a foreign exchange rate is influenced by a foreign exchange intervention. Conversely, the authority decides to intervene depending on the fluctuation of the exchange. When starting from the reduced form, having common factors in endogenous variables is expected. For example, the relationship between income and years of education is usually explained by common exogenous variables, such as ability (IQ).

This section presents the simultaneous equation model with one endogenous variable on the right-hand side and explains the estimation theory of the structural form in the long panel data. We consider the two structural equations:

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i^{(1)} + u_{it}^{(1)} \ , \\ y_{it}^{(2)} &= \beta_1 y_{it}^{(1)} + \gamma_2 y_{it-1}^{(2)} + \alpha_i^{(2)} + u_{it}^{(2)} \ . \end{aligned} \tag{3.2}$$

One of the structural equations of interest to be estimated is called the first structural equation, such as (3.2). The number of endogenous variables on the right-hand side of the first structural equation is equal to $G_2 = 1$, and thus, the first structural equaiton contains $G = 1 + G_2 = 2$ endogenous variables. The reduced form solved for the endogenous variables in period t is as follows:

$$\begin{aligned}
y_{it}^{(1)} &= \frac{1}{1-\beta_1\beta_2} y_{it-1}^{(1)} + \frac{\beta_2}{1-\beta_1\beta_2} y_{it-1}^{(2)} + \frac{\alpha_i^{(1)} + \beta_2\alpha_i^{(2)}}{1-\beta_1\beta_2} + \frac{u_{it}^{(1)} + \beta_2 u_{it}^{(2)}}{1-\beta_1\beta_2} \\
&= \pi_{11} y_{it-1}^{(1)} + \pi_{12} y_{it-1}^{(2)} + \pi_i^{(1)} + v_{it}^{(1)} \ , \\
y_{it}^{(2)} &= \frac{\beta_1}{1-\beta_1\beta_2} y_{it-1}^{(1)} + \frac{1}{1-\beta_1\beta_2} y_{it-1}^{(2)} + \frac{\beta_1\alpha_i^{(1)} + \alpha_i^{(2)}}{1-\beta_1\beta_2} + \frac{\beta_1 u_{it}^{(1)} + u_{it}^{(2)}}{1-\beta_1\beta_2} \\
&= \pi_{21} y_{it-1}^{(1)} + \pi_{22} y_{it-1}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)} \ ,
\end{aligned}$$

where from the discussion in the previous part, all variables ($y_{it-1}^{(1)},\ y_{it}^{(1)},\ y_{it-1}^{(2)},\ y_{it}^{(2)}$) correlate with the individual effect ($\pi_i^{(1)},\ \pi_i^{(2)}$). The endogenous variable on the right-hand side of the first structural equation generally correlates with the following:

$$\begin{aligned}
\mathcal{E}\left[y_{it}^{(2)} u_{it}^{(1)}\right] &= \mathcal{E}\left[v_{it}^{(2)} u_{it}^{(1)}\right] \\
&\neq 0 \ ,
\end{aligned}$$

and the structural error term $u_{it}^{(1)}$, which is called the simultaneity or the endogeneity in period t. That is, the source of endogeneity of the first structural equation is due to the simultaneous equations behind it. In some cases, the variable in period t on the right-hand side may also be exogenous,

$$\mathcal{E}\left[y_{it}^{(2)} u_{it}^{(1)}\right] = 0 \ .$$

The above case occurs when the endogenous variable does not appear because $\beta_1 = 0$, and the reduced form error and the structural error are uncorrelated, that is, $\mathcal{E}[u_{it}^{(1)} u_{it}^{(2)}] = 0$. Then, the structural equation is called a triangular or recursive system, and $y_{it}^{(2)}$ is determined independently of $y_{it}^{(1)}$. We can test whether the variable is exogenous or endogenous.

The above expressions are based on the full information method (system method) that specifies the two structural equations. Moreover, the limited information method specifies and estimates only the first structural equation. The advantage is that considering the specification and identification for the other structural equations is unnecessary. All of the coefficients of the reduced form (π_{11} π_{12}, π_{21}, and π_{22}) are implicitly estimated by some estimator. Hence, we do not usually denote the second reduced form in an empirical analysis. The limited information method simplifies the estimation problem as follows:

$$\begin{aligned}
y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i^{(1)} + u_{it}^{(1)} \ , \\
\mathbf{z}_{it} &= \{\ y_{it-1}^{(1)},\ y_{it-1}^{(2)}\ \} \ , \qquad (3.3)
\end{aligned}$$

where $\mathbf{z}_{it}$ is the list of instrumental variables that includes the predetermined endogenous variables in the period t. By only setting the list of IVs, package software can estimate the first structural equation. The zero constraints or exclusion condition is necessary for the identification of the structural parameter. Part III discusses the tests for endogeneity and identification.

Blundell and Bond (2000) considered a simple structural model with $\beta_1 = 0$ in the panel analysis of a production function.

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i^{(1)} + u_{it}^{(1)} , \tag{3.4}$$

$$y_{it}^{(2)} = \gamma_2 y_{it-1}^{(2)} + \rho\alpha_i^{(1)} + u_{it}^{(2)} , \tag{3.5}$$

where $\mathcal{E}[u_{it}^{(1)} u_{it}^{(2)}] \neq 0$ and $\sigma^2 = \mathcal{E}[(u_{it}^{(1)})^2]$. Akashi and Kunitomo (2012) examined the estimation method of this structural panel model. For a comparison with time series analysis ($N = 1$), let us start with $\alpha_i^{(1)} = 0$, that is, no individual effect exists. Then, how should the first structural equation (3.4) be estimated?

Anderson and Rubin (1949) focused on the marginal likelihood function of only G endogenous variables contained in the first structural equation for the first time. The log-likelihood in this model becomes the following:

$$\mathcal{L} = -\frac{T}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2} \sum_{t=1}^{T} \left[y_{it}^{(1)} - \boldsymbol{\pi}_1' \mathbf{z}_{it},\ y_{it}^{(2)} - \boldsymbol{\pi}_2' \mathbf{z}_{it} \right] \boldsymbol{\Omega}^{-1} \begin{bmatrix} y_{it}^{(1)} - \boldsymbol{\pi}_1' \mathbf{z}_{it} \\ y_{it}^{(2)} - \boldsymbol{\pi}_2' \mathbf{z}_{it} \end{bmatrix} ,$$

where $\boldsymbol{\pi}_1 = (\pi_{11},\ \pi_{12})'$, $\boldsymbol{\pi}_2 = (\pi_{21},\ \pi_{22})'$, and $\boldsymbol{\Omega}$ is the variance-covariance matrix of the reduced form errors. They considered the constrained maximization problem as follows:

$$\max_{\beta_2,\, \gamma_1,\, \boldsymbol{\pi}_1,\, \boldsymbol{\pi}_2,\, \boldsymbol{\Omega}} \quad \mathcal{L} ,$$
$$\text{s.t.} \quad \pi_{11} - \beta_2 \pi_{21} = \gamma_1,\ \pi_{12} - \beta_2 \pi_{22} = 0 .$$

The constraint is relating to the identification of structural parameters, which is obtained by multiplying $\boldsymbol{\beta} = (1, -\beta_2)'$ on the left side of the reduced form. They also derived the concentrated log-likelihood function for β_2 and obtained the LIML estimator $\hat{\beta}_2$ as follows:

$$\min_{\beta_2} \frac{\boldsymbol{\beta}' \mathbf{G} \boldsymbol{\beta}}{\boldsymbol{\beta}' \mathbf{H} \boldsymbol{\beta}} , \tag{3.6}$$

where the $T \times 2$ matrices consist of the following:

$$\underset{2\times 2}{\mathbf{G}} = \mathbf{Y}'(\mathbf{P} - \mathbf{P}_1)\mathbf{Y} ,$$
$$\underset{2\times 2}{\mathbf{H}} = \mathbf{Y}'(\mathbf{I} - \mathbf{P})\mathbf{Y} ,$$

where $\mathbf{Y} = (y_{it}^{(1)},\ y_{it}^{(2)})$, and $\mathbf{P}$ and $\mathbf{P}_1$ are projection matrices generated from $(y_{it-1}^{(1)},\ y_{it-1}^{(2)})$ and $y_{it-1}^{(1)}$, respectively. The derivation is similar to Lemma 2.3, which will be described later.

Under the assumption of a normal distribution, the MLE is obtained as the OLS in regression analysis by minimizing a quadratic form, that is, the sum of squares of residuals. Meanwhile, in the case of structural analysis, they found that the MLE can be obtained by minimizing the ratio of the quadratic forms. For the LIML estimator $\hat{\boldsymbol{\theta}}_{\text{LI}}$ of the parameters $\boldsymbol{\theta}_1 = (\beta_2,\ \gamma_1)'$, the following result holds.

Theorem 2.1 (Anderson and Rubin, 1949, 1950) : *Supposing assumptions (A1) and (A2) hold, then as $T \to \infty$* and $N = 1$,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{\text{LI}} - \begin{bmatrix} \beta_2 \\ \gamma_1 \end{bmatrix}\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1}) \ .$$

The notations and assumptions of Theorem 2.1 overlap with the general structural models as discussed later. Hence, we describe them in Section 3.2. Theorem 2.1 shows that the LIML estimator is consistent and efficient in the time series analysis, and the noncentrality parameter does not appear because individual effects exist. The assumption that the error terms follow a normal distribution is not essential for all MLE estimators mentioned in this work. The same result holds without the normality assumption and the LIML estimator is considered a pseudo-MLE. The LIML estimator can be also derived only by the orthogonal condition $\mathcal{E}[\mathbf{z}_{it}u_{it}^{(1)}] = \mathbf{0}$, which is interpreted as a class of the moment method. The filters and appropriate instrumental variables for the panel analysis should be considered as shown in Part I. Therefore, we slightly improve the original LIML method.

Alvarez and Arellano (2003) examined the estimators of the simplest regression under a long panel data. Meanwhile, we consider the same estimators with the simplest structural equations (3.4) and (3.5). Using the forward filter, the CV estimator is expressed as follows:

$$\tilde{\boldsymbol{\theta}}_{\text{CV}} = \left(\sum_{t=1}^{T-1}\mathbf{X}_t^{(f)'}\mathbf{X}_t^{(f)}\right)^{-1}\sum_{t=1}^{T-1}\mathbf{X}_t^{(f)'}\mathbf{y}_t^{(1,f)} \ ,$$

where $\mathbf{X}_t^{(f)} = (\mathbf{y}_t^{(2,f)},\ \mathbf{y}_{t-1}^{(1,f)})$ is the $N \times 2$ matrix, and $\mathbf{y}_t^{(2,f)}$ is obtained by multiplying the endogenous variable $y_{it}^{(2)}$ by the forward filter. $\mathbf{y}_{t-1}^{(1,f)}$ and $\mathbf{y}_t^{(1,f)}$ are defined in the same way. The AB estimator is based on the GMM method,

$$\tilde{\boldsymbol{\theta}}_{\text{GM}} = \left(\sum_{t=1}^{T-1}\mathbf{X}_t^{(f)'}\mathbf{P}_t\mathbf{X}_t^{(f)}\right)^{-1}\sum_{t=1}^{T-1}\mathbf{X}_t^{(f)'}\mathbf{P}_t\mathbf{y}_t^{(1,f)} \ ,$$

where $\mathbf{P}_t$ is equivalent to the projection matrix provided in (2.11). The instrumental variables become $(y_{it-1}^{(1)},\ y_{it-1}^{(2)})$, and thus, $\mathbf{Z}_t$ is only replaced by the $N \times 2t$ matrix. The LIML estimator should be the minimum solution $\tilde{\boldsymbol{\theta}}_{\text{LI}}$ of the following ratio,

$$\min_{\boldsymbol{\theta}_1} \frac{\boldsymbol{\theta}' \mathbf{G}^{(f)} \boldsymbol{\theta}}{\boldsymbol{\theta}' \mathbf{H}^{(f)} \boldsymbol{\theta}} ,$$

where $\boldsymbol{\theta} = (1, -\boldsymbol{\theta}_1')'$ and $\boldsymbol{\theta}_1 = (\beta_2,\ \gamma_1)'$. If the log-likelihood function is also cocentrated on γ_1, then we have

$$\begin{aligned}
\underset{3\times 3}{\mathbf{G}^{(f)}} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} \mathbf{P}_t \left(\mathbf{y}_t^{(1,f)},\ \mathbf{X}_t^{(f)} \right) , \\
\underset{3\times 3}{\mathbf{H}^{(f)}} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} [\mathbf{I}_N - \mathbf{P}_t] \left(\mathbf{y}_t^{(1,f)},\ \mathbf{X}_t^{(f)} \right) .
\end{aligned}$$

For the simple structural model of Blundel and Bond (2000), the asymptotic results under the long panel data are as follows.

Theorem 2.2 (Akashi and Kunitomo, 2012) : *Let assumptions (A1)-(A3) hold, and suppose that $(v_{it}^{(1)},\ v_{it}^{(2)})$ follows a normal distribution.*

[i] *As $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$\tilde{\boldsymbol{\theta}}_{\text{CV}} - \begin{bmatrix} \beta_2 \\ \gamma_1 \end{bmatrix} \xrightarrow{p} \left[\boldsymbol{\Phi} + \omega_{22} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1,0) \right]^{-1} \begin{bmatrix} (0,1)\boldsymbol{\Omega}\boldsymbol{\beta} \\ 0 \end{bmatrix} .$$

[ii] *Assume $T/N \to c$ $(0 \le c \le 1/2)$ as N and $T \to \infty$. Then,*

$$\tilde{\boldsymbol{\theta}}_{\text{GM}} - \begin{bmatrix} \beta_2 \\ \gamma_1 \end{bmatrix} \xrightarrow{p} \left[\boldsymbol{\Phi} + c\omega_{22} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1,0) \right]^{-1} \begin{bmatrix} c(0,1)\boldsymbol{\Omega}\boldsymbol{\beta} \\ 0 \end{bmatrix} .$$

When $c = 0$, we additionally assume that $0 \le \lim_{N,T\to\infty}(T^3/N) = d_1 < \infty$. Then,

$$\sqrt{NT} \left(\tilde{\boldsymbol{\theta}}_{\text{GM}} - \begin{bmatrix} \beta_2 \\ \gamma_1 \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_0, \sigma^2 \boldsymbol{\Phi}^{-1} \right) ,$$

where

$$\mathbf{b}_0 = \sqrt{d_1} \boldsymbol{\Phi}^{-1} \begin{bmatrix} (0,1)\boldsymbol{\Omega}\boldsymbol{\beta} \\ 0 \end{bmatrix} .$$

[iii] *Assume N and $T \to \infty$ and $T/N \to c$ $(0 \le c \le 1/2)$. Then*

$$\sqrt{NT} \left(\tilde{\boldsymbol{\theta}}_{\text{LI}} - \begin{bmatrix} \beta_2 \\ \gamma_1 \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_c, \sigma^2 \boldsymbol{\Phi}^{-1} + c_* \boldsymbol{\Psi} \right) ,$$

where

$$\begin{aligned}\mathbf{b}_c &= \frac{-\sqrt{c}}{1-c}\mathbf{\Phi}^{-1}\mathbf{\Pi}_{\mathrm{I}}'(\mathbf{I}_2-\mathbf{\Pi})^{-1}\mathbf{\Omega}\boldsymbol{\beta},\\ \mathbf{\Psi} &= |\mathbf{\Omega}|\mathbf{\Phi}^{-1}\begin{pmatrix}1\\0\end{pmatrix}(1,0)\mathbf{\Phi}^{-1},\end{aligned}$$

and $c_* = c/(1-c)$.

The notations and assumptions of Theorem 2.2 are also described in Section 3.2. The assumption of the normal distribution is not essential but simplifies the representation of the asymptotic variance under many instrumental variables.

First, the CV and GMM estimators are not consistent in the structural estimation under the long panel data, which is the important difference from the regression analysis in Part I. Arellano (2003b) also pointed out an order for the bias of the GMM estimator, where we clarify the form of the bias. If we compare the CV and GMM estimators with the OLS and TSLS estimators in the cross-sectional data, then the CV estimator has the simultaneous equation bias.Moreover, the GMM estimator suffers from many instrumetal variables. In the case of Theorem 1.3, a correction of the noncentrality parameter can be considered using the consistency result but cannot be corrected in the structural estimation. Therefore, the structural analysis of long panel data by these well-known estimators is not recommended.

Second, the LIML method can consistently estimate the structural parameters. However, similar to Theorem 1.3, the noncentrality parameter remains because of many instruments and the forward filters. Notably, $\mathbf{b}_0$ of the GMM estimator and $\mathbf{b}_c$ of the LIML estimator are different, where we make that of the LIML estimator strict as a condition. For instance, the GMM method has the noncentrality parameter even under $c = 0$, but the LIML is centered in the case of $c = 0$. As for the asymptotic variance, unlike the regression analysis, the second term $c_*\mathbf{\Psi}$ appears which is the same as the result of Anderson et al. (2010).

In the next section, we consider why these results are obtained from the perspective of many instruments problem in the dynamic panel model.

3.1.1 Many Weak Instruments Problem

Kunitomo (1980) and Anderson et al. (1982) conducted early studies of many instruments problems, which are known as the comparative studies of LIML and TSLS estimators. One of the applications was a large macroeconometric model. In recent years, since the study of Angrist and Krueger (1991), discussions in the field of microeconometrics have been active. Moreover, Andesron et al. (2010) and Kunitomo (2012) reaffirmed the superiority of the LIML estimator. In cross-sectional analysis, the theory of many instruments is sometimes called the large-K theory, but all of the results in this work are set as follows, except for Section 3.7.2,

$$K < \infty \ ,$$

where K is the number of instrumental variables included in the structural equation at period t. For instance, K is equal to 2 as in the simple model of (3.4) and (3.5). Although it can be relaxed by $K \to \infty$, even if the number K is finite, the total number of instrumental variables of long panel data can be $O(KT^2) \to \infty$, as shown in the following example.

Example 2.4 : In the previous part, the reduced form is given by AR(1) model $y_{it}^{(2)} = \pi y_{it-1}^{(2)} + \eta_i + v_{it}$, where the number becomes $K = 1$. However, the filtered variable $y_{it}^{(2,f)}$ becomes the actual endogenous variable. Arellano (2003a, CH. 7) discussed the reduced form, which then becomes similar to the AR(t) model,

$$y_{it}^{(2,f)} = \pi_{tt} y_{it-1}^{(2)} + \pi_{(t-1)t} y_{it-2}^{(2)} + \cdots + \pi_{1t} y_{i0}^{(2)} + \tilde{v}_{it} \ ,$$

where

$$(\pi_{1t},\ \pi_{2t},\ \cdots, \pi_{tt})' = \left(\mathcal{E}\left[\mathbf{y}_{i,-1}^{(2)} \mathbf{y}_{i,-1}^{(2)\prime}\right]\right)^{-1} \mathcal{E}\left[\mathbf{y}_{i,-1}^{(2)} y_i^{(2,f)}\right] \ .$$

The number of instrumental variables in period t is $Kt = t$, and each coefficient depends on t. When t and T are sufficiently large, we have

$$\pi_{st} = O\left(\pi^{t-(s-1)}\right) - O\left(\frac{1}{1+t}\right) \ \longrightarrow \ 0 \ .$$

If period s is separated from period t, then the correlation is reduced due to the nature of AR model, and the instruments can be called the weak instrumental variables. In this estimation, the number of total instruments is $O(T^2)$ and includes many weak instruments. Thus, the LIML method has the robustness to estimate the structural parameter, even if many reduced form parameters π_{st} exist.

We first consider why even if the correct orthogonal condition $\mathcal{E}[z_{it}u_{it}] = 0$ is used, it causes inconsistency when the number of instruments is large. When ignoring the influence of the forward filter and expressing the sampling error,

$$\begin{aligned} \tilde{\theta}_{\text{GM}} - \theta &\propto \frac{1}{NT}\mathcal{E}\left[\mathbf{y}'\mathbf{P}\mathbf{u}\right] \\ &= \frac{\text{tr}(\mathbf{P})}{NT}\mathcal{E}[v_{it}u_{it}] \\ &= \frac{\frac{1}{2}KT^2}{NT}\mathcal{E}[v_{it}u_{it}] \ . \end{aligned} \tag{3.7}$$

If a variable is exogenous, that is, $y_{it} = z_{it}$, then (3.7) becomes 0 because the sum of the orthogonal conditions is zero. However, in the case of an endogenous variable, the sum of the squares of the error term does not become 0. However, in general structural analysis, $\mathrm{tr}(\mathbf{P}) = \mathrm{rank}(\mathbf{P}) = \mathrm{rank}(\mathbf{Z}) < \infty$ due to the property of the projection matrix. If the number of instruments variables is small, then it can be ignored, and the consistency is held. Under many instruments, the ratio r_n of the total number of data and that of instrumental variables converge to nonzero:

$$r_n = \frac{\frac{1}{2}KT^2}{NT} \longrightarrow \frac{K}{2}c \neq 0 \ ,$$

and this causes inconsistency.

We consider the consistency of the LIML estimator in many instruments. When viewed as an M-Estimator, which is obtained by minimizing an objective function, the objective function of the GMM estimator is the numerator of (3.6),

$$\tilde{\boldsymbol{\beta}}^{'}\mathbf{G}\tilde{\boldsymbol{\beta}} \quad \stackrel{p}{\longrightarrow} \quad \boldsymbol{\beta}^{'}\mathbf{G}_0\boldsymbol{\beta} + \frac{K}{2}c\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} \ .$$

As the true value is the minimization point of $\boldsymbol{\beta}^{'}\mathbf{G}_0\boldsymbol{\beta}$, the GMM estimator cannot reach the point because it depends on c in the second term. However, with the LIML objective function,

$$\frac{\tilde{\boldsymbol{\beta}}^{'}\mathbf{G}\tilde{\boldsymbol{\beta}}}{\tilde{\boldsymbol{\beta}}^{'}\mathbf{H}\tilde{\boldsymbol{\beta}}} \quad \stackrel{p}{\longrightarrow} \quad \frac{\boldsymbol{\beta}^{'}\mathbf{G}_0\boldsymbol{\beta}}{\sigma^2} + \frac{K}{2}c \ ,$$

where the relation $\sigma^2 = \boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} > 0$ exists. The second term is canceled by σ^2. Therefore, the minimization point can be reached only by the first term without depending on c, whether the objective function is the quadratic form or the variance ratio is the crucial difference between LIML and GMM (TSLS) methods.[6] Notably, the LIML method of Theorem 2.2 is based on data transformation and is not derived as an exact maximum likelihood estimator, and thus, it is a variance ratio estimator. However, Alvarez and Arellano (2003) and Akashi and Kunimoto (2012, 2015) used the name LIML because of the characteristics of its objective functions. Regarding the asymptotic variance of the LIML estimator, the first term can be improved by increasing the instruments, but the second term $c_*\boldsymbol{\Psi}$ becomes large as the instrumental variables increase. Therefore, the aforementioned discussion of White (1999) is precise, if the number $O(KT^2)$ of instrumental variables is finite; that is, the discussion is limited to short panel data.

The above discussion is common to cross-sectional and time series analysis. In the dynamic panel, we have the additional noncentrality parameter as shown in

[6] Anderson (2005) stated that the TSLS had already been derived in the work of Anderson and Rubin (1949), and why LIML was adopted is discussed.

the theorems. By the effect of the forward filter, even the exogenous variable z_{it} on the right-hand side changes into $z_{it}^{(f)}$, which is an endogenous variable. However, the endogeneity is considered $\mathcal{E}[z_{it}^{(f)}u_{it}^{(f)}] \to 0$ when T is large. Hence, its weak endogeneity disappears asymptotically. In the case of regression analysis, only the weak endogeneity exists, and then, the GMM estimator can maintain consistency. If the weak endogeneity accumulates under many instruments and long panel data, then the noncentrality parameter of the LIML estimator appears as follows:

$$\begin{aligned} b_c &\propto \sqrt{NT}\frac{1}{NT}\mathcal{E}\left[\mathbf{z}^{(f)\prime}\mathbf{P}\mathbf{u}^{(f)}\right] \\ &= \sqrt{NT}O\left(\frac{1}{N}\right) \\ &= O\left(\sqrt{c}\right) \ . \end{aligned}$$

3.2 D-LIML Estimator

In this section, we formulate a general model of the dynamic panel structural equations for empirical analyses and describe the assumptions for the following theorems. This section also shows the asymptotic results of estimation methods in long panel data.

3.2.1 General Model

The first structural equation of the general model is as follows:

$$\begin{aligned} y_{it}^{(1)} &= \alpha_i + \boldsymbol{\beta}_2'\boldsymbol{y}_{it}^{(2)} + \boldsymbol{\gamma}_1'\mathbf{z}_{it}^{(1)} + u_{it} \\ &= \alpha_i + \boldsymbol{\theta}_1'\mathbf{x}_{it} + u_{it} \ , \end{aligned} \tag{3.8}$$

where $y_{it}^{(1)}$ and $\boldsymbol{y}_{it}^{(2)} = (y_{it}^{(g)})$ $(g = 2, \cdots, 1 + G_2)$ are $G = 1 + G_2$ endogenous variables in period t. Hence, for the endogenous variables on the right-hand side,

$$\mathcal{E}\left[\mathbf{y}_{it}^{(2)}u_{it}\right] \neq \mathbf{0} \ .$$

Let $\mathbf{z}_{it}^{(1)}$ be the $K_1 \times 1$ vector of the instrumental variable that appears in the first structural equation. The unknown structural parameters are $G_2 \times 1$ and $K_1 \times 1$ for $\boldsymbol{\beta}_2$ and $\boldsymbol{\gamma}_1$, respectively, and are collectively expressed as $\boldsymbol{\theta}_1$. α_i $(i = 1, \cdots, N)$ are the individual effects, and u_{it} stands for the structural error term assuming $\mathcal{E}[u_{it}] = 0$ and $\mathcal{E}[u_{it}^2] = \sigma^2$.

The reduced form of G endogenous variables $(y_{it}^{(1)}, \cdots,\ y_{it}^{(G)})$ appearing in the first structural equation is given by

$$\begin{aligned}\underset{G\times 1}{\mathbf{y}_{it}} &= \underset{G\times K_1}{\mathbf{\Pi}_{1\cdot}^{'}}\,\mathbf{z}_{it}^{(1)} + \underset{G\times K_2}{\mathbf{\Pi}_{2\cdot}^{'}}\,\mathbf{z}_{it}^{(2)} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \\ &= \mathbf{\Pi}^{'}\mathbf{z}_{it} + \boldsymbol{\pi}_i + \mathbf{v}_{it}\ , \end{aligned} \tag{3.9}$$

where $\mathbf{z}_{it}^{(2)}$ is the instrumental variables that does not appear in the first structural equation and is the $K_2 \times 1$ vector. Then, the number of the instrumental variables $\mathbf{z}_{it} = (\mathbf{z}_{it}^{(1)'},\ \mathbf{z}_{it}^{(2)'})'$ in period t becomes $K_1 + K_2 = K < \infty$. $\boldsymbol{\pi}_i$ is the individual effects of $G \times 1$, and $\mathbf{v}_{it}$ stands for the reduced form error assuming $\mathcal{E}[\mathbf{v}_{it}] = \mathbf{0}$ and $\mathcal{E}[\mathbf{v}_{it}\mathbf{v}_{it}^{'}] = \mathbf{\Omega} > \mathbf{O}$ (a positive definite matrix). For the instrumental variable,

$$\mathcal{E}\left[\mathbf{z}_{it}\mathbf{v}_{it}^{'}\right] = \mathbf{O}\ , \quad \mathcal{E}\left[\mathbf{z}_{it}u_{it}\right] = \mathbf{0}\ ,$$

hold in period t. More precisely, $\mathcal{E}[u_{it}|\mathbf{z}_{it}] = 0$ may be used by the conditional expectation. Notably, $\mathbf{z}_{it}$ includes the endogenous variables $(y_{it-1}^{(g)},\ y_{it-2}^{(g)},\ y_{it-3}^{(g)}, \cdots)$ as the lagged endogenous variables or the exogenous $y_{it}^{(g)} (g \neq 1)$ in period t, which is separately determined by a triangular system. In the reduced form of the dynamic panel model, the instrumental variables are generally correlated with individual effects,

$$\mathcal{E}\left[\mathbf{z}_{it}\boldsymbol{\pi}_i^{'}\right] \neq \mathbf{O}\ .$$

However, in the fixed-effects method, they are not a concern. The instrumental variables mentioned here are different from the instrumental variables used for estimators, because the latter is transformed by some filters.

We look at the relation between the parameters of the first structural equation and those of the reduced form parameter. Then, we divide the coefficients of the reduced form as follows:

$$\underset{G\times(K_1+K_2)}{\mathbf{\Pi}^{'}} = \begin{pmatrix} \boldsymbol{\pi}_{11}^{'} & \boldsymbol{\pi}_{21}^{'} \\ \mathbf{\Pi}_{12}^{'} & \mathbf{\Pi}_{22}^{'} \end{pmatrix} \quad \begin{matrix} \}\ 1 \\ \}\ G_2 \end{matrix}\ ,$$

If we set $\boldsymbol{\beta} = (1\ , -\boldsymbol{\beta}_2^{'})'$, then

$$\mathbf{\Pi}\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} \begin{matrix} \}K_1 \\ \}K_2 \end{matrix}\ , \qquad \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} = \sigma^2\ , \qquad \boldsymbol{\beta}^{'}\boldsymbol{\pi}_i = \alpha_i\ ,$$

must be satisfied, because $y_{it}^{(1)}$ in the first structural equation and that in the reduced form are the same variables. The first K equations include the constraints relating the overidentification. The details of the identification for the structural parameters are discussed in the last section because the test of identification seems more difficult than the estimation theory.

The data generation process of reduced form is necessary for the proof of theoretical analysis but, in practice, we do not have to care about it. The dynamics of G-variate endogenous variables involved in the first structural equation may not be autonomous systems, nor are these necessary. Then, we consider the reduced form as a subset of the G^*-dimensional panel VAR(p) model of all possible endogenous variables,

$$\underset{G^*\times 1}{\mathbf{y}^*_{it}} = \mathbf{\Pi}^*_1 \underset{G^*\times 1}{\mathbf{y}^*_{it-1}} + \cdots + \mathbf{\Pi}^*_p \mathbf{y}^*_{it-p} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \ ,$$

where $G \leq G^*$ by the definition, and each $\mathbf{\Pi}^*_p$ becomes the $G^* \times G^*$ square matrix. Moreover, we turn this panel VAR model into the following extended VAR(1) representation:

$$\underset{K^*\times 1}{\mathbf{z}^*_{it}} = \mathbf{\Pi}^{*'} \underset{K^*\times 1}{\mathbf{z}^*_{it-1}} + \boldsymbol{\pi}^*_i + \mathbf{v}^*_{it} \ , \tag{3.10}$$

where $G^* \leq K^*$ and $\mathbf{\Pi}^{*'}$ are the $K^* \times K^*$ square matrix. We call equation (3.10) the companion reduced form, and the standard case is $K^* = G^* p$, that is,

$$\begin{aligned} \mathbf{z}^*_{it} &= (\mathbf{y}^{*'}_{it},\ \mathbf{y}^{*'}_{it-1},\ \cdots,\mathbf{y}^{*'}_{i,t-(p-1)})' \ , \\ \mathbf{z}^*_{it} &= \begin{pmatrix} \mathbf{\Pi}_1 & \mathbf{\Pi}_2 & \cdots & \mathbf{\Pi}_p \\ \mathbf{I}_p & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_p & \cdots & \mathbf{O} \\ & & \cdots & \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{I}_p \end{pmatrix} \mathbf{z}^*_{it-1} + \boldsymbol{\pi}^*_i + \mathbf{v}^*_{it} \ . \end{aligned} \tag{3.11}$$

In this work, as shown in example 2.5 below, equation (3.10) is made for the minimal representation of the K-variate autonomous system. Then, the VAR(1) representation holds even in the case of $K^* < G^* p$.

Using the selection matrix $\mathbf{J}'_1 = (\mathbf{I}_G,\ \mathbf{O})$ and $\mathbf{J}'$ whose elements are 1 or 0, let the endogenous and exogenous variables of (3.9) be

$$\mathbf{y}_{it} = \mathbf{J}'_1 \mathbf{z}^*_{it} \ , \quad \mathbf{z}_{it} = \mathbf{J}' \mathbf{z}^*_{it-1} \ ,$$

where the representative subscript of $\mathbf{z}_{it}$ is t but is uncorrelated with $\mathbf{v}_{it}$ by definition. For convenience, the first G rows of $\mathbf{z}^*_{it}$ are the endogenous variables $\mathbf{y}_{it}$. Then, the relation $\mathbf{J}'_1 \mathbf{\Pi}^{*'} \mathbf{J} = \mathbf{\Pi}'$ exists because $\mathbf{J}_1 \mathbf{\Pi}^{*'} \mathbf{z}^*_{it-1} = \mathbf{\Pi}' \mathbf{z}_{it}$, and $\mathbf{J}'\mathbf{J} = \mathbf{I}_K$. Put $\boldsymbol{\mu}_i = (\mathbf{I} - \mathbf{\Pi}^{*'})^{-1} \boldsymbol{\pi}^*_i$. Then, we have a state-space representation:

$$\begin{aligned} \mathbf{z}^*_{it} &= \mathbf{w}_{it} + \boldsymbol{\mu}_i \ , \\ \underset{K^*\times 1}{\mathbf{w}_{it}} &= \mathbf{\Pi}^{*'} \mathbf{w}_{it-1} + \mathbf{v}^*_{it} \ . \end{aligned} \tag{3.12}$$

The following assumptions are made for the structural estimation in long panel data.[7]

(A1) $\{\mathbf{v}_{it}^*\}$ $(i = 1, \cdots, N;\ t = 1, \cdots, T)$ are i.i.d. across time and individuals and independent of $\mathbf{z}_{i0}^*$ with $\mathcal{E}[\mathbf{v}_{it}^*] = \mathbf{0}$, $\mathcal{E}[\mathbf{v}_{it}^*\mathbf{v}_{it}^{*'}] = \mathbf{\Omega}^*$, and $\mathcal{E}[\|\mathbf{v}_{it}^*\|^8]$ exsit. For some $\mathbf{J}_*$,

$$\underset{K^*\times K^*}{\mathbf{J}_*^{'}\mathbf{\Omega}^*\mathbf{J}_*} = \begin{pmatrix} \underset{G^*\times G^*}{\mathbf{\Omega}_*} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} , \tag{3.13}$$

where $\mathbf{\Omega}_* > \mathbf{O}$. All roots of $|\mathbf{\Pi}^{*'} - \ell\mathbf{I}| = 0$ satisfy the stationarity condition $|\ell_k| < 1$ $(k = 1, \cdots, K^*)$.

(A2) The initial observation satisfies $\mathbf{y}_{i0}^* = (\mathbf{I} - \mathbf{\Pi}^{*'})^{-1}\boldsymbol{\pi}_i^* + \mathbf{w}_{i0}$ $(i = 1, ..., N)$, where $\mathbf{w}_{i0} = \sum_{s=0}^{\infty}(\mathbf{\Pi}^{*'})^s\mathbf{v}_{i,-s}^*$.

(A3) $\boldsymbol{\mu}_i$ are i.i.d. across individuals with $\mathcal{E}[\boldsymbol{\mu}_i] = \mathbf{0}$, $\mathcal{E}[\boldsymbol{\mu}_i\boldsymbol{\mu}_i^{'}] = \mathbf{\Omega}_\mu^*$ with the finite moments up to fourth order and is independent of $\{\mathbf{v}_{it}^*\}$.

These assumptions correspond to those of the regression analysis in the previous part. In fixed-effects estimation, assumption (A3) is unnecessary for the results presented in this work. Except for some proof of theorems, we do not need to define the companion reduced form. As in (3.3), only the setting of the first structural equation and the instrumental variables under the limited information method is necessary for empirical analysis.

In the following example, we can confirm that the reduced form of thr dynamic panel model corresponds to a subsystem of the panel VAR model.

Example 2.5 : The case of the model of Theorem 2.2 holds that $\mathbf{J} = \mathbf{I}_2$ by $\mathbf{\Pi}^{'} = \mathbf{\Pi}^{*'}$, and the reduced form is a two-dimensional panel VAR(1) model through (3.4) and (3.5).

As for a slightly more general model,

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} x_{it-1} + \alpha_i + u_{it} , \\ \mathbf{z}_{it} &= \{\ y_{it-1}^{(1)},\ x_{it-1},\ y_{it-1}^{(2)},\ y_{it-2}^{(2)}\ \} , \end{aligned}$$

where $K_1 = K_2 = 2$. We suppose that the following three-dimensional panel

[7]The existence of the eighth moment of (A1) is made for Theorem 2.2, 2.3, and the other theorems need only that of the fourth moments.

VAR(2) model is behind. Its companion's reduced form becomes

$$\begin{bmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ x_{it} \\ y_{it-1}^{(2)} \\ x_{it-1} \end{bmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{15} & 0 \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{25} & 0 \\ 0 & 0 & \pi_{33} & \pi_{35} & \pi_{36} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ x_{it-1} \\ y_{it-2}^{(2)} \\ x_{it-2} \end{bmatrix} + \begin{bmatrix} \pi_i^{(1)} \\ \pi_i^{(2)} \\ \pi_i^{(3)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_{it}^{(1)} \\ v_{it}^{(2)} \\ \epsilon_{it} \\ 0 \\ 0 \end{bmatrix}, \quad (3.14)$$

where the coefficient matrix of the reduced form $\mathbf{\Pi}'$ corresponds to the upper left 2×4 matrix:

$$\mathbf{\Pi}' = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \pi_{15} \\ \pi_{21} & \pi_{22} & \pi_{23} & \pi_{25} \end{pmatrix}.$$

As all information on the structural parameter is included in this coefficient matrix, only $\mathbf{\Pi}'$ needs to be estimated. Although the reduced form simultaneously determines the endogenous variable $(y_{it}^{(1)},\ y_{it}^{(2)})$ in period t, $\mathbf{\Pi}'$ may not be a square matrix if several exogenous variables exist, and the dynamics are not determined and not autonomous after period $(t+1)$. Thus, x_{it-2} is irrelevant to the reduced form but is added for the VAR(1) representation.

If $x_{it} = y_{it}^{(3)}$ and $\epsilon_{it} = v_{it}^{(3)}$, then $G = 2$ and $G^* = 3$. Hence, $x_{it-1} = y_{it-1}^{(3)}$ in the first structural equation is a lagged endogenous variable $y_{it}^{(3)}$ that does not appear in the structural equation. That is, the predetermined variables are not limited to those of the G endogenous variables that appear in the structural equation. If $x_{it} = y_{it+1}^{(3)}$ and $\mathcal{E}[v_{it}^{(1)} v_{it}^{(3)}] = \mathcal{E}[v_{it}^{(2)} v_{it}^{(3)}] = 0$, then the data generating process of x_{it} means a triangular system. Hence, $x_{it-1} = y_{it}^{(3)}$ in the first structural equation is an exogenous variable in period t. The test of whether $y_{it}^{(3)}$ is exogenous and the model selection of the reduced form are examined in the next part.

As for the standard VAR(1) representation,

$$\begin{bmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ x_{it} \\ y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ x_{it-1} \end{bmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} & 0 & \pi_{15} & 0 \\ \pi_{21} & \pi_{22} & \pi_{23} & 0 & \pi_{25} & 0 \\ 0 & 0 & \pi_{33} & 0 & \pi_{35} & \pi_{36} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{bmatrix} y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ x_{it-1} \\ y_{it-2}^{(1)} \\ y_{it-2}^{(2)} \\ x_{it-2} \end{bmatrix} + \begin{bmatrix} \pi_i^{(1)} \\ \pi_i^{(2)} \\ \pi_i^{(3)} \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_{it}^{(1)} \\ v_{it}^{(2)} \\ \epsilon_{it} \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

where the fourth row and column may be excluded because $y_{it-2}^{(1)}$ does not directly affect the variables in period t. Then, the minimum expression becomes (3.14), and we call the expression of (3.14) the companion reduced form.

3.2.2 D-LIML and Other Estimators to the General Model

Considering a fixed-effects estimation for the general model, the first structural equation applied the forward filter is given as follows:

$$y_{it}^{(1,f)} = \boldsymbol{\beta}_2' \boldsymbol{y}_{it}^{(2,f)} + \boldsymbol{\gamma}_1' \boldsymbol{z}_{it-1}^{(1,f)} + u_{it}^{(f)} ,$$

Applying $\mathbf{D}_f$, the filtered data are obtained:

$$\underset{(T-1)\times 1}{\mathbf{y}_i^{(1,f)}} = (\ y_{it}^{(1,f)}\ ) , \quad \underset{(T-1)\times G_2}{\mathbf{Y}_i^{(2,f)}} = (\ \mathbf{y}_{it}^{(2,f)'}\ ) , \quad \underset{(T-1)\times K_1}{\mathbf{Z}_i^{(1,f)}} = (\ \mathbf{z}_{it}^{(1,f)'}\ ) .$$

The data reorganized in each period t are expressed as follows:

$$\underset{N\times 1}{\mathbf{y}_t^{(1,f)}} = (y_{it}^{(1,f)}) , \quad \underset{N\times G_2}{\mathbf{Y}_t^{(2,f)}} = (\mathbf{y}_{it}^{(2,f)}) , \quad \underset{N\times K_1}{\mathbf{Z}_t^{(1,f)}} = (\mathbf{z}_{it}^{(1,f)}) .$$

The variable on the right-hand side of the first structural equation is summarized by the $N \times (G_2 + K_1)$ matrix:

$$\mathbf{X}_t^{(f)} = \left(\mathbf{Y}_t^{(2,f)}, \ \mathbf{Z}_t^{(1,f)'} \right) .$$

The LIML method studied in Alonso-Borrego and Arellano (1999) is based on the following two $(G + K_1) \times (G + K_1)$ matrices,

$$\begin{aligned} \mathbf{G}^{(f)} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} \mathbf{P}_t \left(\mathbf{y}_t^{(1,f)}, \ \mathbf{X}_t^{(f)} \right) , \\ \mathbf{H}^{(f)} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} [\mathbf{I}_N - \mathbf{P}_t] \left(\mathbf{y}_t^{(1,f)}, \ \mathbf{X}_t^{(f)} \right) , \end{aligned}$$

For $\boldsymbol{\theta} = (1, \ -\boldsymbol{\theta}_1')'$, the estimator corresponds to the minimization point $\tilde{\boldsymbol{\theta}}_{\text{LI}} = (\tilde{\boldsymbol{\beta}}_{\text{LI}}', \ \tilde{\boldsymbol{\gamma}}_{\text{LI}}')'$ of

$$\mathcal{VR} = \frac{\boldsymbol{\theta}' \mathbf{G}^{(f)} \boldsymbol{\theta}}{\boldsymbol{\theta}' \mathbf{H}^{(f)} \boldsymbol{\theta}} .$$

The general case has one difference in terms of the instrumental variables, which are included in the estimators. In the case of the general model, we have

$$\underset{K\times 1}{\mathcal{E}\left[\mathbf{z}_{it} u_{it}^{(f)} \right]} = \mathbf{0} , \ (t = 1, \cdots, T-1) .$$

The number of orthogonal conditions increases as period t becomes larger. However, for the projection matrix $\mathbf{P}_t = \mathbf{Z}_t(\mathbf{Z}_t'\mathbf{Z}_t)^{-1}\mathbf{Z}_t$, $\mathbf{Z}_t = (\mathbf{z}_{it})$ cannot always be $N \times Kt$. As in example 2.5, if the model consists of the AR(2) process, then

$(y_{it-1}^{(2)},\ y_{it-2}^{(2)},\ y_{it-3}^{(2)}\ ,\cdots)$ and $(y_{it-2}^{(2)},\ y_{it-3}^{(2)},\ y_{it-4}^{(2)}\ ,\cdots)$ overlap in a set of instrumental variables in a certain period t, or redundant instruments exist. That is, the rank of $\mathbf{Z}_t$ is reduced, and thus, we have to set $\mathbf{Z}_t$ as $N \times G^* t$ by selecting the G^* different series. In the case of example 2.5, G^* is equal to 3 from $(y_{i\cdot}^{(1)},\ y_{i\cdot}^{(2)},\ x_{i\cdot})$, and the number of total instrumental variables becomes $O(G^* T^2)$.

We consider another estimation method as follows. In Part I, we introduced the instrumental variables applied the backward filter, and Hayakawa (2006b) obtained the same results as Theorem 1.5 based on a GMM estimator. As the reduced form is implicitly estimated even in the structural estimation, the optimal instrumental variables are expected to improve the LIML estimator. Akashi and Kunitomo (2015) investigated the following fixed-effects estimation.

For the $(G + K_1) \times (G + K_1)$ matrices,

$$\begin{aligned}
\mathbf{G}^{(f,b)} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} \mathbf{P}_t^{(b)} \left(\mathbf{y}_t^{(1,f)},\ \mathbf{X}_t^{(f)}\right) , \\
\mathbf{H}^{(f,b)} &= \sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{y}_t^{(1,f)'} \\ \mathbf{X}_t^{(f)'} \end{pmatrix} \left[\mathbf{I}_N - \mathbf{P}_t^{(b)}\right] \left(\mathbf{y}_t^{(1,f)},\ \mathbf{X}_t^{(f)}\right) ,
\end{aligned}$$

let $\hat{\boldsymbol{\theta}}_{\mathrm{DL}} = (\hat{\boldsymbol{\beta}}_{\mathrm{DL}}^{'},\ \hat{\boldsymbol{\gamma}}_{\mathrm{DL}}^{'})'$ be the minimization point of the follwing:

$$\mathcal{VR}_1 = \frac{\boldsymbol{\theta}' \mathbf{G}^{(f,b)} \boldsymbol{\theta}}{\boldsymbol{\theta}' \mathbf{H}^{(f,b)} \boldsymbol{\theta}} \ . \tag{3.15}$$

In this work, we call it D-LIML estimator. The difference from $\tilde{\boldsymbol{\theta}}_{LI}$ is that the D-LIML estimator uses the projection matrix $\mathbf{P}_t^{(b)} = \mathbf{Z}_t^{(b)} (\mathbf{Z}_t^{(b)'} \mathbf{Z}_t^{(b)})^{-1} \mathbf{Z}_t^{(b)}$, and the number of orthogonal conditions is the same for each t,

$$\mathcal{E}\underset{K \times 1}{\left[\mathbf{z}_{it}^{(b)} u_{it}^{(f)}\right]} = \mathbf{0} \ .$$

The instrumental variable is applied the backward filter $\mathbf{D}_b$,

$$\mathbf{z}_{it-1}^{(b)} = \left[\mathbf{z}_{it-1} - \frac{1}{t}(\mathbf{z}_{it-2} + \cdots + \mathbf{z}_{i0} + \mathbf{z}_{i,-1})\right] ,$$

where $\mathbf{z}_{i(-1)}$ is included to simplify the notation regarding the range of subscript. Then, the instrumental variable matrix $\mathbf{Z}_t^{(b)} = (\mathbf{z}_{it}^{(b)})$ is $N \times K$, and thus, the number of total instrumental variables is reduced to $O(KT)$.

If minimizing the numerator of variance ratio $\mathcal{VR}_1$,

$$\mathcal{Q}_1 = \boldsymbol{\theta}' \mathbf{G}^{(f,b)} \boldsymbol{\theta} \ ,$$

then the GMM estimator (D-GMM) $\hat{\boldsymbol{\theta}}_{\mathrm{DG}} = (\hat{\boldsymbol{\beta}}_{\mathrm{DG}}^{'},\ \hat{\boldsymbol{\gamma}}_{\mathrm{DG}}^{'})'$ is obtained as the minimization point. The explicit forms of these estimators are given in the section of numerical experiments.

We prepare the notations to state the results of the asymptotic theory. Regarding the moment matrix of instrumental variables,

$$\underset{K^*\times K^*}{\mathbf{\Gamma}_0} = \mathcal{E}\left[\mathbf{w}_{it-1}\mathbf{w}'_{it-1}\right] = \sum_{h=0}^{\infty}(\mathbf{\Pi}^{*'})^h\mathbf{\Omega}^*\mathbf{\Pi}^{*h} > \mathbf{O}\ , \tag{3.16}$$

because $\mathbf{\Omega}_* > \mathbf{O}$ (cf. Anderson (1971, Ch. 5)). Then, $\mathbf{\Omega} > \mathbf{O}$ and $\sigma^2 > 0$ hold. The leading term of the asymptotic variance becomes the same as that of the usual LIML estimator,

$$\underset{(G_2+K_1)\times(G_2+K_1)}{\mathbf{\Phi}} = \mathbf{\Pi}_{\mathrm{I}}'\mathbf{J}'\mathbf{\Gamma}_0\mathbf{J}\mathbf{\Pi}_{\mathrm{I}}\ , \qquad \underset{(G_2+K_1)\times K}{\mathbf{\Pi}_{\mathrm{I}}'} = \begin{pmatrix} \mathbf{\Pi}_{12}' & \mathbf{\Pi}_{22}' \\ \mathbf{I}_{K_1} & \mathbf{O} \end{pmatrix}\ . \tag{3.17}$$

If $\mathrm{rank}(\mathbf{\Pi}_{22}) = G_2$, then $\mathbf{\Phi} > \mathbf{O}$, which can be tested in the last section. The noncentrality parameter may appear. From the previous discussion, the reason is the influence of many instruments based on $T/N \to c$. If depending on $N/T \to d$, then thw reason is the problem related to the initial value as shown in a later section. Both noncentrality parameters are proportional to the following $\boldsymbol{\rho}^*$:

$$\begin{aligned} \boldsymbol{\rho}^* &= \mathbf{\Phi}^{-1}\mathbf{\Pi}_{\mathrm{I}}'\mathbf{J}'(\mathbf{I}-\mathbf{\Pi}^{*'})^{-1}\mathbf{\Omega}^*\mathbf{J}_1\boldsymbol{\beta}\ , \\ \boldsymbol{\rho}_0 &= \mathbf{\Phi}^{-1}\begin{bmatrix} \mathbf{J}_2'\mathbf{\Omega}\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}\ , \end{aligned}$$

where $\mathbf{J}_2' = (\mathbf{0},\ \mathbf{I}_{G_2})$.

In the structural estimation for the general model (3.8), the asymptotic results of $\tilde{\boldsymbol{\theta}}_{\mathrm{LI}}$, $\tilde{\boldsymbol{\theta}}_{\mathrm{DG}}$, and $\tilde{\boldsymbol{\theta}}_{\mathrm{DL}}$ in long panel data are as follows.

Theorem 2.3 (Akashi and Kunitomo, 2015) : *Suppose assumptions (A1) and (A2) and that* $\mathbf{v}_{it}$ *follows a normal distribution.*

[i] *Provided* $0 \le G^* \lim_{N,T\to\infty}(T/N) < 1$ *and assumption (A3), then as both* N *and* $T \to \infty$,

$$\sqrt{NT}\left(\tilde{\boldsymbol{\theta}}_{\mathrm{LI}} - \boldsymbol{\theta}_1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{b}_c, \sigma^2\mathbf{\Phi}^{-1} + c_*\mathbf{\Psi}\right)\ ,$$

where

$$\begin{aligned} \mathbf{b}_c &= -\sqrt{\frac{G^*}{2}}\frac{\sqrt{c}}{1-c}\boldsymbol{\rho}^*\ , \\ \mathbf{\Psi} &= \mathbf{\Phi}^{-1}\mathbf{J}_{12}\left[\sigma^2\mathbf{\Omega} - \mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}\right]\mathbf{J}_{12}'\mathbf{\Phi}^{-1}, \end{aligned}$$

$c_* = c/(1-c)$, *and* $\mathbf{J}_{12}' = (\mathbf{J}_2,\ \mathbf{O})$.

[ii] *Suppose* $c_1 = K/N > 0$ *or* N *is fixed. Then, as* $T \to \infty$,

$$\tilde{\boldsymbol{\theta}}_{\text{DG}} - \boldsymbol{\theta}_1 \quad \xrightarrow{p} \quad \left[\boldsymbol{\Phi} + c_1 \begin{pmatrix} \mathbf{J}_2'\boldsymbol{\Omega}\mathbf{J}_2 & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\right]^{-1} \begin{bmatrix} c_1\mathbf{J}_2'\boldsymbol{\Omega}\boldsymbol{\beta} \\ \mathbf{O} \end{bmatrix} ,$$

As for $c_1 = 0$ *or* $N \to \infty$, *provided* $0 \le \lim_{N,T\to\infty}(T/N) = c < \infty$, *then*

$$\sqrt{NT}\left(\tilde{\boldsymbol{\theta}}_{\text{DG}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(\mathbf{b}_{1\cdot 0}, \sigma^2\boldsymbol{\Phi}^{-1}\right) ,$$

where

$$\mathbf{b}_{1\cdot 0} = K\sqrt{c}\boldsymbol{\rho}_0 .$$

[iii] *For* $0 \le c_1 < 1$, *then, as* $T \to \infty$,

$$\sqrt{NT}\left(\tilde{\boldsymbol{\theta}}_{\text{DL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1} + c_{1*}\boldsymbol{\Psi}\right) ,$$

where $c_{1*} = c_1/(1-c_1)$.

Similar to Theorem 2.2, the assumption of normality is to express the asymptotic variance concisely. These results do not depend on the parameter of the individual effect $\boldsymbol{\Omega}_\mu^*$. The difference between the assumptions of [i] and those of [ii] and [iii] is that the former must be the double asymptotics whereas the latter can be $T \to \infty$ only. Hence, the definitions of $c = \lim T/N$ and $c_1 = \lim K/N$ are also different.

First, we compare [i] with [iii], which are the results of the LIML methods. The result of [i] is reduced to Theorem 2.2 when $K = 2$, but the correction of the noncentrality parameter becomes difficult because $\boldsymbol{\Pi}^*$ must be estimated. Moreover, in a general model, the data sequence may be constrained by the following:

$$G^*T < N \quad \Rightarrow \quad \frac{T}{N} < \frac{1}{G^*} ,$$

to define the projection matrix in period T, and thus, it cannot be provided for any long panel data. To compare relative efficiency, set $c_{1*} = 0$ or the double asymptotics. Then, the difference in the asymptotic covariance matrices is given by the following:

$$(\sigma^2\boldsymbol{\Phi}^{-1} + c_*\boldsymbol{\Psi}) - \sigma^2\boldsymbol{\Phi}^{-1} \ge \mathbf{O} .$$

Therefore, the D-LIML estimator of [iii] is relatively efficient.

Second, we compare the D-GMM estimator of [ii] with the D-LIML estimator of [iii]. The LIML method always does not have a noncentrality parameter. Although the GMM estimator is consistent in the double asymptoitcs, if N is regarded as fixed, then it becomes inconsistent. As for the approximation of the asymptotic distribution in a finite sample, the LIML estimator would be better because the second term $c_{1*}\mathbf{\Psi}$ does not ignore the value of K/N when N is fixed. Therefore, we conclude that the D-LIML estimator of [iii] has better asymptotic properties than that of [i] and [ii]. The finite sample properties of D-LIML estimator are expected to be better than those of the D-GMM estimator, and in fact, it will be.

There is a remark on the D-LIML estimator. The ratio of the number of instruments to that of the data should satisfy $r_{1n} = K(T-1)/NT \simeq K/N < 1$. Although $K < N$ does not seem to be restricted in a cross-sectional analysis, it is somewhat puzzling because it cannot be used in a time-series analysis ($N = 1,\ T \to \infty$), which is a special case of panel analysis. This is not a problem with the property of the LIML method, but the usage of the orthogonal conditions can be further improved. This problem is reconsidered and improved in a later section.

3.3 Transformed LIML Estimator

The transformed maximum likelihood method discussed by Hsiao (2014, Ch. 4) is another different approach from the fixed-effects estimation. Hsiao et al. (2002), Binder et al. (2005), Hayakawa and Pesaran (2015), and Hsiao and Zhang (2015) examined the transformed MLE with the regression analysis. In addition, Hsiao and Zhou (2015) investigated the structural analysis. Although this method is an exact maximum likelihood estimator, the approach is also different from the random-effects MLE, which assumes the identical distribution for individual effects. In previous studies, the finite sample properties show that the transformed method is better than the estimator of Arellano and Bond (1991) for the reduced form, and the estimator of Akashi and Kunitomo (2015) for the structural equation. We reconsider the transformed maximum likelihood estimator in the following sections.

Considering the reduced form of the AR(1) model again, if we take the firest-difference in (2.1), then the individual effect η_i disappears.

$$\Delta y_{it} = \pi \Delta y_{it-1} + \Delta v_{it}\ , \quad (t = 2,\ 3,\ \cdots, T)\ .$$

However, the right-hand side of Δy_{i1} in period $t = 1$ is the problem because Δy_{i0} cannot be observed. Through repeated substitution,

$$\begin{aligned}\Delta y_{i1} &= \pi^s \Delta y_{i1-s} + \sum_{h=0}^{s-1} \pi^h \Delta v_{i1-h} \\ &= \pi^s \Delta y_{i1-s} + \epsilon_{i1}\ ,\end{aligned}$$

where $s < \infty$ can be allowed. However, for the sake of simplicity, we put $s \to \infty$ similar to assumption (a2). Then,

$$\mathcal{E}[\Delta y_{i1}] = 0 \ , \quad \mathcal{V}ar[\Delta y_{i1}] = \frac{2\omega}{1+\pi} = \omega_1 \ ,$$

and the correlations of the error term become

$$\mathcal{E}[\epsilon_{i1}\Delta v_{i2}] = -\omega \ , \quad \mathcal{E}[\epsilon_{i1}\Delta v_{it}] = 0 \ , \ (t \geq 3) \ .$$

Hsiao et al. (2002) discussed some types of the data generation process for Δy_{i1}, and stated that even if the form of variance is different it can be expressed as the free parameter ω_1. Then, the correlation structure for the entire period is determined, and thus, the joint distribution of

$$\begin{aligned}\underset{T\times 1}{\Delta \mathbf{y}_i^*} &= (\Delta y_{i1}, \ \Delta y_{i2}, \ \cdots, \Delta y_{iT})' \\ &= (\Delta y_{i1}, \ \Delta \mathbf{y}_i')' \ ,\end{aligned}$$

is also obtained under the assumption of a normal distribution. The Jacobian of the transformation form $\Delta \mathbf{y}_i^*$ to the following error vector is unity,

$$\underset{T\times 1}{\Delta \mathbf{v}_i^*} = (\Delta y_{i1}, \ \Delta v_{i2}, \ \cdots, \Delta v_{iT})' \ .$$

The variance-covariance matrix of the error terms becomes

$$\mathbf{\Omega}_\Delta = \omega \mathbf{W}_0 \ ,$$

where

$$\underset{T\times T}{\mathbf{W}_0} = \begin{pmatrix} \omega_0 & -1 & 0 & \cdots & 0 \\ -1 & & & & \\ 0 & & \mathbf{W} & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix} .$$

$\mathbf{W}$ is given by (2.6), and $\omega_0 = \omega_1/\omega$ is redefined. Then, the log-likelihood function for the joint distribution of unconditional $\Delta \mathbf{y}_i^*$ is as follows:

$$\mathcal{L}_\Delta = -\frac{N}{2} \log |\mathbf{\Omega}_\Delta| - \frac{1}{2} \sum_{i=1}^{N} \Delta \mathbf{v}_i^{*'} \mathbf{\Omega}_\Delta^{-1} \Delta \mathbf{v}_i^* \ , \tag{3.18}$$

except for the constant part, where

$$\underset{T\times 1}{\Delta \mathbf{v}_i^*} = (\Delta y_{i1}, \ \Delta y_{i2} - \pi \Delta y_{i1}, \ \cdots, \Delta y_{iT} - \pi \Delta y_{iT-1})' \ .$$

Maximizing with respect to the unknown parameter $(\pi,\ \omega,\ \omega_0)$, the transformed MLE $\hat{\pi}_{\text{TM}}$ (T-MLE) is obtained. A feature is that it is the exact likelihood function after the data transformation, which can be called a fixed-effects MLE, and does not depend on the individual effect η_i. The first element Δy_{i1} of $\Delta \mathbf{v}_i^*$ that is treated as an error term should be observed, and its variance is estimated by the free parameter ω_1. If we use the pseudo likelihood function of $\Delta \mathbf{y}_i$ without the initial distribution of Δy_{i1}, then the objective function is equivalent to Lemma 2.1, as wil be described later:

$$\min_{\pi} \ \sum_{i=1}^{N} (\mathbf{y}_i - \pi \mathbf{y}_{i,-1})' \mathbf{Q}_{T-1} (\mathbf{y}_i - \pi \mathbf{y}_{i,-1}) \ .$$

That is, the CV estimator is obtained. The T-MLE has the asymptotic normality,

$$\sqrt{N}\,(\hat{\pi}_{\text{TM}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}(\ 0,\ \frac{\omega}{\phi_{0T}}\ )\ ,$$

where the asymptotic variance is

$$\phi_{0T} = \mathcal{E}\left[\begin{pmatrix} 0, & \Delta \mathbf{y}_{i,-1}' \end{pmatrix} \mathbf{W}_0^{-1} \begin{pmatrix} 0 \\ \Delta \mathbf{y}_{i,-1} \end{pmatrix} \right] \ .$$

Then, the following holds in the short panel data.

Theorem 2.4 (Hsiao et al., 2002) : *Supposing assumptions (a1)-(a3) hold, then as $N \to \infty$ and T is fixed,*

$$\frac{\omega}{\phi_T} \geq \frac{\omega}{\phi_{0T}}\ .$$

That is, T-MLE is relatively efficient than the AB estimator in Part I. Hsiao et al. (2002) suggested in the numerical experiment that the gain of efficiency by estimating initial values is large in short panel data. The result should be held in the limit of $T \to \infty$, that is, the gain would also be more efficient in long panel data. As $\omega/\phi_T \to 1 - \pi^2$, T-MLE may also reache the lower bound of efficiency. Although we would like to confirm it in the asymptotic theory ($T \to \infty$), the structure of the log-likelihood function is given by as follows:

$$\begin{aligned} |\mathbf{\Omega}_\Delta| &= \omega^{2T}(1+T(\omega_0-1)) \ , \\ \mathbf{\Omega}_\Delta^{-1} &= \frac{1}{\omega}\mathbf{W}_0^{-1} \\ &= \tfrac{1}{\omega(1+T(\omega_0-1))}\begin{pmatrix} T & T-1 & \cdots & 2 & 1 \\ T-1 & (T-1)\omega_0 & \cdots & 2\omega_0 & \omega_0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 2\omega_0 & \cdots & 2((T-2)\omega_0-(T-3)) & (T-2)\omega_0-(T-3) \\ 1 & \omega_0 & \cdots & (T-2)\omega_0-(T-3) & (T-1)\omega_0-(T-2) \end{pmatrix} . \end{aligned} \tag{3.19}$$

Thus, structure is highly nonlinear, and derivation becomes complicated as it is.

3.3.1 Long Difference

Grassetti (2011) provided another useful representation of the transformed maximum likelihood method:

$$\mathbf{D}_\ell \mathbf{y}_i^* = \mathbf{y}_i - y_{i0}\boldsymbol{\iota} \ , \tag{3.20}$$

where $\mathbf{y}_i^* = (y_{i0},\ y_{i1},\ ,\cdots, y_{iT})'$ is $(T+1)\times 1$ including the initial value y_{i0}, and

$$\underset{T\times(T+1)}{\mathbf{D}_\ell} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ & & & \cdots & \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix} .$$

The relations with the transformed method using the first-difference are given by as follows:

$$\begin{aligned} \mathbf{D}_\ell \mathbf{y}_i^* &= \mathbf{L}\Delta\mathbf{y}_i^* \\ &= (\mathbf{L}\mathbf{D}_{T+1})\mathbf{y}_i^* \ , \end{aligned}$$

where $\mathbf{D}_{T+1}$ is the $T\times(T+1)$ first-difference matrix of (2.4) and $\mathbf{L}$ is the cumulative matrix, which is a lower triangular matrix, and is nonsingular:

$$\underset{T\times T}{\mathbf{L}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ & & & \cdots & \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} . \tag{3.21}$$

Therefore, the first-difference, including Δy_{i1}, is also represented by the difference from the initial value y_{i0}, and then, (3.20) is called the long difference.

Consider the reduced form of AR(1), for period $t \geq 1$, an identity is formulated as follows:

$$y_{it} - y_{i0} = \pi(y_{it-1} - y_{i0}) + (-y_{i0} + \pi y_{i0}) + \eta_i + v_{it} \ .$$

Let $y_{it}^{(\ell)}$ be the data applied the long difference,

$$\begin{aligned} y_{it}^{(\ell)} &= \pi y_{it-1}^{(\ell)} + (-y_{i0} + \pi y_{i0}) + \eta_i + v_{it} \\ &= \pi y_{it-1}^{(\ell)} - (1-\pi)\left(w_{i0} + \frac{\eta_i}{1-\pi}\right) + \eta_i + v_{it} \\ &= \pi y_{it-1}^{(\ell)} + \xi_i + v_{it} \ , \end{aligned}$$

where we note that $y_{i0}^{(\ell)} = 0$. The second equation uses the state space representation of (2.2) so that the individual effect η_i disappears. However, ξ_i appears instead of η_i ,

$$\begin{aligned} \xi_i &= -(1-\pi)w_{i0} \\ &= -(1-\pi)\sum_{s=0}^{\infty} \pi^s v_{i,-s} \ , \end{aligned}$$

where ξ_i is invariant for $t \geq 1$ so that the subscript becomes only i. That is, the long difference eliminates the original individual effect η_i but uses an artificial individual effect ξ_i as an error correction. Importantly, ξ_i can assume a random-effect which follows the identical distribution.

The transformation $\Delta \mathbf{y}_i^* = \mathbf{D}_{T+1}\mathbf{y}_i^*$ is the shift by $\mathbf{y}_i - y_{i0}\boldsymbol{\iota}$ because of (3.20), and it remains $T \times 1$. Meanwhile, the filtered data $\mathbf{D}_T\mathbf{y}_i$ are $(T-1)\times 1$. The transformed maximum likelihood method is invariant to a regular transformation $\mathbf{T}$ as follows.

Lemma 2.1 (Hsiao et al., 2002) : *For $\mathbf{D}_{T+1}\mathbf{y}_i^*$, the log likelihood function $\mathcal{L}_0$ of the transformed data $\mathbf{T}\mathbf{D}_{T+1}\mathbf{y}_i^*$ becomes*

$$\mathcal{L}_0 = -N \log|\mathbf{T}| + \mathcal{L}_\Delta \ .$$

Therefore, we need to consider only the first-difference matrix $\mathbf{D}_{T+1}$ on how to eliminate the individual effect. If $\mathbf{T}$ does not depend on unknown parameters of interest, then we can select a retransformation $\mathbf{T}$ that is easy to calculate. In the case of the long difference, $\mathbf{T} = \mathbf{L}$, and $\mathbf{L}$ does not depend on the unknown parameters. From Lemma 2.1, we conclude that the long difference is another expression of the transformed maximum likelihood method.

We consider the log-likelihood function $\mathcal{L}_0$. For the long difference, η_i can be elimonated even with $-y_{i1}\boldsymbol{\iota}$. However, if the initial value $-y_{i0}\boldsymbol{\iota}$ is used, then $\mathcal{E}[\xi_i v_{it}] = 0$ $(t \geq 1)$, which is the simplest. Then,

$$\underset{T\times T}{\boldsymbol{\Omega}_{\xi v}} = \omega_\xi \boldsymbol{\iota}\boldsymbol{\iota}' + \omega \mathbf{I}_T \ ,$$

where $\omega_\xi = \mathcal{V}ar[\xi_i]$, which is equivalent to the correlation structure of the random-effects MLE and is easy to handle. The log-likelihood function is given by

$$\mathcal{L}_0 = -\frac{N}{2}\log|\mathbf{\Omega}_{\xi v}| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i^{(\ell)} - \pi\mathbf{y}_{i,-1}^{(\ell)})'\mathbf{\Omega}_{\xi v}^{-1}(\mathbf{y}_i^{(\ell)} - \pi\mathbf{y}_{i,-1}^{(\ell)}) \ ,$$

where $\mathbf{y}_i^{(\ell)} = (y_{i1}^{(\ell)},\ \cdots,\ y_{iT}^{(\ell)})'$ and $\mathbf{y}_{i,-1}^{(\ell)} = (0,\ \cdots,\ y_{iT-1}^{(\ell)})'$. Moreover, when maximized with respect to the unknown parameter $(\pi,\ \omega,\ \omega_\xi)$, the transformed MLE $\hat{\pi}_{\text{TM}}$(T-MLE) is obtained. This log-likelihood is simpler than (3.18), Grassetti (2011) pointed out the advantage that it can be calculated through the random-effects routine of existing packages.

Next, consider the case of structural estimation. For a simple structural model presented in the previous section, Hsiao and Zhou (2015) proposed a transformed LIML estimator using the expression by the long difference:

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} \ , \\ y_{it}^{(2)} &= \pi_{21} y_{it-1}^{(1)} + \pi_{22} y_{it-1}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)} \ , \end{aligned} \tag{3.22}$$

where $y_{it}^{(2)}$ is represented as the reduced form by the limited information method. With the long difference,

$$\begin{aligned} y_{it}^{(1,\ell)} &= \beta_2 y_{it}^{(2,\ell)} + \gamma_1 y_{it-1}^{(1,\ell)} + \xi_{ui} + u_{it} \ , \\ y_{it}^{(2,\ell)} &= \pi_{21} y_{it-1}^{(1,\ell)} + \pi_{22} y_{it-1}^{(2,\ell)} + \xi_i^{(2)} + v_{it} \ , \end{aligned}$$

where $y_{it}^{(g,\ell)} = y_{it}^{(g)} - y_{i0}^{(g)}$ $(g = 1,\ 2\ ;\ t \geq 0)$. $(\xi_{ui},\ \xi_i^{(2)})$ turns out to be a random individual effect generated from the initial state. The notations are regarded as $\boldsymbol{\pi}_2 = (\pi_{21},\ \pi_{22})$,

$$\underset{2\times 2}{\mathbf{\Omega}_u} = \mathcal{E}\left[(u_{it},\ v_{it}^{(2)})'(u_{it},\ v_{it}^{(2)})\right] \ , \quad \underset{2\times 2}{\mathbf{\Omega}_{u\xi}} = \mathcal{E}\left[(\xi_{ui},\ \xi_i^{(2)})'(\xi_{ui},\ \xi_i^{(2)})\right] \ ,$$

and then, the correlation structure is as follows:

$$\underset{2T\times 2T}{\mathbf{\Omega}_{\xi u}} = \mathbf{\Omega}_{u\xi} \otimes \boldsymbol{\iota}\boldsymbol{\iota}' + \mathbf{\Omega}_u \otimes \mathbf{I}_T \ .$$

The log-likelihood function under the limited information method is given by the following:

$$\mathcal{L}_1 = -\frac{N}{2}\log|\mathbf{\Omega}_{\xi u}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_{ui}^{(\ell)\prime}\mathbf{\Omega}_{\xi u}^{-1}\mathbf{v}_{ui}^{(\ell)} \ ,$$

where

$$\underset{2T\times 1}{\mathbf{v}_{ui}^{(\ell)}} = \begin{bmatrix} \mathbf{y}_i^{(1,\ell)} - \mathbf{X}_i^{(\ell)}\boldsymbol{\theta}_1 \\ \mathbf{y}_i^{(2,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}\boldsymbol{\pi}_2 \end{bmatrix} \ , \quad \underset{T\times 2}{\mathbf{X}_i^{(\ell)}} = \left(\mathbf{y}_i^{(2,\ell)},\ \mathbf{y}_{i,-1}^{(1,\ell)}\right) \ , \quad \underset{T\times 2}{\mathbf{Y}_{i,-1}^{(\ell)}} = \left(\mathbf{y}_{i,-1}^{(1,\ell)},\ \mathbf{y}_{i,-1}^{(2,\ell)}\right) \ .$$

They provide the exact LIML estimator of the fixed effect method. If we set the parameters as follows:

$$\boldsymbol{\phi} = (\boldsymbol{\theta}_1', \ \boldsymbol{\pi}_2')' \ , \quad \boldsymbol{\omega}_* = (\text{vec}(\boldsymbol{\Omega}_u)', \ \text{vec}(\boldsymbol{\Omega}_{u\xi})')' \ ,$$

then $\hat{\boldsymbol{\phi}}_{\text{TL}}$(T-LIML) is obtained by maximizing $\mathcal{L}_1$ with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\omega}_*$.

Theorem 2.5 (Hsiao and Zhou, 2015) : *Supposing assumptions (A1) and (A2) hold and that* $(v_{it}^{(1)}, \ v_{it}^{(2)})$ *follows a normal distribution, then as* $N \to \infty$ *or* $T \to \infty$ *or both,*

$$\sqrt{NT}\left(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(\mathbf{0}, \ -(\mathbf{H}_{\phi\phi} - \mathbf{H}_{\phi\omega}\mathbf{H}_{\omega\omega}^{-1}\mathbf{H}_{\phi\omega}')^{-1} \right) \ ,$$

where

$$\mathbf{H}_{\phi\phi} = \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\right] \ , \ \mathbf{H}_{\phi\omega} = \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\phi} \partial \boldsymbol{\omega}_*'}\right] \ , \ \mathbf{H}_{\omega\omega} = \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\omega}_* \partial \boldsymbol{\omega}_*'}\right] \ .$$

As $(\xi_{ui}, \ \xi_i^{(2)})$ also follows a normal distribution due to the normality of the error term, the asymptotic variance-covariance matrix of $(\hat{\boldsymbol{\phi}}_{\text{TL}}, \ \hat{\boldsymbol{\omega}}_{*\text{TL}})$ is given by the inverse of the information matrix, which becomes a simple structure. This result is desirable because the noncentrality parameter is zero even if a nonnormality assumption exists. Considering that T-LIML estimator is an exact maximum likelihood method, the score functions corresponding to the orthogonal conditions become a finite number which is equal to that of unknown parameters. Thus, the many instruments problem does not occur. In the next section, we investigate the results of Hsiao and Zhou (2015) in more detail.

3.3.2 Asymptotic Variance When $T \to \infty$

This section clarifies the asymptotic variance for the structural parameter $\boldsymbol{\theta}_1$ of interest within $\hat{\boldsymbol{\phi}}_{\text{TL}}$ and compares it with other estimators. Before that, we derive the asymptotic variance for the reduced form AR (1) model using the long difference and confirm that the conjecture relating to Theorem 2.4 is correct. Hsiao and Zhang (2015) derived the T-MLE with the first-difference, and Hsiao and Zhou (2016) derived the T-LIML using the long difference when $\boldsymbol{\Omega}_{\xi v}$ is given. We consider the following assumptions for the case when $\boldsymbol{\Omega}_{\xi v}$ is not given.

We generally refer to the log-likelihood functions of the transformed maximum likelihood method as $\mathcal{L}$. In deriving the maximum likelihood estimator, the following is obtained by the Taylor series:

$$\sqrt{NT}(\hat{\boldsymbol{\psi}}_{\text{TL}} - \boldsymbol{\psi}) = -\left(\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\psi}^* \partial \boldsymbol{\psi}^{*'}}\right)^{-1} \frac{1}{\sqrt{NT}}\frac{\partial \mathcal{L}}{\partial \boldsymbol{\psi}} \ , \tag{3.23}$$

where $\boldsymbol{\psi}$ denotes all parameters appearing in the transformed maximum likelihood estimator. For instance, $\boldsymbol{\psi} = (\boldsymbol{\phi}^{'},\ \boldsymbol{\omega}_*^{'})'$ holds in the case of Theorem 2.5 and $\boldsymbol{\psi}^*$ is a mean value between the estimator $\boldsymbol{\psi}^*$ and the true value $\boldsymbol{\psi}$. The following are made for the asymptotic results of the transformed methods.

(A3) [i] As N and T tend to infinity,

$$\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\psi}^* \partial \boldsymbol{\psi}^{*'}} \quad \xrightarrow{p} \quad \mathbf{H}_{\psi\psi} = \lim_{N,\, T\to\infty} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'}\right] \ .$$

[ii] As T tends to infinity, $(1/NT)\mathcal{L}$ converges to a nonstchastic function which attains a unique global maximum at $\boldsymbol{\psi}$.

The transformed maximum likelihood estimator shows that some off-diagonal elements of $\mathbf{H}_{\psi\psi}$ are zero. Thus we do not need to derive all elements of $\mathbf{H}_{\psi\psi}$ under assumption [i]. Then, we focus on the inverse matrix of the diagonal block of $\mathbf{H}_{\psi\psi}$. Although the asymptotic property of the transformed maximum likelihood method may be obtained under $N < \infty$, the law of large numbers then does not hold for the terms related to the random effects ξ_i ($i = 1, \ \cdots, \ N$). Thus, the assumption [i] is not sufficient. The second assumption [ii] is made for the case of $N < \infty$

The results of Hsiao and Zhou (2016) also hold for $\hat{\pi}_{\text{TM}}$ when the long difference of (3.20) is applied.

Theorem 2.6 : *Supposing assumptions (a1), (a2), and [i] of (A3) hold, then as N and T tend to infinity,*

$$\sqrt{NT}\,(\hat{\pi}_{\text{TM}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}(0, 1-\pi^2) \ .$$

This result is equivalent to Theorem 1.5 in Part I, that is, T-MLE attains the lower bound of efficiency. Therefore, a consequence here is that the IV estimator or T-MLE is the desirable methods for regression analyses in long panel data.

To derive the asymptotic variance for the structure estimation of (3.22), we simplify the model as $\pi_{21} = 0$ because $\beta_1 = 0$. Furthermore, we have to consider the parameterization of the structural parameter $\boldsymbol{\theta}_1 = (\beta_2,\ \gamma_1)'$ regarding the maximization of the log-likelihood function. In the work of Hsiao and Zhou (2015), the representation of the first structural equation is given by the following:

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} \ .$$

Let us call this the first formulation. If we substitute the reduced form of $y_{it}^{(2)} = \pi_{22} y_{it-1}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)}$, then

$$y_{it}^{(1)} = \beta_2 \pi_{22} y_{it-1}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \pi_i^{(1)} + v_{it}^{(1)} \ . \tag{3.24}$$

For instance, this formulation was used by Hahn (2002), and we refer to it as the second formulation. A difference between the formulations whether the right-hand side is represented by an endogenous variable. The first formulation is natural, but the second one can easily derive the asymptotic variance. We confirm in the following example that the asymptotic variance is invariant by these formulations.

Example 2.6 : We illustrate the asymptotic variance of the first formulation, using the simplest simultaneous equations model under the cross-sectional data $(T = 1)$.

$$\begin{aligned} y_i^{(1)} &= \beta y_i^{(2)} + u_i \ , \\ y_i^{(2)} &= \pi z_i + v_i^{(2)} \ , \end{aligned}$$

which is further simplified by $\mathbf{\Omega} = \mathbf{I}_2$ for the error terms of the reduced form. Then, the variance-covariance matrix $\mathcal{E}[(u_i, v_i^{(2)})'(u_i, v_i^{(2)})] = \mathbf{\Omega}_u$ of the structural and reduced form error terms becomes the following:

$$\mathbf{\Omega}_u^{-1} = \begin{pmatrix} 1+\beta^2 & -\beta \\ -\beta & 1 \end{pmatrix}^{-1} \ .$$

Regarding the Hessian of the log-likelihood function, we put $\boldsymbol{\omega}_- = \mathrm{vec}(\mathbf{\Omega}_u^{-1})$ and $\boldsymbol{\phi} = (\beta, \pi)'$ because of the invariance for parameter transformations of MLE. Then, the following are obtained:

$$
\begin{aligned}
\underset{2\times 2}{\mathbf{H}_{\phi\phi}} &= \mathcal{E}\left[\frac{1}{N}\frac{\partial^2 L}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\right] \\
&= \begin{pmatrix} -\pi^2\mathcal{E}[z_i^2]-1 & -\beta\pi\mathcal{E}[z_i^2] \\ -\beta\pi\mathcal{E}[z_i^2] & -(1+\beta^2)\mathcal{E}[z_i^2] \end{pmatrix}, \\
\underset{2\times 4}{\mathbf{H}_{\phi\omega}} &= \mathcal{E}\left[\frac{1}{N}\frac{\partial^2 L}{\partial \boldsymbol{\phi}\partial \boldsymbol{\omega}_-}\right] \\
&= \begin{pmatrix} -\beta & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
\underset{4\times 4}{\mathbf{H}_{\omega\omega}} &= \mathcal{E}\left[\frac{1}{N}\frac{\partial^2 L}{\partial \boldsymbol{\omega}_-\partial \boldsymbol{\omega}_-}\right] \\
&= \frac{1}{2}\begin{pmatrix} -(1+\beta^2)^2 & \beta(1+\beta^2) & \beta(1+\beta^2) & -\beta^2 \\ \beta(1+\beta^2) & -\beta^2 & -(1+\beta^2) & \beta \\ \beta(1+\beta^2) & -(1+\beta^2) & -\beta^2 & \beta \\ -\beta^2 & \beta & \beta & -1 \end{pmatrix}.
\end{aligned}
$$

In the second formulation, the (1,1) element of $\mathbf{H}_{\phi\phi}$ is $-\pi^2\mathcal{E}[z_i^2]$. Hence, -1 seems to be different from the first formulation. Notably, $\mathbf{H}_{\phi\omega} = \mathbf{O}$ in the second formulation, whereas the first formulation is not a zero matrix because of endogeneity. The asymptotic variance-covariance matrix of $\boldsymbol{\phi}$ becomes the 2×2 submatrix of the 6×6 inverse matrix. As shown in Theorem 2.5, we must evaluate the following:

$$
-(\mathbf{H}_{\phi\phi} - \mathbf{H}_{\phi\omega}\mathbf{H}_{\omega\omega}^{-1}\mathbf{H}_{\phi\omega}')^{-1}. \tag{3.25}
$$

Finding only the upper left 3×3 matrix of $\mathbf{H}_{\omega\omega}^{-1}$ is sufficient, because $\mathbf{H}_{\phi\omega}$ contains zeros as its elements. Using the formula for the submatrix of an inverse matrix,

$$
\begin{aligned}
(\mathbf{I}_3, \mathbf{0})\mathbf{H}_{\omega\omega}^{-1}\begin{pmatrix} \mathbf{I}_3 \\ \mathbf{0}' \end{pmatrix} &= \begin{pmatrix} -\frac{1}{2}(1+2\beta^2) & \frac{1}{2}\beta & \frac{1}{2}\beta \\ \frac{1}{2}\beta & 0 & -\frac{1}{2} \\ \frac{1}{2}\beta & -\frac{1}{2} & 0 \end{pmatrix}^{-1} \\
&= -2\begin{pmatrix} 1 & \beta & \beta \\ \beta & \beta^2 & 1+\beta^2 \\ \beta & 1+\beta^2 & \beta^2 \end{pmatrix}.
\end{aligned}
$$

Therefore, we have that

$$
\mathbf{H}_{\phi\omega}\mathbf{H}_{\omega\omega}^{-1}\mathbf{H}_{\phi\omega}' = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.
$$

That is, the variance-covariance matrix of each formulation for $\boldsymbol{\phi}$ is the same as in (3.25). Therefore, the asymptotic variance for $\theta_1 = \beta$, which is the (1,1) element, is also the same.

In panel estimation, the covariance $\boldsymbol{\Omega}_{u\xi}$ of the individual effect must be considered, but $\mathbf{H}_{\phi\omega} = \mathbf{O}$ holds asymptotically. We adopt the second formulation for deriviation using the long difference:

$$\begin{aligned} y_{it}^{(1,\ell)} &= \beta_2\pi_{22}y_{it-1}^{(2,\ell)} + \gamma_1 y_{it-1}^{(1,\ell)} + \xi_i^{(1)} + v_{it}^{(1)} \ , \\ y_{it}^{(2,\ell)} &= \pi_{22}y_{it-1}^{(2,\ell)} + \xi_i^{(2)} + v_{it}^{(2)} \ . \end{aligned}$$

Then, we correct the notation slightly,

$$\underset{2\times 2}{\boldsymbol{\Omega}_\xi} = \mathcal{E}\left[(\xi_i^{(1)},\ \xi_i^{(2)})'(\xi_i^{(1)},\ \xi_i^{(2)})\right] \ , \quad \underset{2T\times 2T}{\boldsymbol{\Omega}_{\xi v}} = \boldsymbol{\Omega}_\xi \otimes \boldsymbol{\iota}\boldsymbol{\iota}' + \boldsymbol{\Omega} \otimes \mathbf{I}_T \ . \tag{3.26}$$

We replace the log-likelihood function $\mathcal{L}_1$ of the limited information method with

$$\mathcal{L}_2 = -\frac{N}{2}\log|\boldsymbol{\Omega}_{\xi v}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_i^{(\ell)\prime}\boldsymbol{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)} \ , \tag{3.27}$$

where

$$\underset{2T\times 1}{\mathbf{v}_i^{(\ell)}} = \begin{bmatrix} \mathbf{y}_i^{(1,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}(\ \gamma_1,\ \beta_2\pi_{22})' \\ \mathbf{y}_i^{(2,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}\boldsymbol{\pi}_2 \end{bmatrix} \ , \quad \boldsymbol{\pi}_2 = (0,\ \pi_{22})' \ .$$

For the T-LIML estimators obtained by maximizing $\boldsymbol{\phi}$ and $(\mathrm{vec}(\boldsymbol{\Omega})',\ \mathrm{vec}(\boldsymbol{\Omega}_\xi)')'$, the following holds for the estimator of the parameters of interest $\hat{\boldsymbol{\theta}}_{\mathrm{TL}} = (\hat{\beta}_{2\mathrm{TL}},\ \hat{\gamma}_{1\mathrm{TL}})'$.

Theorem 2.7 : *Supposing assumptions (A1), (A2), and (A3) hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{\mathrm{TL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1}) \ .$$

This result clarifies the expression of Theorem 2.5. When $T \to \infty$, the normality assumption of the error term would not be necessary. We compare it with the D-LIML estimator of Theorem 2.2 [iii]. Although the result of Theorem 2.2 is for a general model, the form of the asymptotic variance does not change in the case of (3.22). Under the double asymptotics, the D-LIML and T-LIML estimators do not have the noncentrality parameter and have the same asymptotic variance $\sigma^2\boldsymbol{\Phi}^{-1}$. However, when N is fixed, we conclude that

$$(\sigma^2\boldsymbol{\Phi}^{-1} + c_{1*}\boldsymbol{\Psi}) - \sigma^2\boldsymbol{\Phi}^{-1} \geq \mathbf{O} \ ,$$

where $c_1 = K/N = 2/N$. That is, the T-LIML estimator is more efficient than the D-LIML estimator and is expected to have better finite sample properties with fixed N.

3.4 Extension of the T-LIML Estimator to General Models

From the discussions in the previous section the T-LIML estimator is the best method, but two issues should be to consider as far as the author knowns. Weakly exogenous variables and AR(2) models are considered the first extensions to the model. Previous studies on the transformed maximum likelihood method focused on AR(1) or VAR(1) models, and strongly exogenous variables are assumed. Thus, we extend the model as follows.

3.4.1 Weak Exogeneity

In this section, we consider why strong exogeneity is necessary and demonstrate the estimation method under weak exogeneity using a simple model. Let us consider an additional exogenous variable z_{it}, which is not the lagged endogenous variable of $(y_{it}^{(1)},\ y_{it}^{(2)})$:

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} z_{it} + \alpha_i + u_{it} \ .$$

The reduced form is as follows:

$$\begin{aligned} y_{it}^{(1)} &= \pi_{11} y_{it-1}^{(1)} + \pi_{12} y_{it-1}^{(2)} + \pi_{13} z_{it} + \pi_i^{(1)} + v_{it}^{(1)} \ , \\ y_{it}^{(2)} &= \pi_{21} y_{it-1}^{(1)} + \pi_{22} y_{it-1}^{(2)} + \pi_{23} z_{it} + \pi_i^{(2)} + v_{it}^{(2)} \ . \end{aligned}$$

In vector representation,

$$\begin{aligned} \underset{2\times 1}{\mathbf{y}_{it}} &= \underset{2\times 3}{\mathbf{\Pi}'} \mathbf{z}_{it} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \\ &= \underset{2\times 2}{\mathbf{\Pi}_{12}'} \mathbf{y}_{it-1} + \boldsymbol{\pi}_3 z_{it} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \ , \end{aligned}$$

where $\boldsymbol{\pi}_3 = (\pi_{13},\ \pi_{23})'$. The linear process is assumed for the exogenous variable:

$$z_{it} = \eta_i + \sum_{h=0}^{\infty} \theta_h \epsilon_{it-h} \ , \quad \sum_{h=0}^{\infty} |\theta_h| < \infty \ , \tag{3.28}$$

To eliminate the individual effect $\boldsymbol{\pi}_i$, the long difference is taken with respect to $\mathbf{y}_{i0}$:

$$\mathbf{y}_{i0} = (\mathbf{I}_2 - \mathbf{\Pi}_{12}' L)^{-1} \boldsymbol{\pi}_3 z_{i0} + (\mathbf{I}_2 - \mathbf{\Pi}_{12}' L)^{-1} \boldsymbol{\pi}_i + (\mathbf{I}_2 - \mathbf{\Pi}_{12}' L)^{-1} \mathbf{v}_{i0} \ ,$$

where L is the lag operator such that $L\mathbf{y}_{it} = \mathbf{y}_{it-1}$.

$$\begin{aligned}\mathbf{y}_{it}^{(\ell)} &= \mathbf{\Pi}_{12}^{'}\mathbf{y}_{it-1}^{(\ell)} + \boldsymbol{\pi}_3 z_{it} + \boldsymbol{\xi}_i + \mathbf{v}_{it} \\ &= \mathbf{\Pi}_{12}^{'}\mathbf{y}_{it-1}^{(\ell)} + \boldsymbol{\pi}_3 z_{it} + (\boldsymbol{\xi}_{i,z} + \boldsymbol{\xi}_{i,v}) + \mathbf{v}_{it} ,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\xi}_i &= \boldsymbol{\pi}_i - (\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'})\mathbf{y}_{i0} \\ &= -(\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'})(\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'}L)^{-1}\boldsymbol{\pi}_3 z_{i0} - (\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'})(\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'}L)^{-1}\mathbf{v}_{i0} \\ &= \boldsymbol{\xi}_{i,z} + \boldsymbol{\xi}_{i,v} .\end{aligned}$$

The individual effect disappears at the second equality from the following fact:

$$(\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'})(\mathbf{I}_2 - \mathbf{\Pi}_{12}^{'}L)^{-1}\boldsymbol{\pi}_i = \boldsymbol{\pi}_i .$$

For simplicity, suppose that we can observe $\boldsymbol{\xi}_{i,z}$.[8] If not observed, z_{it} and $\boldsymbol{\xi}_{i,z}$ correlate, thereby causing endogeneity. If observed, then the strong exogeneity is also necessary for the following orthogonal condition,

$$\mathcal{E}[z_{it}(\boldsymbol{\xi}_{i,v} + \mathbf{v}_{it})] = \mathbf{0} .$$

For instance, if v_{it} and ϵ_{is} are independent for all (t, s), then

$$\mathcal{E}[z_{it}\mathbf{v}_{is}] = \mathbf{0} , \quad (t, s = 1, \cdots T) .$$

However, this assumption is strong for the dynamic panel model. We would like to consider the variable z_{it} as a weakly exogenous one,

$$\mathcal{E}[z_{it}\mathbf{v}_{it}] = \mathbf{0} , \quad \mathcal{E}[z_{it}\mathbf{v}_{is}] \neq \mathbf{0} , \quad (s \leq t-1) .$$

That is, this variable is uncorrelated in period t but is allowed to correlate with the past $\mathbf{v}_{is}$. As the reduced form is considered a panel VAR model in this work, the exogenous variables in period t generally become weakly exogenous, which are generated by a triangular system. If we change the model of example. 2.5 into

$$z_{it} = \phi_1 z_{it-1} + \phi_2 y_{it-1}^{(2)} + \eta_i + \epsilon_{it} , \quad \mathcal{E}[\mathbf{v}_{it}\epsilon_{it}] = \mathbf{0} ,$$

then this bariable is weakly exogenous such that the feedback loop with $y_{it-1}^{(2)}$ exists. The representation in MA(∞) process is given as follows:

$$\begin{aligned}z_{it} &= \mathbf{e}_3^{'}(\mathbf{I}_3 - \mathbf{\Pi}^{*'})^{-1}\boldsymbol{\pi}_i^* + \mathbf{e}_3^{'}\sum_{h=0}^{\infty}(\mathbf{\Pi}^{*'})^h \mathbf{v}_{it-1-h}^* , \\ v_{it-1}^{(3)} &= \phi_2 v_{it-1}^{(2)} + \epsilon_{it} ,\end{aligned}$$

[8] In previous studies, the regression $\boldsymbol{\xi}_{i,z}$ on $\bar{z}_i = (1/T)\sum_t z_{it}$ is adopted.

where $\mathbf{v}^*_{it-1} = (\mathbf{v}'_{it-1},\ v^{(3)}_{it-1})'$ and $\mathbf{e}_3 = (0,\ 0,\ 1)'$. Thus, we can confirme that the variable is weakly exogenous.

Next, we consider the estimation method. In the first place, the lagged endogenous variable that applied the long difference has the following properties,

$$\begin{aligned} \mathcal{E}[\mathbf{y}^{(\ell)}_{it-1}(\boldsymbol{\xi}_{i,v} + \mathbf{v}_{it})'] &= \mathcal{E}[\mathbf{y}^{(\ell)}_{it-1}\boldsymbol{\xi}'_{i,v}] \\ &\neq \mathbf{O}\ . \end{aligned}$$

The orthogonal condition is not satisfied because of the correlation with the initial value. However, the transformed maximum likelihood estimator is the consistent estimation because one-to-one correspondence exists between the observed data and the error terms as a T-variate system:

$$\left\{\mathbf{y}^{(\ell)}_{i0},\ \mathbf{y}^{(\ell)}_{i1}, \cdots ,\ \mathbf{y}^{(\ell)}_{iT}\right\} \ \rightleftarrows\ \left\{\boldsymbol{\xi}_{i,v},\ \mathbf{v}_{i1},\ , \cdots ,\ \mathbf{v}_{iT}\right\}\ .$$

If the exogenous variable z_{it} is also treated such as a lagged endogenous variable, then parameters can be estimated consistently using the long difference $z^{(\ell)}_{it}$. For this purpose, we consider the companion reduced form that extends the two reduced forms of (3.28) into three equations. Notably, z_{it} is not an endogenous variable in period t, so that it must be included in the right-hand side in the reduced form. To express the companion reduced form, we introduce the lead variable z_{it+1}:

$$\begin{aligned} y^{(1)}_{it} &= \pi_{11}y^{(1)}_{it-1} + \pi_{12}y^{(2)}_{it-1} + \pi_{13}z_{it} + \pi^{(1)}_i + v^{(1)}_{it}\ , \\ y^{(2)}_{it} &= \pi_{21}y^{(1)}_{it-1} + \pi_{22}y^{(2)}_{it-1} + \pi_{23}z_{it} + \pi^{(2)}_i + v^{(2)}_{it}\ , \\ z_{it+1} &= \phi_1 z_{it} + \phi_2 y^{(2)}_{it} + \eta_i + \epsilon_{it}\ . \end{aligned} \tag{3.29}$$

We would like to summarize the left-hand side as $\mathbf{z}^*_{it} = (y^{(1)}_{it},\ y^{(2)}_{it},\ z_{it+1})'$, but $y^{(2)}_{it}$ remains on the right-hand side.[9] By transposing this to the left-hand side and solving it or substituting (3.29) into $y^{(2)}_{it}$, the companion reduced form is obtained:

$$\begin{bmatrix} y^{(1)}_{it} \\ y^{(2)}_{it} \\ z_{it+1} \end{bmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \phi_2\pi_{21} & \phi_2\pi_{22} & \phi_1 + \phi_2\pi_{23} \end{pmatrix} \begin{bmatrix} y^{(1)}_{it-1} \\ y^{(2)}_{it-1} \\ z_{it} \end{bmatrix} + \begin{bmatrix} \pi^{(1)}_i \\ \pi^{(2)}_i \\ \pi^{(3)}_i \end{bmatrix} + \begin{bmatrix} v^{(1)}_{it} \\ v^{(2)}_{it} \\ v^{(3)}_{it} \end{bmatrix}.$$

In the vector representation defined in the previous section,

$$\mathbf{z}^*_{it} = \underset{3\times 3}{\boldsymbol{\Pi}^{*'}}\mathbf{z}^*_{it-1} + \boldsymbol{\pi}^*_i + \mathbf{v}^*_{it}\ ,$$

[9]In example 2.5, the lead variable becomes $z_{it+1} = \pi_{33}z_{it} + \pi_{35}y^{(2)}_{it-1} + \pi_{36}z_{it-1} + \cdots$, and thus, the endogenous variable $y^{(2)}_{it}$ does not appear on the right-hand side.

where the original coefficient of reduced form $\mathbf{\Pi}^{'}$ is invariant. Using $\mathbf{\Pi}^{*'}$, the MA expression of (3.28) is obtained. Applying the long difference, we have the following:

$$\mathbf{z}_{it}^{*(\ell)} = \mathbf{\Pi}^{*'}\mathbf{z}_{it-1}^{*(\ell)} + \boldsymbol{\xi}_i^* + \mathbf{v}_{it}^* \ , \tag{3.30}$$

where $z_{it-1}^{*(\ell)} = z_{it} - z_{i1}$.

From the above discussions, if we add an exogenous variable following a panel VAR(1) model, then the exogenous variable should be treated as a lagged endogenous variable. Moreover, we consider the companion reduced form, which is applied the long difference. On the contrary, only applying the long difference is not sufficient, and $\mathbf{\Pi}^*$ should be estimated instead of the original $\mathbf{\Pi}$. The same is true when adding multiple exogenous variables. By setting a lead variables on the left side, we extend the dimension from G to G^*, regarded as a natural extension.

The D-LIML estimator implicitly estimates only $\mathbf{\Pi}$, but the T-LIML has to estimate $\mathbf{\Pi}^*$. We consider the effect of estimating $\mathbf{\Pi}^*$ on the estimation of structural parameters. Although z_{it} is the exogenous variable in period t, the lead variable z_{it+1} is endogenous as follows:

$$\mathcal{E}[v_{it}^{(3)}\mathbf{v}_{it}] = \boldsymbol{\omega}_{13} \neq \mathbf{0} \ , \qquad \underset{3\times 3}{\mathbf{\Omega}_*} = \begin{pmatrix} \mathbf{\Omega} & \boldsymbol{\omega}_{13} \\ \boldsymbol{\omega}_{13}^{'} & \omega_{33} \end{pmatrix} .$$

However, this variable can be considered as an endogenous variable that does not appear in the first structural equation, that is,

$$z_{it+1} = y_{it}^{(3)} \ .$$

If the reduced form of the endogenous variable that does not appear is estimated, then what happens to the asymptotic variance or the efficiency for the structural parameters of interest?

Theorem 2.8 : *Supposing Assumptions (A1), (A2), and (A3) and that the exogenous variables $\mathbf{z}_{it-1}^*$ are common, then the structure of asymptotic variance-covariance matrix is invariant, that is, $\sigma^2\mathbf{\Phi}^{-1}$.*

In a simultaneous equations model, the error terms of the reduced form are generally correlated between equations. Hence, this model is the seemingly an unrelated regression model (SUR) by Zellner (1962). When estimated with common instrumental variables, the efficiency for the coefficient of reduced form $\mathbf{\Pi}$ does not change even if estimated as a system (cf. Amemiya (1985)). Meanwhile, in the estimation of the structural parameter $\boldsymbol{\theta}_1$, estimating the companion reduced form through the transformed maximum likelihood method has no effect if the estimation is conducted for endogenous variables that do not appear. In addition, $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$ is affected by $\boldsymbol{\Omega}$ but not by $v_{it}^{(3)}$. If we added the structural equations of the additional endogenous variables, then we could obtain a full information maximum likelihood method.

3.4.2 Higher Order VAR

Hsiao et al. (2002) stated that when a variable has a higher order than the AR(2) structure, the transformed maximum likelihood method can be applied, but it becomes complicated. In the case of the VAR(1) model, the estimation method is a natural extension even if the dimension of the variables increases as shown in the previous section. However, the case of VAR(2) model cannot be called a natural extension in the expression by the long difference. Therefore, the generalization for the higher orders seems more difficult than that for the additional exogenous variables.

In the following example, we consider how to apply the transformed maximum likelihood method when the AR(2) model is included.

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} y_{it-2}^{(1)} + \alpha_i + u_{it} . \tag{3.31}$$

The reduced form is as follows:

$$\begin{aligned} y_{it}^{(1)} &= \pi_{11} y_{it-1}^{(1)} + \pi_{12} y_{it-1}^{(2)} + \pi_{13} y_{it-2}^{(1)} + \pi_i^{(1)} + v_{it}^{(1)} , \\ y_{it}^{(2)} &= \pi_{22} y_{it-1}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)} . \end{aligned}$$

For simplicity, we suppose that the reduced form of $y_{it}^{(2)}$ is AR(1), and only $y_{it}^{(1)}$ is VAR(2). In vector representation, using the companion reduced form, the following is formulated:

$$\begin{bmatrix} y_{it}^{(1)} \\ y_{it}^{(2)} \\ y_{it-1}^{(1)} \end{bmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ 0 & \pi_{22} & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} y_{it-1}^{(1)} \\ y_{it-1}^{(2)} \\ y_{it-2}^{(1)} \end{bmatrix} + \begin{bmatrix} \pi_i^{(1)} \\ \pi_i^{(2)} \\ 0 \end{bmatrix} + \begin{bmatrix} v_{it}^{(1)} \\ v_{it}^{(2)} \\ 0 \end{bmatrix} ,$$

where the third equation is the identity $y_{it-1}^{(1)} = y_{it-1}^{(1)}$. In the case of the AR(p) model, we suppose that data are available from $(y_{i0}, \cdots, y_{i,1-p})$. The initial value eliminating the individual effect π_i can be either $y_{i,0}$ or $y_{i,-1}$, which is adopt as follows:

$$\begin{bmatrix} y_{it}^{(1)} - y_{i0}^{(1)} \\ y_{it}^{(2)} - y_{i0}^{(2)} \\ y_{it-1}^{(1)} - y_{i,-1}^{(1)} \end{bmatrix} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ 0 & \pi_{22} & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{bmatrix} y_{it-1}^{(1)} - y_{i0}^{(1)} \\ y_{it-1}^{(2)} - y_{i0}^{(2)} \\ y_{it-2}^{(1)} - y_{i,-1}^{(1)} \end{bmatrix} + \begin{bmatrix} \xi_i^{(1)} \\ \xi_i^{(2)} \\ \xi_i^{(01)} \end{bmatrix} + \begin{bmatrix} v_{it}^{(1)} \\ v_{it}^{(2)} \\ 0 \end{bmatrix} .$$

In vector representation such as (3.30),

$$\mathbf{z}_{it}^{*(\ell)} = \mathbf{\Pi}^{*'} \mathbf{z}_{it-1}^{*(\ell)} + \boldsymbol{\xi}_i^* + \mathbf{v}_{it}^* \ , \tag{3.32}$$

where the vector of initial values in the long difference $\mathbf{z}_{it}^{*(\ell)} = \mathbf{z}_{it}^* - \mathbf{z}_{i0}^*$ is given as follows:

$$\mathbf{z}_{i0}^* = \left(y_{i0}^{(1)},\ y_{i0}^{(2)},\ y_{i,-1}^{(1)} \right)' \ , \tag{3.33}$$

and $\mathbf{z}_{i0}^{*(\ell)} = \mathbf{0}$. As for $\boldsymbol{\xi}_i^*$,

$$\begin{aligned} \boldsymbol{\xi}_i^* &= \left(\xi_i^{(1)},\ \xi_i^{(2)},\ \xi_i^{(01)} \right)' \\ &= -(\mathbf{I} - \mathbf{\Pi}^{*'})\mathbf{w}_{i0} \ , \end{aligned}$$

is the same as before. Using the long difference, the third identity of (3.32) slightly changes into

$$\left(y_{it-1}^{(1)} - y_{i,-1}^{(1)} \right) = \left(y_{it-1}^{(1)} - y_{i0}^{(1)} \right) + \xi_i^{(01)} \ . \tag{3.34}$$

Unlike time series analyses, the new individual effect $\xi_i^{(01)}$ is unknown, so that this identity should be included in the likelihood function. As the representation of extended VAR(1) is obtained, we consider whether the likelihood function forms the same as before. In the transformed maximum likelihood method, the expression is simplified using the long difference because it can be the following random-effects model:

$$\underset{3T \times 3T}{\vec{\mathbf{J}}_T' \mathbf{\Omega}_{\xi v}^* \vec{\mathbf{J}}_T} = \begin{pmatrix} \mathbf{\Omega}^* & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{\Omega}^* & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & \mathbf{O} & \mathbf{\Omega}^* \end{pmatrix} + \begin{pmatrix} \mathbf{\Omega}_\xi^* & \mathbf{\Omega}_\xi^* & \cdots & \mathbf{\Omega}_\xi^* \\ \mathbf{\Omega}_\xi^* & \mathbf{\Omega}_\xi^* & \cdots & \mathbf{\Omega}_\xi^* \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Omega}_\xi^* & \cdots & \mathbf{\Omega}_\xi^* & \mathbf{\Omega}_\xi^* \end{pmatrix} ,$$

where $\vec{\mathbf{J}}_T'$ is a sort matrix. Then, the expression of the inverse matrix becomes simple, and the following result can be applied,

$$\begin{aligned} \mathbf{\Omega}_{\xi v}^{-1} &= (\mathbf{\Omega}_\xi \otimes \boldsymbol{\iota}\boldsymbol{\iota}' + \mathbf{\Omega} \otimes \mathbf{I}_T)^{-1} \\ &= \mathbf{\Omega}^{-1} \otimes \mathbf{Q}_T + (\mathbf{\Omega} + T\mathbf{\Omega}_\xi)^{-1} \otimes \mathbf{J}_T \ , \end{aligned} \tag{3.35}$$

where $\mathbf{J}_T = (1/T)\boldsymbol{\iota}\boldsymbol{\iota}'$.

However, if a structure of AR(2) is included, then the variance-covariance matrix for the error terms of the companion reduced form is given by

$$\underset{3\times 3}{\mathbf{\Omega}^*} = \begin{pmatrix} \underset{2\times 2}{\mathbf{\Omega}_*} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} ,$$

where the notation in (3.13) is used. From the assumption $\mathbf{\Omega}_* = \mathbf{\Omega} > \mathbf{O}$, but $\mathbf{\Omega}^*$ is clearly singular. If $T = 1$, then $\vec{\mathbf{J}}_T' \mathbf{\Omega}_{\xi v}^* \vec{\mathbf{J}}_T$ is nonsingular because it is the sum of the positive and semi-invertible matrices. However, if $T \geq 2$, then it is singular:

$$|\mathbf{\Omega}_{\xi v}^*| = |\vec{\mathbf{J}}_T' \mathbf{\Omega}_{\xi v}^* \vec{\mathbf{J}}_T| = 0 \ .$$

Under the singularity $\mathbf{z}_{it}^{*(\ell)}$ follows a degenerated normal distribution, so that the likelihood function cannot be constructed as it is (cf. Anderson (2003)). That is, if one of the G^* variables has the structure of AR(2), then the simple expression of (3.35) cannot be obtained, which is also clear because (3.34) holds for any period t. In other words, $T-1$ equations are redundant, and (3.34) is equivalent to the following initial conditions.

$$y_{i0}^{(1)} = y_{i,-1}^{(1)} + \xi_i^{(01)} \ .$$

In the case of the AR(1) model, the effect of the initial values can be replaced with a new individual effect by the long difference, but in the case of the AR(2) model, another initial condition needs to be added.

Excluding the redundant conditions, the log-likelihood function including AR(2) structure of (3.32) is given by the following:

$$\mathcal{L}_2 = -\frac{N}{2} \log |\mathbf{\Omega}_{\xi v}^*| - \frac{1}{2} \sum_{i=1}^{N} \mathbf{v}_i^{(\ell)\prime} \mathbf{\Omega}_{\xi v}^{*-1} \mathbf{v}_i^{(\ell)} \ , \qquad (3.36)$$

where the initial conditions are as follows:

$$\underset{(1+2T)\times 1}{\mathbf{v}_i^{(\ell)}} = \begin{bmatrix} y_{i0}^{(1)} - y_{i,-1}^{(1)} \\ \mathbf{y}_i^{(1,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}(\ \gamma_{11},\ \beta_2\pi_{22},\ \gamma_{12})' \\ \mathbf{y}_i^{(2,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}\boldsymbol{\pi}_2 \end{bmatrix} \ , \quad \boldsymbol{\pi}_2 = (0,\ \pi_{22},\ 0)' \ ,$$

and

$$\underset{2T\times 3}{\mathbf{Y}_{i,-1}^{(\ell)}} = \left(\mathbf{y}_{i,-1}^{(1,\ell)},\ \mathbf{y}_{i,-1}^{(2,\ell)},\ \mathbf{y}_{i,-2}^{(1,\ell)}\right) \ , \quad \underset{T\times 1}{\mathbf{y}_{i,-2}^{(1,\ell)}} = \left(y_{i,t-2}^{(1,\ell)}\right) \ .$$

Notably, $y_{i,t-2}^{(1,\ell)} = y_{i,t-2}^{(1)} - y_{i,-1}^{(1)}$. Regarding the variance-covariance matrix,

$$\underset{(1+2T)\times(1+2T)}{\mathbf{\Omega}_{\xi v}^*} = \begin{pmatrix} \omega_{\xi 0} & \boldsymbol{\omega}_{\xi\iota}' \\ \boldsymbol{\omega}_{\xi\iota} & \mathbf{\Omega}_{\xi v} \end{pmatrix} \ ,$$

where

$$\begin{aligned} \omega_{\xi 0} &= \mathcal{E}\left[(\xi_i^{(01)})^2\right] \ , \\ \underset{1\times 2T}{\boldsymbol{\omega}_{\xi\iota}'} &= \mathcal{E}\left[\xi_i^{(01)}(\xi_i^{(1)}\boldsymbol{\iota}',\ \xi_i^{(2)}\boldsymbol{\iota}')\right] = \boldsymbol{\omega}_{0\xi}' \otimes \boldsymbol{\iota}' \ . \end{aligned}$$

$\mathbf{\Omega}_{\xi v}$ is the same as the variance-covariance matrix in (3.26) of the VAR(1) model. If the following holds,

$$\begin{aligned}\omega_{*\xi 0} &= \omega_{\xi 0} - \boldsymbol{\omega}_{\xi\iota}^{'}\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\omega}_{\xi\iota} \\ &\neq 0\ ,\end{aligned}$$

then the inverse matrix becomes the following from (6.18) and (6.3):

$$\mathbf{\Omega}_{\xi v}^{*-1} = \frac{1}{\omega_{*\xi 0}}\begin{pmatrix} 1 & -\boldsymbol{\omega}_{\xi\iota}^{'}\mathbf{\Omega}_{\xi v}^{-1} \\ -\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\omega}_{\xi\iota} & \omega_{*\xi 0}\mathbf{\Omega}_{\xi v}^{-1} + \mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\omega}_{\xi\iota}\boldsymbol{\omega}_{\xi\iota}^{'}\mathbf{\Omega}_{\xi v}^{-1}\end{pmatrix}\ .$$

The inverse matrix is more complicated than $\mathbf{\Omega}_{\xi v}^{-1}$ because the bordered matrix has to include the additional initial conditions. However, as the long difference is applied, the dependence of T on each element becomes uniform through $\mathbf{\Omega}_{\xi v}^{-1}$ as compared with the inverse matrix in (6.3). The, the determinant becomes the following:

$$|\mathbf{\Omega}_{\xi v}^{*}| = \omega_{*\xi 0}|\mathbf{\Omega}_{\xi v}|\ .$$

.

The model of (3.31) is the case when only one initial condition is added. From the above discussions, if the number of exogenous variables or the order of AR model increases by one, then the dimension of $\mathbf{\Omega}_{\xi v}^{*}$ must also increase by one. In the following, we give the T-LIML estimation method for the general structural model in Section 3.2. To simplify the expression, the order of the VAR model is $p(\geq 2)$ using the standard representation in (3.11).

$$\mathcal{L}_2 = -\frac{N}{2}\log|\mathbf{\Omega}_{\xi v}^{*}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_i^{(\ell)'}\mathbf{\Omega}_{\xi v}^{*-1}\mathbf{v}_i^{(\ell)}\ , \tag{3.37}$$

where

$$\underset{(G^*(p-1)+G^*T)\times 1}{\mathbf{v}_i^{(\ell)}} = \begin{bmatrix} \Delta\mathbf{y}_{i0}^{*} \\ \vdots \\ \Delta\mathbf{y}_{i,-(p-2)}^{*} \\ \mathbf{y}_i^{(1,\ell)} - \mathbf{Z}_{i,-1}^{(\ell)}\left((\boldsymbol{\gamma}_1^{'} + \boldsymbol{\beta}_2^{'}\mathbf{\Pi}_{12}^{'}),\ \boldsymbol{\beta}_2^{'}\mathbf{\Pi}_{22}^{'}\right)^{'} \\ \mathbf{y}_i^{(2,\ell)} - \mathbf{Z}_{i,-1}^{*(\ell)}\mathbf{\Pi}^{*}\mathbf{e}_2 \\ \vdots \\ \mathbf{y}_i^{(G^*,\ell)} - \mathbf{Z}_{i,-1}^{*(\ell)}\mathbf{\Pi}^{*}\mathbf{e}_{G^*} \end{bmatrix}\ .$$

the only gth element of $\mathbf{e}_g$ $(g = 2,\ \cdots, G^*)$ is 1 and 0, otherewise.

$$\underset{T\times(K_1+K_2)}{\mathbf{Z}_{i,-1}^{(\ell)}} = \left(\mathbf{z}_{it}^{(1,\ell)'}, \mathbf{z}_{it}^{(2,\ell)'}\right)\ ,\quad \underset{T\times K^*}{\mathbf{Z}_{i,-1}^{*(\ell)}} = \left(\mathbf{z}_{it-1}^{*(\ell)'}\right)\ .$$

.

Regarding the exogenous variables in period t, the lead variables of $G+1, \cdots, G^*$ must also be included in addition to the original G variables of the first structural equation. The initial conditions represented by the first-difference before $t = 0$ should be added, and the number becomes $G^*(p-1)$ in total. In the general model, the log-likelihood is expressed by the first-difference and the long difference. Notably, for the long differences, each element of the initial value has to be in the representation of the extended VAR(1), as shown in (3.33). That is,

$$\mathbf{z}_{i0}^* = \left(y_{i0}^{(1)}, \cdots, y_{i0}^{(G^*)}, \cdots, y_{i,-(p-1)}^{(1)}, \cdots, y_{i,-(p-1)}^{(G^*)}\right)' .$$

The variance-covariance matrix is given as follows:

$$\underset{(G^*(p-1)+G^*T)^2}{\boldsymbol{\Omega}_{\xi v}^*} = \begin{pmatrix} \boldsymbol{\Omega}_{\xi 0} & \boldsymbol{\Omega}_{\xi\iota}' \\ \boldsymbol{\Omega}_{\xi\iota} & \boldsymbol{\Omega}_{*\xi v} \end{pmatrix} ,$$

where $\boldsymbol{\xi}_i^{(0)} = (\Delta\mathbf{y}_{i0}^{*'}, \cdots, \Delta\mathbf{y}_{i,-(p-2)}^{*'})'$,

$$\begin{aligned} \underset{G^*(p-1)\times G^*(p-1)}{\boldsymbol{\Omega}_{\xi 0}} &= \mathcal{E}\left[\boldsymbol{\xi}_i^{(0)}\boldsymbol{\xi}_i^{(0)'}\right] , \\ \underset{G^*(p-1)\times G^*T}{\boldsymbol{\Omega}_{\xi\iota}'} &= \boldsymbol{\Omega}_{0\xi}' \otimes \boldsymbol{\iota}' , \end{aligned}$$

and

$$\underset{G^*T\times G^*T}{\boldsymbol{\Omega}_{*\xi v}^{-1}} = \underset{G^*\times G^*}{\boldsymbol{\Omega}_*^{-1}} \otimes \mathbf{Q}_T + (\boldsymbol{\Omega}_* + T\boldsymbol{\Omega}_{*\xi})^{-1} \otimes \mathbf{J}_T .$$

The upper left submatrix of $\boldsymbol{\Omega}_*$ is the $G \times G$ matrix $\boldsymbol{\Omega}$. From (6.18), the inverse matrix becomes

$$\boldsymbol{\Omega}_{\xi v}^{*-1} = \begin{pmatrix} \boldsymbol{\Omega}_{*\xi 0}^{-1} & -\boldsymbol{\Omega}_{*\xi 0}^{-1}\boldsymbol{\Omega}_{\xi\iota}'\boldsymbol{\Omega}_{*\xi v}^{-1} \\ -\boldsymbol{\Omega}_{*\xi v}^{-1}\boldsymbol{\Omega}_{\xi\iota}\boldsymbol{\Omega}_{*\xi 0}^{-1} & \boldsymbol{\Omega}_{*\xi v}^{-1} + \boldsymbol{\Omega}_{*\xi v}^{-1}\boldsymbol{\Omega}_{\xi\iota}\boldsymbol{\Omega}_{*\xi 0}^{-1}\boldsymbol{\Omega}_{\xi\iota}'\boldsymbol{\Omega}_{*\xi v}^{-1} \end{pmatrix} ,$$

where $\boldsymbol{\Omega}_{\xi 0}^* = \boldsymbol{\Omega}_{\xi 0} - \boldsymbol{\Omega}_{\xi\iota}'\boldsymbol{\Omega}_{*\xi v}^{-1}\boldsymbol{\Omega}_{\xi\iota}$. As for the determinant,

$$|\boldsymbol{\Omega}_{\xi v}^*| = |\boldsymbol{\Omega}_{*\xi 0}||\boldsymbol{\Omega}_{\xi v}| .$$

Although the parameters of interest is $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1')'$, the maximization procedure has to be performed using the following parameters:

$$\{\boldsymbol{\theta}_1, \boldsymbol{\Pi}_{12}, \boldsymbol{\Pi}_{22}, \boldsymbol{\Pi}^*\mathbf{e}_{G+1}, \cdots, \boldsymbol{\Pi}^*\mathbf{e}_{G^*}, \boldsymbol{\Omega}_{\xi 0}, \boldsymbol{\Omega}_{0\xi}, \boldsymbol{\Omega}_*, \boldsymbol{\Omega}_{*\xi}\} .$$

Thus, in the case of the general model, the likelihood function of the T-LIML estimator becomes complicated and the number of parameters to be estimated can be quite large.

3.5 D-LIML Estimator Revisited

In the case of the general model, T-LIML estimation is not a natural extension of the VAR(1) model, and the calculation becomes complicated. Moreover, as the D-LIML is based on orthogonal conditions, this estimator has the advantage that the calculation does not change even with the general model. However, as shown in Section 3.3, if N is fixed, then the D-LIML estimator is inferior in asymptotic efficiency to the T-LIML estimator. In the following, we discuss that the D-LIML and T-LIML estimators are asymptotically equivalent under $T \to \infty$ by slightly modifying the projection matrix of the D-LIML estimator.

3.5.1 Improving the Projection Matrix

In the estimation method of Arellano and Bond (1995), the matrix of instrumental variables contains zeros as shown in (3.11). This corresponds to the orthogonal condition being considered in each period t and may be called the sequential moment condition. Akashi and Kunitomo (2015) applied the instrumental variables with the backward filtere, but its construction is the same as the sequential moment condition. That is,

$$\mathcal{E}\left[\mathbf{z}_{it}^{(b)} u_{it}^{(f)}\right] = \underset{K\times 1}{\mathbf{0}} ,$$

for $t = 1; \cdots , T-1$. If all periods are collectively represented by a matrix, then

$$\mathcal{E}\left[\mathbf{Z}_{*}^{(b)\prime} \mathbf{u}_{*}^{(f)}\right] = \underset{K(T-1)\times 1}{\mathbf{0}} ,$$

where

$$\underset{K(T-1)\times N(T-1)}{\mathbf{Z}_{*}^{(b)\prime}} = \begin{pmatrix} \mathbf{Z}_1^{(b)\prime} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{Z}_2^{(b)\prime} & \cdots & \mathbf{O} \\ \vdots & \ddots & \vdots & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{Z}_{T-1}^{(b)\prime} \end{pmatrix} , \quad \underset{K\times N}{\mathbf{Z}_t^{(b)\prime}} = \left(\mathbf{z}_{1t}^{(b)}, \cdots , \mathbf{z}_{Nt}^{(b)}\right) ,$$

and

$$\underset{N(T-1)\times 1}{\mathbf{u}_{*}^{(f)}} = \left(\mathbf{u}_1^{(f)\prime}, \ \cdots , \ \mathbf{u}_{T-1}^{(f)\prime}\right)' .$$

Unlike regression analyses, structural estimation is generally overidentified ($K > G_2 + K_1$). Thus, a projection matrix is used in the objective function. If evaluated by the true value $\boldsymbol{\theta}$, then the numerator of the objective function of the D-LIML estimator is as follows:

$$\begin{aligned} \mathbf{u}_{*}^{(f)\prime} \mathbf{P}_{*}^{(b)} \mathbf{u}_{*}^{(f)} &= \mathbf{u}_{*}^{(f)\prime} \mathbf{Z}_{*}^{(b)\prime} \left(\mathbf{Z}_{*}^{(b)\prime} \mathbf{Z}_{*}^{(b)}\right)^{-1} \mathbf{Z}_{*}^{(b)\prime} \mathbf{u}_{*}^{(f)} \\ &= \sum_{t=1}^{T-1} \mathbf{u}_t^{(f)\prime} \mathbf{P}_t^{(b)} \mathbf{u}_t^{(f)} , \end{aligned}$$

where $\mathbf{P}_*^{(b)}$ is a diagonal matrix in the case of sequential moment condition, so the objective function is expressed as the sum of the quadratic form similar to Alvarez and Arellano (2003). From Theorem 2.3-[iii], the asymptotic variance of the D-LIML estimator is as follows:

$$\sigma^2 \mathbf{\Phi}^{-1} + c_{1*} \mathbf{\Psi} \ . \tag{3.38}$$

In structural estimation, the first term $\sigma^2 \mathbf{\Phi}^{-1}$ can be smaller if many instrumental variables are used, but the second term $c_* \mathbf{\Psi}$ becomes larger on the contrary. Regarding $c_{1*} = c_1/(1-c_1)$, as discussed in Section 3.1.1, c_1 is the ratio such that

$$\begin{aligned} c_1 &= \lim_{T\to\infty} \frac{\text{rank}(\mathbf{P}_*^{(b)})}{n} \\ &= \frac{K}{N} \ , \end{aligned}$$

which depends on the total number of orthogonal conditions and the total data $n = NT$. If we consider reducing only the second term, then the number of orthogonal conditions should be decreased. Therefore, instead of the sequential moment condition, we consider the orthogonal condition added over the entire period:

$$\mathcal{E}\left[\sum_{t=1}^{T-1} \mathbf{z}_{it}^{(b)} u_{it}^{(f)}\right] = \underset{K\times 1}{\mathbf{0}} \ . \tag{3.39}$$

When expressed collectively as a matrix,

$$\begin{aligned} \mathcal{E}\left[\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)}\right] &= \mathcal{E}\left[\left(\mathbf{Z}_1^{(b)\prime}, \ \cdots \mathbf{Z}_{T-1}^{(b)\prime}\right)\mathbf{u}_*^{(f)}\right] \\ &= \underset{K\times 1}{\mathbf{0}} \ , \end{aligned} \tag{3.40}$$

where in the case of arranging data in each i, we have the following:

$$\underset{K\times N(T-1)}{\mathbf{Z}^{(b)\prime}} = \left(\mathbf{Z}_1^{(b)\prime}, \cdots, \mathbf{Z}_N^{(b)\prime}\right) \ , \quad \underset{K\times(T-1)}{\mathbf{Z}_i^{(b)\prime}} = \left(\mathbf{z}_{i0}^{(b)}, \cdots, \mathbf{z}_{i(T-1)}^{(b)}\right) \ ,$$

and

$$\underset{N(T-1)\times 1}{\mathbf{u}^{(f)}} = \left(\mathbf{u}_1^{(f)\prime}, \ \cdots, \ \mathbf{u}_N^{(f)\prime}\right)' \ .$$

Then, the projection matrix is given as follows:

$$\mathbf{P}^{(b)} = \mathbf{Z}^{(b)}(\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)})^{-1}\mathbf{Z}^{(b)\prime},$$

where the construction of $\mathbf{Z}^{(b)}$ becomes simple such that the vectors $\mathbf{z}_{it}^{(b)}$ are rearranged into n. In addition, each instrumental variable is the same as that of

Akashi and Kunitomo (2015). The numerator of the objective function is based on the K orthogonal conditions in (3.40) compared with (3.15), that is,

$$\mathbf{u}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)} \ .$$

Then,[10]

$$\begin{aligned} c_2 &= \lim_{T\to\infty} \frac{\mathrm{rank}(\mathbf{P}^{(b)})}{n} \\ &= \lim_{T\to\infty} \frac{K}{NT} \\ &= 0 \ . \end{aligned}$$

Therefore, the second term $\mathbf{\Psi}$ may disappear.[11] However, we need to confirm that the efficiency of the first term $\sigma^2\mathbf{\Phi}$ is not reduced by the summed orthogonal condition in (3.39).

We redefine the D-LIML estimator using the projection matrix above. The data are rearranged into $n = N(T-1)$ pieces,

$$\underset{N(T-1)\times 1}{\mathbf{y}^{(1,f)}} = \left(\mathbf{y}_1^{(1,f)\prime}, \cdots, \mathbf{y}_N^{(1,f)\prime}\right)^{\prime} , \qquad \underset{(G_2+K_1)\times N(T-1)}{\mathbf{X}^{(f)\prime}} = \left(\mathbf{X}_1^{(f)\prime}, \cdots, \mathbf{X}_N^{(f)\prime}\right) ,$$

where

$$\underset{1\times(T-1)}{\mathbf{y}_i^{(1,f)\prime}} = \left(y_{i1}^{(1,f)}, \cdots, y_{i(T-1)}^{(1,f)}\right), \qquad \underset{(G_2+K_1)\times(T-1)}{\mathbf{X}_i^{(1,f)\prime}} = \left(\mathbf{x}_{i1}^{(f)}, \cdots, \mathbf{x}_{i(T-1)}^{(f)}\right),$$

and

$$\mathbf{x}_{it}^{(f)\prime} = \left(\mathbf{y}_{it}^{(2,f)\prime}, \ \mathbf{z}_{it}^{(1,f)\prime}\right) \ .$$

For the two $(G+K_1)\times(G+K_1)$ matrices,

$$\mathbf{G}_n^{(f,b)} = \begin{pmatrix} \mathbf{y}^{(1,f)\prime} \\ \mathbf{X}^{(f)\prime} \end{pmatrix} \mathbf{P}^{(b)} \left(\mathbf{y}^{(1,f)}, \mathbf{X}^{(f)}\right)$$

and

$$\mathbf{H}_n^{(f,b)} = \begin{pmatrix} \mathbf{y}^{(1,f)\prime} \\ \mathbf{X}^{(f)\prime} \end{pmatrix} [\mathbf{I}_n - \mathbf{P}^{(b)}] \left(\mathbf{y}^{(1,f)}, \mathbf{X}^{(f)}\right) .$$

Let $\hat{\boldsymbol{\theta}}_{\mathrm{DL}} = (\hat{\boldsymbol{\beta}}_{\mathrm{DL}}^{\prime}, \ \hat{\boldsymbol{\gamma}}_{\mathrm{DL}}^{\prime})^{\prime}$ be the minimization point of

$$\mathcal{VR}_2 = \frac{\boldsymbol{\theta}^{\prime}\mathbf{G}_n^{(f,b)}\boldsymbol{\theta}}{\boldsymbol{\theta}^{\prime}\mathbf{H}_n^{(f,b)}\boldsymbol{\theta}} \ . \tag{3.41}$$

[10]Rigorously speaking, the expectation of the numerator becomes $\mathcal{E}[\mathbf{u}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)}/n] \neq \sigma^2 K/n$. This detail is described in Section 3.7.2.

[11]Hsiao (2014) and Hsiao and Zhou (2015) also discussed why the efficiency decreases in sequential moment conditions.

Then, unlike the D-LIML estimator $\tilde{\boldsymbol{\theta}}_{\mathrm{DL}}$ in Theorem 2.3-[iii], the following holds for the new D-LIML estimator $\hat{\boldsymbol{\theta}}_{\mathrm{DL}}$.

Theorem 2.9 : *Supposing assumptions (A1) and (A2) hold, then, as $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{\mathrm{DL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Phi}^{-1}) \ .$$

As the asymptotic variance-covariance matrix is the same as the first term of (3.38), this result improves Theorem 2.3 of Akashi and Kunitomo (2015). Moreover, the revised D-LIML and T-LIML estimators are asymptotically equivalent in long panel data without depending on N because the result is the same as that of Theorem 2.7.

The following also holds for the GMM estimator based on (3.39).

Corollary 2.1 : *The D-GMM estimator $\hat{\boldsymbol{\theta}}_{DG}$ obtained by minimizing the following objective function*

$$\mathcal{Q}_2 = \boldsymbol{\theta}' \mathbf{G}_n^{(f,b)} \boldsymbol{\theta}$$

is asymptotically equivalent to $\hat{\boldsymbol{\theta}}_{DL}$.

We summarize the results regarding the asymptotic efficiency of the LIML estimators. In the works of Alvarez and Arellano (2003) and Akashi and Kunitomo (2012), the ratio r_n/n of the number of instruments to the total data $n = NT$ is presented as follows:

$$\frac{K}{2}\frac{T(T-1)}{NT} \quad \longrightarrow \quad c = \frac{K}{2} \lim_{N,T\to\infty} \frac{T}{N} \ ,$$

where K is the number of instrumental variables for the structural model in period t. In the work of Akashi and Kunitomo (2015), the asymptotically optimal instrumental variables are used, and the number is reduced such that

$$\frac{K(T-1)}{NT} \quad \longrightarrow \quad c_1 = \frac{K}{N} \ .$$

If $N \to \infty$, then c_1 becomes zero. Furthermore, when the sequential moment condition is improved such as in Theorem 2.9, we obtain the following:

$$\frac{K}{NT} \quad \longrightarrow \quad c_2 = 0 \ .$$

As the first term $\sigma^2\mathbf{\Phi}^{-1}$ of the asymptotic variance is common to the three estimation methods, the third estimation method gives the best result. Meanwhile, if the normal equations of the T-LIML estimator are regarded as the orthogonal conditions, then the number of the conditions is equal to that of estimated parameters. The number is finite, and this method does not cause incidental parameter problems. In Theorem 2.9, the number of orthogonal conditions is also K, so that the D-LMIL and T-LIML estimators have the same asymptotic property. In addition, when the ratio r_n/n is zero, the many instruments problem does not occur. Hence, the LIML and GMM estimators can also be asymptotically equivalent. That is, if $K < \infty$, then the number of orthogonal conditions can be finite even in long panel data. The comparison between the LIML and GMM estimators under $K \to \infty$ is examined in a later section.

3.5.2 Relation between T-LIML and D-LIML Estimators

As the objective functions have a similar form, we confirm that the asymptotic properties of the T-LIML and D-LIML are equivalent. That is, the variance ratio of D-LIML is an asymptotic approximation of the concentrated log-likelihood function of the transformed method.

For simplicity, we consider a VAR(1) model with $G = G_*$. The log-likelihood function $\mathcal{L}_2$ of (3.37) represented by the long difference becomes

$$\mathcal{L}_2 = -\frac{N}{2}\log|\mathbf{\Omega}_{\xi v}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_i^{(\ell)'}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)} \; .$$

By applying the inverse transformation such as Lemma 2.1, the function is returned to the expression by the first-difference:

$$\mathcal{L}_{2\Delta} = -\frac{N}{2}\log|\mathbf{\Omega}_{2\Delta}| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_i^{*'}\mathbf{\Omega}_{2\Delta}^{-1}\Delta\mathbf{v}_i^{*} \; .$$

where

$$\begin{aligned}
\underset{GT\times GT}{\mathbf{\Omega}_{2\Delta}} &= \begin{pmatrix} \mathbf{\Omega}_1 & -\mathbf{\Omega} & \mathbf{O} & \cdots & \mathbf{O} \\ -\mathbf{\Omega} & & & & \\ \mathbf{O} & & \mathbf{\Omega}_{\Delta} & & \\ \vdots & & & & \\ \mathbf{O} & & & & \end{pmatrix} , \\
\underset{G\times G}{\mathbf{\Omega}_1} &= \mathcal{E}[\Delta\mathbf{y}_{i1}\Delta\mathbf{y}_{i1}^{'}] \; , \\
\underset{GT\times 1}{\Delta\mathbf{v}_i^{*}} &= (\Delta\mathbf{y}_{i1}^{'},\ \Delta\mathbf{v}_{it}^{'})^{'} \; .
\end{aligned}$$

This function is a multivariate version of the transformed maximum likelihood method in Section 3.3. This log-likelihood is devided into the pseudo likelihood $\mathcal{L}_{2\cdot 0}$ and the term $\mathcal{R}_0$, in which the initial value $\Delta\mathbf{y}_{i1}$ is included:

$$\mathcal{L}_{2\Delta} = \mathcal{L}_{2\cdot 0} + \mathcal{R}_0 \ ,$$

where

$$\begin{aligned} \mathcal{L}_{2\cdot 0} &= -\frac{N}{2}\log|\boldsymbol{\Omega}_\Delta| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_i'\boldsymbol{\Omega}_\Delta^{-1}\Delta\mathbf{v}_i \ , \qquad (3.42) \\ \underset{G(T-1)\times G(T-1)}{\boldsymbol{\Omega}_\Delta} &= \begin{pmatrix} 2\boldsymbol{\Omega} & -\boldsymbol{\Omega} & \cdots & \mathbf{O} \\ -\boldsymbol{\Omega} & 2\boldsymbol{\Omega} & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & -\boldsymbol{\Omega} & 2\boldsymbol{\Omega} \end{pmatrix} . \end{aligned}$$

$\mathcal{L}_{2\cdot 0}$ resembles a conditional likelihood function given the initial value $\Delta\mathbf{y}_{i1}$. The first-difference of the initial value is an endogenous variable that correlates with the first-difference of the error term, so it is not the conditional likelihood function.[12] Thus, we call $\mathcal{L}_{2\cdot 0}$ the pseudo log-likelihood function of the transformed method.

As for the D-LIML estimator, we replace the projection matrix $\mathbf{P}^{(b)}$ of $\mathcal{L}_2$ with

$$\mathbf{P}^{(f)} = \mathbf{Z}^{(f)}(\mathbf{Z}^{(f)\prime}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(f)\prime} \ .$$

We define the objective function expressed only by the forward filter as follows:

$$\mathcal{VR}_{2\cdot 0} = \frac{\boldsymbol{\theta}'\mathbf{G}_n^{(f,f)}\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{H}_n^{(f,f)}\boldsymbol{\theta}} \ ,$$

where the instrumental variable matrix that applied the forward filter is

$$\underset{K\times N(T-1)}{\mathbf{Z}^{(f)\prime}} = \left(\mathbf{z}_{11}^{(f)}, \cdots, \mathbf{z}_{1T-1}^{(f)}, \cdots, \mathbf{z}_{N1}^{(f)}, \cdots, \mathbf{z}_{NT-1}^{(f)}\right) \ .$$

Then, the following holds.

Lemma 2.2 : *Suppose assumptions (A1), (A2), and (A3) hold.*

[i] *Maximizing the concentrated pseudo log-likelihood function of $\boldsymbol{\theta}_1$ is identical to minimizing the variance ratio:*

$$\underset{\boldsymbol{\theta}_1}{\operatorname{argmax}} \ \mathcal{L}_{2\cdot 0} = \underset{\boldsymbol{\theta}_1}{\operatorname{argmin}} \ \mathcal{VR}_{2\cdot 0} \ .$$

[12] Although it is a random-effects model, the correct conditional likelihood method given the initial value is examined by Alvarez and Arellano (2003).

[ii] *As $T \to \infty$ and N is fixed*

$$\begin{aligned} \text{plim argmax } \mathcal{L}_2 &= \text{plim argmin } \mathcal{VR}_{2\cdot 0} \\ &= \text{plim argmin } \mathcal{VR}_2 \ , \end{aligned}$$

and these estimators have the same asymptotic distribution.

The estimator obtained by the pseudo log-likelihood $\mathcal{L}_{2\cdot 0}$ is the pseudo T-LIML estimator $\breve{\boldsymbol{\theta}}_{\text{PL}}$. Lemma 2.2-[i] states that this estimator is derived from the minimization of the concentrated pseudo log-likelihood function $\mathcal{VR}_{2\cdot 0}$. When N is fixed, the result of [ii] implies the following:

$$\frac{1}{n} \max_{\boldsymbol{\theta}_1} \mathcal{L}_2 \quad \xrightarrow{p} \quad \text{const.} - \log(1 + \frac{1}{n} \min_{\boldsymbol{\theta}_1} \mathcal{VR}_2) \ .$$

That is, the concentrated log-likelihood function of T-LIML is asymptotically equal to $\mathcal{VR}_2$, which is the objective function of the D-LIML estimator, and thus, they have the same asymptotic distribution. Another implication of [ii] is as follows. When N is fixed, the variance $\boldsymbol{\Omega}_\xi$ of the random-effects or $\boldsymbol{\Omega}_1$ of the initial value cannot be consistently estimated (cf. Hsiao (2014)). However, the structural estimator for $\boldsymbol{\theta}_1$ is unaffected and is consistent such as Theorem 2.7.

Next, we confirm the properties of the pseudo T-LIML estimator $\breve{\boldsymbol{\theta}}_{\text{PL}}$, which is the case when the initial value $\Delta \mathbf{y}_{i1}$ is not used in the transformed maximum likelihood method and is expressed as follows:

$$\mathbf{J}_0^{'} \mathbf{D}_{T+1} \mathbf{y}_i^* = \mathbf{D}_T \mathbf{y}_i \ ,$$

where $\mathbf{J}_0 = (\mathbf{0},\ \mathbf{I}_{T-1})$.[13] Hsiao and Zhou (2015) showed that the T-LIML estimator without initial values has the noncentrality parameter whose order is $O(d^{\frac{1}{2}})$. We clarify the noncentrality parameter in the following. From Lemma 2.2, the pseudo T-LIML estimator is the D-LIML estimator based on the incorrect orthogonal condition:

$$\mathcal{E}\left[\mathbf{z}_{it-1}^{(f)} u_{it}^{(f)}\right] \neq \mathbf{0} \ .$$

To state the results first, as discussed in Section 3.1.1, this endogeneity is weakened as $T \to \infty$, so that the pseudo T-LIML estimator is consistent but has the noncentrality parameter $-\sqrt{d}\boldsymbol{\rho}^*$ of (3.43). Similar to Hahn and Kuersteiner (2002), the noncentrality parameter can be consistently estimated by other statistics, and the corrected estimator becomes

$$\breve{\breve{\boldsymbol{\theta}}}_{\text{PL}} = \breve{\boldsymbol{\theta}}_{\text{PL}} + \frac{1}{T} \breve{\boldsymbol{\rho}}^* \ ,$$

[13] Hsiao and Zhou (2015) considered $\mathbf{J}_0^{'} \mathbf{L} \mathbf{D}_{T+1} \mathbf{y}_i^*$ without the initial value $\mathbf{y}_{i1}^{(\ell)}$ expressed by the long difference, which is slightly different from our formulation.

where the detail of correction term $\breve{\boldsymbol{\rho}}^*$ is given in the proof of Theorem 2.10. The following holds for the general model of (3.8).

Theorem 2.10 : *Supposing assumptions (A1) and (A2) hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

[i] *Under $N/T \to 0 \le d < \infty$,*

$$\sqrt{NT}\left(\breve{\boldsymbol{\theta}}_{\mathrm{PL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{b}_d, \sigma^2 \boldsymbol{\Phi}^{-1}) ,$$

where

$$\mathbf{b}_d = -\sqrt{d}\boldsymbol{\rho}^* . \tag{3.43}$$

[ii] *Under $N/T \to 0 \le d < \infty$,*

$$\sqrt{NT}\left(\breve{\breve{\boldsymbol{\theta}}}_{\mathrm{PL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Phi}^{-1}) .$$

The corrected estimation of [ii] must estimate $\boldsymbol{\Pi}^*$ compared with the D-LIML estimation and has the constraint on $d < \infty$. Hence, the corrected estimation is not recommended.

From the results so far, the following implications are obtained. First, asymptotic efficiency does not change without using the initial value, but the noncentrality parameter occurs depending on $d = \lim N/T$. For instance, a time series analysis ($N = 1$) does not have the noncentrality parameter because $d = 0$. Thus, the dynamic panel analysis is more sensitive to the initial values. The reasen is that the bias of $O(1/T)$ accumulates as N increases. Second, the noncentrality parameters depending on d are due to the initial value, and those depending on c are caused by many instruments. For instance, the CV estimator in Theorem 1.3 has the noncentrality parameters depending on d because its objective function is the same as the pseudo-log-likelihood function in (3.42). The noncentrality parameter based on the instrumental variables also depends on K, and the bias of $O(1/N)$ may accumulate as T increases in the case of the sequential orthogonal conditions.

3.6 Lower Bounds of Asymptotic Efficiency

In the dynamic panel structure model, we have examined several estimators for the structural parameter $\boldsymbol{\theta}_1$. The lower bound for efficinecy under the long panel data shoul be considered a criterion for deciding which estimation method to use.

First, we extend the result of Theorem 1.6 of Hahn and Kuersteiner (2002) to the dynamic structural model. Hahn (2002), a related study, which investigated the effeciency lower bound using the Hajek-type convolution theorem when many instruments in a structural model of cross-sectional data has many instruments. In a panel model, if the individual effect α_i is regarded as the fixed-effects, then these become the incidental parameters under $N \to \infty$. Thus, the Cramer-Rao lower bound for $\boldsymbol{\theta}_1$ may not be evident because the incidental parameters α_i $(i = 1, \cdots, N)$ exist. Similar to Hahn and Kuersteiner (2002), we make the following assumptions,

(A4) $(1/N)\sum_{i=1}^{N} \boldsymbol{\mu}_i' \boldsymbol{\mu}_i = O(1)$ and $\mathbf{v}_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$.

For estimation methods to the structural parameter of the general model (3.8), the following holds.

Theorem 2.11 : *Supposing assumptions (A1), (A2), and (A4) hold, then as N and T tend to infinity, the asymptotic distribution of any regular estimator of* $\boldsymbol{\theta}_1$ *cannot be more concentrated than* $\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1})$.

The lower bound is $\sigma^2\boldsymbol{\Phi}^{-1}$ which appears in the theorems and does not depend on incidental parameters such as the result of regression analysis. That is, the bound is the same as when $N < \infty$. Although the theorem does not state whether it is attainable, our results suggest the following.

Corollary 2.2 : *Under assumptions (A1), (A2), and (A4), the T-LIML, D-LIML, and D-GMM estimators are asymptotically efficient.*

The T-LIML, D-LIML and D-GMM estimators here correspond to Theorem 2.7, Theorem 2.9, and Corollary 2.1, respectively. The asymptotic normality of these estimators does not depend on normality of the error term $\mathbf{v}_{it}$, but whether the form of the lower bound is $\sigma^2\boldsymbol{\Phi}^{-1}$ depends on the normality assumption (A4). Other estimators attain the lower bound, but the additional constraint on T/N or N/T is then necessary. Therefore, the conclusion so far is that the estimation methods of Corollary 2.2 have the desirable properties in the structural estimation in long panel data.

The above results need the assumption that the error term follows a normal distribution. Second, let us consider another approach to show the lower bound without the normality assumption. Anderson et al.(2010) investigated an estimator $\boldsymbol{\phi}(\mathbf{G}, \mathbf{H})$ based on the sufficient statistics for the structural parameter $\boldsymbol{\theta}_1$, where $\mathbf{G}$ and $\mathbf{H}$ correspond to the sufficient statistics in the cross-sectional or time series data under the normality assumption. In the panel analysis, the asymptotic sufficient statistics of the transformed maximum likelihood estimator can be interpreted as $\mathbf{G}_n^{(f,f)}$ and $\mathbf{H}_n^{(f,f)}$ by Lemma 2.2. The transformed method does not depend on the individual effects, but the pseudo T-LIML estimator may have the noncentrality parameter when N is fixed. Although the noncentrality parameter is irrelevant to the lower bound, we replace $\mathbf{G}_n^{(f,f)}$ with $\mathbf{G}_n^{(b,f)}$ to obtain simple results, where $\mathbf{G}_n^{(b,f)}$ satisfies the orthogonal condition in (3.40). We consider the class of the estimators as follows:

$$\hat{\boldsymbol{\theta}}_1 = \boldsymbol{\phi}(\frac{1}{n}\mathbf{G}_n^{(b,f)}) \; .$$

For simplicity, set $\boldsymbol{\gamma}_1 = \mathbf{0}$ or $\boldsymbol{\theta}_1 = \boldsymbol{\beta}_2$, that is, the exogenous variable does not appear in the first structural equation:

$$y_{it}^{(1)} \;\; = \;\; \boldsymbol{\beta}_2' \mathbf{y}_{it}^{(2)} + \alpha_i + u_{it} \; .$$

Moreover the reduced form is supposed as the VAR(1) model.

($\mathbf{A4}'$) [i] $\boldsymbol{\phi}(.)$ is the consistent estimator of $\boldsymbol{\theta}$ for any N and $T \to \infty$. [ii] $\boldsymbol{\phi}(.)$ is a continuously differentiable function that does not depend on n, and the first order derivative is bounded in the neighborhood of any true value $\boldsymbol{\theta}$. [iii] For any T, $\boldsymbol{\phi}(\mathbf{G}_{T0}^{(b,f)}) = \boldsymbol{\beta}_2$, where

$$\begin{aligned} \mathbf{G}_{T0}^{(b,f)} &= \mathbf{G}_{10}^{(b,f)\prime} \left(\mathbf{G}_{20}^{(b,f)}\right)^{-1} \mathbf{G}_{10}^{(b,f)} \; , \\ \mathbf{G}_{10}^{(b,f)} &= \frac{1}{T-1}\sum_{t=1}^{T-1} \mathcal{E}\left[\mathbf{z}_{it}^{(b)} \mathbf{y}_{it}^{(f)\prime}\right] \; , \\ \mathbf{G}_{20}^{(b,f)} &= \frac{1}{T-1}\sum_{t=1}^{T-1} \mathcal{E}\left[\mathbf{z}_{it}^{(b)} \mathbf{z}_{it}^{(b)\prime}\right] \; . \end{aligned}$$

Assumptions [i] and [ii] are equivalent to those of Anderson et al.(2010). Assumption [iii] is related to the possibility of the asymptotics under $N \to \infty$ or $T \to \infty$, and the meaning of this assumption is explained by the following example.

Example 2.7 : Suppose $N \to \infty$ and $T = 2$, and consider the TSLS estimator with the following objective function:

$$\min_{\mathbf{b}} \quad \mathbf{b}'\mathcal{E}[\mathbf{y}_{it}^{(f)}\mathbf{z}_{it}^{(b)\prime}] \left(\mathcal{E}[\mathbf{z}_{it}^{(b)}\mathbf{z}_{it}^{(b)\prime}]\right)^{-1} \mathcal{E}[\mathbf{z}_{it}^{(b)}\mathbf{y}_{it}^{(f)\prime}]\mathbf{b} \; .$$

This solution is immediately obtained by the true value $\mathbf{b} = \boldsymbol{\beta}$, because the orthogonal condition of $\mathcal{E}[\mathbf{z}_{it}^{(b)} u_{it}^{(f)}] = \mathbf{0}$ is then satisfied. Although the expectations are unknown, they are approximated by a law of large numbers. Similarly, they are also consistently estimated under $N \to \infty$ and $T < \infty$. Therefore, the meaning of assumption [iii] is to consider a class of consistent estimators even in the short panel data.

Under the assumptions, the result for the simple case of Anderson et al. (2010) holds.

Theorem 2.12 : *Supposing assumptions (A1), (A2), and (A4$'$) hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$\sqrt{NT}\left(\boldsymbol{\phi}(\ \frac{1}{n}\mathbf{G}_n^{(b,f)}\ ) - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{\Phi}^{-1})\ ,$$

for any function $\boldsymbol{\phi}$.

This asymptotic variance is the lower bound in the sense that any $\boldsymbol{\phi}$ cannot be smaller than $\sigma^2\mathbf{\Phi}^{-1}$, which is the same as that of Theorem 2.11. As $K < \infty$ is assumed here, the usual bound appears, but importantly, this approach can obtain the lower bound even under $K \to \infty$. Then, Anderson et al. (2010) showed that the lower bound becomes larger such as (3.38), and the LIML estimator can attain the bound. Although the D-GMM estimator, which is based only on $\mathbf{G}_n^{(b,f)}$, can be consistent and attain the bound under $K < \infty$, the next section considers the setting $K \to \infty$ in the dynamic panel model.

3.7 Incidental Parameters Problem Revisited

Under standard assumptions, the T-LIML, D-LIML, and D-GMM estimators have the desirable properties. This section relaxes the standard assumptions and investigates the two incidental parameter problems. First, Anderson and Hsiao (1981) raised the problem of the initial value in Section 2.2, which is related to the robustness of the T-LIML estimator for the initial value. Second, we compare the D-LIML and D-GMM estimators under the large-K theory and show the superiority of the D-LIML estimator.

3.7.1 Robustness of T-LIML for the Initial Condition

The finite sample properties of the T-LIML estimator are better than those of other methods using the filters as shown in the next section. One of the reasons is that the number of total data decreases from NT to $N(T-1)$ when filtered.

However, in order for the transformed maximum likelihood estimator to use all data, assuming the random-effects on the initial values is important. Hence, we consider the asymptotic properties of the T-LIML estimator when the initial values are incidental parameters.

Regarding the setting of the initial value, we consider the case when the maximum likelihood method falls into the incidental parameters problem as discussed by Anderson and Hsiao (1981). That is, the initial state w_{i0}, which is not observable unlike y_{i0}, is heterogeneous even if individual effects are removed. The structural model is the one considered in Section 3.3.1, expressed by the long difference,

$$\begin{aligned} y_{it}^{(1,\ell)} &= \beta_2 \pi_{22} y_{it-1}^{(2,\ell)} + \gamma_1 y_{it-1}^{(1,\ell)} + \xi_i^{(1)} + v_{it}^{(1)} \ , \\ y_{it}^{(2,\ell)} &= \pi_{22} y_{it-1}^{(2,\ell)} + \xi_i^{(2)} + v_{it}^{(2)} \ . \end{aligned}$$

The estimation method is based on (3.27). That is, even if the initial values are incidental parameters, we regarded it as a random-effects model. Therefore, we examine the property of the maximum likelihood method under the misspecification as in the work of White (1982).

$$y_{it}^{(2,\ell)} = \pi_{20} + \pi_{22} y_{it-1}^{(2,\ell)} + \xi_i^{(2)} + v_{it}^{(2)} \ ,$$

unlike Hsiao and Zhou (2015), the estimator is made without including the constant term π_{20} for absorbing the initial value. Instead of assumption (A2) for the initial value, we consider the initial state $\mathbf{w}_{i0}$ $(i = 1, \cdots, N)$ as incidental parameters; that is, $\boldsymbol{\xi}_i$ is also set as an incidental parameter.

(**A2**$'$) [i] Let $\boldsymbol{\Omega}_\xi \geq \mathbf{O}$ be the parameter space. Suppose that $\|\boldsymbol{\xi}_i\|$ $(i = 1, \cdots, N)$ are bounded and that

$$\underset{2\times 2}{\bar{\boldsymbol{\Omega}}_N} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\xi}_i \boldsymbol{\xi}_i' \quad \underset{N\to\infty}{\longrightarrow} \quad \bar{\boldsymbol{\Omega}}_\xi \ ,$$

where the (1,1) and (2,2) elements of $\bar{\boldsymbol{\Omega}}_\xi$ are not zero.

[ii] $\mathbf{H}_{\psi\psi}$ of (A3) exists when evaluated at $\bar{\boldsymbol{\Omega}}_\xi$.

Although $\omega_{\xi,11} \geq 0$ and $\omega_{\xi,22} \geq 0$ by definition, we assume $\bar{\omega}_{\xi,11} \neq 0$ and $\bar{\omega}_{\xi,22} \neq 0$ to avoid a corner solution.[14] $|\bar{\boldsymbol{\Omega}}_\xi| > 0$ generally holds, but if the model has a constant term $\boldsymbol{\xi}_i = \boldsymbol{\xi}$ for all i, then $|\bar{\boldsymbol{\Omega}}_\xi| = 0$. The case of starting from the common initial state is included:

$$\mathbf{w}_{i0} = \mathbf{w}_0 \implies \bar{\boldsymbol{\Omega}}_\xi = (\mathbf{I}_2 - \boldsymbol{\Pi}')\mathbf{w}_0\mathbf{w}_0'(\mathbf{I}_2 - \boldsymbol{\Pi}) \ .$$

[14]If short panel data $(T < \infty)$, $\bar{\boldsymbol{\Omega}}_N$ would be an interior point.

When the initial states $\mathbf{w}_{i0}$ $(i = 1, \cdots, N)$ are the incidental parameters, the following holds.

Theorem 2.13 : *Supposing assumptions (A1) and (A2') hold, then as both N and T tend to infinity,*

$$\hat{\boldsymbol{\theta}}_{1\mathrm{TL}} \xrightarrow{p} \boldsymbol{\theta}_1 , \quad \hat{\boldsymbol{\Omega}}_{\xi\mathrm{TL}} \xrightarrow{p} \bar{\boldsymbol{\Omega}}_\xi .$$

If $|\bar{\boldsymbol{\Omega}}_\xi| > 0$, *then*

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{\mathrm{TL}} - \boldsymbol{\theta}_1\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1}) .$$

Even if the initial values are incidental parameters, the result is the same as that of Theorem 2.7. Therefore, the T-LIML estimator can be said to be robust to the incidental parameters of the initial value. However, when $\bar{\boldsymbol{\Omega}}_\xi$ is a semidefinite matrix, a more detailed examination would be required.

This result may have issues. First, the maximum likelihood method can be a consistent estimator without a noncentrality parameter. In the work of Anderson and Hsiao (1981), estimating many initial states causes the incidental parameters problem, which is the motivation to develop the instrumental variable method. However, the maximum likelihood method works well under the misspecification, thereby providing a solution to the problem. Second, the asymptotic variance of Theorem 2.7 does not depend on $\bar{\boldsymbol{\Omega}}_\xi$. That is, whether the initial value is random or an incidental parameter does not affect the asymptotic efficiency in long panel data. Third, including a constant term is not necessary. The transformed maximum likelihood method can include a constant term to absorb the effect of the initial value. When the constant term π_{20} is added, then we have the following:

$$\hat{\pi}_{20\mathrm{TL}} \xrightarrow{p} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_i^{(2)} .$$

That is, the constant term estimates the average value of the individual effect that depends on the initial state. Meanwhile, the T-LIML estimator can absorb the influence by the average of the second moment $\bar{\boldsymbol{\Omega}}_\xi$ without the constant term.

The transformed maximum likelihood method is an approach that combines the fixed-effects and random-effects estimations and has the form of the random-effects MLE under the assumption that the initial value is random. Notably, our result can be interpreted as the robustness of the random effects-MLE itself to the incidental parameters. Hence, we reconsider why we need to care about the random-effects or the fixed-effects. Binder et al. (2005) discussed the settings of several random-effects. A nonrandom-effects is not to follow the same distribution independently, and the following examples clarify the several cases.

Example 2.8 : When the original individual effect η_i does not follow the identical distribution, the following two cases can be considered.

[i] The expectation varies between individuals with zero variance,

$$\mathcal{E}[\eta_i] = \eta_i \ .$$

That is, N constant terms or incidental parameters exist.

[ii] The individual effect have a distribution but the variance is not homogeneous,

$$\mathcal{V}ar[\eta_i^2] = \omega_i > 0 \ ,$$

then, the effects depend on the parameters of N distributions. However, a counter example is such that

$$\eta_i \ \sim \mathcal{N}(\mu_i,\ \omega_i) \ , \quad \mu_i \ \sim \ \mathcal{N}(\mu,\ \omega_\mu) \ , \quad \omega_i \ \sim \ \chi_1^2 \ ,$$

and μ_i and ω_i are independent. If $(\mu_i,\ \omega_i)$ is conditioned, then the distribution is not identical, but the unconditional expectation and variance are given as follows:

$$\mathcal{E}[\eta_i] = \mu \ , \quad \mathcal{V}ar[\eta_i] = \omega_\mu + 1 \ .$$

That is, a random-effects model follows the identical distribution with only two parameters $(\mu,\ \omega_\mu)$. When parameter μ_i or ω_i is constant, it is not a random-effects model.

The followings are related to the assumption of independence.

[iii] The individual effect correlates with other variables. When correlated with $\mathbf{z}_{it}$ on the right-hand side,

$$\mathcal{E}[\eta_i \mathbf{z}_{it}] \neq \mathbf{0} \ ,$$

or with the error term,

$$\mathcal{E}[\eta_i v_{it}] = \rho_{it} \ , \quad (t = 1,\ \cdots,\ T) \ ,$$

then the model can include NT parameters.

[iv] The effects are not independent between individuals,

$$\mathcal{E}[\eta_i \eta_j] = \omega_{ij} \ , \quad (i,\ j = 1,\ \cdots,\ N) \ ,$$

then, the order of parameters can be $O(N^2)$.

[v] Although the distribution is identical, the moments $\mathcal{E}[\eta_i]$ or $\mathcal{E}[\eta_i^2]$ does not exist.

Even in the above cases, no problem with the T-LIML and D-LIML estimators exists because they eliminate individual effects η_i. The case of [i] is related to the setting of Anderson and Hsiao (1981) or the incidental parameters problem of the initial states ξ_i, if η_i is regreded as ξ_i. In relation to [ii], Hayakawa and Pesaran (2015) suggested that the transformed maximum likelihood method is robust when the error term of the reduced form is not identical, $\mathcal{V}ar[v_{it}^{(2)}] = \omega_{2i}$. Even when [i] and [ii] are combined, the T-LIML method without adding a constant term consistently estimates the second moment $(1/N)\sum_i(\eta_i^2 + \omega_i)$ and is not affected by the incidental parameters. We may call $\bar{\mathbf{\Omega}}_\xi$ the averaged parameter for the incidental parameters. If the random-effects MLE consistently estimates a finite number of averaged parameters, then it can be a consistent estimator without the homogeneous assumption of the individual effect. The fixed-effects estimation eliminates the individual effect by subtraction, but the random-effects estimation can be interpreted as adding up the individual effects to eliminate the individual effect.

Thus, in linear panel data models, the problem with the random-effects estimation may not be significant for the heterogeneity of individual effects, but the correlations between individual effects and other variables may be crucial such as the case of [iii]. In empirical analyses, Hausmann test often suggests a correlation between the individual effect η_i and the explanatory variable z_{it}. In the dynamic model, of course, a correlation exists because $z_{it} = y_{it-1}$, but the random-effects MLE is consistent. The reason is that the AR structure completely describes the data generation process in the likelihood function. Therefore, the correlation is also not the problem, but whether the correlation between all variables $\mathbf{z}_{it}$ and the individual effect η_i can be specified and added to the likelihood function is important. For instance, Hsiao and Zhou (2015) suggested that the random-effects MLE is biased when the individual effect correlates with the exogenous variable $\mathbf{z}_{it}$, which does not follow an AR model. The formulation of the random-effects model requires additional attention to the endogeneity biases based on the individual effects.

3.7.2 Robustness of D-LIML for Large-K Asymptotics

This section closely compares the D-LIML estimator of Theorem 2.9 with the D-GMM estimator of Corollary 2.1. From the numerical experiments in the next section, their finite sample properties can be quite different. For instance, when $N = 100$ and $T = 25$, if the number of instrumental variables is $K_n = 2$, then the finite sample properties are similar. However, if $K_n = 10$, then the empirical distribution of the D-GMM estimator seems to have a noncentrality parameter. The results of Theorem 2.9 cannot explain this phenomenon because the LIML and GMM estimators have the same asymptotic distribution under the standard assumptions.

As shown in example 2.3, many instrumental variables are available in some cases. The advantages of panel analysis is that the total data $n = NT$ is large, so that the number of explanatory variables K_n can be increased. However, K_n/n is usually considered as almost zero, that is, $K_n/n = 10/2500 \simeq 0$ in the above example. Therefore, for long panel data,

$$\frac{K_n}{NT} \quad \longrightarrow \quad c_2 = 0 \ ,$$

may be natural even in $K_n \to \infty$. However, the following is possible,

$$\frac{(K_n)^2}{NT} \quad \longrightarrow \quad d_2 \neq 0 \ . \tag{3.44}$$

Then, the asymptotic properties differ between the LIML and GMM estimators. In the previous example, $K_n^2/n = 100/2500$ may be better not regarded as zero. In the following, we explain the difference in finite sample properties by using the large-K theory. In the long panel data, a sufficient condition for $d_2 \neq 0$ is as follows:

$$\frac{K_n}{T} \quad \longrightarrow \quad c_3 > 0 \ , \qquad \frac{T}{N} \quad \longrightarrow \quad c > 0 \ .$$

Then, $d_2 = c_3^2 c > 0$ is obtained, and the assumption of $c_3 \geq 0$ is relatively weak than the standard assumption $c_3 = 0$. That is, if the ratio K_n/T in the long panel is not regarded as 0, then the situation of $c_2 = 0$ and $d_2 > 0$ most likely occurs. This asymptotics is more accurate, so the differences in the finite sample properties can be explained.

Considering the simple structural model in Section 3.1,

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} \ , \qquad (3.45) \\ y_{it}^{(2)} &= \pi_{22} y_{it-1}^{(2)} + \sum_{k=3}^{K_{2n}+2} \pi_{2k,n} y_{it-1}^{(k)} + \pi_i^{(2)} + v_{it}^{(2)} \ , \\ \pi_{2k,n} &= \frac{\pi_{2k}}{\sqrt{K_{2n}}} \ , \end{aligned}$$

where the many weak instrumental variables are added. Then, the reduced form of the first structural equation becomes

$$y_{it}^{(1)} = \pi_{11} y_{it-1}^{(1)} + \pi_{12} y_{it-1}^{(2)} + \sum_{k=3}^{K_{2n}+2} \pi_{1k,n} y_{it-1}^{(k)} + \pi_i^{(1)} + v_{it}^{(1)} \ ,$$

$y^{(1)}_{it-1}$ ($K_1 = 1$) appears in the first structural equation, and the number of instrumental variables that do not appear becomes $K_2 = 1 + K_{2n}$. $y^{(k)}_{it-1}$ ($k = 3, \cdots, K_n = K_{2n} + 2$) are K_{2n} instrumental variables that can increase, and we assume AR(1) models as follows:

$$y^{(k)}_{it} = \pi_{3k} y^{(k)}_{it-1} + \pi^{(k)}_{i} + v^{(k)}_{it} \ , \ (k = 3, \ \cdots, \ K_n) \tag{3.46}$$

and

$$K_n = 2 + K_{2n} \quad \longrightarrow \quad \infty \ .$$

Thus, incidental prameters as the coefficients $\pi_{1k,n}$ and $\pi_{2k,n}$ ($k = 3, \ \cdots, \ K_n$) exist.

We consider the setting under the large-K asymptotics in the dynamic structural panel model, which is different from the previous sections or the previous studies in cross-sectional analysis. First, the reduced form becomes the high dimensional VAR(1) model (cf. Davis et al. (2016)). Then, the number of potential endogenous variables is also large, so the stationarity condition should be confirmed. To be a covariance stationary process under $K_{2n} \to \infty$, the following is required,

$$\mathcal{V}ar\left[y^{(2)}_{it}\right] = O(1) \ .$$

If $y^{(k)}_{it-1}$ ($k = 3, \ \cdots, \ K_n$) are mutually independent and

$$\pi_{2k,n} = O\left(K_{2n}^{-\frac{1}{2}}\right) \ , \tag{3.47}$$

then the variance becomes bounded, which can be interpreted that the contribution of each coefficient is small when the number of explanatory variables is large.[15]

Second, the dimension of the projection matrix is large, and we have to evaluate the effects of the filters in the panel data analysis. For instance, we need to evaluate the following quantity,

$$\frac{1}{n}\mathbf{v}^{(2)\prime}\mathbf{P}^{(b)}\bar{\mathbf{u}}_{T\sim} = \underbrace{\left(\frac{1}{n}\mathbf{v}^{(2)\prime}\mathbf{Z}^{(b)}\right)}_{1\times K_n} \underbrace{\left(\frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)}\right)^{-1}}_{K_n\times K_n} \underbrace{\left(\frac{1}{n}\mathbf{Z}^{(b)\prime}\bar{\mathbf{u}}_{T\sim}\right)}_{K_n\times 1} \ .$$

As $K_n \to \infty$, evaluating each of the three terms by a law of large numbers such as the estimator in Theorem 2.9, is not possible, which is a difficulty of the large-K asymptotics. Moreover, if the error terms have heterogeneous variances such as Chao et al. (2012) and Kunitomo and Akashi (2010), or a serial correlation

[15] The case of the cross-sectional analysis is usually expressed on the order of $\pi_{2k,n} = O(N^{-\frac{1}{2}})$. As $K_n = O(N)$, it has the same meaning.

such as Alvarez and Arellano (2003) and Akashi and Kunitomo (2015), then the calculation becomes complicated. By the influence due to the filter,

$$\underset{n\times 1}{\bar{\mathbf{u}}_{T\sim}} = \left(\frac{\bar{u}_{it,T}}{f_t} \right) \ , \quad \bar{u}_{it,T} = \frac{1}{T-t+1}\left(u_{it} + \cdots + u_{iT}\right) \ ,$$

are heterogeneous and have a series correlation. To reduce the calculation, we consider the technical conditions for (3.44),

$$N = O(T^{\frac{1}{2}}) \ , \quad K_{2n} = O(T^{\frac{3}{4}}) \ ,$$

and then, $K_{2n}^2/n = O(1)$ holds again.

Third, the model has weakly exogenous variables under the large-K asymptotics. The weakly exogenous variables $(y_{it-1}^{(1)},\ y_{it-1}^{(2)})$ are included in the $n \times K_n$ instrumental variable matrix $\mathbf{Z}^{(b)}$, and many terms correlate with $\mathbf{u} = (u_{it})$. Then, for the conditional expectations given $\mathbf{Z}^{(b)}$,

$$\begin{aligned} \mathcal{E}\left[\mathbf{u}^{'}\mathbf{P}^{(b)}\mathbf{u}|\mathbf{Z}^{(b)}\right] &\neq \sigma^2 \mathrm{tr}(\mathbf{P}^{(b)}) \\ &= \sigma^2 K_n \ . \end{aligned}$$

In cross-sectional analysis, $\mathbf{P}^{(b)}$ can be treated as constant so that the first equality holds, but more careful evaluation is required in the dynamic model. However, if $K < \infty$ or the sequential moment conditions, then it does not matter.

The objective functions of the D-LIML and D-GMM estimators are the same as Theorem 2.9 and Corollary 2.1, respectively. That is,

$$\mathcal{VR}_2 = \frac{\boldsymbol{\theta}^{'}\mathbf{G}_n^{(f,b)}\boldsymbol{\theta}}{\boldsymbol{\theta}^{'}\mathbf{H}_n^{(f,b)}\boldsymbol{\theta}} \ ,$$

and

$$\mathcal{Q}_2 = \boldsymbol{\theta}^{'}\mathbf{G}_n^{(f,b)}\boldsymbol{\theta} \ .$$

Instead of (2.12), we use the following:

$$b_t = f_t \ ,$$

this is not essential but simplifies the evaluation of the filters.

Instead of assumption (A1), we make the following assumption.

(A1$'$) [i] $\{\mathbf{v}_{it}^*\}$ $(i = 1, \cdots, N;\ t = 1, \cdots, T)$ are i.i.d. across time and individuals, normally distributed random variables, and independent of $\mathbf{z}_{i0}^*$ with $\mathcal{E}[\mathbf{v}_{it}^*] = \mathbf{0}$ and $\mathcal{E}[\mathbf{v}_{it}^*\mathbf{v}_{it}^{*'}] = \boldsymbol{\Omega}^*$, where

$$\underset{K^*\times K^*}{\boldsymbol{\Omega}^*} = \begin{pmatrix} \underset{2\times 2}{\boldsymbol{\Omega}} & \mathbf{O}^{'} \\ \mathbf{O} & \omega_3 \mathbf{I}_{K_{2n}} \end{pmatrix} \ .$$

All absolute values of π_{11}, π_{22}, and $\pi_{3k} = \pi_{33}$ are less than 1, and

$$\pi_{2k,n} = \frac{\pi_2}{\sqrt{K_{2n}}} \ , \quad (k = 3, \ \cdots , \ K_n) \ .$$

[ii] $N = O(T^{\frac{1}{2}})$, $K_{2n} = O(T^{\frac{3}{4}})$, and $K_{2n}^2/n \to d_2$ as N, T, and K_{2n} tend to infinity.

From the assumption of $\mathbf{\Omega}^*$, the K_{2n} variables are strongly exogenous instrumental variables and are mutually independent, so that the order of (3.47) is supposed. As for the required stationary condition, π_2 is not included in the condition unlike the assumption in (A1). However, π_2 affects the asymptotic variance through $\mathbf{\Pi}_{1n}$ as follows:

$$\begin{aligned} \mathbf{\Phi}^* &= \mathbf{\Pi}_{\mathrm{I}n}^{'} \mathbf{\Gamma}_{0n} \mathbf{\Pi}_{\mathrm{I}n} \\ &= \mathbf{\Pi}_{1\mathrm{I}}^{'} \mathbf{\Gamma}_1 \mathbf{\Pi}_{1\mathrm{I}} + \mathbf{\Pi}_{1n}^{'} \mathbf{\Gamma}_{1n} \mathbf{\Pi}_{1\mathrm{I}} + \mathbf{\Pi}_{1\mathrm{I}}^{'} \mathbf{\Gamma}_{1n}^{'} \mathbf{\Pi}_{1n} + \sigma_3^2 \mathbf{\Pi}_{1n}^{'} \mathbf{\Pi}_{1n} \ , \end{aligned}$$

where the notations are given in the proof. The simplification of the coefficients is as follows:

$$\pi_{2k} = \pi_2 \ , \quad \pi_{3k} = \pi_{33} \ , \quad \mathcal{V}ar[v_{it}^{(k)}] = \omega_3 \ , \quad (k = 3, \ \cdots , \ K_n) \ ,$$

these may not be essential. The following is the result under the triple asymptotics that N, T, and K_{2n} go to infinity.

Theorem 2.14 : *Supposing assumptions (A1') and (A2) hold, then $K_{2n}/n \to c_2 = 0$, $K_{2n}^2/n \to d_2$, and*

[i]

$$\sqrt{NT} \left(\hat{\boldsymbol{\theta}}_{\mathrm{DG}} - \boldsymbol{\theta}_1 \right) \quad \xrightarrow{d} \quad \mathcal{N} \left(\mathbf{b}_{2\cdot 0}, \sigma^2 \mathbf{\Phi}^{*-1} \right) \ ,$$

where

$$\mathbf{b}_{2\cdot 0} = \sqrt{d_2} \boldsymbol{\rho}_0 \ .$$

[ii]

$$\sqrt{NT} \left(\hat{\boldsymbol{\theta}}_{\mathrm{DL}} - \boldsymbol{\theta}_1 \right) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{\Phi}^{*-1}) \ .$$

Unlike Theorem 2.9 and Corollary 2.1, the asymptotic distributions of the D-LIML and D-GMM estimators are not the same. The D-GMM estimator has the noncentrality parameter, whhereas the D-LIML estimator is still centered. $\mathbf{\Phi}^*$ of the asymptotic variance corresponds to the first term so far, and the second term $\mathbf{\Psi}$ does not appear because $c_2 = 0$. The D-LIML estimation is robust because the same result as the usual asymptotics can be obtained even if the explanatory variables increase. Theorem 2.14 can explain the difference between the D-LIML and D-GMM estimators in the finite sample properties as shown in the next section.

3.8 Simulation

This section compares the finite sample properties of various estimation methods. Following Akashi and Kunitomo (2012, 2015), we use the cumulative empirical distribution, which is the most informative. Although the normalization of $\boldsymbol{\beta} = (1,\ -\boldsymbol{\beta}_2)$ is commonly used in econometric analyses, the exact moments of the LIML estimator may not exist. Thus, the comparison by the mean square error is meaningless and should be based on such as the median of the empirical distribution.[16]

First, the calculation method for each LIML estimator is shown. Although the LIML estimator can be obtained numerically by maximizing the objective function such as $-\mathcal{VR}_2$, the following simple procedures are usually taken. We represent the LIML estimators from Theorem 2.3 to Theorem 2.10 as a generic $\hat{\boldsymbol{\theta}}_1 = (\hat{\boldsymbol{\beta}}_2^{'}, \hat{\boldsymbol{\gamma}}_1^{'})^{'}$, and then, $(1,\ -\hat{\boldsymbol{\theta}}_1^{'})^{'}$ becomes an eigenvector from the first-order condition of minimization,

$$[\mathbf{G} - \lambda_n \mathbf{H}] \begin{bmatrix} 1 \\ -\hat{\boldsymbol{\theta}}_1 \end{bmatrix} = \mathbf{0} \ , \tag{3.48}$$

λ_n is the smallest root of the following eigenvalue equation,[17]

$$|\mathbf{G} - l\mathbf{H}| = 0 \ ,$$

where $\mathbf{G} = \mathbf{G}^{(f)}$, $\mathbf{G}^{(f,b)}$, $\mathbf{G}_n^{(f,b)}$, and $\mathbf{G}_n^{(f,f)}$ for Theorems 2.2, 2.3, 2.9 (2.14), and 2.10, respectively. $\mathbf{H}$ is defined in the same way, for example $\mathbf{H} = \mathbf{H}_n^{(f,b)}$ for Theorem 2.9. The minimum eigenvalue λ_n is also related to the minimum value, and if the generic variance ratio is expressed as $\mathcal{VR}_n$, then

$$\lambda_n = \min_{\boldsymbol{\theta}_1} \ \mathcal{VR}_n \ .$$

Once the minimum eigenvalue λ_n is obtained, by solving (3.48) for $\hat{\boldsymbol{\theta}}_1$, the LIML estimator is calculated as follows:

$$\underset{(G_2+K_1)\times 1}{\hat{\boldsymbol{\theta}}_1} = \left(\mathbf{J}_{01}^{'}\mathbf{G}\mathbf{J}_{01} - \lambda_n \mathbf{J}_{01}^{'}\mathbf{H}\mathbf{J}_{01}\right)^{-1} \left(\mathbf{J}_{01}^{'}\mathbf{G}\mathbf{e}_1 - \lambda_n \mathbf{J}_{01}^{'}\mathbf{H}\mathbf{e}_1\right) \ ,$$

where $\mathbf{J}_{01}^{'} = (\mathbf{0},\ \mathbf{I}_{G_2+K_1})$ is the $(G_2 + K_1) \times (G + K_1)$ selection matrix, and $\mathbf{e}_1 = (1,\ 0,\ \cdots,\ 0)^{'}$ is the $(G + K_1)$ vector. The corresponding GMM estimator can be obtained by putting $\lambda_n = 0$.

---

[16] However, Anderson (2010) observed a moment under the natural normalization of $\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta} = \sigma^2 = 1$.

[17] In the case of Ox, the generalized eigenvectors and eigenvalues are obtained by eigensymgen(mG, mH, &vlambda, &vtheta). As for a usual command, the smallest root is obtained by $|\mathbf{H}^{-1/2'}\mathbf{G}\mathbf{H}^{-1/2} - l\mathbf{I}| = 0$. Ox is provided by Jurgen A. Doornik.

The T-LIML estimator is obtained by maximizing the log-likelihood function for all parameters. For example, in the case of (3.36), the T-LIML estimator is maximized with respect to π_{22}, $\mathbf{\Omega}$, and $\mathbf{\Omega}_\xi$ in addition to the parameter $\boldsymbol{\theta}_1$ of interest. Hsiao et al. (2002) and Han and Phillips (2013) noted that in a nearly nonstationary process, the behavior can considerably change.[18]

The following settings compare the estimators based on the cumulative standard normal distribution, which requires the same standardization for all estimators. We represent each estimator as $\hat{\boldsymbol{\theta}}_1^{[j]}$ $(j = 1, \cdots, J)$, for example, $\hat{\boldsymbol{\theta}}_1^{[1]} = \hat{\boldsymbol{\theta}}_{\mathrm{DL}}$ or $\hat{\boldsymbol{\theta}}_1^{[2]} = \tilde{\boldsymbol{\theta}}_{\mathrm{DG}}$. Then, the estimator of the k-th parameter $\boldsymbol{\theta}_1^{'}\boldsymbol{e}_k$ $(k = 1, \cdots, G_2+K_1)$ becomes $\hat{\boldsymbol{\theta}}_1^{'}\boldsymbol{e}_k$ $(k = 1, \cdots, G_2 + K_1)$. The number of endogenous variables on the right side is $G_2 = 1$; for example, we have that $\boldsymbol{\theta}_1^{'}\boldsymbol{e}_1 = \beta_2$ or $\boldsymbol{\theta}_1^{'}\boldsymbol{e}_2 = \gamma_2$. The asymptotic variance of many estimators becomes $\sigma^2\mathbf{\Phi}^{-1}$ in this work. Using the D-LIML estimator $\hat{\boldsymbol{\theta}} = (1, -\hat{\boldsymbol{\theta}}_{\mathrm{DL}}^{'})^{'}$ of Theorem 2.9, we define the consistent variance estimator for the standardization as follows,

$$\begin{aligned}\hat{\mathbf{V}} &= \left(\frac{1}{n}\hat{\boldsymbol{\theta}}^{'}\mathbf{H}_n^{(f,b)}\hat{\boldsymbol{\theta}}\right)\left(\frac{1}{n}\mathbf{J}_{01}^{'}\mathbf{G}_n^{(f,b)}\mathbf{J}_{01}\right)^{-1} \\ &\xrightarrow{p} \sigma^2\mathbf{\Phi}^{-1} \,.\end{aligned}$$

The empirical distribution of the standardized statistic t_{jk} for the k-th parameter of a certain j-th estimation method is summarized in each figure,

$$t_{jk} = \frac{\sqrt{n}}{\boldsymbol{e}_k^{'}\hat{\mathbf{V}}\boldsymbol{e}_k}\left(\hat{\boldsymbol{\theta}}_1^{[j]} - \boldsymbol{\theta}_1^{[j]}\right)^{'}\boldsymbol{e}_k \,, \tag{3.49}$$

where $n = N(T-2)$ because the D-LIML estimator uses the forward and backward filters. If the asymptotic variance of the j-th estimator is actually $\sigma^2\mathbf{\Phi}^{-1}$, then $t_{jk} \xrightarrow{d} \mathcal{N}(0, 1)$; that is, it becomes a t-test statistic.

Regarding Theorem 2.2, Akashi and Kunitomo (2012) already compared the finite sample properties in detail, so that they are omitted. Akashi and Kunitomo (2015) compared Theorem 2.2 with Theorem 2.3 and found that the D-LIML estimator of Theorem 2.3 significantly improved the finite sample properties. In the following, the finite sample properties of the efficient estimators such as the D-LIML and T-LIML estimators are investigated. We confirm that the proposed estimators in this work, which are not based on the sequential moment conditions, further improve those of Akashi and Kunitomo (2012, 2015).

Design 2.1 : This setting uses the simplest model in the works of Blundell and

[18]The maximization is based on Ox's BFGS algorithm by imposing the stationarity constraint of $\gamma_1 = 2\gamma/(1+\gamma^2)$ (cf. Bhargava and Sargan (1983)).

Bond (2000) and Akashi and Kunitomo (2012). The number of repetitions is $R = 3000$ times.

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} \ , \\ y_{it}^{(2)} &= \gamma_2 y_{it-1}^{(2)} + \rho \alpha_i + v_{it} \ , \end{aligned}$$

where the coefficients are as follows, $(\beta_2,\ \gamma_1,\ \gamma_2) = (0.5,\ 0.3,\ 0.3)$. The error term follows the normal distribution with zero mean and $(\omega_{11},\ \omega_{22},\ \omega_{12}) = (1,\ 1,\ 0.3)$. The individual effect α_i also follows $\mathcal{N}(0,1)$, where $\rho = 1$. The stationary and identification conditions are $|\gamma_1| < 1$, $|\gamma_2| < 1$, and $\gamma_2 \neq 0$. We start with $(y_{it}^{(1)},\ y_{it}^{(2)}) = (0,\ 0)$ and discard $T_{-} = 10$ times before the initial value to approximate the stationarity.

In the following, $R = 3000$ and $T_{-} = 10$ are the same, the error term and individual effect are based on the same normal distribution, and the stationary condition (A1) is satisfied. However, in Design 2.2, the initial value is accurate.

Figures 1-3 show the empirical distributions of the D-LIML (D-LIML'15 in the figure), D-GMM (D-GMM'15), D-LIML (D-LIML), and T-LIML (T-LIML) estimators, which correspond to Theorems 2.3-[ii], 2.3-[iii], 2.9, and 2.7, respectively. Notably, in a just identified case such as this design, the D-GMM estimator of Corollary 2.1 is numerically equal to the D-LIML estimator of Theorem 2.9.

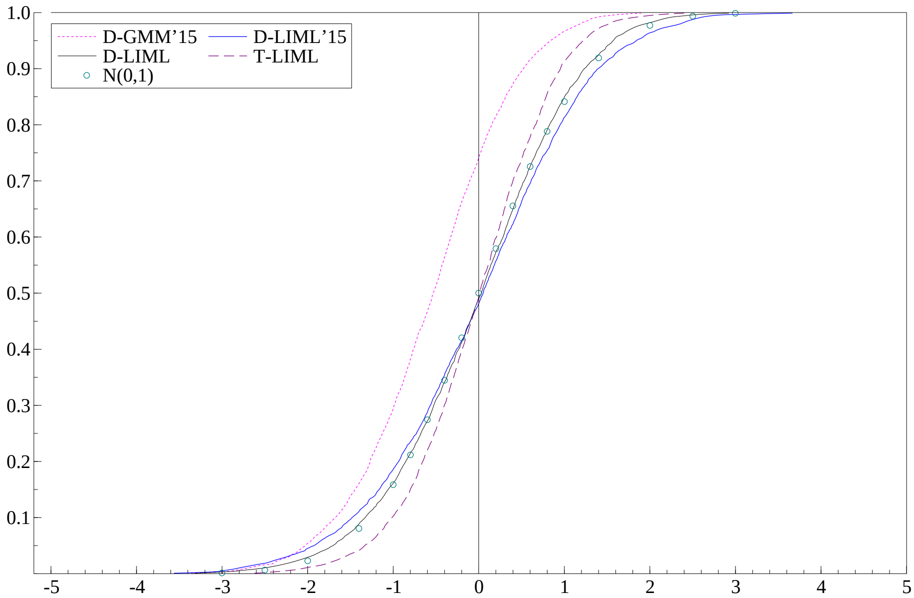

Fig. 1: Design 2.1 ($\beta_2,\ N = 100,\ T = 25$)

First, Figure 1 shows that the D-LIML estimator is more efficient than the D-LIML'15 estimator because the empirical distribution is shrunk. This result is

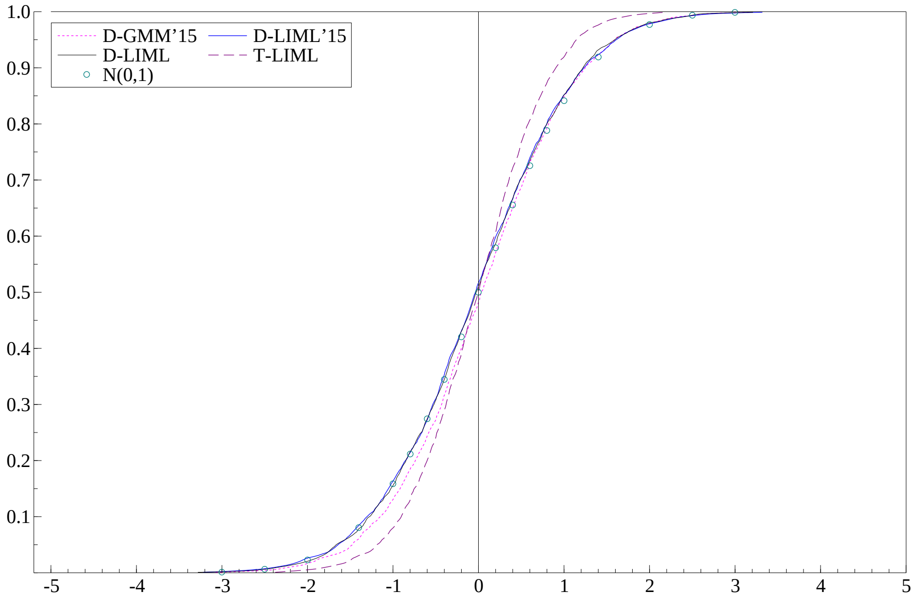

Fig. 2: Design 2.1 ($\gamma_1,\ N = 100,\ T = 25$)

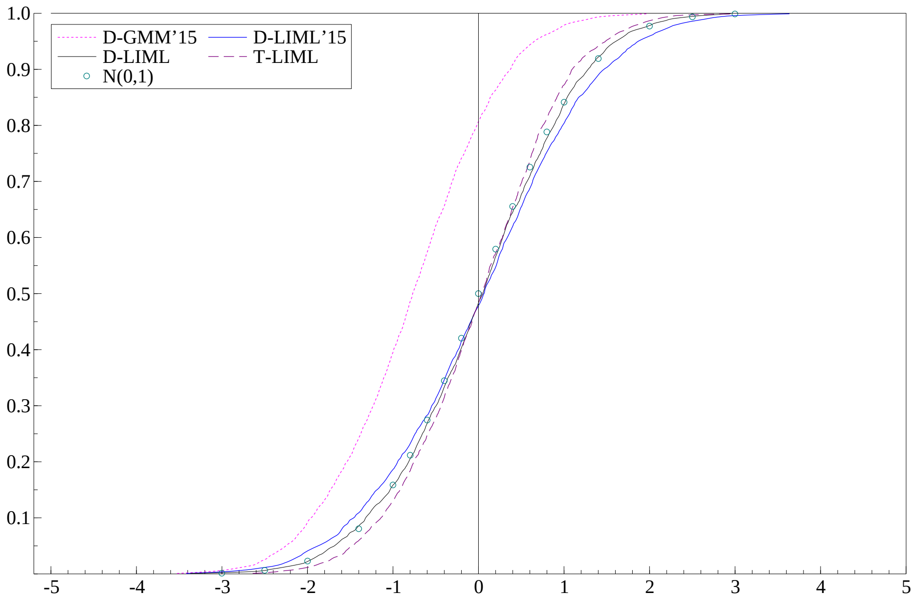

Fig. 3: Design 2.1 ($\beta_2,\ N = 100,\ T = 50$)

consistent with the fact that the second term $\mathbf{\Psi}$ of the variance appears in Theorem 2.3, whereas the D-LIML estimator of Theorem 2.9 is the efficient estimator. The noncentrality parameter of D-GMM'15 estimator is confirmed similar to the result of Akashi and Kunitomo (2015). Hsiao and Chou (2015) also showed that the T-LIML estimator is better than the D-LIML'15 estimator.

Second, although the D-LIML and T-LIML estimators are asymptotically equivalent, the T-LIML seems to be more efficient because the empirical distribution of the latter shrinks. The reason may be that there is no loss occurs because of the forward and backward filters. Figure 2 shows the case of γ_1, but the difference is not as large as that in Figure 1, so the results for the case of β_2 are mainly shown in the following. Figure 3 shows the case of $T = 50$, and the convergence in distribution can be confirmed. As $\hat{\mathbf{V}}$ is used for the standardization, the empirical distribution of D-LIML is the closest to the standard normal distribution. Therefore, if viewed as the t-test statistic, then the case of D-LIML has less size distortion.

Design 2.2 : We confirm that the T-LIML estimator can estimate the AR(2) model and the case when the initial values are incidental parameters. The panel VAR(2) model in Section 3.4.2 can be expressed in state-space representations:

$$\begin{aligned} y_{it}^{(1)} &= w_{it}^{(1)} + \mu_i^{(1)} \ , \\ y_{it}^{(2)} &= w_{it}^{(2)} + \mu_i^{(2)} \ , \end{aligned} \tag{3.50}$$

and

$$\begin{aligned} w_{it}^{(1)} &= \beta_2 w_{it}^{(2)} + \gamma_{11} w_{it-1}^{(1)} + \gamma_{12} w_{it-2}^{(1)} + u_{it} \ , \\ w_{it}^{(2)} &= \gamma_2 w_{it-1}^{(2)} + v_{it} \ , \end{aligned}$$

where $(\beta_2,\ \gamma_{11},\ \gamma_{12},\ \gamma_2) = (0.5,\ 0.3,\ 0.3,\ 0.3)$, and the values of $\mathbf{\Omega}$ are the same as those in Design 2.1. The individual effect is $\boldsymbol{\mu}_i = (\mathbf{I}_2 - \mathbf{\Pi}^{'})^{-1}\boldsymbol{\pi}_i$ as shown in (3.12), where $\boldsymbol{\pi}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. The initial value vector $(w_{i0}^{(1)},\ w_{i(-1)}^{(1)},\ w_{i0}^{(2)})$ follows $\mathcal{N}(\mathbf{0}, \mathbf{I}_3)$.

When the AR(2) model is included, the likelihood function $\mathcal{L}_2$ is calculated using the long difference and first-difference for the initial value as shown in (3.36). Figure 4 suggests that the T-LIML estimator is more efficient similar to the previous design.

Next, we consider the case when the initial states are incidental parameters using the VAR(1) model. Expressed in state-space representation to set the initial value, the observation equation of $(y_{it}^{(1)},\ y_{it}^{(2)})$ is the same as that in (3.50):

$$\begin{aligned} w_{it}^{(1)} &= \beta_2 w_{it}^{(2)} + \gamma_{11} w_{it-1}^{(1)} + u_{it} \ , \\ w_{it}^{(2)} &= \gamma_2 w_{it-1}^{(2)} + v_{it} \ , \end{aligned}$$

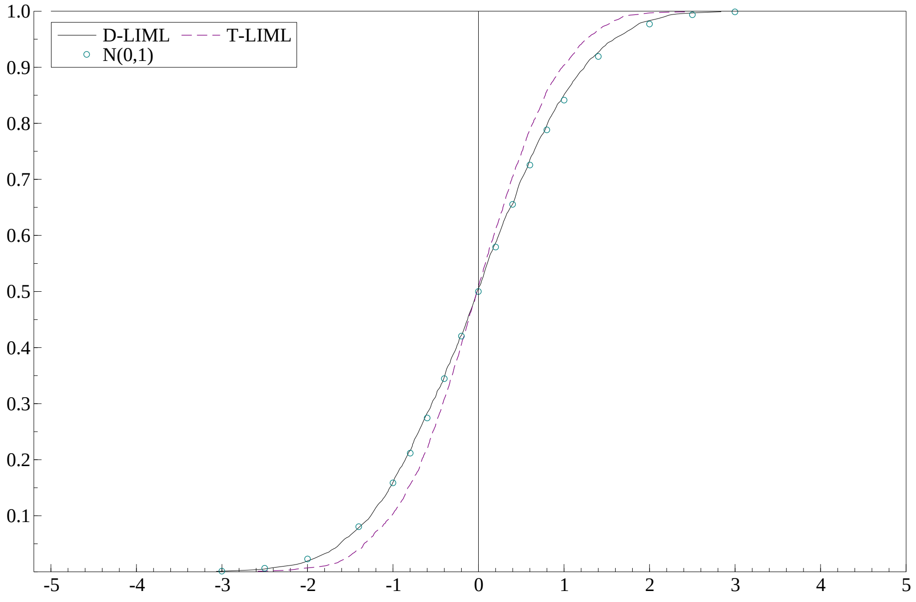

Fig. 4: Design 2.2 ($\beta_2,\ N = 100,\ T = 25$)

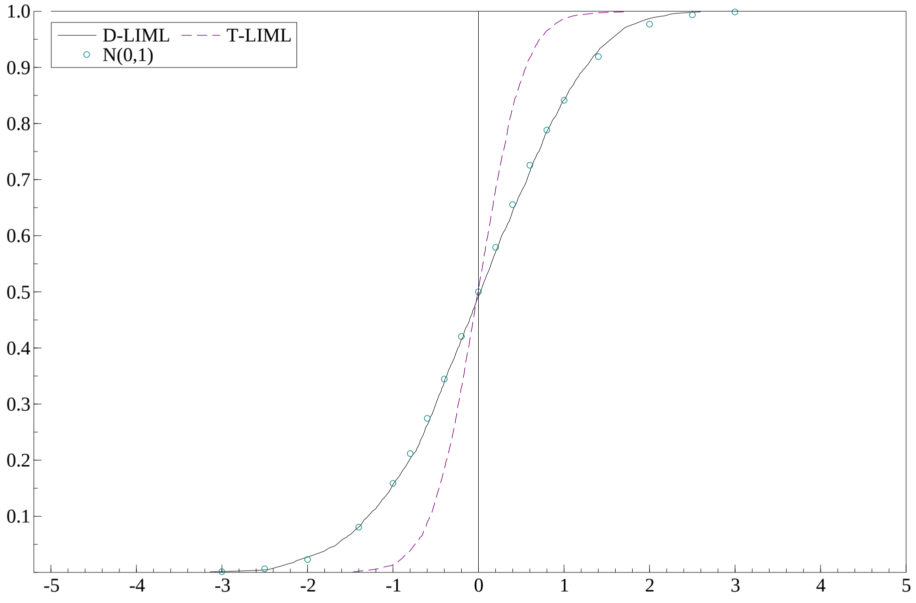

Fig. 5: Design 2.2 ($\beta_2,\ N = 100,\ T = 25$)

where $(\beta_2,\ \gamma_{11},\ \gamma_2)$ and $\boldsymbol{\mu}_i$ are the same as those in the VAR (2) model above. However, the initial state $(w_{it}^{(1)},\ w_{it}^{(2)})$ is generated by

$$\begin{aligned} w_{i0}^{(1)} &\sim \mathcal{N}(-1, 3)\ , \\ w_{i0}^{(2)} &\sim \mathcal{N}(3, 2)\ . \end{aligned}$$

Notably, the same realization is used as the initial value for all repetitions $R = 3000$, and thus, it can be regarded as the incidental parameters. Even if we try other fixed-effects such as $w_{i0}^{(1)} = -1 + \log(i)/N$, a similar result is obtained.

From Figure 5, even if the initial values are incidental parameters, the D-LIML estimator is consistent. Moreover, as the empirical distribution of T-LIML is centered at the origin, the noncentrality parameter does not appear although the initial state is not random-effects. The finite sample properties of the T-LIML estimator are also better than those of the D-LIML estimator.

Design 2.3 : We consider the more general model and the large-K model.

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} x_{it} + \alpha_i + u_{it}\ , \\ y_{it}^{(2)} &= \gamma_{21} y_{it-1}^{(2)} + \gamma_{22} y_{it-2}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)}\ , \\ x_{it} &= \gamma_{31} x_{it-1} + \gamma_{32} x_{it-2} + \pi_i^{(3)} + v_{it}^{(3)}\ , \end{aligned}$$

where the structural parameters are as follows $(\beta_2,\ \gamma_{11},\ \gamma_{12},\ \gamma_{21},\ \gamma_{22}) = (0.5,\ 0.3,\ 0.3,\ 0.3,\ 0.1)$ and $(\gamma_{31},\ \gamma_{32}) = (0.3,\ 0.1)$ for x_{it}. The values of $\boldsymbol{\Omega}$ are the same as those in Design 2.1, but $v_{it}^{(3)}$ with unit variance is independent of $\mathbf{v}_{it}$, so x_{it} is exogenous in period t. As for the individual effect, $\boldsymbol{\pi}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$.

Although the finite sample properties of the T-LIML estimator are the best in the previous designs, the calculation becomes quite complicated in the case of the general model, so the results without the T-LIML estimator are given. Moreover, the procedure for calculating the D-LIML estimator is easy even with the general model. For instance, with the package software such as EViews, the D-LIML estimator can be obtained by the procedure of the original LIML estimator for cross-sectional data. We generate the filtered data using $\mathbf{D}_f$ in Section 2.4 and $\mathbf{D}_b$ in Section 2.5. Then, we consider the data as the cross-sectional data consisting of $n = N(T-1)$ samples, that is,

$$\underset{n\times(G+K_1+K)}{\left(\mathbf{y}^{(1,f)},\ \mathbf{X}^{(f)},\ \mathbf{Z}^{(b)}\right)}\ .$$

The D-LIML estimator is obtained by specifying $(G + K_1)$ variables in the first structural equation and K variables as the instrumental variables.

From the reduced form of $(y_{it}^{(1)},\ y_{it}^{(2)})$, the use of $(y_{it-1}^{(1,b)},\ y_{it-1}^{(2,b)},\ y_{it-2}^{(2,b)}, x_{it}^{(b)}\ )$ as instrumental variables is sufficient, where $K = 4$ and $K_2 = 2$. Figure 6 shows that the D-LIML estimator is a further improvement over D-LIML'15 compared with Design 2.1. Moreover, the empirical distributions of D-LIML and D-GMM are almost the same. The result corresponds to Theorem 2.9 and Corollary 2.1 because $c_2 = K/n = 4/2500$ can be regarded as almost zero.

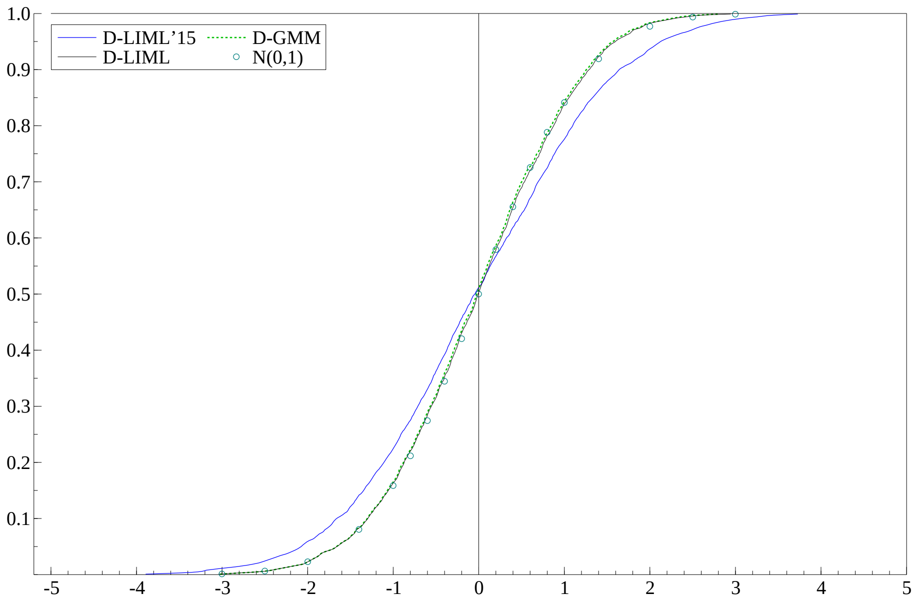

Fig. 6: Design 2.3 ($\beta_2,\ N = 100,\ T = 25$)

Next, we consider the design of the large-K asymptotics. The setting is the same as that in (3.45) to (3.46) in Section 3.7.2:

$$
\begin{aligned}
y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} , \\
y_{it}^{(2)} &= \pi_{22} y_{it-1}^{(2)} + \sum_{k=3}^{12} \pi_{2k,n} y_{it-1}^{(k)} + \pi_i^{(2)} + v_{it}^{(2)} ,
\end{aligned}
$$

where $(\beta_2,\ \gamma_1,\ \pi_{22}) = (0.5,\ 0.5,\ 0.3)$. Regarding $k = 3,\ \cdots,\ 12$, the common values are used such that $\pi_{2k,n} = 0.1$, $\pi_{3k} = 0.5$, and $\omega_3 = 1$. The individual effect is the same as in Design 2.1, but the setting of the error term is $(\omega_{11},\ \omega_{22},\ \omega_{12}) = (1,\ 1,\ -0.3)$. In other words, $K = 2 + 10 = 12$ and $c_2 = 12/2500$. Although c_2 is also almost zero, $d_2 = 144/2500$ may be better regarded as nonzero. As indicated by Theorem 2.14, Figure. 7 shows that the D-LIML estimator is still centered at the origin, whereas the D-GMM estimator has the noncentrality parameter $\mathbf{b}_{2\cdot 0}$ depending on d_2.

We summarize the results of the estimation theory in Part II. We introduced that the GMM estimator of Arellano and Bond (1991) may not perform well in Part I, but in the structural estimation, this estimator becomes inconsistent

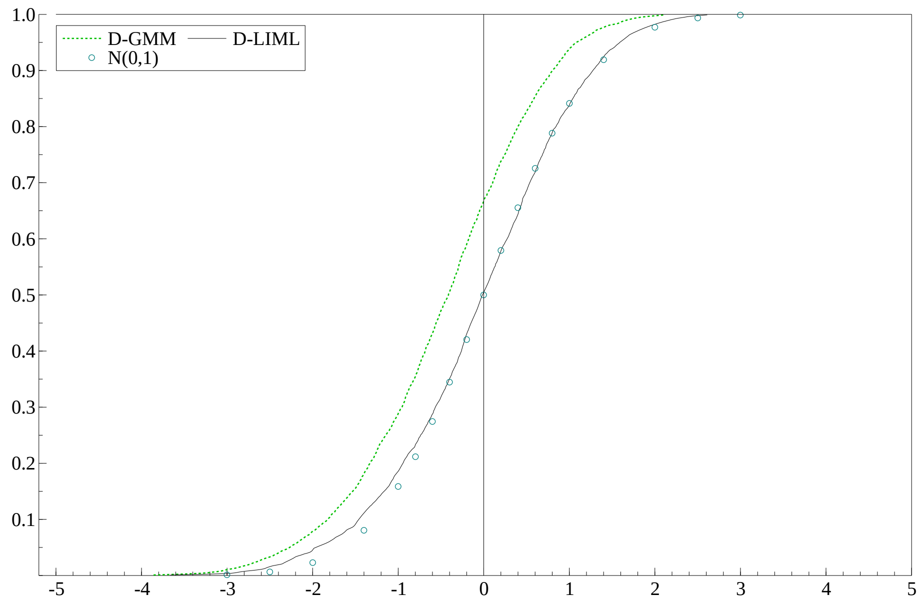

Fig. 7: Design 2.3 ($\beta_2,\ N = 100,\ T = 25$)

in long panel data, which is more remarkable. The T-LIML, D-LIML, and D-GMM estimators are shown efficient so that the estimators based on the sequential moment conditions mean inefficient. The finite sample property of the T-LIML estimator is the best, and the D-LIML estimator is more robust than the D-GMM estimator. Hence, we would like to recommend the T-LIML and D-LIML estimators of Theorems 2.7 and 2.9. However, as the T-LIML estimator is the exact maximum likelihood method, the calculation is complicated for a general model. Therefore, the D-LIML estimator, which approximates the T-LIML estimator and is easy to implement, is practical. In Part III, we examine the hypothesis testing for the panel structural estimation using the proposed D-LIML estimator.

4 Part III: Tests of Structural Analysis

This part discusses the hypothesis testing and the specification of the dynamic structural panel model from the viewpoint of empirical analyses. Unlike regression analyses, structural models require some additional procedures. In particular, we consider testing exogeneity, model selections by information criterion, and rank tests for identification. We mainly use the variance ratio as the statistics, which is the objective function of the practical D-LIML estimator. The test of the significance for each coefficient is also important in empirical analyses, for instance,

$$\mathrm{H}_0 :\ \beta_2 = 0\ .$$

The t-test statistic has already been given by (3.49), which is based on the D-LIML estimator of Theorem 2.9.

For the hypothesis test of structural parameters, the variance ratio of the LIML estimator has been called the AR test statistic since Anderson and Rubin (1949). In the following, the test statistic is composed entirely of the variance ratio of the D-LIML estimator. The advantage of the LIML method is that it can discuss the estimation and testing in a unified manner, and we may refer to as the panel AR test statistic in a broad sense.

4.1 Overidentification Tests

To estimate the structural parameters, the instrumental variables in period t must satisfy the orthogonal condition,

$$\mathrm{H}_0 : \mathcal{E}\left[\mathbf{z}_{it}u_{it}\right] = \mathbf{0} . \tag{4.1}$$

If the specification is correct, then $u_{it} = \mathbf{v}_{it}^{'}\boldsymbol{\beta}$, so that the condition is satisfied. When the D-LIML estimatior is used, the orthogonality after removing the individual effect implies the following:

$$\mathrm{H}_0 : \mathcal{E}\left[\mathbf{z}_{it}^{(b)}u_{it}^{(f)}\right] = \mathbf{0} ,$$

where $\mathbf{z}_{it}$ consists of the lagged endogenous variables before period $t-1$, such as $y_{t-1}^{(1)}$ and $y_{t-2}^{(1)}$ or the exogenous variables in period t. For instance, Arellano and Bond (1991) considered the case when the error term follows a moving average process, then, the orthogonal condition is not satisfied.

On the basis of D-LIML estimator, we check whether the candidates of instrumental variables are the predetermined variables, satisfing (4.1). Let

$$\lambda = \min_{\boldsymbol{\theta}_1} \ \mathcal{VR}_2 , \tag{4.2}$$

be the panel AR test statistic, which is the same as the minimum eigenvalue in Section 3.5.1 by (4.2). Put $n = N(T-2)$, and the following holds under the null hypothesis.

Theorem 3.1 : *Suppose assumptions (A1), (A2), and $K_2 > G_2$ hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$n\lambda \ \xrightarrow{d} \ \chi^2_{K_2-G_2} ,$$

under null hypothesis (4.1) .

If the conditions are not satisfied, then the estimation becomes inconsistent. Therefore, this test is crucial for the structural models. The AR statistic follows a chi-square distribution because the D-LIML estimation does not have a noncentricity parameter even in long panel data. Similar to the overidentification tests in cross-sectional analyses, the degree of freedom is equal to that of the overidentification in period t, that is, $K-(G_2+K_1)=K_2-G_2$. Hence, this AR test can be called the overidentification test of all instruments. In the empirical analysis, not rejecting the null hypothesis at 5% or 1% significance levels is desirable.

4.1.1 Testing Exogeneity

In the structural panel model, we suspect $\mathbf{y}_{it}^{(2)}$ as the endogeous variables on the right-hand side, or not all variables may be endogenous. Thus, whether it is endogenous or not is interesting, which is known as the Wu-Hausman specification test and is a special case of the overidentification test (cf. Wooldridge (2002)).

Example 3.1 : Let $y_{it}^{(1)}$ and $y_{it}^{(21)}$ be growth rates of GDP and government expenditure, respectively,

$$y_{it}^{(1)} = \alpha_i + \boldsymbol{\beta}_2'\mathbf{y}_{it}^{(22)} + (\gamma_{21}y_{it}^{(21)} + \boldsymbol{\gamma}_1'\mathbf{z}_{it}^{(1)}) + u_{it} \ . \tag{4.3}$$

If an original budget is implemented as it is, then $y_{it}^{(21)}$ should be an exogenous variable in period t. Meanwhile, if a supplementary budget is passed in period t, then $y_{it}^{(21)}$ is determined simultaneously with the GDP growth rate; that is, it becomes an endogenous variable. We are interested in whether it is statistically endogenous.

Similar to (4.3) in this example, we check whether some $\mathbf{y}_{it}^{(21)}$ of $\mathbf{y}_{it}^{(2)}$ is endogenous variable,

$$\begin{pmatrix} \mathbf{y}_{it}^{(2)} \\ \mathbf{z}_{it}^{(1)} \\ \mathbf{z}_{it}^{(2)} \end{pmatrix} \begin{matrix} \}G_2 \\ \}K_1 \\ \}K_2 \end{matrix} \ , \qquad \begin{pmatrix} \mathbf{y}_{it}^{(22)} \\ \mathbf{y}_{it}^{(21)} \\ \mathbf{z}_{it}^{(1)} \\ \mathbf{z}_{it}^{(2)} \end{pmatrix} \begin{matrix} \}G_{22} \\ \}G_{21} \\ \}K_1 \\ \}K_2 \end{matrix} \ .$$

where $G_2 = G_{22}+G_{21}$. Assuming that all instruments $\mathbf{z}_{it}$ satisfy the orthogonality condition and that $\mathbf{y}_{it}^{(21)}$ is an exogenous variable, we consider the null hypothesis:

$$\mathrm{H}_0 : \ \mathcal{E}\left[\mathbf{y}_{it}^{(21)}u_{it}\right] = \mathbf{0} \ . \tag{4.4}$$

Then, the G_{21} variables appear on the G_{22} reduced forms, that is, $\mathbf{y}_{it}^{(21)}$ is included in the projection matrix of instrumental variables. Moreover, $\mathbf{y}_{it}^{(21)}$ appears in the first structural equation, and it is treated in the same way as the K_1 variables; that is, the number of parameters $\boldsymbol{\theta}_1$ to be estimated is still $G_2 + K_1 = G_{22} + G_{21} + K_1$, whereas the number of instrumental variables usually increases from K to $G_{21} + K$. The case of the K_2 variables decrease is illustrated in the section of numerical experiments.

$\mathbf{Z}_1^{(b)} = (\mathbf{Y}^{(21,b)}, \mathbf{Z}^{(b)})$ is the $n \times (G_{21} + K)$ instrumental variable matrix when $\mathbf{y}_{it}^{(21)}$ is added, and

$$\lambda_1 = \min_{\boldsymbol{\theta}_1} \ \mathcal{VR}_2 \tag{4.5}$$

is the AR test statistic generated using $\mathbf{Z}_1^{(b)}$.

Hayashi (2000) discussed the relation between the overidentification test and the Wu-Hausman test statistic expressed in the form of the difference, $n\lambda_1 - n\lambda$. When considering the form of the difference, we suppose that the denominator $\hat{\sigma}^2$ of λ in (4.2) is replaced with $\bar{\sigma}^2$ of λ_1 in (4.5) for convenience.

Under the null hypotheses, the following holds as an exogeneity test.

Theorem 3.2 : *Supposing assumptions (A1), (A2), and $K_2 > G_{22}$ hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

$$n\lambda_1 \ \xrightarrow{d} \ \chi^2_{K_2 - G_{22}} \ , \tag{4.6}$$

and

$$n\lambda_1 - n\lambda \ \xrightarrow{d} \ \chi^2_{G_{21}} \ , \tag{4.7}$$

under null hypotheses (4.1) and (4.4) .

Assuming that the exogeneity of $\mathbf{z}_{it}$ is guaranteed by the overidentification test in the first step, then, the exogeneity tests in (4.6) or (4.7) should be performed as the second step. If the chi-square statistic becomes large because of the additional $\mathbf{y}_{it}^{(21)}$, then null hypothesis (4.4) would be rejected, or $\mathbf{y}_{it}^{(21)}$ is regarded as the endogenous variable.

4.1.2 Many Instruments for Just Identified Case

In the just identification case, the structural model cannot be generally tested, because the degree of overidentification becomes $K_2 - G_2 = 0$. Hence, the chi-square distribution degenerates on the origin. However, if we start with a simple structural model in empirical analyses, then $K_2 - G_2 = 1 - 1$ can occur.

In another context, Hayakawa (2014) considered the overidentification test using the diagonalized projection matrix. Then, we notice that the degree of overidentification increases. Moreover, Lee and Okui (2012) and Anatolyev (2013) investigated the overidentification tests under many instruments. In their argument, the chi-square test statistic based on the GMM estimator diverges. However, the panel AR test statistic based on the D-LIML estimator in Theorem 2.3 is robust even in many instruments;

$$\tilde{\lambda} = \min_{\boldsymbol{\theta}_1} \ \mathcal{VR}_1 \ .$$

This estimator is also made by the sequential moment conditions as described in Section 3.5.1, that is, the projection matrix is diagonalized. In the case of overidentification, the test of Theorem 3.1 based on the efficient D-LIML estimation is desirable. However, in the case of just identification, the D-LIML estimator of Theorem 2.3 can be applied. The statistic under many instruments becomes the sum of many chi-square distributions, so that it converges to a normal distribution. In the dynamic panel, the case when $c_1 = K/N$ converges to zero is also important. However, if regarded as $c_1 = 0$, then the expression of $\sqrt{n}\tilde{\lambda}$ degenerates. Therefore, we use another expression of normalization in the following theorem.

Under null hypothesis (4.1), t-tests can be constructed by the panel AR test statistic even in the case of $K_2 = G_2$.

Theorem 3.3 : *Supposing assumptions (A1), (A2), and $K_2 \geq G_2$ hold, then as $T \to \infty$,*

[i] *N is fixed or $c_1 > 0$. If $\mathbf{v}_{it}^*$ follows a normal distribution, then*

$$t_c = \frac{\sqrt{n}(\tilde{\lambda} - c_1)}{\sqrt{2c_{1*}}} \ \xrightarrow{d} \ \mathcal{N}(0,1) \ ,$$

where $c_1 = K/N$ and $c_{1*} = c_1/(1 - c_1)$.

[ii] *N tends to infinity or $c_1 = 0$. If $N/T \to 0 \leq d < \infty$, then*

$$t_0 = \frac{n\tilde{\lambda} - d_T}{\sqrt{2d_T}} \ \xrightarrow{d} \ \mathcal{N}(0,1) \ ,$$

where $d_T = KT - (G_2 + K_1)$.

Similar to the chi-square test, the one-sided test should be conducted in the right tail area. As shown in the numerical experiments, t_c and t_0 are numerically almost the same value, so t_0 is recommended. For the result of [ii], $-(G_2 + K_1)$ is not necessary for the asymptotics in $KT \to \infty$. However, as shown in the numerical experiments, this finite sample correction, which is based on the degree of freedom in the case of $KT < \infty$, has a considerable effect.

4.2 Simulation

We check the finite sample properties from Theorems 3.1 to 3.3 with the following settings.

Design 3.1 : In the case of Theorem 3.1,

$$\begin{aligned} y_{it}^{(1)} &= \beta_{21} y_{it}^{(2)} + \beta_{22} y_{it}^{(3)} + \gamma_{11} y_{it-1}^{(1)} + \alpha_i + u_{it} \ , \\ y_{it}^{(2)} &= \pi_{21} y_{it-1}^{(2)} + \pi_{22} y_{it-2}^{(2)} + \pi_i^{(2)} + v_{it}^{(2)} \ , \\ y_{it}^{(3)} &= \pi_{31} y_{it-1}^{(3)} + \pi_{32} y_{it-2}^{(3)} + \pi_i^{(3)} + v_{it}^{(3)} \ , \end{aligned}$$

where the numerical setting is the same as that in Design 2.3. However, $x_{it} = y_{it}^{(3)}$ is considered an endogenous variable in period t, and then, $G_2 = 2$. Examine the weak exogeniety of the $K = 5$ instrumental variables $(y_{it-1}^{(1,b)},\ y_{it-1}^{(2,b)},\ y_{it-2}^{(2,b)},\ y_{it-1}^{(3,b)},\ y_{it-2}^{(3,b)})$. The degree of overidentification becomes $K_2 - G_2 = 4 - 2$. Figure 8 shows the empirical cumulative distribution of $n\lambda$, which follows the chi-square distribution with two degrees of freedom.

The second step is the exogeneity test for $x_{it} = y_{it}^{(3)}$ using the $n\lambda_1$ of (4.6). x_{it} is an exogenous variable in period t,

$$y_{it}^{(1)} = \beta_{21} y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} x_{it} + \alpha_i + u_{it} \ ,$$

From the reduced form of $(y_{it}^{(1)},\ y_{it}^{(2)})$, the degree of overidentification $K_2 - G_{22} = 2 - 1$ is obtained by the $K = 4$ instrumental variables $(y_{it-1}^{(1,b)},\ y_{it-1}^{(2,b)},\ y_{it-2}^{(2,b)},\ x_{it}^{(b)})$.

If we know that $(y_{it-1}^{(3)},\ y_{it-2}^{(3)})$ does not appear in reduced form, the K_2 variables are reduced. From Figure 9, the statistic certainly follows the chi-square distribution with one degree of freedom

Alternatively, by using the instrumental variables $(y_{it-1}^{(1,b)},\ y_{it-1}^{(2,b)},\ y_{it-2}^{(2,b)},\ x_{it}^{(b)},\ y_{it-1}^{(3,b)},\ y_{it-2}^{(3,b)})$ in which x_{it} is simply added, null hypothesis can be tested by (4.6) where $K_2 - G_{22} = 4 - 1$.

Design 3.2 : Let us examine the finite sample properties of Theorem 3.3 using the just identification model:

$$\begin{aligned} y_{it}^{(1)} &= \beta_2 y_{it}^{(2)} + \gamma_1 y_{it-1}^{(1)} + \alpha_i + u_{it} \ , \\ y_{it}^{(2)} &= \gamma_2 y_{it-1}^{(2)} + \rho\alpha_i + v_{it} \ , \end{aligned}$$

where $G_2 - K_2 = 1 - 1 = 0$, and the numerical setting is the same as that in Design 2.1.

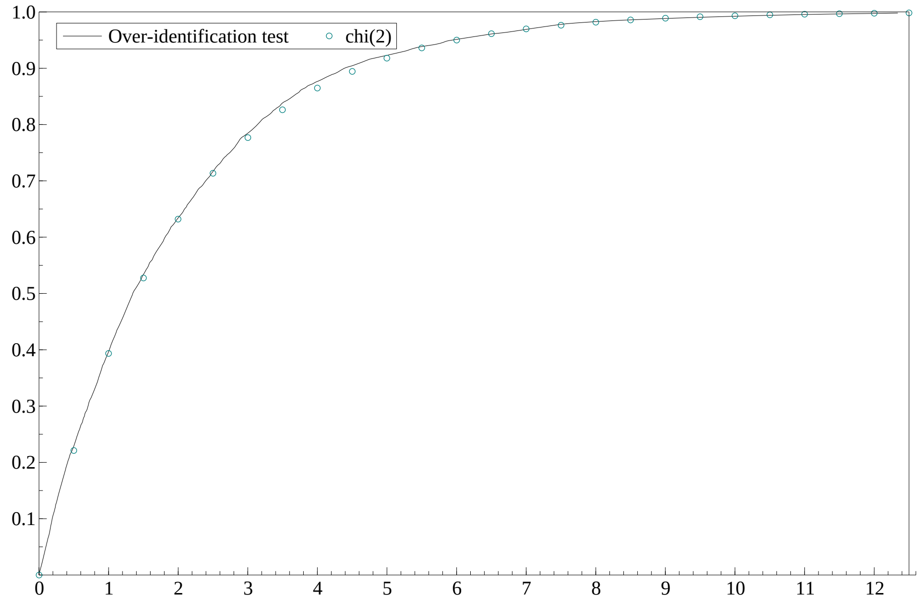

Fig. 8: Design 3.1 ($N = 100, T = 25$)

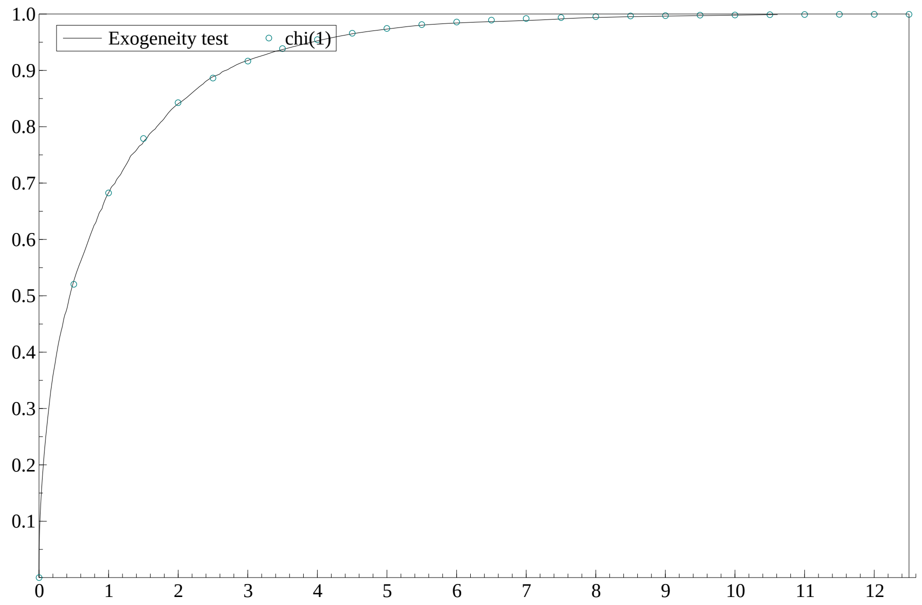

Fig. 9: Design 3.1 ($N = 100, T = 25$)

Notably, t_c is numerically proportional to t_0 without finite sample correction of $d_T = K(T-2)$,

$$t_c = t_0 \times \left(\frac{1}{1 - K/N} \right)^{-\frac{1}{2}} .$$

Therefore, when N is large, the empirical distributions are almost the same and overlap in Figure 10. When t_0 is applied the finite sample correction (t_0 adjusted in the figure),

$$\begin{aligned} d_T &= K(T-2) - (G_2 + K_1) \\ &= 2(T-2) - 2 , \end{aligned}$$

it is quite effective, and the approximation to the standard normal distribution is more accurate. When making this finite sample correction to t_c, $c_1 = K/N$ should be replaced with $[K(T-2) - (G_2 + K_1)]/n$, but the result is omitted because the empirical distribution almost overlaps with t_0 adjusted.

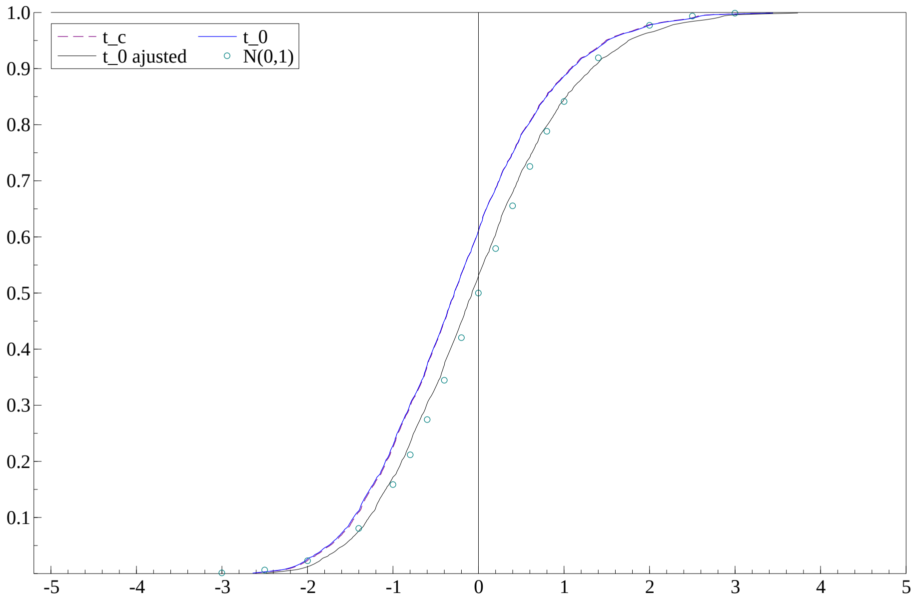

Fig. 10: Design 3.2 ($N = 100, T = 25$)

4.3 Panel Information Criterion for the Reduced Form

Under the limited information method, the structural equaiton is formulated a priori from an economic theory,

$$y_{it}^{(1)} \quad = \quad \boldsymbol{\beta}_2' \mathbf{y}_{it}^{(2)} + \boldsymbol{\gamma}_1' \mathbf{z}_{it}^{(1)} + \alpha_i + u_{it} ,$$

but determining the K_2 variables that do not appear in the first structural equation is difficult. The problem can be reduced to the order selection for a panel VAR model, which is the reduced form of $\mathbf{y}_{it} = (y_{it}^{(1)},\ \mathbf{y}_{it}^{(2)'})'$. The implications of the selection for the reduced form are the following. First, we are interested in what are the true instrumental variables. Second, the number of instrumental variables affects the efficiency of the estimation for the structural parameter $\boldsymbol{\theta}_1$. In general, if few instrumental variables exist, then the efficiency is reduced. From the discussion of large-K asymptotics, adding redundant instrumental variables may also cause a large varicance.

Andrews and Lu (2001) considered the selection criteria that simultaneously conduct the overidentification test and model selection in a dynamic panel model. Morimune and Sawa (1980) considered the selection rule for the structural equation that specifies the correct model with a high probability using an F distribution. In the following, we show the consistency of model selection based on the information criterion developed by Akaike (1974). However, as the exact likelihood function in fixed-effects estimation or the transformed maximum likelihood function is complicated, we consider using the objective function of the D-LIML estimator.

To make the informaton critria, we first confirm multiple true espressions of a reduced form.

Example 3.2 : Let $x_{it} = \pi_3 x_{it-1} + v_{it}^{(3)}$ be the exogenous variable in period t. Then, the reduced form has the following two expressions, which are observationally equivalent,

$$y_{it}^{(g)} = \pi_{1g} y_{it-1}^{(1)} + \pi_{2g} x_{it} + \pi_i^{(g)} + v_{it}^{(g)} \tag{4.8}$$
$$= \pi_{1g} y_{it-1}^{(1)} + \pi_{3g} x_{it-1} + \pi_i^{(g)} + (v_{it}^{(g)} + v_{it}^{(3)})\ ,\ (g = 1,\ 2)\ . \tag{4.9}$$

As $\mathcal{V}ar[v_{it}^{(g)}] < \mathcal{V}ar[v_{it}^{(g)} + v_{it}^{(3)}]$, (4.9) is inferior in the explanatory power, so we should select (4.8). Thus, we can consider selecting the expression with the smallest $\mathrm{tr}(\boldsymbol{\Omega})$ among the true expressions, where $\boldsymbol{\Omega}$ is the variance-covariance matrix of the error terms.

For the reduced form consisting of G endogeous variables, let

$$\mathbf{z}_{it} = \{z_{it}^{[1]},\ z_{it}^{[2]},\ \cdots,\ z_{it}^{[K]}\}\ ,$$

be the list of the instrumental variables that construct the true reduced form. The conditions for the true representation are given as follows:

$$\sum_{k=1}^{K} |\pi_{gk}| \neq 0\ ,\quad (g = 1, \cdots, G)\ .$$

That is, an instrumental variable $z_{it}^{[k]}$ should be included in at least one of the G reduced forms, and then, its coefficient is not 0.

Now we consider the difference from the usual order selection of the VAR(p) model. In the case of the VAR model, the common order p can be selected for all the G-dimensional variables. However, the reduced form is not necessarily in the form of VAR(p) as shown in Example 2.5. Selecting a different order for each variable is more practical, and the models are then non-nested, that is, $K \neq G \times p$ is allowed. If a candidate of $K_{\{1\}}$ instrumental variables is represented as $\mathbf{z}_{it}^{\{1\}} (\neq \mathbf{z}_{it})$, then the cases can be divided into exclusion:

$$\mathbf{z}_{it} \subset \mathbf{z}_{it}^{\{1\}} \ , \quad \mathbf{z}_{it} \not\subset \mathbf{z}_{it}^{\{1\}} \ .$$

Notably, even if variables that are included in $\mathbf{z}_{it}$ and not included in $\mathbf{z}_{it}^{\{1\}}$ rxist, that is, the omitted variables, $K_{\{1\}} \geq K$ is possible.

Following Schwarz (1978), we set the penalty term as the order of $\log n$. K is the number of instrumental variables of a candidate, and then, a panel information criteria (PIC) is given as follows:

$$\begin{aligned} \text{PIC}_1 &= \sum_{g=1}^{G} \left(\hat{\omega}_{gg} + K \frac{\log n}{n} \right) \\ &= \text{tr}(\hat{\mathbf{\Omega}}) + GK \frac{\log n}{n} \ . \end{aligned}$$

To select the true model without being affected by individual effects, we estimate $\mathbf{\Omega}$ from the residuals of the IV estimator of Theorem 1.5. Each candidate can be easily estimated as follows:

$$\begin{aligned} \hat{\mathbf{\Omega}} &= \frac{1}{n} \mathbf{Y}^{(f)\prime} \mathbf{Q}' \mathbf{Q} \mathbf{Y}^{(f)} \ , \\ \mathbf{Q} &= \mathbf{I}_n - \mathbf{Z}^{(f)} (\mathbf{Z}^{(b)\prime} \mathbf{Z}^{(f)})^{-1} \mathbf{Z}^{(b)\prime} \ , \end{aligned}$$

where $n = N(T-2)$ and the IV estimator is $\hat{\mathbf{\Pi}}_{\text{IV}} = (\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(b)\prime}\mathbf{Y}^{(f)}$. Then, $\mathbf{Q}$ remains idempotent but becomes an asymmetric matrix. For instance, $\mathbf{Z}^{(f)}$ and $\mathbf{Z}^{(b)}$ become $n \times K$ matrices if generated by $\mathbf{z}_{it}$, and $n \times K_{\{1\}}$ if generated by $\mathbf{z}_{it}^{\{1\}}$. In the case of regression analysis we have $G = 1$, and $G \geq 2$ corresponds to a reduced form of the structural analysis.

(A5) [i] All candidates of instrumental variables are a subset of $\mathbf{z}_{it}^{*}$ generated by (3.10). The rank of $n \times K^{*}$ matrix $\mathbf{W} = (\mathbf{w}_{it-1})$ is K^{*}.

[ii] The rank of $\mathbf{\Pi}$ is G .

Assumption [i] means that we do not search for candidates in variables that are multicollinearity, where the constant term must not be included because of the double filters. Assumption [ii] may be interpreted as not being the multicollinearity among the endogenous variables.

From the following results, $\mathrm{PIC}_{1,0}$ based on the true instrumental variables $\mathbf{z}_{it}$ is asymptotically minimized. Therefore, we search for the list of instrumental variables that minimize PIC_1.

Theorem 3.4 : *Supposing assumptions (A1), (A2), and (A5)-[i] hold, then as* $T \to \infty$*, regardless of* N *is fixed or tends to infinity, provided* $0 \le d < \infty$*,*

$$\Pr(\ \mathrm{PIC}_{1,0} < \mathrm{PIC}_1\ )\ \ \overset{p}{\longrightarrow}\ \ 1\ ,$$

for any PIC_1 *based on* $\mathbf{z}_{it}^{\{1\}} \neq \mathbf{z}_{it}$.

Moreover, we can consider the log-likelihood function of the T-LIML estimator as an information criteria. From (6.37) of Lemma 2.2, the pseudo log-likelihood function that approximates the transformation likelihood method has the following relation:

$$\mathcal{L}_{2\cdot 0} \propto -\log((1+\lambda)|\hat{\mathbf{\Omega}}|)\ . \tag{4.10}$$

Minimizing the eigenvalue λ is used for the specification of the structural equation in the next section. Hence, depending only on the term of generalized variance $|\mathbf{\Omega}|$ in the first step is appropriate for the reduced form. Thus, another PIC for the reduced form is as follows:

$$\mathrm{PIC}_2 = \log(|\hat{\mathbf{\Omega}}|) + GK\frac{\log n}{n}\ ,$$

that is, it is almost the same as the Schwarz information criterion for the usual VAR model.

For $\mathrm{PIC}_{2,0}$ based on the true instrumental variable $\mathbf{z}_{it}$, the same result as PIC_1 holds.

Theorem 3.5 : *Supposing assumptions (A1), (A2), and (A5) hold, then as* $T \to \infty$*, regardless of* N *is fixed or tends to infinity, provided* $0 \le d < \infty$*,*

$$\Pr(\ \mathrm{PIC}_{2,0} < \mathrm{PIC}_2\ )\ \ \overset{p}{\longrightarrow}\ \ 1\ .$$

for any PIC_2 *on the basis of* $\mathbf{z}_{it}^{\{1\}} \neq \mathbf{z}_{it}$.

From the rank condition $\mathrm{rank}(\mathbf{\Pi}_{2\cdot}) = G_2 < G$ in the next section, assumption (A5)-[ii] implicitly requires $K_1 \ge 1$.

The theorems implicitly assume that $\mathbf{z}_{it}$ satisfies the orthogonality condition. If a candidate mistakenly contains an endogenous variable, then the consistency of model selection is generally not guaranteed. Thus, the overidentification test in the previous section is important. However, the inverse problem that not all instrumental variables are explanatory just because they satisfy the orthogonality condition exists. Therefore, performing the overidentification test and the selection of the reduced form would be better.

4.4 Simulation

We suppose the first structural equation of $G = 2$ under the limited information method as follows:

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_{11} y_{it-1}^{(1)} + \gamma_{12} y_{it-1}^{(2)} + \alpha_i + u_{it} .$$

We assume that the following second reduced form and the form of the exogenous variable are unknown to an econometrician,

$$\begin{aligned} y_{it}^{(2)} &= \pi_{21} y_{it-1}^{(2)} + \pi_{22} y_{it-2}^{(2)} + \pi_{23} x_{it} + \pi_i^{(2)} + v_{it}^{(2)} , \\ x_{it} &= \pi_{31} x_{it-1} + \pi_{32} x_{it-2} + + \pi_i^{(3)} + v_{it}^{(3)} , \end{aligned}$$

where $\beta_2 = \gamma_{11} = 0.5$, $\gamma_{12} = \pi_{21} = \pi_{23} = \pi_{31} = 0.3$, $\pi_{22} = 0.2$, $\pi_{32} = 0.1$, $\omega_{gg} = 1$ $(g = 1,\ 2,\ 3)$, and $\omega_{12} = 0.3$. The individual effect is $\boldsymbol{\pi}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$. We consider the case when the specified K_1 variables $(y_{it-1}^{(1)},\ y_{it-1}^{(2)})$ are included in all candidates.

The next model-0 is the true reduced form and the number of instrumental variables is $K = 4$, that is, two K_2 variables exist, $(y_{it-2}^{(2)},\ x_{it})$. The reduced forms $(g = 1,\ 2)$ are as follows:

$$y_{it}^{(g)} = \pi_{g1} y_{it-1}^{(1)} + \pi_{g2} y_{it-1}^{(2)} + \pi_{g3} y_{it-2}^{(2)} + \pi_{g4} x_{it} + \pi_i^{(g)} + v_{it}^{(g)} .$$

The following five models are the incorrectly specified reduced form, and each of them has a different candidate $\mathbf{z}_{it}^{\{1\}}$.

Model-1 is a smaller model than model-0, and the omitted variable is $y_{it-2}^{(2)}$ $(K_{\{1\}} = 3)$:

$$y_{it}^{(g)} = \pi_{g1} y_{it-1}^{(1)} + \pi_{g2} y_{it-1}^{(2)} + \pi_{g3} x_{it} + \pi_i^{(g)} + v_{it}^{(g)} .$$

Model-2 is also a smaller model with the omitted variable $y_{it-2}^{(2)}$. However, the irrelevant variable $y_{it-2}^{(1)}$ $(K_{\{1\}} = 4)$ exists:

$$y_{it}^{(g)} = \pi_{g1} y_{it-1}^{(1)} + \pi_{g2} y_{it-2}^{(1)} + \pi_{g3} y_{it-1}^{(2)} + \pi_{g4} x_{it} + \pi_i^{(g)} + v_{it}^{(g)} .$$

Model-3 is another expression of the true reduced form:

$$y_{it}^{(g)} = \pi_{g1}y_{it-1}^{(1)} + \pi_{g2}y_{it-1}^{(2)} + \pi_{g3}y_{it-2}^{(2)} + \pi_{g4}x_{it-1} + \pi_{g5}x_{it-2} + \pi_i^{(g)} + v_{it}^{(g)} ,$$

where x_{it} is substituted ($K_{\{1\}} = 5$).

Model-4 is a larger model than model-0:

$$y_{it}^{(g)} = \pi_{g1}y_{it-1}^{(1)} + \pi_{g1}y_{it-2}^{(1)} + \pi_{g3}y_{it-1}^{(2)} + \pi_{g4}y_{it-2}^{(2)} + \pi_{g5}x_{it} + \pi_i^{(g)} + v_{it}^{(g)} ,$$

where $y_{it-2}^{(1)}$ is added ($K_{\{1\}} = 5$).

Model-5 is also a larger model:

$$y_{it}^{(g)} = \pi_{g1}y_{it-1}^{(1)} + \pi_{g1}y_{it-2}^{(1)} + \pi_{g3}y_{it-1}^{(2)} + \pi_{g4}y_{it-2}^{(2)} + \pi_{g5}y_{it-3}^{(2)} + \pi_{g6}x_{it} + \pi_i^{(g)} + v_{it}^{(g)} ,$$

where $y_{it-2}^{(1)}$ and $y_{it-3}^{(2)}$ are added ($K_{\{1\}} = 6$).

Table 1 summarizes the ratio that a model has the minimum value of information criterion, where $N = 50$, and the number of repetitions $R = 3000$ times. PIC_1

Table 1: Percentages of model selection

		model-0	model-1	model-2	model-3	model-4	model-5
$T = 15$	PIC_1	69.9	4.47	0.07	0.00	22.7	2.87
	PIC_2	66.8	11.3	0.27	0.03	17.8	3.80
$T = 25$	PIC_1	87.4	0.00	0.00	0.00	12.2	0.40
	PIC_2	89.9	0.00	0.00	0.00	9.50	0.57
$T = 50$	PIC_1	97.4	0.00	0.00	0.00	2.60	0.00
	PIC_2	98.9	0.00	0.00	0.00	1.01	0.07

and PIC_2 show similar properties, and as T increases, the selection rate of model-0 increases. Thus, the consistency of model selection can be confirmed. The model-3 is another true expression of the reduced form, but as discussed in Example 3.2, this model is the most difficult to be selected. There is a slight possibility that a larger model would be selected. However, in the larger model, the coefficients of irrelevant variables are consistently estimated to be zero, and the effect on efficiency would be limited than selecting a smaller model. Therefore, choosing a slightly larger model would not be too much of a problem.

4.5 Specification Test for a Structural Equation

In the last section, we discuss the specification of the structural model. The identification problem for the structural parameter $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_2', \ \boldsymbol{\gamma}_1')'$ should be discussed first because it is the condition for consistent estimations and relating hypotheses testing. However, this case is not often discussed in textbooks recently and may be difficult to understand, so we consider it last. The panel structural equations in the linear simultaneous equation model are given as follows:

$$\underset{G\times G}{\mathbf{B}}\mathbf{y}_{it} = \underset{G\times K}{\mathbf{\Gamma}}\mathbf{z}_{it} + \underset{G\times 1}{\boldsymbol{\alpha}_i} + \mathbf{u}_{it} \ . \tag{4.11}$$

When represented as a linear model in this way, an economic model has many expressions in by multiplying it by an arbitrary regular $G \times G$ matrix $\mathbf{T}$ on the left. However, the coefficient of the reduced form $\mathbf{\Pi}^{'} = \mathbf{B}^{-1}\mathbf{T}^{-1}\mathbf{T}\mathbf{\Gamma}$ is uniquely determined as the solution, which can be estimated by the moments of data. Therefore, the question is whether one can select a meaningful structural expression from the reduced form. A condition familiar with the economic theory would be the traditional zero constraints (exclusion condition) since the Cowles Commission (cf. Hsiao and Zhou (2015)).

Example 3.3 : Reconsider the production function in Example. 2.1. In the case of $G = 3$,

$$y_{it}^{(1)} = \beta_{21}y_{it}^{(2)} + \beta_{22}y_{it}^{(3)} + \alpha_i + u_{it} \ ,$$

where the endogenous variables are the logarithmic values of production, labor, and capital. Usually, only the endogenous variables appear in the production function so that the zero constraints are satisfied.

The following example is regarding the demand and supply functions,

$$\begin{aligned} y_{it}^{(1)} &= \beta_{21}y_{it}^{(2)} + \gamma_1 z_{it}^{(1)} + \alpha_i^{(1)} + u_{it}^{(1)} \ , \\ y_{it}^{(1)} &= \beta_{22}y_{it}^{(2)} + \gamma_2 z_{it}^{(2)} + \alpha_i^{(2)} + u_{it}^{(2)} \ , \end{aligned}$$

$y_{it}^{(1)}$ and $y_{it}^{(2)}$ are the quantity and price of a good in some region i, respectively. Which equation is the demand function is determined by the zero constraints. If $z_{it}^{(1)}$ is the consumption tax rate and $z_{it}^{(2)}$ is the corporate tax rate, then $z_{it}^{(2)}$ does not appear in the first equation. Therefore, the first equation can be regarded as the demand function. Meanwhile, the equilibrium does not change even if multiplied by a certain transformation $\mathbf{T}$. However, a coefficient $\beta^* = \beta_{21}+\beta_{22}$ of transformed expression, which does not satisfy a zero constraint, has no economic meaning.

In this work, we call the dynamic structural panel model and do not refer to the structural panel VAR model because the former uses zero constraints. Another constraint is supposed in the structural VAR model, and there is also the position that zero constraints are ad hoc (cf. Amisano and Giannini (1996)). Therefore, examining the zero constraint using the data is desirable; that is, a specification for the structural model should be tested. In the linear model, the identification condition is reduced to the simple rank condition.

First, we confirm the three concepts of normalization, zero constraints, and identification. Notably, the normalization and zero constraints are neither necessary nor sufficient conditions for identification. Generally, the formulation of the limited information method can be expressed as follows:

$$\begin{aligned} \beta_1 y_{it}^{(1)} &= \boldsymbol{\beta}_2^{'} \mathbf{y}_{it}^{(2)} + \boldsymbol{\gamma}_1^{'} \mathbf{z}_{it}^{(1)} + \boldsymbol{\gamma}_2^{'} \mathbf{z}_{it}^{(2)} + \alpha_i + u_{it} \,, \\ \mathbf{y}_{it}^{(2)} &= \boldsymbol{\Pi}_{12}^{'} \mathbf{z}_{it}^{(1)} + \boldsymbol{\Pi}_{22}^{'} \mathbf{z}_{it}^{(2)} + \boldsymbol{\pi}_i^{(2)} + \mathbf{v}_{it}^{(2)} \,. \end{aligned} \tag{4.12}$$

This method has an advantage that only identifying the parameter of the first structural equation needs to be considered. In the first structural equation (4.12), $y_{it}^{(1)}$ has a coefficient β_1, and thus, the scale of each coefficient is not determined so that some normalization for the coefficients is necessary.

Example 3.4 : Express (3.1) of the utility function in Example 2.3 as follows:

$$\beta_1^* y_{it}^{(1)} = \beta_2^* y_{it}^{(2)} + \gamma_1^* z_t^{(1)} + \gamma_2^* z_t^{(2)} + u_{it}^{(1)*} \,.$$

If we divide both sides by β_1^*, then

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_1 z_t^{(1)} + \gamma_2 z_t^{(2)} + u_{it}^{(1)} \,, \tag{4.13}$$

where $\beta_2 = \beta_2^*/\beta_1^*$ and $u_{it}^{(1)} = u_{it}^{(1)*}/\beta_1^*$. However, if $\beta_1^* = 0$, then the expression of (4.13) cannot exist. When the first good is not purchased, it is possible that $\beta_2^* = 0$.

As $\sigma^2 > 0$, the natural normalization is known that $(\boldsymbol{\beta}/\sigma)^{'}\boldsymbol{\Omega}(\boldsymbol{\beta}/\sigma) = 1$ (cf. Anderson and Rubin (1949)). Anderson and Kunitomo (1992, 1994) provided a general discussion of the overidentification test and identification based on the natural normalization in detail. However, in this work, we have derived the estimators by the conventional normalization, which is often used in the applied analysis (cf. Amemiya (1985)). Let us suppose the conventional normalization such that $\beta_1 \neq 0$ in (4.12); that is, the following first element is 1,

$$\boldsymbol{\beta}^{'} = (1,\ -\boldsymbol{\beta}_2^{'}) \,. \tag{4.14}$$

Second, we consider the zero constraints on the coefficients of the exogenous variables in the first structural equation (4.12). Let us confirm that the zero constraints are not unique and a candidate of the structural expressions.

Example 3.5 : Give the following structural equation a zero constraint,

$$y_{it}^{(1)} = \beta_2 y_{it}^{(2)} + \gamma_1 z_{it}^{(1)} + \gamma_2 z_{it}^{(2)} + \alpha_i + u_{it} \ ,$$

where the reduced form is as follows:

$$\mathbf{y}_{it} = \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix} \mathbf{z}_{it} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \ .$$

From the discussion of Lemma 3.1 below, we have $\beta_2 = -0.5$ and $\gamma_1 = 2.5$ in the case of the zero constraint $\gamma_2 = 0$. If another zero constraint is $\gamma_1 = 0$, then $\beta_2 = 2$ and $\gamma_2 = -3$. When two structural equations exist as in Example 3.3, the different zero constraints correspond to the change of sign of β_2 with the demand and supply functions.

Finally, we consider the identification of the parameters in the first structural equation. The zero constraints are expressed in the vector as follows:

$$\mathrm{H}_0 : \ \boldsymbol{\gamma}_2 = \mathbf{0} \ ,$$

and an econometrician specifies them by (4.12). That is, suppose that $\mathbf{z}_{it}^{(2)}$ of K_2 variables that does not appear in the first structural equation exists. We check the following notation for the rank condition,

$$\begin{aligned} \mathbf{y}_{it} &= \mathbf{\Pi}' \mathbf{z}_{it} + \boldsymbol{\pi}_i + \mathbf{v}_{it} \ , \\ \mathbf{\Pi}' &= \begin{pmatrix} \mathbf{\Pi}'_{1.} \ \mathbf{\Pi}'_{2.} \end{pmatrix} = \underset{(1+G_2)\times(K_1+K_2)}{\begin{pmatrix} \boldsymbol{\pi}'_{11} & \boldsymbol{\pi}'_{21} \\ \mathbf{\Pi}'_{12} & \mathbf{\Pi}'_{22} \end{pmatrix}} \ . \end{aligned}$$

Hsiao (1983) investigated the identification of the dynamic structural model based on the likelihood function. In a panel model, the question is whether the individual effect does not affect the identification of the structural parameters. Bhargava (1991) examined the identification condition for the dynamic panel structural model under $N \to \infty$.

The next lemma is derived under the conventional normalization (4.14) and indicates that the condition of identification is determined independently of individual effects under the long panel data.

Lemma 3.1 : *Suppose (A1) and (A2). If* $\mathcal{E}[\boldsymbol{\pi}_i^* \mathbf{v}_{it}^{*'}] = \mathbf{O}$ *or* $T \to \infty$ *in the case of* $\mathcal{E}[\boldsymbol{\pi}_i^* \mathbf{v}_{it}^{*'}] \neq \mathbf{O}$*, then*

[i] *The following expressions are equivalent:*

$$\boldsymbol{\gamma}_2 = \mathbf{0} \;\Leftrightarrow\; \boldsymbol{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}\ .$$

[ii] *For some $\boldsymbol{\gamma}_2 = \mathbf{0}$, the necessary and sufficient condition that $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_2^{'},\ \boldsymbol{\gamma}'_1)^{'}$ is uniquely determined from $\boldsymbol{\Pi}$ is as follows:*

$$\mathrm{rank}(\boldsymbol{\Pi}_{2\cdot}) = \mathrm{rank}(\boldsymbol{\Pi}_{22})\ , \quad \mathrm{rank}(\boldsymbol{\Pi}_{22}) = G_2\ .$$

In other words, the usual rank condition for identification is obtained even in the dynamic structural panel model. In the previous sections, we introduced the estimators using many instruments, but those instrumental variables do not contribute to the identification; that is, the K_2 variables in each period t are important.

If $\boldsymbol{\beta}$ satisfies $\boldsymbol{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}$, then the zero constraint is correct. This condition means $\mathrm{rank}(\boldsymbol{\Pi}_{2\cdot}) = \mathrm{rank}(\boldsymbol{\Pi}_{22})$. Then, we could suggest that the reduced form is reduced from the structural equations that satisfy the zero constraint. Moreover, some zero constraints can have multiple solutions $(\boldsymbol{\beta}_{[1]},\ \boldsymbol{\beta}_{[2]},\ \cdots)$. Although the multiple expressions in the structural equation are allowed, the unique expression is desirable. Then, the condition becomes $\mathrm{rank}(\boldsymbol{\Pi}_{22}) = G_2$, which are called the rank conditions for the identification, and its necessary condition is called the order condition:

$$K_2 \geq G_2\ .$$

In empirical analyses, the discussion of identification is often completed by checking only the order condition. Notably, the rank conditions of Lemma 3.1 have been implicitly assumed in Parts II and III. However, some hypotheses can be tested from the panel data, so performing the rank test as discussed below would be desirable.

One of the advantages of the LIML method is that the identification problem is reduced to the eigenvalue problem using the objective function, where the representation (4.2) of the D-LIML estimator is slightly changed. First, we use the form that is concentrated to $\boldsymbol{\beta}_2$ instead of $\boldsymbol{\theta}_1$ such as the original concentrated log-likelihood function of Anderson and Rubin (1949) in Section 3.1.

$$\mathcal{VR}_{2\cdot 1} = \frac{\boldsymbol{\beta}^{'}\mathbf{G}_{n1}^{(f,b)}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\mathbf{H}_{n1}^{(f,b)}\boldsymbol{\beta}}\ ,$$

where $\boldsymbol{\beta} = (1,\ -\boldsymbol{\beta}_2^{'})^{'}$,

$$\begin{aligned}\mathbf{G}_{n1}^{(f,b)} &= \mathbf{Y}^{(f)'}(\mathbf{P}^{(b)} - \mathbf{P}_1^{(b)})\mathbf{Y}^{(f)}\ ,\\ \mathbf{H}_{n1}^{(f,b)} &= \frac{1}{n}\mathbf{Y}^{(f)'}(\mathbf{I}_n - \mathbf{P}^{(b)})\mathbf{Y}^{(f)}\ ,\end{aligned}$$

and $\mathbf{Y}^{(f)} = (\mathbf{y}_{it}^{(f)})$ is the $n \times G$ matrix consisting only of endogenous variables,

$$\mathbf{P}_1^{(b)} = \mathbf{Z}_1^{(b)} \left(\mathbf{Z}_1^{(b)\prime} \mathbf{Z}_1^{(b)} \right)^{-1} \mathbf{Z}_1^{(b)\prime} ,$$

where $\mathbf{Z}_1^{(b)} = (\mathbf{z}_{it}^{(1,b)})$ is the $N \times K_1$ matrix consisting of K_1 variables.

Second, $\mathbf{H}_{n1}^{(f,b)}$ is normalized by $(1/n)$ in advance, and the representation of $n\lambda.$ is changed to $\lambda.$ in the following theorem.[19]

$$\lambda_{\cdot 1} = \min_{\boldsymbol{\beta}_2} \ \mathcal{VR}_{2\cdot 1}$$

is equal to the minimum eigenvalue of the following eigenvalue equation,

$$\left| \mathbf{G}_{n1}^{(f,b)} - \ell \mathbf{H}_{n1}^{(f,b)} \right| = 0 .$$

We express the eigenvalues of this equation in ascending order as follows:

$$0 \leq \lambda_{\cdot 1} \leq \lambda_{\cdot 2} \leq \cdots .$$

If the eigenvalue is evaluated at the true value, then the eigenvector $\boldsymbol{\beta}$ corresponding to the eigenvalue 0 satisfies $\mathbf{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}$. However, if zeros of the eigenvalues overlap, then multiple solutions $(\boldsymbol{\beta}_{[1]},\ \boldsymbol{\beta}_{[2]},\ \cdots)$ exist. On the contrary, if the minimum eigenvalue is positive, then no $\boldsymbol{\beta}$ satisfies $\mathbf{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}$ or the zero constraints are incorrectly specified. Therefore, the following theorem intuitively corresponds to searching for the number of eigenvalues close to zero.

As the rank conditions are divided into two, we prepare another notation,

$$\left| \mathbf{J}_2' \mathbf{G}_{n1}^{(f,b)} \mathbf{J}_2 - \ell \mathbf{J}_2' \mathbf{H}_{n1}^{(f,b)} \mathbf{J}_2 \right| = 0 ,$$

where $\mathbf{J}_2' = (\mathbf{0},\ \mathbf{I}_{G_2})$ and express these eigenvalues as $0 \leq \lambda_{21} \leq \lambda_{22} \leq \cdots$ in ascending order.

The following result is a rank test of the dynamic structure panel model in long panel data.

Theorem 3.6 : *Supposing assumptions (A1), (A2), and $K_2 > G_2$ hold, then as $T \to \infty$, regardless of N is fixed or tends to infinity,*

[i] *Under* $\mathrm{H}_0 : \mathrm{rank}(\mathbf{\Pi}_{2\cdot}) = G_* < G$,

$$\sum_{g=1}^{G-G_*} \lambda_{\cdot g} \quad \xrightarrow{d} \quad \chi^2_{(G-G_*)(K_2-G_*)} .$$

[19] The reason is that $|\mathbf{G} - \ell(1/n)\mathbf{H}| = |\mathbf{G} - \ell^*\mathbf{H}| = 0$, where $\ell = n\ell^*$.

Under H_0 : $\mathrm{rank}(\mathbf{\Pi}_{2\cdot}) \leq G_* = G_2$, *there exists* q *such that* $\sum_{g=1}^{G-G_*} \lambda_{\cdot g} = \lambda_{\cdot 1} \leq q$ *a.s., and*

$$q \quad \xrightarrow{d} \quad \chi^2_{(G-G_*)(K_2-G_*)} = \chi^2_{K_2-G_2} \ .$$

[ii] *Under* $\mathrm{H}_0 : \mathrm{rank}(\mathbf{\Pi}_{22}) = G_{2*} < G_2$,

$$\sum_{g=1}^{G_2-G_{2*}} \lambda_{2g} \quad \xrightarrow{d} \quad \chi^2_{(G_2-G_{2*})(K_2-G_{2*})} \ .$$

Under H_0 : $\mathrm{rank}(\mathbf{\Pi}_{22}) \leq G_{2*} = G_2 - 1$, *there exists* q_2 *such that* $\sum_{g=1}^{G_2-G_{2*}} \lambda_{2g} = \lambda_{21} \leq q_2$ *a.s., and*

$$q_2 \quad \xrightarrow{d} \quad \chi^2_{(G_2-G_{2*})(K_2-G_{2*})} = \chi^2_{K_2-G_2+1} \ .$$

We notice that only the two panel AR test statistics $\lambda_{\cdot 1}$ and λ_{21} are eventually used, which are the minimum eigenvalues of two eigenvalue equations.

First, we consider the following procedure:

$$\mathrm{H}_0 : \ \mathrm{rank}(\mathbf{\Pi}_{22}) \leq G_2 - 1 \quad \text{v.s.} \quad \mathrm{H}_1 : \ \mathrm{rank}(\mathbf{\Pi}_{22}) = G_2 \ .$$

This hypothesis test corresponds to the work of Koopmans and Hood (1953). If the null hypothesis is true, then multiple solutions $\boldsymbol{\beta}$ exist. As q_2 follows the chi-square distribution with $K_2 + G_2 - 1$ degree of freedom, this rank test becomes a conservative test with an actual size smaller than the nominal size α%. Using the critical values of α% and λ_{21}, rejecting the null hypothesis is desirable.

Second, if the above null hypothesis is rejected, then we consider that $\mathrm{rank}(\mathbf{\Pi}_{2\cdot})$ is either G_2 or $G = G_2 + 1$. Therefore,

$$\mathrm{H}_0 : \ \mathrm{rank}(\mathbf{\Pi}_{2\cdot}) = G_2 \quad \text{v.s.} \quad \mathrm{H}_1 : \ \mathrm{rank}(\mathbf{\Pi}_{2\cdot}) = G_2 + 1 \ .$$

This hypothesis test corresponds to the study of Anderson and Rubin (1949). If the alternative hypothesis is true, then $\boldsymbol{\beta}$ does not exist. Under the null hypothesis the rank is reduced to G_2. Hence, $\boldsymbol{\beta}$ is unique. Then, $\lambda_{\cdot 1}$ follows the chi-square distribution with $K_2 - G_2$ degrees of freedom, and accepting the null hypothesis is desirable. The panel AR statistics $\lambda_{\cdot 1}$ and $n\lambda$ of Theorem 3.1 bring the same result. Therefore, the null hypothesis of the overidentification test is also accepted if this rank test is accepted.

In sum, the rank conditions are expressed as follows:

$$\mathrm{rank}(\mathbf{\Pi}_{2\cdot}) = G_2 = \mathrm{rank}(\mathbf{\Pi}_{22}) \ .$$

This condition is confirmed by the above procedure. Then, the specification is justified in the sense that the zero constraints are correct as a structural expression, and the unique parameter vector $(\boldsymbol{\beta}_2', \ \boldsymbol{\gamma}_1', \ \boldsymbol{\gamma}_2' = \mathbf{0}')$ exists.

4.6 Simulation

We confirm the finite sample properties of Theorem 3.6. The numbers of variables are $G = 3$, $G_2 = 2$, and $K_1 = 0$. That is, the right-hand side is only endogenous variables with the zero constraints such as

$$y_{it}^{(1)} = \beta_{21} y_{it}^{(2)} + \beta_{22} y_{it}^{(3)} + \alpha_i + u_{it} \ ,$$

where $(\beta_{21},\ \beta_{22}) = (0.7,\ 0.3)$. The reduced form has $K = 4$ exogenous variables, for $g = 2,\ 3$,

$$y_{it}^{(g)} = \pi_{g1} y_{it-1}^{(2)} + \pi_{g2} y_{it-2}^{(2)} + \pi_{g3} y_{it-1}^{(3)} + \pi_{g4} y_{it-2}^{(3)} + \pi_i^{(g)} + v_{it}^{(g)} \ .$$

As for the error term $\omega_{gg} = 1, \quad \omega_{gh} = 0.3 \ (g \neq h)$, the individual effect is $\boldsymbol{\pi}_i \sim \mathcal{N}(\mathbf{0},\ \mathbf{I}_3)$. $\mathbf{\Pi} = \mathbf{\Pi}_{2\cdot}$ holds because $K = K_2$.

Design 3.3 : This design considers the case when the rank condition of Theorem 3.6 [i] is satisfied. The coefficients of reduced form for $y_{it}^{(1)}$ are as follows:

$$\boldsymbol{\pi}_{21}^{'} = (0.27,\ 0.07,\ 0.23,\ 0.03) \ ,$$

as for $g = 2,\ 3$,

$$\mathbf{\Pi}_{22}^{'} = \begin{pmatrix} 0.3 & 0.1 & 0.2 & 0.0 \\ 0.2 & 0.0 & 0.3 & 0.1 \end{pmatrix} \ ,$$

and then, $\text{rank}(\mathbf{\Pi}_{22}) = 2$. Hence, $\boldsymbol{\pi}_{21}$ is a linear combination by $\mathbf{\Pi}_{22}$, and the null hypothesis $\text{rank}(\mathbf{\Pi}_{2\cdot}) = G_2 = 2$ is satisfied; that is, $(\beta_{21},\ \beta_{22}) = (0.7,\ 0.3)$ is identified.

Figure 11 shows the empirical distribution of $\lambda_{\cdot 1}$ under the null hypothesis, which follows the chi-square distribution with $K_2 - G_2 = 4 - 2$ degree of freedom.

Design 3.4 : In this setting, we check the finite sample properties of Theorem 3.6 [ii], and the null hypothesis is that the rank of $\mathbf{\Pi}_{22}$ is reduced. In the case of $\text{rank}(\mathbf{\Pi}_{22}) = 1$, we set the values as follows:

$$\mathbf{\Pi}_{22}^{'} = \begin{pmatrix} 0.3 & 0.1 & 0.2 & 0.0 \\ 0.15 & 0.05 & 0.1 & 0.0 \end{pmatrix} \ .$$

Figure 12 shows the empirical distribution of λ_{21}, which follows the chi-square distribution with $(G_2 - G_{2*})(K_2 - G_{2*}) = 1 \times 3$ degree of freedom. In the case of $\text{rank}(\mathbf{\Pi}_{22}) = 0$, we have the following:

$$\mathbf{\Pi}_{22}^{'} = \mathbf{O} \ .$$

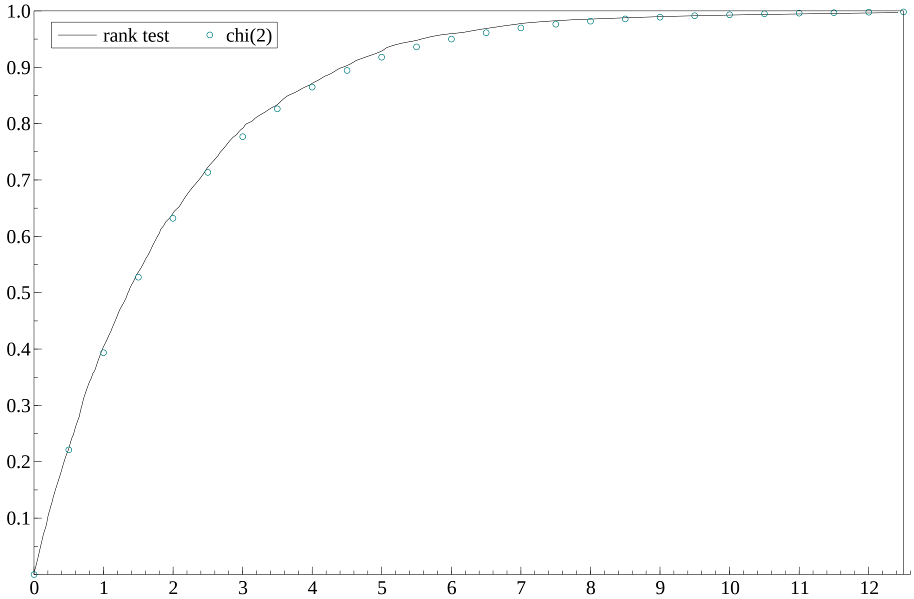

Fig. 11: Design 3.3 ($G_* = 2,\ N = 100,\ T = 25$)

Figure 13 shows the empirical distribution of the sum $(\lambda_{21} + \lambda_{22})$, which follows the chi-square distribution with $(G_2 - G_{2*})(K_2 - G_{2*}) = 2 \times 4$ degrees of freedom.

Although Figures 12 and 13 show the cases when the null hypothesis is the equality, in the empirical analysis, we should suppose the inequality $\text{rank}(\mathbf{\Pi}_{22}) \leq 1$; that is, $\text{rank}(\mathbf{\Pi}_{22})$ is 0 or 1. Then, the conservative test using only the minimum eigenvalue λ_{21} is performed. Figure 14 shows the empirical distribution of λ_{21} in the case of $\text{rank}(\mathbf{\Pi}_{22}) = 0$ again with q_2. The empirical distribution is actually on the left side of the chi-square distribution with $(K_2 - G_2 + 1) = 3$ degree of freedom. Thus, the experiment suggests that $\lambda_{21} \leq q_2 \sim \chi^2_3$ as shown in Theorem 3.6 [ii].

In Part III, we discussed the tests and specifications of the structural panel model, which are conducted by the panel AR test statistics based on the D-LIML estimator in Theorem 2.9. The overidentification and exogeneity tests, model selection by the information criterion, and rank tests can be constructed without being affected by individual effects or long panels. Thus, these procedures could provide a profound structural analysis. In addition, the GMM method has difficulties in performing the overidentification test in the case of the just identified case, whereas the LIML method can conduct the test even under the many instruments and the identification test using the eigenvalues.

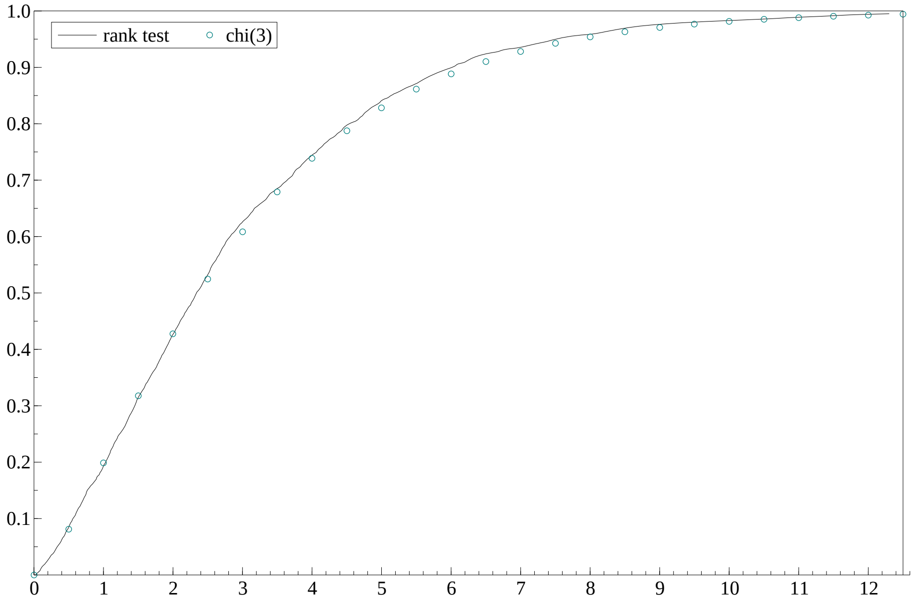

Fig. 12: Design 3.4 ($G_{2*} = 1,\ N = 100,\ T = 25$)

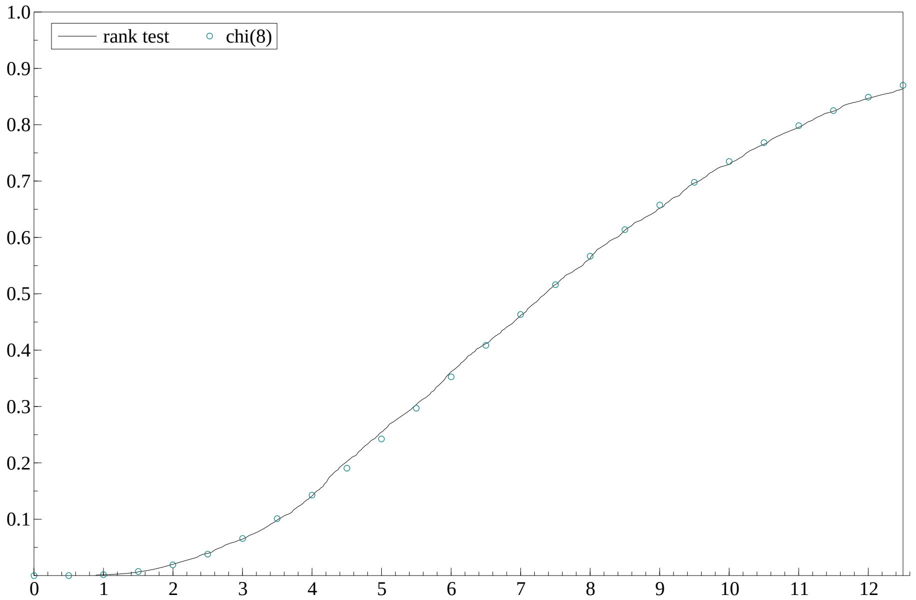

Fig. 13: Design 3.4 ($G_{2*} = 0,\ N = 100,\ T = 25$)

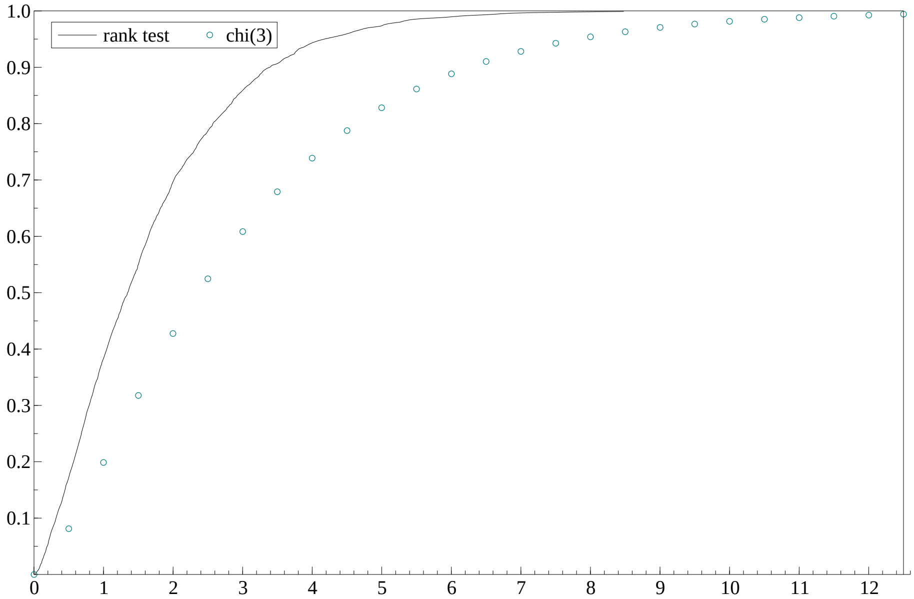

Fig. 14: Design 3.4 ($G_{2*} = 0,\ N = 100,\ T = 25$)

5 Conclusions

As existing estimation methods may not work in long panel data, this study aims to discuss the estimators that can be used in the dynamic structural panel model. Although many simple models exist in the theoretical analyses, this work considered the general models for empirical analyses. We have summarized the results of previous studies and proposed the estimation and testing procedures, focusing on the usefulness of the LIML methods such as the T-LIML and D-LIML estimators.

We showed that T-LIML estimation is robust to the incidental parameters problem of the initial values indicated by Anderson and Hsiao (1981). Although the T-LIML estimator is the best finite sample property, the calculation is complicated in the general model. Therefore, we proposed the asymptotically equivalent D-LIML estimator. This estimator is based on the doubly filters and the variance ratio in the study of Anderson and Rubin (1949), which is originally the concentrated log-likelihood function. D-LIML estimation is practical in estimation and hypothesis test and robust in the dynamic panel model under the large-K asymptotics, which is developed by Kunitomo (1980). We hope that the results of the LIML methods would contribute to structural panel analysis.

6 Appendix: Proofs

Proof of Theorem 2.6 : We first show the following:

$$\lim_{N,\,T\to\infty} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_0}{\partial\pi\partial\omega}\right] = 0\ , \quad \lim_{N,\,T\to\infty} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_0}{\partial\pi\partial\omega_\xi}\right] = 0\ . \tag{6.1}$$

We use the following relations,

$$\mathbf{y}_{i,-1}^{(\ell)} = \mathbf{w}_i - \mathbf{w}_{i0}\ , \quad \mathbf{v}_i^{(\ell)} = \boldsymbol{\xi}_i + \mathbf{v}_i\ , \tag{6.2}$$

where $\mathbf{w}_i = (w_{i0}, ..., w_{iT-1})'$, $\mathbf{w}_{i0} = w_{i0}\boldsymbol{\iota}$, $\boldsymbol{\xi}_i = \xi_i\boldsymbol{\iota}$, and $\boldsymbol{\iota}$ is the $T\times 1$ vector whose elements are unity.

$$(\mathbf{A}+\mathbf{b}\mathbf{b}')^{-1} = \mathbf{A}^{-1} - \frac{1}{1+\mathbf{b}'\mathbf{A}^{-1}\mathbf{b}}\mathbf{A}^{-1}\mathbf{b}\mathbf{b}'\mathbf{A}^{-1}\ , \tag{6.3}$$

using this formula, it follows that

$$\begin{aligned}\boldsymbol{\Omega}_{\xi v}^{-1} &= (\omega_\xi\boldsymbol{\iota}\boldsymbol{\iota}' + \omega\mathbf{I}_T)^{-1}\\ &= \frac{1}{\omega}\left(\mathbf{I}_T - \psi_T\boldsymbol{\iota}\boldsymbol{\iota}'\right)\ ,\end{aligned}$$

where ψ_T is defined by

$$\psi_T = \frac{\omega_\xi}{T\omega_\xi+\omega}\ . \tag{6.4}$$

The derivative for ω evaluated at the true value is given by

$$\begin{aligned}&\mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_0}{\partial\pi\partial\omega}\right]\\ =\ &\mathcal{E}\left[\frac{1}{T}\frac{\partial\left(\mathbf{y}_{i,-1}^{(\ell)\prime}\boldsymbol{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right)}{\partial\omega}\right]\\ =\ &-\frac{1}{T\omega}\mathcal{E}\left[\mathbf{y}_{i,-1}^{(\ell)\prime}\boldsymbol{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right]\\ &+\frac{\omega_\xi}{T\omega(T\omega_\xi+\omega)^2}\mathcal{E}\left[\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i + \mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i - \mathbf{w}_{i0}'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i - \mathbf{w}_{i0}'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i\right]\ ,\end{aligned} \tag{6.5}$$

where the second equality is from (6.2) and the following:

$$\frac{\partial\boldsymbol{\Omega}_{\xi v}^{-1}}{\partial\omega} = -\frac{1}{\omega}\boldsymbol{\Omega}_{\xi v}^{-1} + \frac{\omega_\xi}{\omega(T\omega_\xi+\omega)^2}\boldsymbol{\iota}\boldsymbol{\iota}'\ . \tag{6.6}$$

The fact that for any T the first term of (6.5) becomes zero is shown by (6.7) below. The second term of (6.5) is shown as $O(T^2/T^3)$ and converges to zero. The derivative for ω_ξ is

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_0}{\partial\pi\partial\omega_\xi}\right] &= \mathcal{E}\left[\frac{1}{T}\frac{\partial\left(\mathbf{y}_{i,-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right)}{\partial\omega_\xi}\right] \\
&= -\frac{1}{T\omega}\frac{(T\omega_\xi+\omega)-\omega_\xi T}{(T\omega_\xi+\omega)^2}\mathbf{y}_{t-1}^{(\ell)\prime}\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_\xi \\
&= -\frac{1}{T\omega}\frac{\omega}{(T\omega_\xi+\omega)^2}\mathbf{y}_{t-1}^{(\ell)\prime}\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_\xi\ .
\end{aligned}$$

This term also converges to zero because it is proportional to the second term of (6.6). Therefore, (6.1) is obtained. As the Hessian of (A3)-[i] is the block diagonal matrix, we have that

$$\sqrt{NT}(\hat{\pi}_{\text{TM}}-\pi) = -\frac{1}{h_\pi}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)} + o_p(1)\ ,$$

where

$$h_{\pi\pi} = \lim_{N,\,T\to\infty}\mathcal{E}\left[\frac{1}{NT}\frac{\partial\mathcal{L}_0}{\partial\pi\partial\pi}\right]\ .$$

We confirm the following:

$$\begin{aligned}
&\mathcal{E}\left[\frac{1}{\sqrt{T}}\mathbf{y}_{t-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right] \\
= \ &\frac{1}{\sqrt{T}}\mathcal{E}\left[-\frac{\psi_T}{\omega}\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i + \frac{1}{\omega}(\mathbf{w}_i'\boldsymbol{\xi}_i - \psi_T\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i) - \frac{1}{\omega}(\mathbf{w}_0'\boldsymbol{\xi}_i - \psi_T\mathbf{w}_{0i}'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i)\right] \\
= \ &0\ .
\end{aligned} \tag{6.7}$$

The sum of the terms whose expectations are nonzero becomes zero. For the sum of the second and third terms, using $\boldsymbol{\xi}_i = -(1-\pi)\mathbf{w}_{i0}$,

$$\begin{aligned}
&\mathcal{E}\left[\frac{1}{\omega}(\mathbf{w}_i'\boldsymbol{\xi}_i - \psi_T\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i) - \frac{1}{\omega}(\mathbf{w}_{i0}'\boldsymbol{\xi}_i - \psi_T\mathbf{w}_0'\boldsymbol{\iota}\boldsymbol{\iota}'\boldsymbol{\xi}_i)\right] \\
= \ &-\frac{1-\pi}{1-\pi^2}(1-\psi_T T)\left(\frac{1-\pi^T}{1-\pi}-T\right) \\
= \ &-\frac{1-\pi}{1-\pi^2}\frac{\omega}{\omega+\omega_\xi T}\left(\frac{1-\pi^T}{1-\pi}-T\right)\ .
\end{aligned}$$

The expectation of the first term of (6.7) is given by

$$\begin{aligned}
\mathcal{E}\left[-\frac{\psi_T}{\omega}\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i\right] &= -\frac{\psi_T}{1-\pi}\left((T-1)-\pi\frac{1-\pi^{T-1}}{1-\pi}\right) \\
&= -\frac{1-\pi}{1-\pi^2}\frac{\omega}{\omega+\omega_\xi T}\left((T-1)-\pi\frac{1-\pi^{T-1}}{1-\pi}\right)\ ,
\end{aligned}$$

where the second equality is from the following by assumption *(a2)*,

$$\begin{aligned} \omega_\xi &= \mathcal{V}ar[-(1-\pi)w_{i0}] \\ &= \frac{(1-\pi)^2}{1-\pi^2}\omega \ . \end{aligned}$$

Thus, from the following relation,

$$\left(\frac{1-\pi^T}{1-\pi} - T\right) = -\left((T-1) - \pi\frac{1-\pi^{T-1}}{1-\pi}\right) \ ,$$

the sum of (6.7) becomes zero.

Regarding the generalized Lindeberg-Feller condition (cf. Phillips and Moon (1999), Hahn and Kuersteiner (2002)), for any i and T the following sufficient condition holds by assumption (*a1*),

$$\mathcal{E}\left[\left(\frac{1}{\sqrt{T}}\mathbf{y}_{i,-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right)^4\right] < \infty \ .$$

Therefore, the asymptotic normality holds,

$$\sqrt{NT}\,(\hat{\pi}_{\text{TM}} - \pi) \quad \xrightarrow{d} \quad \mathcal{N}(\,0,\ \frac{g_\pi}{h_{\pi\pi}^2}\,) \ ,$$

where

$$g_\pi = \lim_{N,\,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\mathcal{E}\left[\left(\mathbf{y}_{i,-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right)^2\right] \ .$$

$h_{\pi\pi}$ and g_π are as follows:

$$\begin{aligned} h_{\pi\pi} &= -\lim_{T\to\infty}\mathcal{E}\left[\frac{1}{T}\mathbf{y}_{t-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{y}_{t-1}^{(\ell)\prime}\right] \\ &= -\lim_{T\to\infty}\frac{1}{T}\mathcal{E}\left[\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i - 2\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_{i0} + \mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_0\right] \ . \end{aligned} \tag{6.8}$$

For the first term,

$$\begin{aligned} \frac{1}{T}\mathcal{E}\left[\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i\right] &= \mathcal{E}\left[\frac{1}{T\omega}\mathbf{w}_i'\mathbf{w}_i - \frac{\psi_T}{\omega}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w_{it-1}\right)^2\right] \\ &\to \frac{1}{\omega}\frac{\omega}{1-\pi^2} \ . \end{aligned}$$

As $\psi_T = O(1/T)$ and $(\sum_{t=1}^T w_{it-1}/\sqrt{T})^2 = O_p(1)$, this second term converges in probability to zero. As for the third term of (6.8),

$$\begin{aligned} \frac{1}{T}\mathcal{E}\left[\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_{i0}\right] &= \frac{1}{T}\mathcal{E}[Tw_{i0}^2 - \psi_T(Tw_{i0})^2] \\ &= (1-\psi_T T)\mathcal{E}[w_{i0}^2] \\ &\to 0 \ , \end{aligned}$$

where the second equality is from that

$$1-\psi_T T = \frac{\omega}{T\omega_\xi + \omega} = O(\frac{1}{T}) .$$

Therefore, the second term of (6.8) also converges to zero by repeatedly using the Cauchy-Schwarz inequality (hereinafter, abbreviated as CS),

$$\begin{aligned} \mathcal{E}\left[\frac{1}{T}\left|(\mathbf{\Omega}_{\xi v}^{-\frac{1}{2}}\mathbf{w}_i)'(\mathbf{\Omega}_{\xi v}^{-\frac{1}{2}}\mathbf{w}_{i0})\right|\right] &\leq \mathcal{E}\left[\left(\frac{1}{T}\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i\right)^{\frac{1}{2}}\left(\frac{1}{T}\mathbf{w}_0'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_{i0}\right)^{\frac{1}{2}}\right] \\ &\leq \left(\mathcal{E}\left[\frac{1}{T}\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i\right]\mathcal{E}\left[\frac{1}{T}\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_{i0}\right]\right)^{\frac{1}{2}} . \end{aligned}$$

g_π is given by

$$\begin{aligned} g_\pi &= \lim_{T\to\infty}\mathcal{E}\left[\frac{1}{T}\left(\mathbf{y}_{i,-1}^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\right)^2\right] \\ &= \lim_{T\to\infty}\frac{1}{T}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)})^2 - 2\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)}\mathbf{v}_i^{(\ell)\prime}\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i + (\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)})^2\right] . \end{aligned} \tag{6.9}$$

For the first term,

$$\frac{1}{T}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)})^2\right] = \frac{1}{T}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i)^2 + 2\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i\boldsymbol{\xi}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_i + (\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\xi}_i)^2\right] . \tag{6.10}$$

This first term converges to the following:

$$\begin{aligned} \frac{1}{T}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i)^2\right] &= \frac{1}{T\omega^2}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{v}_i)^2 - 2\phi_T\mathbf{w}_i'\mathbf{v}_i\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i + \phi_T^2(\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i)^2\right] \\ &\to \frac{1}{\omega^2}\mathcal{E}[w_{it-1}^2 v_{it}^2] = \frac{1}{1-\pi^2} , \end{aligned}$$

because the third term $(1/T)\phi_T^2\mathcal{E}[(\mathbf{w}_i'\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i)^2] = O(1/T^3)O(T^2) \to 0$ by $\phi_T = O(1/T)$, and the second term also converges to zero by the CS inequality. As for the third term of (6.10),

$$\begin{aligned} \frac{1}{T}\mathcal{E}\left[(\mathbf{w}_i'\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\xi}_i)^2\right] &= \frac{1}{T\omega^2}\mathcal{E}\left[\left(\xi_i\sum_{t=1}^T w_{it-1} - \phi_T\xi_i T\sum_{t=1}^T w_{it-1}\right)^2\right] \\ &= \frac{1}{\omega^2}(1-\phi_T T)^2\mathcal{E}\left[\xi_i^2\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T w_{it-1}\right)^2\right] \\ &\to 0 , \end{aligned}$$

where $(1-\phi_T T)^2 = O(1/T^2)$. From the above, the first term of (6.9) converges to $\mathcal{E}[w_{it-1}^2 v_{it}^2]$.

Finally, we evaluate the third term of (6.9),

$$\frac{1}{T}\mathcal{E}\left[(\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i^{(\ell)})^2\right] = \frac{1}{T}\mathcal{E}\left[(\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i)^2 + 2\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{v}_i\boldsymbol{\xi}_i'\mathbf{\Omega}_{\xi v}^{-1}\mathbf{w}_{i0} + (\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\xi}_i)^2\right] .$$

This first term is $O(1/T)$ under similar arguments. Regarding the third term,

$$\begin{aligned}\frac{1}{T}\mathcal{E}\left[(\mathbf{w}_{i0}'\mathbf{\Omega}_{\xi v}^{-1}\boldsymbol{\xi}_i)^2\right] &= \frac{1}{T\omega^2}(T-\phi_T T^2)^2\mathcal{E}\left[w_{i0}^2\xi_i^2\right] \\ &= O(\frac{1}{T}) ,\end{aligned}$$

the second term converges to zero by the CS inequality, and thus, the third term of (6.9) converges to zero. Using similar arguments, the second term of (6.9) also converges to zero. Therefore, $(g_\pi/h_{\pi\pi}^2)$ becomes $1-\pi^2$. □

Proof of Theorem 2.7 : We first consider the case when $N<\infty$. From the result of Lemma 2.2-[ii],

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{TL}}-\boldsymbol{\phi}) = \sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{PL}}-\boldsymbol{\phi}) + o_p(1) .$$

Therefore, the T-LIML estimator is asymptotically equivalent to the pseudo T-LIML estimator and the result of $d=0$ holds for Theorem 2.10.

Next, consider the case when $N\to\infty$. As $\mathbf{\Omega}_\xi$ is consistently estimated, the Hessian can be evaluated at the true value. We show the following:

$$\lim_{N,\,T\to\infty}\mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_2}{\partial\boldsymbol{\phi}\partial\boldsymbol{\omega}'}\right] = \underset{3\times 4}{\mathbf{O}} , \qquad \lim_{N,\,T\to\infty}\mathcal{E}\left[\frac{1}{NT}\frac{\partial^2\mathcal{L}_2}{\partial\boldsymbol{\phi}\partial\boldsymbol{\omega}_\xi'}\right] = \underset{3\times 4}{\mathbf{O}} , \tag{6.11}$$

where $\boldsymbol{\phi} = (\beta_2,\gamma_1,\pi_{22})$, $\boldsymbol{\omega} = \text{vec}(\mathbf{\Omega})$, and $\boldsymbol{\omega}_\xi = \text{vec}(\mathbf{\Omega}_\xi)$. Following Hsiao and Zhou (2015) we have the following expression:

$$\begin{aligned}\underset{2T\times 2T}{\mathbf{\Omega}_{\xi v}^{-1}} &= (\mathbf{\Omega}_\xi\otimes\boldsymbol{\iota}\boldsymbol{\iota}' + \mathbf{\Omega}\otimes\mathbf{I}_T)^{-1} \\ &= \mathbf{\Omega}^{-1}\otimes\mathbf{Q}_T + \mathbf{\Psi}_T^{-1}\otimes\mathbf{J}_T ,\end{aligned}$$

where

$$\underset{2\times 2}{\mathbf{\Psi}_T} = \mathbf{\Omega} + T\mathbf{\Omega}_\xi , \quad \underset{T\times T}{\mathbf{J}_T} = \frac{1}{T}\boldsymbol{\iota}\boldsymbol{\iota}' ,$$

and we express the elements of the inverse matrix as follows:

$$\mathbf{\Omega}^{-1} = \begin{pmatrix}\omega^{11} & \omega^{12}\\ \omega^{21} & \omega^{22}\end{pmatrix} , \quad \mathbf{\Psi}_T^{-1} = \begin{pmatrix}\psi_T^{11} & \psi_T^{12}\\ \psi_T^{21} & \psi_T^{22}\end{pmatrix} .$$

Regarding the score function of the reduced form,

$$
\begin{aligned}
s_{11,i} &= \omega^{11}\mathbf{y}_{i,-1}^{(1,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(1)} + \omega^{12}\mathbf{y}_{i,-1}^{(1,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(2)} + \ \psi_T^{11}\mathbf{y}_{i,-1}^{(1,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \psi_T^{12}\mathbf{y}_{i,-1}^{(1,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)} \ , \\
s_{12,i} &= \omega^{11}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(1)} + \omega^{12}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(2)} + \ \psi_T^{11}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \psi_T^{12}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)} \ , \\
s_{22,i} &= \omega^{21}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(1)} + \omega^{22}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(2)} + \ \psi_T^{21}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \psi_T^{22}\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)} \ ,
\end{aligned}
$$

where $\mathbf{y}_{i,-1}^{(g,\ell)'} = (0, y_{i1}^{(g,\ell)}, ..., y_{iT-1}^{(g,\ell)})'$ and $\mathbf{v}_i^{(g,\ell)} = \boldsymbol{\xi}_i^{(g)} + \mathbf{v}_i^{(g)}$ $(g = 1,\ 2)$. Meanwhile, for the structural equation,

$$
\frac{\partial \mathcal{L}_2}{\partial \beta_2} = \sum_{i=1}^{N} \pi_{22} s_{12,i} \ , \quad \frac{\partial \mathcal{L}_2}{\partial \gamma_1} = \sum_{i=1}^{N} s_{11,i} \ , \quad \frac{\partial \mathcal{L}_2}{\partial \pi_{22}} = \sum_{i=1}^{N} \beta_2 s_{12,i} + s_{22,i} \ .
$$

We show the following as an example:

$$
\lim_{N,\,T\to\infty} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \beta_2 \partial \omega_{11}}\right] = 0 \ .
$$

It follows that

$$
\begin{aligned}
& \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \beta_2 \partial \omega_{11}}\right] \\
=\ & \frac{\pi_{22}}{T}\mathcal{E}\left[\left(\mathbf{y}_{i,-1}^{(2,\ell)'}, \mathbf{0}'\right)\frac{\partial \boldsymbol{\Omega}_{\xi v}^{-1}}{\partial \omega_{11}}\left(\mathbf{v}_i^{(1,\ell)'},\ \mathbf{v}_i^{(2,\ell)'}\right)'\right] \\
=\ & \frac{\pi_{22}}{T}\mathcal{E}\left[\left(\mathbf{y}_{i,-1}^{(2,\ell)'}, \mathbf{0}'\right)\left(\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \omega_{11}} \otimes \mathbf{Q}_T + \frac{\partial \boldsymbol{\Psi}_T^{-1}}{\partial \omega_{11}} \otimes \mathbf{J}_T\right)\left(\mathbf{v}_i^{(1,\ell)'},\ \mathbf{v}_i^{(2,\ell)'}\right)'\right] \ .
\end{aligned}
$$

For the derivative with respect to $\boldsymbol{\Omega}^{-1}$,

$$
\begin{aligned}
\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \omega_{11}} &= -\frac{\omega_{22}}{|\boldsymbol{\Omega}|}\boldsymbol{\Omega}^{-1} + \frac{1}{|\boldsymbol{\Omega}|}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \dot{\omega}_{11} & \dot{\omega}_{12} \\ \dot{\omega}_{21} & \dot{\omega}_{22} \end{pmatrix} \ (say,) \ .
\end{aligned}
$$

As for the derivative with respect to $\boldsymbol{\Psi}_T^{-1}$, put the following:

$$
\begin{aligned}
\boldsymbol{\Psi}_T^{-1} &= \frac{1}{T}\left(\frac{1}{T}\boldsymbol{\Omega} + \boldsymbol{\Omega}_\xi\right)^{-1} \\
&= \frac{1}{T}(\boldsymbol{\Omega}_T)^{-1} \ . \qquad (6.12)
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{\partial \boldsymbol{\Psi}_T^{-1}}{\partial \omega_{11}} &= \frac{1}{T}\left[-\frac{\omega_{T,22}}{T|\boldsymbol{\Omega}_T|}\boldsymbol{\Omega}_T^{-1} + \frac{1}{|\boldsymbol{\Omega}_T|}\begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{T} \end{pmatrix}\right] \\
&= \begin{pmatrix} \dot{\psi}_T^{11} & \dot{\psi}_T^{12} \\ \dot{\psi}_T^{21} & \dot{\psi}_T^{22} \end{pmatrix} \ (say,) \ ,
\end{aligned}
$$

where $|\mathbf{\Omega}_T| > 0$ because $\mathbf{\Omega}_T$ is the sum of the positive definite matrices. In fact,

$$\mathbf{\Omega}_\xi = (\mathbf{I}_2 - \mathbf{\Pi}^{'})\mathcal{E}[\mathbf{w}_{i0}\mathbf{w}_{i0}^{'}](\mathbf{I}_2 - \mathbf{\Pi}) > \mathbf{O} \ .$$

From the above,

$$\begin{aligned} & \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \beta_2 \partial \omega_{11}}\right] \\ = \ & \pi_{22}\left(\frac{\dot{\omega}^{11}}{T}\mathcal{E}\left[\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(1)}\right] + \frac{\dot{\omega}^{12}}{T}\mathcal{E}\left[\mathbf{y}_{i,-1}^{(2,\ell)'}\mathbf{Q}_T\mathbf{v}_i^{(2)}\right]\right) \\ & +\pi_{22}\left(\frac{\dot{\psi}_T^{11}}{T^2}\mathcal{E}\left[\mathbf{y}_{i,-1}^{(2,\ell)'}\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i^{(1,\ell)}\right] + \frac{\dot{\psi}_T^{12}}{T^2}\mathcal{E}\left[\mathbf{y}_{i,-1}^{(2,\ell)'}\boldsymbol{\iota}\boldsymbol{\iota}'\mathbf{v}_i^{(2,\ell)}\right]\right) \\ = \ & O\left(\frac{1}{T}\right) + O\left(\frac{T^2}{T^4}\right) \ , \end{aligned}$$

because $\dot{\psi}_T^{11}$ and $\dot{\psi}_T^{12}$ are $O(1/T^2)$, and the following elements are $O(1)$,

$$\begin{aligned} & \mathcal{E}\left[\begin{pmatrix} \mathbf{y}_{i,-1}^{(1,\ell)'} \\ \mathbf{y}_{i,-1}^{(2,\ell)'} \end{pmatrix}\mathbf{Q}_T\left(\mathbf{v}_i^{(1)}, \ \mathbf{v}_i^{(2)}\right)\right] \\ = \ & -\left(\mathbf{I}_2 - \mathbf{\Pi}^{'}\right)^{-2}\left[\left(1 - \frac{1}{T}\right)\mathbf{I}_2 - \mathbf{\Pi}^{'} + \frac{1}{T}(\mathbf{\Pi}^{'})^T\right]\mathbf{\Omega} \ . \end{aligned}$$

By using similar aruguments for the other elements, the former of (6.11) is verified. For the latter of (6.11), we show that

$$\lim_{N,\, T\to\infty}\mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \beta_2 \partial \omega_{\xi,11}}\right] = 0 \ .$$

It follows that

$$\begin{aligned} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \beta_2 \partial \omega_{\xi,11}}\right] &= \frac{\pi_{22}}{T}\mathcal{E}\left[\left(\mathbf{y}_{i,-1}^{(2,\ell)'}, \mathbf{0}^{'}\right)\left(\frac{\partial \mathbf{\Psi}_T^{-1}}{\partial \omega_{\xi,11}} \otimes \mathbf{J}_T\right)\left(\mathbf{v}_i^{(1,\ell)'}, \ \mathbf{v}_i^{(2,\ell)'}\right)^{'}\right] \\ &= O\left(\frac{T^2}{T^3}\right) \ . \end{aligned}$$

For the derivative,

$$\frac{\partial \mathbf{\Psi}_T^{-1}}{\partial \omega_{\xi,11}} = \frac{1}{T}\left[-\frac{\omega_{T,22}}{|\mathbf{\Omega}_T|}\mathbf{\Omega}_T^{-1} + \frac{1}{|\mathbf{\Omega}_T|}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right] ,$$

because each element is $O(1/T)$. Other elements of the latter in (6.11) are also $O(1/T)$ under similar aruguments. Thus, when N and T go to infinity,

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}) = -\mathbf{H}_{\phi\phi}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{s}_i^{(\ell)} + o_p(1) \ ,$$

where $\mathbf{s}_i^{(\ell)} = (\pi_{22}s_{12,i},\ s_{11,i},\ \beta_2 s_{12,i} + s_{22,i})'$ and

$$\begin{aligned}
\mathbf{H}_{\phi\phi} &= \lim_{N,\,T\to\infty} \mathcal{E}\left[\frac{1}{NT}\frac{\partial^2 \mathcal{L}_2}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'}\right] \\
&= -\lim_{N,\,T\to\infty} \mathcal{E}\left[\frac{1}{T}\begin{pmatrix} \pi_2 \mathbf{y}_{i,-1}^{(2,\ell)'} & \mathbf{0}' \\ \mathbf{y}_{i,-1}^{(1,\ell)'} & \mathbf{0}' \\ \beta \mathbf{y}_{i,-1}^{(2,\ell)'} & \mathbf{y}_{i,-1}^{(2,\ell)'} \end{pmatrix} \boldsymbol{\Omega}_{\xi v}^{-1} \begin{pmatrix} \pi_2 \mathbf{y}_{i,-1}^{(2,\ell)} & \mathbf{y}_{i,-1}^{(1,\ell)} & \beta \mathbf{y}_{i,-1}^{(2,\ell)} \\ \mathbf{0} & \mathbf{0} & \mathbf{y}_{i,-1}^{(2,\ell)} \end{pmatrix}\right] .
\end{aligned}$$

The expectation of the score function vanishes:

$$\mathcal{E}\left[\mathbf{s}_i^{(\ell)}\right] = \mathbf{0} ,$$

because Hsiao and Zhou (2015) show that for any T,

$$\mathcal{E}[s_{11,i}] = \mathcal{E}[s_{12,i}] = \mathcal{E}[s_{22,i}] = 0 .$$

As for the score function such as $\boldsymbol{\Omega}_\xi$, the proof is the same as that of Theorem 2.13.

Consider the (1,1) element of $\mathbf{H}_{\phi\phi}$ as an example,

$$\begin{aligned}
\left[\mathbf{y}_{i,-1}^{(2,\ell)'},\ \mathbf{0}'\right] \boldsymbol{\Omega}_{\xi v}^{-1} \begin{bmatrix} \mathbf{y}_{i,-1}^{(2,\ell)} \\ \mathbf{0} \end{bmatrix} &= \left[(\mathbf{w}_i^{(2)} - \mathbf{w}_{i0}^{(2)})',\ \mathbf{0}'\right] \boldsymbol{\Omega}_{\xi v}^{-1} \begin{bmatrix} \mathbf{w}_i^{(2)} - \mathbf{w}_{i0}^{(2)} \\ \mathbf{0} \end{bmatrix} \\
&= (\mathbf{w}_i^{(2)} - \mathbf{w}_{i0}^{(2)})' \left(\omega^{11}\mathbf{Q}_T + \psi_T^{11}\mathbf{J}_T\right)(\mathbf{w}_i^{(2)} - \mathbf{w}_{i0}^{(2)}) \\
&= \mathbf{w}_i^{(2)'}\boldsymbol{\Omega}_{\xi v}^{11}\mathbf{w}_i^{(2)} - 2\mathbf{w}_i^{(2)'}\boldsymbol{\Omega}_{\xi v}^{11}\mathbf{w}_{i0}^{(2)} + \mathbf{w}_{i0}^{(2)'}\boldsymbol{\Omega}_{\xi v}^{11}\mathbf{w}_{i0}^{(2)} ,
\end{aligned} \tag{6.13}$$

where

$$\boldsymbol{\Omega}_{\xi v}^{11} = \omega^{11}\mathbf{Q}_T + \psi_T^{11}\mathbf{J}_T .$$

First, we show that the third term of (6.13) converges in probability to zero.

$$\begin{aligned}
\frac{1}{T}\mathcal{E}\left[\mathbf{w}_{i0}^{(2)'}\boldsymbol{\Omega}_{\xi v}^{11}\mathbf{w}_{i0}^{(2)}\right] &= \frac{1}{T}\mathcal{E}\left[\mathbf{w}_{i0}^{(2)'}(\omega^{11}\mathbf{I}_T - (\omega^{11} - \psi_T^{11})\mathbf{J}_T)\mathbf{w}_{i0}^{(2)}\right] \\
&= w^{11}\mathcal{E}[w_{i0}^{(2)2}] - (\omega^{11} - \psi_T^{11})\mathcal{E}[w_{i0}^{(2)2}] \\
&= \psi_T^{11}\mathcal{E}[w_{i0}^{(2)2}] ,
\end{aligned}$$

where $\psi_T^{11} = O(1/T)$ is negligible due to (6.4). Regarding the first term of (6.13),

$$\begin{aligned}
\frac{1}{T}\mathcal{E}\left[\mathbf{w}_i^{(2)'}\boldsymbol{\Omega}_{\xi v}^{11}\mathbf{w}_i^{(2)}\right] &= \mathcal{E}\left[\frac{\omega^{11}}{T}\mathbf{w}_i^{(2)'}\mathbf{w}_i^{(2)} - \frac{\omega^{11} - \psi_T^{11}}{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_{it-1}^{(2)}\right)^2\right] \\
&\to \omega^{11}\mathcal{E}[w_{it-1}^{(2)2}] .
\end{aligned}$$

Therefore, the (1,1) element converges to $\omega^{11}\pi_{22}^2\mathcal{E}[w_{it-1}^{(2)2}]$. Then, we obtain that

$$\mathbf{H}_{\phi\phi} = -\mathcal{E}\left[\begin{pmatrix} \pi_{22}w_{it-1}^{(2)} & 0 \\ w_{it-1}^{(1)} & 0 \\ \beta_2 w_{it-1}^{(2)} & w_{it-1}^{(2)} \end{pmatrix} \mathbf{\Omega}^{-1} \begin{pmatrix} \pi_{22}w_{it-1}^{(2)} & w_{it-1}^{(1)} & \beta_2 w_{it-1}^{(2)} \\ 0 & 0 & w_{it-1}^{(2)} \end{pmatrix}\right] . \quad (6.14)$$

Second, we clarify the limit of the sum of squares $\mathbf{G}_\phi$,

$$\mathbf{G}_\phi = \lim_{N,\,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N} \mathcal{E}\left[\mathbf{s}_i^{(\ell)}\mathbf{s}_i^{(\ell)\prime}\right] . \quad (6.15)$$

For instance, the (1,1) element is given by

$$\frac{1}{T}\mathcal{E}\left[(\pi_{22}s_{12,i})^2\right] = \frac{1}{T}\mathcal{E}\left[\pi_{22}^2\left(\mathbf{y}_{i,-1}^{(2,\ell)\prime}\mathbf{\Omega}_{\xi v}^{11}\mathbf{v}_i^{(\ell,1)} + \mathbf{y}_{i,-1}^{(2,\ell)\prime}\mathbf{\Omega}_{\xi v}^{12}\mathbf{v}_i^{(\ell,2)}\right)^2\right] ,$$

where another expression of $\mathbf{\Omega}_{\xi v}^{11}$ is as follows:

$$\begin{aligned} \mathbf{\Omega}_{\xi v}^{11} &= \omega^{11}\left(\mathbf{I}_T - \varphi_T^{11}\boldsymbol{\iota}\boldsymbol{\iota}'\right) , \\ \varphi_T^{11} &= \frac{\omega^{11} - \phi_T^{11}}{T\omega^{11}} . \end{aligned}$$

Then, $\varphi_T^{11} = O(1/T)$ and

$$1 - \varphi_T^{11}T = \frac{\phi_T^{11}}{\omega^{11}} = O\left(\frac{1}{T}\right) ,$$

that is, the order is the same as that in (6.4). Then, under the similar arguments of Theorem 2.6,

$$\begin{aligned} \frac{1}{T}\mathcal{E}\left[(\pi_{22}s_{12,i})^2\right] &\to \pi_{22}^2\mathcal{E}\left[\left(w_{it-1}^{(2)},\ 0\right)\mathbf{\Omega}^{-1}(\mathbf{v}_{it}\mathbf{v}_{it}')\mathbf{\Omega}^{-1}\begin{pmatrix} w_{it-1}^{(2)} \\ 0 \end{pmatrix}\right] \\ &= \omega^{11}\pi_{22}^2\mathcal{E}[w_{it}^{(2)2}] . \end{aligned}$$

Using similar arguments for other elements, we have

$$\mathbf{G}_\phi = -\mathbf{H}_{\phi\phi} .$$

Moreover, from the results of (6.14) and (6.15), the terms related to the long-difference can be asymptotically ignored as the remainder terms,

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}) = -\mathbf{H}_{\phi\phi}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{s}_{it} + o_p(1) ,$$

where

$$\underset{3\times 1}{\mathbf{s}_{it}} = \begin{bmatrix} (\pi_{22} w_{it-1}^{(2)},\ 0) \\ (w_{it-1}^{(1)},\ 0) \\ (\beta_2 w_{it-1}^{(2)},\ w_{it-1}^{(2)}) \end{bmatrix} \mathbf{\Omega}^{-1} \mathbf{v}_{it} \ .$$

If the fourth moment of the error term exists, then the generalized Lindeberg-Feller condition holds. Therefore,

$$\sqrt{NT}\left(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}\right) \quad \xrightarrow{d} \quad \mathcal{N}(\ \mathbf{0},\ \mathbf{H}_{\phi\phi}^{-1} \mathbf{G}_{\phi} \mathbf{H}_{\phi\phi}^{-1}\ ) \ .$$

The variance-covariance matrix for the structural parameter $\boldsymbol{\theta}_1 = (\beta_2,\ \gamma_1)'$ is obtained as the 2×2 submatrix of the upper left of the 3×3 matrix $-\mathbf{H}_{\phi\phi}^{-1}$. After some calculation,

$$\underset{2\times 2}{\left(\frac{1}{\sigma^2}\mathbf{\Phi}\right)^{-1}} = \left(\frac{\omega^{11}\omega^{22} - \omega^{12}\omega^{21}}{\beta_2^2\omega^{11} + 2\beta_2\omega^{12} + \omega^{22}}\mathbf{\Phi}\right)^{-1} , \tag{6.16}$$

$$\mathbf{\Phi} = \begin{pmatrix} 0 & \pi_{22} \\ 1 & 0 \end{pmatrix} \mathcal{E}\left[\begin{pmatrix} w_{it-1}^{(1)2} & w_{it-1}^{(1)} w_{it-1}^{(2)} \\ w_{it-1}^{(2)} w_{it-1}^{(1)} & w_{it-1}^{(2)2} \end{pmatrix}\right] \begin{pmatrix} 0 & 1 \\ \pi_{22} & 0 \end{pmatrix} ,$$

where the first equality holds because $\sigma^2 = \boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}$ and

$$\mathbf{\Omega} = \frac{1}{|\mathbf{\Omega}^{-1}|} \begin{pmatrix} \omega^{22} & -\omega^{12} \\ -\omega^{21} & \omega^{11} \end{pmatrix} .$$

Thus, we obtain the desired result. □

Proof of Theorem 2.8 : For the companion reduced form of (3.30), we put $z_{it+1} = y_{it}^{(3)}$,

$$\underset{3\times 3}{\mathbf{\Pi}^*} = (\boldsymbol{\pi}_1,\ \boldsymbol{\pi}_2,\ \boldsymbol{\pi}_3) \ ,$$

and

$$\underset{3\times 3}{\mathbf{\Omega}_*} = \begin{pmatrix} \mathbf{\Omega} & \boldsymbol{\omega}_{13} \\ \boldsymbol{\omega}_{13}' & \omega_{33} \end{pmatrix} , \quad \mathcal{E}[\mathbf{v}_{it} v_{it}^{(3)}] = \boldsymbol{\omega}_{13} \ , \quad \mathcal{E}[(v_{it}^{(3)})^2] = \omega_{33} \ .$$

Then, the log-likelihood function is given by

$$\mathcal{L}_2 = -\frac{N}{2}\log|\mathbf{\Omega}_{*\xi v}| - \frac{1}{2}\sum_{i=1}^{N} \mathbf{v}_i^{(\ell)'} \mathbf{\Omega}_{*\xi v}^{-1} \mathbf{v}_i^{(\ell)} \ ,$$

where

$$\underset{3T\times 1}{\mathbf{v}_i^{(\ell)}} = \begin{bmatrix} \mathbf{y}_i^{(1,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)} \left(\gamma_{11} + \beta_2\pi_{21},\ \beta_2\pi_{22},\ \gamma_{12} + \beta_2\pi_{23} \right)' \\ \mathbf{y}_i^{(2,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}\boldsymbol{\pi}_2 \\ \mathbf{y}_i^{(3,\ell)} - \mathbf{Y}_{i,-1}^{(\ell)}\boldsymbol{\pi}_3 \end{bmatrix} .$$

As the two-dimensional panel VAR of Theorem 2.7 is just replaced by the three-dimensional panel VAR, the following holds as $T \to \infty$ from the discussion of Theorem 2.7 and Theorem 2.12,

$$\sqrt{NT}\left(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}\right) \quad \xrightarrow{d} \quad \mathcal{N}(\ \mathbf{0},\ -\mathbf{H}_{\phi\phi}^{-1}\ ) ,$$

where

$$\begin{aligned} \underset{9\times 1}{\boldsymbol{\phi}} &= \left(\boldsymbol{\theta}_1',\ \boldsymbol{\pi}_2',\ \boldsymbol{\pi}_3'\right)' , \\ \boldsymbol{\theta}_1 &= \left(\beta_2,\ \gamma_{11},\ \gamma_{12}\right)' , \end{aligned}$$

and

$$\begin{aligned} \underset{9\times 9}{-\mathbf{H}_{\phi\phi}} &= \mathcal{E}\left[\begin{pmatrix} \mathbf{W}_{12} & \mathbf{0} \\ \mathbf{O} & \mathbf{w}_{it-1} \end{pmatrix} \boldsymbol{\Omega}_*^{-1} \begin{pmatrix} \mathbf{W}_{12}' & \mathbf{O}' \\ \mathbf{0}' & \mathbf{w}_{it-1}' \end{pmatrix}\right] , \\ \underset{6\times 2}{\mathbf{W}_{12}} &= \begin{pmatrix} \pi_{21}w_{it-1}^{(1)} + \pi_{22}w_{it-1}^{(2)} + \pi_{23}w_{it-1}^{(2)} & 0 \\ w_{it-1}^{(1)} & 0 \\ w_{it-1}^{(3)} & 0 \\ \beta_2\mathbf{w}_{it-1} & \mathbf{w}_{it-1} \end{pmatrix} , \\ \underset{3\times 1}{\mathbf{w}_{it-1}} &= \left(w_{it-1}^{(1)},\ w_{it-1}^{(2)},\ w_{it-1}^{(3)}\right)' . \end{aligned}$$

For the asymptotic normality:

$$\sqrt{NT}\left(\hat{\boldsymbol{\theta}}_{\text{TL}} - \boldsymbol{\theta}_1\right) \quad \xrightarrow{d} \quad \mathcal{N}(\ \mathbf{0},\ \sigma^2\boldsymbol{\Phi}^{-1}\ ) ,$$

we show that the asymptotic variance-covariance matrix of the structural parameters becomes

$$\left(\mathbf{I}_3, \mathbf{O}\right)\left(-\mathbf{H}_{\phi\phi}\right)^{-1}\left(\mathbf{I}_3, \mathbf{O}\right)' = \sigma^2\boldsymbol{\Phi}^{-1} ,$$

and that this expression satisfies the following using the notations of (3.16) and (3.17),

$$\begin{aligned} \underset{3\times 3}{\boldsymbol{\Gamma}_0} &= \mathcal{E}\left[\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\right] , \quad w_{it-1}^{(3)} = z_{it} - \mu_i^{(3)} , \\ \underset{(1+2)\times 3}{\boldsymbol{\Pi}_{\text{I}}'} &= \begin{pmatrix} (\pi_{21},\ \pi_{23}) & \pi_{22} \\ \mathbf{I}_2 & \mathbf{0} \end{pmatrix} . \end{aligned}$$

That is, $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$ is invariant, if only the exogenous variable $w_{it-1}^{(3)}$ is added compared with Theorem 2.7. Consider the variance-covariance matrix $-\mathbf{H}_{12}^{-1}$ when the recued form of $y_{it}^{(3)}$ is not estimated,

$$-\underset{6\times 6}{\mathbf{H}_{12}} = \mathcal{E}\left[\mathbf{W}_{12}\boldsymbol{\Omega}^{-1}\mathbf{W}_{12}'\right] .$$

From the proof of Theorem 2.12, it follows that

$$(\mathbf{I}_3, \mathbf{O})\,(-\mathbf{H}_{12})^{-1}\,(\mathbf{I}_3, \mathbf{O})' = \sigma^2\boldsymbol{\Phi}^{-1} .$$

Therefore, if the following holds, then the theorem is verified,

$$(\mathbf{I}_6, \mathbf{O})\,\mathbf{H}_{\phi\phi}^{-1}\,(\mathbf{I}_6, \mathbf{O})' = \mathbf{H}_{12}^{-1} . \tag{6.17}$$

We show this euqlity in the following. Put

$$\boldsymbol{\Omega}_*^{-1} = \begin{pmatrix} \boldsymbol{\Omega}^{11} & \boldsymbol{\omega}^{13} \\ \boldsymbol{\omega}^{13'} & \omega^{33} \end{pmatrix} .$$

The formula of the inverse matrix for a symmetric partitioned matrix is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1} & -(\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1} \\ & (\mathbf{C}-\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix} . \tag{6.18}$$

For the left-hand side of (6.17), we have the following expression:

$$\begin{aligned} &(\mathbf{I}_6, \mathbf{O})\,\mathbf{H}_{\phi\phi}^{-1}\,(\mathbf{I}_6, \mathbf{O})' \\ = \;& (\mathbf{I}_6, \mathbf{O}) \begin{pmatrix} \mathcal{E}\left[\mathbf{W}_{12}\boldsymbol{\Omega}^{11}\mathbf{W}_{12}'\right] & \mathcal{E}\left[\mathbf{W}_{12}\boldsymbol{\omega}^{13}\mathbf{w}_{it-1}'\right] \\ \mathcal{E}\left[\mathbf{w}_{it-1}\boldsymbol{\omega}^{13'}\mathbf{W}_{12}'\right] & \mathcal{E}\left[\mathbf{w}_{it-1}\omega^{33}\mathbf{w}_{it-1}'\right] \end{pmatrix}^{-1} (\mathbf{I}_6, \mathbf{O})' \\ = \;& \left(\mathcal{E}[\mathbf{W}_{12}\boldsymbol{\Omega}^{11}\mathbf{W}_{12}'] - \mathcal{E}[\mathbf{W}_{12}\boldsymbol{\omega}^{13}\mathbf{w}_{it-1}']\left(\mathcal{E}[\mathbf{w}_{it-1}\omega^{33}\mathbf{w}_{it-1}']\right)^{-1}\mathcal{E}[\mathbf{w}_{it-1}\boldsymbol{\omega}^{13'}\mathbf{W}_{12}']\right)^{-1} . \end{aligned}$$

As for the right-hand side of (6.17), $\boldsymbol{\Omega}$ is represented by the element of $\boldsymbol{\Omega}_*^{-1}$ using (6.18),

$$\mathbf{H}_{12}^{-1} = \left(\mathcal{E}\left[\mathbf{W}_{12}\left(\boldsymbol{\Omega}^{11} - \frac{1}{\omega^{33}}\boldsymbol{\omega}^{13}\boldsymbol{\omega}^{13'}\right)\mathbf{W}_{12}'\right]\right)^{-1} .$$

Hence, it is sufficient to show that

$$\mathcal{E}\left[\mathbf{W}_{12}\boldsymbol{\omega}^{13}\mathbf{w}_{it-1}'\right]\left(\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\right)^{-1}\mathcal{E}\left[\mathbf{w}_{it-1}\boldsymbol{\omega}^{13'}\mathbf{W}_{12}'\right] = \mathcal{E}\left[\mathbf{W}_{12}\boldsymbol{\omega}^{13}\boldsymbol{\omega}^{13'}\mathbf{W}_{12}'\right] . \tag{6.19}$$

Note the following relations:

$$\underset{6\times 1}{\mathbf{W}_{12}\boldsymbol{\omega}^{13}} = \boldsymbol{\Theta}_\omega \mathbf{w}_{it-1} ,$$

$$\underset{6\times 3}{\boldsymbol{\Theta}_\omega} = \begin{pmatrix} \omega_1^{13}\pi_{21} & \omega_1^{13}\pi_{22} & \omega_1^{13}\pi_{23} \\ \omega_1^{13} & 0 & 0 \\ 0 & 0 & \omega_1^{13} \\ & (\omega_1^{13}\beta_2 + \omega_2^{13})\mathbf{I}_3 & \end{pmatrix} ,$$

where $\boldsymbol{\omega}^{13} = (\omega_1^{13},\ \omega_2^{13})'$. Therefore, the equality of (6.19) holds. □

Proof of Theorem 2.9 : We first show that

$$\frac{1}{NT}\mathbf{G}_n^{(f,b)} \xrightarrow{p} \mathbf{G}_0 = \boldsymbol{\Theta}_{\mathrm{I}}'\boldsymbol{\Phi}\boldsymbol{\Theta}_{\mathrm{I}} , \quad \boldsymbol{\Phi} > \mathbf{O} , \tag{6.20}$$

and

$$\frac{1}{NT}\mathbf{H}_n^{(f,b)} \xrightarrow{p} \mathbf{H}_0 = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix} , \tag{6.21}$$

where

$$\underset{(G_2+K_1)\times(G+K_1)}{\boldsymbol{\Theta}_{\mathrm{I}}} = (\boldsymbol{\theta},\ \mathbf{I}_{G_2+K_1}) ,$$

$(\mathbf{y}^{(1,f)}, \mathbf{Y}^{(2,f)}, \mathbf{Z}^{(1,f)}) = (\mathbf{y}^{(1,f)}, \mathbf{X}^{(f)})$ and $\mathbf{X}^{(f)}$ is the $N(T-1)\times(G_2+K_1)$ matrix.

For $k = 1, \cdots, K$ and $g = 1, \cdots, (1+G_2+K_1)$, we have that

$$\begin{aligned} & \frac{1}{NT}\mathbf{e}_g'(\mathbf{y}^{(1,f)'}, \mathbf{X}^{(f)'})\mathbf{Z}^{(b)}\mathbf{e}_k \\ = & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}(y_{it}^{(1,f)}, \mathbf{x}_{it}^{(f)'})\mathbf{e}_g\mathbf{z}_{it-1}^{(b)'}\mathbf{e}_k \\ = & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1} w_{it}^{[g]}w_{it-1}^{[k]} - (1-f_t)w_{it}^{[g]}w_{it-1}^{[k]} - f_t\bar{w}_{it-1,0}^{[k]}w_{it}^{[g]} - w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]} + \bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]} , \end{aligned} \tag{6.22}$$

where $\mathbf{e}_k = (0, \cdots, 1, \cdots, 0)'$ whose k-th element is only unity,

$$\begin{aligned} \bar{w}_{it-1,0}^{[k]} &= \frac{1}{t}(w_{it-2}^{[k]} + \cdots + w_{i-1}^{[k]}) , \\ \tilde{w}_{it,T}^{[g]} &= \frac{f_t}{T-t}(w_{it+1}^{[g]} + \cdots + w_{iT}^{[g]}) , \end{aligned} \tag{6.23}$$

and for $g = 1, \cdots, (1+G_2)$,

$$w_{it}^{[g]} = \mathbf{e}_g'(\boldsymbol{\Pi}'\mathbf{J}'\mathbf{w}_{it-1} + \mathbf{v}_{it}) ,$$

but for $g = 1, \cdots, K_1$, we set $w_{it}^{[g]} = w_{it-1}^{[g]}$.

Regarding the last term of (6.22),

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{NT}\sum_{i=1}^{N}|\sum_{t=1}^{T-1}\bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]}|\right] &= \mathcal{E}\left[\frac{1}{T}|\sum_{t=1}^{T-1}\bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]}|\right] \\
&\le \frac{1}{T}\sum_{t=1}^{T-1}\sqrt{\mathcal{E}[\bar{w}_{it-1,0}^{[k]2}]}\sqrt{\mathcal{E}[\bar{w}_{it,T}^{[g]2}]} \\
&= \frac{1}{T}\sum_{t=1}^{T-1}O\left(\sqrt{\frac{1}{t}}\right)O\left(\sqrt{\frac{1}{(T-t)}}\right) \\
&\le O\left(\frac{\log T}{T}\right) ,
\end{aligned}$$

where the second and fourth inequalities are based on the CS inequality. That is, this term converges in the 1th mean to zero as $T \to \infty$. From similar arguments, the third and fourth terms of (6.22) are $O(\sqrt{T}/T)$. The second term is also evaluated as $O(\sqrt{T}/T)$ using the relation that $(1-c_t)^2 \le 1/(T-t+1)$. Therefore, we obtain

$$\frac{1}{NT}(\mathbf{y}^{(1,f)'}, \mathbf{X}^{(f)'})\mathbf{Z}^{(b)} \xrightarrow{p} \mathbf{\Theta}_{\rm I}'\mathbf{\Pi}_{\rm I}'\mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}'] ,$$

where

$$\underset{(G_2+K_1)\times K}{\mathbf{\Pi}_{\rm I}'} = \begin{pmatrix} \mathbf{\Pi}_2' \\ \mathbf{I}_{K_1}\ \mathbf{O} \end{pmatrix} = \begin{pmatrix} \mathbf{\Pi}_{12}' & \mathbf{\Pi}_{22}' \\ \mathbf{I}_{K_1} & \mathbf{O} \end{pmatrix} ,$$

because $\mathcal{E}[\mathbf{v}_{it}\mathbf{w}_{it-1}'] = \mathbf{O}$. Note that for the rank of $\mathbf{\Pi}_{\rm I}$,

$$\text{rank}(\mathbf{\Pi}_{\rm I}') = G_2 + K_1 .$$

If there is a $(G_2 + K_1) \times 1$ non-zero vector $(\mathbf{a}_2', \mathbf{a}_1')'$ such that

$$\mathbf{\Pi}_2\mathbf{a}_2 + \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix}\mathbf{a}_1 = \mathbf{0} ,$$

then $\mathbf{\Pi}_{22}\mathbf{a}_2 = \mathbf{0}$. Considering the rank condition of identification, $\mathbf{a}_2$ must be $\mathbf{0}$, and thus, $\mathbf{I}_{K_1}\mathbf{a}_1 = \mathbf{0}$. But it contradicts $(\mathbf{a}_2', \mathbf{a}_1')' \neq \mathbf{0}$. Therefore,

$$\mathbf{\Phi} = \mathbf{\Pi}_{\rm I}'\mathbf{J}'\mathcal{E}\left[\mathbf{w}_{it-1}\mathbf{w}_{it-1}'\right]\mathbf{J}\mathbf{\Pi}_{\rm I} > \mathbf{O} .$$

Moreover, using the similar arguments of (6.22),

$$\frac{1}{NT}\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)} \xrightarrow{p} \mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J} .$$

From the above results, the convergence of (6.20) is shown.

Next, we show the convergence of $\mathbf{H}_n^{(f,b)}$. For $g,\ h = 1,\ \cdots,\ (1 + G_2 + K_1)$,

$$\begin{aligned} \frac{1}{NT}\mathbf{e}_g'(\mathbf{y}^{(1,f)'},\mathbf{X}^{(f)'})(\mathbf{y}^{(1,f)},\mathbf{X}^{(f)})\mathbf{e}_h &= \frac{1}{NT}\sum_{i=1}^{N}\mathbf{w}_i^{[g]'}\mathbf{Q}_T\mathbf{w}_i^{[h]} \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}^{[g]}w_{it}^{[h]} - \frac{1}{N}\sum_{i=1}^{N}\bar{w}_i^{[g]}\bar{w}_i^{[h]} \\ &\xrightarrow{p} \mathbf{e}_g'\mathbf{\Theta}_{\mathrm{I}}'\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{I}}\mathbf{e}_h + \mathbf{e}_g'\begin{pmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\mathbf{e}_h\ , \end{aligned}$$

where $\mathbf{w}_i^{[g]} = (w_{i1}^{[g]}, ..., w_{iT}^{[g]})'$ and $\bar{w}_i^{[g]} = (1/T)\sum_t w_{it}^{[g]}$. Therefore, (6.21) is obtained by using the result of convergence for $\mathbf{G}_n^{(f,b)}$.

Solving the first-order condition of the minimization for (4.2), the sampling error of the D-LIML estimator is given by

$$\begin{aligned} \sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1) &= \left(\frac{1}{NT}\mathbf{X}^{(f)'}\mathbf{P}^{(b)}\mathbf{X}^{(f)} - \frac{\lambda}{NT}\mathbf{X}^{(f)'}(\mathbf{I} - \mathbf{P}^{(b)})\mathbf{X}^{(f)'}\right)^{-1} \\ &\times \left(\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} - \frac{\sqrt{NT}\lambda}{NT}\mathbf{X}^{(f)'}(\mathbf{I} - \mathbf{P}^{(b)})\mathbf{u}^{(f)}\right)\ , \end{aligned}$$

where

$$\begin{aligned} \lambda &= \min_{\boldsymbol{\theta}_1} \mathcal{VR}_1\ , \\ \underset{n\times 1}{\mathbf{u}^{(f)}} &= (\mathbf{u}_1^{(f)'},\ \cdots,\ ,\mathbf{u}_N^{(f)'})'\ . \end{aligned}$$

Consider the convergence of λ. From the continuity of the minimum eigenvalue λ, we have the following determinant:

$$\left|\mathbf{\Theta}_{\mathrm{I}}'\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{I}} - \operatorname*{plim}_{T\to\infty}\lambda\begin{pmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\right| = 0\ , \tag{6.24}$$

where zero is a solution due to nonsingularity of $\mathbf{\Theta}_{\mathrm{I}}'\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{I}}$. If there is a solution such that $\operatorname*{plim}_{T\to\infty}\lambda < 0$, then

$$\left|\mathbf{\Theta}_{\mathrm{I}}'\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{I}} - \operatorname*{plim}_{T\to\infty}\lambda\begin{pmatrix} \mathbf{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}\right| > 0\ .$$

This is because that for any $(1 + G_2 + K_1)$ non-zero vector $(\mathbf{a}_2', \mathbf{a}_1')'$ we have

$\mathbf{a}_2'\mathbf{\Omega}\mathbf{a}_2 > 0$. In the case when $\mathbf{a}_2 = \mathbf{0}$ and $\mathbf{a}_1 \neq \mathbf{0}$, we also have

$$\begin{aligned}
(\mathbf{0}', \mathbf{a}_1')\mathbf{\Theta}_{\mathrm{I}}'\mathbf{\Phi}\mathbf{\Theta}_{\mathrm{I}} \begin{pmatrix} \mathbf{0} \\ \mathbf{a}_1 \end{pmatrix} &= (\mathbf{0}', \mathbf{a}_1')\mathbf{\Phi} \begin{pmatrix} \mathbf{0} \\ \mathbf{a}_1 \end{pmatrix} \\
&= \mathbf{a}_1'(\mathbf{I}_{K_1}, \mathbf{O})\mathbf{J}'\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}']\mathbf{J} \begin{pmatrix} \mathbf{I}_{K_1} \\ \mathbf{O} \end{pmatrix} \mathbf{a}_1 \\
&> 0 \, .
\end{aligned}$$

This violates (6.24), and thus, $\operatorname*{plim}_{T\to\infty} \lambda = 0$. Furthermore, from the definition of λ,

$$0 \leq \sqrt{NT}\lambda \;\leq\; \frac{\frac{\sqrt{NT}}{NT}\boldsymbol{\theta}'\mathbf{G}_n^{(f,b)}\boldsymbol{\theta}}{\frac{1}{NT}\boldsymbol{\theta}'\mathbf{H}_n^{(f,b)}\boldsymbol{\theta}} \, .$$

For the numerator,

$$\begin{aligned}
\frac{\sqrt{NT}}{NT}\boldsymbol{\theta}'\mathbf{G}_n^{(f,b)}\boldsymbol{\theta} &= \left(\frac{\mathbf{u}^{(f)\prime}\mathbf{Z}^{(b)}}{\sqrt{NT}}\right)\left(\frac{\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)}}{NT}\right)^{-1}\left(\frac{\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)}}{NT}\right) \\
&= O_p(1) \times o_p(1) \, ,
\end{aligned}$$

that is, $\sqrt{n}\lambda$ converges in probability to zero. From the above, the sampling error of the D-LIML estimator is asymptotically equivalent to that of the D-GMM estimator of Corollary 2.1,

$$\begin{aligned}
\sqrt{NT}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1) &= \left(\frac{1}{NT}\mathbf{X}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{X}^{(f)}\right)^{-1}\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)} + o_p(1) \\
&= \mathbf{\Phi}^{-1}\left(\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)}\right) + o_p(1) \, .
\end{aligned}$$

As the dimension K is finite, each term converges as follows:

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)} &= \frac{1}{NT}\mathbf{X}^{(f)\prime}\mathbf{Z}^{(b)}\left(\frac{1}{NT}\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)}\right)^{-1}\frac{1}{\sqrt{NT}}\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)} \\
&= \frac{1}{\sqrt{NT}}\mathbf{\Pi}_{\mathrm{I}}'\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)} + o_p(1) \, .
\end{aligned}$$

We consider the effects of the forward filter and the convergence in distribution as follows:

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\mathbf{e}_k'\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\mathbf{z}_{it-1}^{(b)\prime}\mathbf{e}_k u_{it}^{(f)\prime} \\
&= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}(w_{it-1}^{[k]}u_{it} - (1-f_t)w_{it-1}^{[k]}u_{it} - w_{it-1}^{[k]}\tilde{u}_{it,T} - \bar{w}_{it-1,0}^{[k]}u_{it}^{(f)}) \, ,
\end{aligned} \tag{6.25}$$

where

$$\tilde{u}_{it,T} = \frac{f_t}{T-t}(u_{it+1} + \cdots + u_{iT}) \ .$$

First, it holds that $\mathcal{E}[\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)}] = \mathbf{0}$. For the second term of (6.25),

$$\begin{aligned}
\mathcal{V}ar\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T-1}(1-f_t)w_{it-1}^{[k]}u_{it}\right] &= \mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}(1-f_t)w_{it-1}^{[k]}u_{it}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T-1}(1-f_t)^2\mathcal{V}ar[w_{it-1}^{[k]}u_{it}] \\
&= O\left(\frac{\log T}{T}\right) \ .
\end{aligned}$$

As for the third term,

$$\begin{aligned}
\mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}w_{it-1}^{[k]}\tilde{u}_{it,T}\right] &= \frac{1}{T}\sum_{s=1}^{T-1}\sum_{t=1}^{T-1}\mathcal{E}[w_{is-1}^{[k]}\mathcal{E}_{t-1}[\tilde{u}_{is,T}\tilde{u}_{it,T}]w_{it-1}^{[k]}] \\
&= \frac{1}{T}\sum_{s=1}^{T-1}\sum_{t=1}^{T-1}\frac{\sigma^2 f_s f_t}{T-s}\mathcal{E}[w_{is-1}^{[k]}w_{it-1}^{[k]}] \\
&= O\left(\frac{\log T}{T}\right) \ .
\end{aligned}$$

Finally, regarding the fourth term,

$$\begin{aligned}
\mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\bar{w}_{it-1,0}^{[k]}u_{it}^{(f)}\right] &= \frac{\sigma^2}{T}\sum_{t=1}^{T-1}\mathcal{E}[\bar{w}_{it-1,0}^{[k]2}] \\
&= O\left(\frac{\log T}{T}\right) \ .
\end{aligned}$$

Therefore, we obtain the following:

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\mathbf{\Pi}_{\text{I}}^{\prime}\mathbf{Z}^{(b)\prime}\mathbf{u}^{(f)} &= \frac{1}{\sqrt{NT}}\mathbf{\Pi}_{\text{I}}^{\prime}\mathbf{J}^{\prime}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\mathbf{w}_{it-1}u_{it} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(\mathbf{0},\ \sigma^2\mathbf{\Phi}) \ ,
\end{aligned}$$

where the asymptotic normality is based on the standard central limit theorem for autoregressive processes (cf. Anderson (1978)). □

Proof of Lemma 2.2 : [i] We first show that the pseudo log-likelihood function

$\mathcal{L}_{2\cdot 0}$ can be also expressed by using the forward filter. Put $\Delta\mathbf{v}_{*i} = (\Delta\mathbf{v}_{*i}^{(1)'},\ \cdots,\ \Delta\mathbf{v}_{*i}^{(G)'})'$, where

$$\underset{(T-1)\times 1}{\Delta\mathbf{v}_{*i}^{(g)}} = \left(\Delta v_{i2}^{(g)},\ ,\cdots,\ \Delta v_{iT}^{(g)}\right)' \ .$$

Let $\vec{\mathbf{J}}_{T-1}$ be the matrix such that $\vec{\mathbf{J}}'_{T-1}\Delta\mathbf{v}_{*i} = \Delta\mathbf{v}_i$; that is, $\vec{\mathbf{J}}'_{T-1}$ sorts $\Delta\mathbf{v}_{*i}$. Then,

$$\mathcal{L}_{2\cdot 0} = -\frac{N}{2}\log|\mathcal{E}[\mathbf{v}_{*i}\mathbf{v}_{*i}']| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_{*i}'\left(\mathcal{E}[\mathbf{v}_{*i}\mathbf{v}_{*i}']\right)^{-1}\Delta\mathbf{v}_{*i}\ ,$$

because $\vec{\mathbf{J}}'_{T-1}$ has the following properties:

$$(\vec{\mathbf{J}}'_{T-1})^{-1} = \vec{\mathbf{J}}_{T-1}\ ,\quad |\vec{\mathbf{J}}_{T-1}| = 1\ .$$

From $\mathbf{D}_f$ of (2.8),

$$(\mathbf{D}_T\mathbf{D}_T')^{-\frac{1}{2}}\Delta\mathbf{v}_{*i}^{(g)} = \mathbf{v}_i^{(g,f)}\ .$$

Hence, using the following transformation:

$$\underset{G(T-1)\times G(T-1)}{\mathbf{T}} = \mathbf{I}_G\otimes(\mathbf{D}_T\mathbf{D}_T')^{-\frac{1}{2}}\ ,$$

we obtain

$$\begin{aligned}\mathcal{L}_{2\cdot 0} &= -\frac{N}{2}\log|\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}\right]| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{T}\Delta\mathbf{v}_{*i})'\left(\mathbf{T}\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}'\right]\mathbf{T}'\right)^{-1}\mathbf{T}\Delta\mathbf{v}_{*i} \\ &= N\log|\mathbf{T}| - \frac{N}{2}\log|\mathcal{E}\left[\mathbf{v}_{*i}^{(f)}\mathbf{v}_{*i}^{(f)'}\right]| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_{*i}^{(f)'}\left(\mathcal{E}\left[\mathbf{v}_{*i}^{(f)}\mathbf{v}_{*i}^{(f)'}\right]\right)^{-1}\mathbf{v}_{*i}^{(f)}\ .\end{aligned}$$

If we use $\vec{\mathbf{J}}'_{T-1}$ again, then

$$\mathcal{L}_{2\cdot 0} = N\log|\mathbf{T}| + \mathcal{L}_f\ ,$$

where

$$\begin{aligned}\mathcal{L}_f &= -\frac{N}{2}\log|\mathbf{I}_{T-1}\otimes\mathbf{\Omega}| - \frac{1}{2}\sum_{i=1}^{N}\mathbf{v}_i^{(f)'}(\mathbf{I}_{T-1}\otimes\mathbf{\Omega})^{-1}\mathbf{v}_i^{(f)} \\ &= -\frac{n}{2}\log|\mathbf{\Omega}| - \frac{1}{2}\mathrm{tr}\left((\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\mathbf{\Pi})'(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\mathbf{\Pi})\mathbf{\Omega}^{-1}\right)\ ,\end{aligned}$$

and

$$\begin{aligned}\underset{G(T-1)\times 1}{\mathbf{v}_i^{(f)}} &= \left(\mathbf{v}_{i1}^{(f)'},\ \cdots,\ \mathbf{v}_{i(T-1)}^{(f)'}\right)'\ , \\ \underset{N(T-1)\times G}{\mathbf{Y}^{(f)}} &= \left(\mathbf{y}^{(1,f)}, \mathbf{Y}^{(2,f)}\right)\ , \\ \underset{N(T-1)\times(K_1+K_2)}{\mathbf{Z}^{(f)}} &= \left(\mathbf{Z}^{(1,f)},\ \mathbf{Z}^{(2,f)}\right)\ .\end{aligned}$$

Therefore, considering $\mathcal{L}_f$ is sufficient for maximizing $\mathcal{L}_{2\cdot 0}$. We show that the concentrated likelihood function of $\mathcal{L}_{2\cdot 0}$ becomes $\mathcal{VR}_{2\cdot 0}$ based on the following:

$$\begin{aligned}
\mathbf{G}_n^{(f,f)} &= \begin{pmatrix} \mathbf{y}^{(1,f)\prime} \\ \mathbf{X}^{(f)\prime} \end{pmatrix} \mathbf{P}^{(f)} \left(\mathbf{y}^{(1,f)}, \mathbf{X}^{(f)}\right) , \\
\mathbf{H}_n^{(f,f)} &= \begin{pmatrix} \mathbf{y}^{(1,f)\prime} \\ \mathbf{X}^{(f)\prime} \end{pmatrix} [\mathbf{I}_n - \mathbf{P}^{(f)}] \left(\mathbf{y}^{(1,f)}, \mathbf{X}^{(f)}\right) .
\end{aligned} \tag{6.26}$$

We maximize $\mathcal{L}_f$ with respect to $\{\boldsymbol{\beta}_2, \boldsymbol{\Pi}_{22}, \boldsymbol{\Pi}_{1\cdot}, \boldsymbol{\Omega}\}$ and consider the concentrated log-likelihood function for $\boldsymbol{\beta}_2$, where

$$\underset{(K_1+K_2)\times(1+G_2)}{\boldsymbol{\Pi}} = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{22}\boldsymbol{\beta}_2 & \boldsymbol{\Pi}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Pi}_{1\cdot} \\ \boldsymbol{\Pi}_{2\cdot} \end{pmatrix} .$$

From the exclusion restrictions, the structural coefficient $\boldsymbol{\gamma}_1$ becomes

$$\boldsymbol{\gamma}_1 = \boldsymbol{\pi}_{11} - \boldsymbol{\Pi}_{12}\boldsymbol{\beta}_2 , \tag{6.27}$$

but this is not a constraint in maximization.

The following proof is almost the same as that of Morimune (1984, Appen.) up to the derivation of (6.37). Although $\mathbf{z}_{it}^{(f)}$ is not a valid instrumental variable, it does not affect the derivation of the concentrated log-likelihood function. By solving the first-order condition for $\boldsymbol{\Omega}$, we have

$$\boldsymbol{\Omega} = \frac{1}{n}(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\boldsymbol{\Pi})'(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\boldsymbol{\Pi}) . \tag{6.28}$$

If we substitute the above equation into the log-likelihood function, then maximizing the log-likelihood is equivalent to minimizing the following with respect to $(\boldsymbol{\beta}_2, \boldsymbol{\Pi}_{22}, \boldsymbol{\Pi}_{1\cdot})$,

$$|\boldsymbol{\Omega}| = \left|\frac{1}{n}(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\boldsymbol{\Pi})'(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\boldsymbol{\Pi})\right| . \tag{6.29}$$

The derivative of determinant for a nonsingular matrix $\mathbf{A}$ is as follows (cf. Abadir and Mangus, 2005),

$$\frac{\partial|\mathbf{A}|}{\partial x} = |\mathbf{A}|\mathrm{tr}\left(\mathbf{A}'^{-1}\frac{\partial \mathbf{A}}{\partial x}\right) .$$

Then, from the derivative $|\boldsymbol{\Omega}|$ of (6.29) with respect to $\boldsymbol{\Pi}_{1\cdot}$, the first-order condition becomes

$$\boldsymbol{\Pi}_{1\cdot} = (\mathbf{Z}^{(1,f)\prime}\mathbf{Z}^{(1,f)})^{-1}\mathbf{Z}^{(1,f)\prime}(\mathbf{Y}^{(f)} - \mathbf{Z}^{(2,f)}\boldsymbol{\Pi}_{2\cdot}) . \tag{6.30}$$

Similarly, for $\mathbf{\Pi}_{22}$ we obtain the first-order condition,

$$\mathbf{Z}^{(2,f)\prime}(\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\mathbf{\Pi})\mathbf{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} = \mathbf{0} .$$

$\mathbf{Z}^{(f)}\mathbf{\Pi}$ included in the above equation can be decomposed into $\mathbf{Z}^{(1,f)}\mathbf{\Pi}_{1\cdot}+\mathbf{Z}^{(2,f)}\mathbf{\Pi}_{22}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})$. By substituting (6.30),

$$\mathbf{\Pi}_{22} = (\mathbf{Z}_0^{(2,f)\prime}\mathbf{Z}_0^{(2,f)})^{-1}\mathbf{Z}_0^{(2,f)\prime}\mathbf{Y}^{(f)}\mathbf{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix}(\mathbf{\Sigma}^{22})^{-1} , \tag{6.31}$$

where $\mathbf{P}_1^{(f)} = \mathbf{Z}^{(1,f)}(\mathbf{Z}^{(1,f)\prime}\mathbf{Z}^{(1,f)})^{-1}\mathbf{Z}^{(1,f)\prime}$, $\mathbf{Z}_0^{(2,f)} = (\mathbf{I}_n - \mathbf{P}_1^{(f)})\mathbf{Z}_2^{(f)}$, and $\mathbf{\Sigma}^{22} = (\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\mathbf{\Omega}^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})'$ is the submatrix of the following:

$$\begin{aligned} \mathbf{\Sigma}^{-1} &= \begin{pmatrix} \sigma^{11} & \boldsymbol{\sigma}^{12\prime} \\ \boldsymbol{\sigma}^{21} & \mathbf{\Sigma}^{22} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0}' \\ \boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{pmatrix}\mathbf{\Omega}^{-1}\begin{pmatrix} 1 & \boldsymbol{\beta}_2' \\ \mathbf{0} & \mathbf{I}_{G_2} \end{pmatrix} . \end{aligned}$$

Conversely, from the inverse matrix on the right-hand side, $\mathbf{\Sigma}$ can be also expressed as follows:

$$\begin{aligned} \mathbf{\Sigma} &= \begin{pmatrix} \sigma^2 & \boldsymbol{\sigma}_{12}' \\ \boldsymbol{\sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} \\ &= \begin{pmatrix} 1 & -\boldsymbol{\beta}_2' \\ \mathbf{0} & \mathbf{I}_{G_2} \end{pmatrix}\begin{pmatrix} \omega_{11} & \boldsymbol{\omega}_{12}' \\ \boldsymbol{\omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{pmatrix} , \end{aligned}$$

where $\sigma^2 = \mathcal{E}[u_{it}^2]$ corresponds to the error term of the first structural equation. We rewrite the residual by (6.30) and (6.31). For $\mathbf{P}_0^{(f)} = \mathbf{Z}_0^{(2,f)}(\mathbf{Z}_0^{(2,f)\prime}\mathbf{Z}_0^{(2,f)})^{-1}\mathbf{Z}_0^{(2,f)\prime}$,

$$\begin{aligned} \mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\mathbf{\Pi} &= \mathbf{Y}^{(f)} - \mathbf{Z}^{(f,1)}\mathbf{\Pi}_{1\cdot} - \mathbf{Z}^{(f,2)}\mathbf{\Pi}_{22}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \\ &= (\mathbf{I}_n - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)} - \mathbf{P}_0^{(f)}\mathbf{Y}^{(f)}\mathbf{\Omega}^{-1}(\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \\ &= (\mathbf{I}_n - \mathbf{P}^{(f)})\mathbf{Y}^{(f)} + \mathbf{P}_0^{(f)}\mathbf{Y}^{(f)}\left[\mathbf{I}_G - \mathbf{\Omega}^{-1}\begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix}(\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\right] , \end{aligned}$$

the third equality is based on the following orthogonal decomposition,

$$\begin{aligned} \mathbf{P}^{(f)} &= \mathbf{Z}^{(f)}(\mathbf{Z}^{(f)\prime}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(f)\prime} \\ &= \mathbf{Z}^{(f)}\mathbf{C}^{(f)}[(\mathbf{Z}^{(f)}\mathbf{C}^{(f)})'\mathbf{Z}^{(f)}\mathbf{C}^{(f)}]^{-1}(\mathbf{Z}^{(f)}\mathbf{C}^{(f)})' \\ &= \mathbf{P}_1^{(f)} + \mathbf{P}_0^{(f)} , \end{aligned}$$

where $\mathbf{C}^{(f)} = (\mathbf{Z}^{(f)'}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(f)'}(\mathbf{Z}^{(1,f)}, \mathbf{Z}_0^{(2,f)})$. To simplify the second term, we use the relation that

$$
\begin{aligned}
& \left[\mathbf{I}_G - \mathbf{\Omega}^{-1} \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} (\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\right] \\
= & \begin{pmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}' \\ \boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{pmatrix} \left[\mathbf{I}_G - \mathbf{\Omega}^{-1} \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix} (\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\right] \\
= & \begin{pmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta}_2 & \mathbf{I}_{G_2} \end{pmatrix} \begin{pmatrix} (1, \mathbf{0}') - \boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) \\ \mathbf{O} \end{pmatrix} \\
= & \begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix} \Big((1, \mathbf{0}') - \boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\Big) \\
= & \boldsymbol{\beta} \left((1, \mathbf{0}') + \frac{1}{\sigma^2}\boldsymbol{\sigma}_{12}'(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\right) \\
= & \boldsymbol{\beta}\frac{1}{\sigma^2}(\sigma^2 + \boldsymbol{\omega}_{12}'\boldsymbol{\beta}_2 - \boldsymbol{\beta}_2'\mathbf{\Omega}_{22}\boldsymbol{\beta}_2, \mathbf{0}' + \boldsymbol{\omega}_{12}' - \boldsymbol{\beta}_2\mathbf{\Omega}_{22}) \\
= & \frac{1}{\sigma^2}\boldsymbol{\beta}(\omega_{11} - \boldsymbol{\omega}_{12}'\boldsymbol{\beta}_2, \boldsymbol{\omega}_{12}' - \boldsymbol{\beta}_2\mathbf{\Omega}_{22}) \\
= & \frac{\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega}}{\boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta}} ,
\end{aligned}
$$

where $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}_2')'$ and the fourth equality is from $-\boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1} = (1/\sigma^2)\boldsymbol{\sigma}_{12}$. This is because that

$$
\boldsymbol{\sigma}_{12} = -\frac{1}{\sigma^{11} - \boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1}\boldsymbol{\sigma}^{12}}\boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1} = -\sigma^2\boldsymbol{\sigma}^{12'}(\mathbf{\Sigma}^{22})^{-1} .
$$

Then, the residual is simplified as follows:

$$
\mathbf{Y}^{(f)} - \mathbf{Z}^{(f)}\mathbf{\Pi} = (\mathbf{I}_n - \mathbf{P}^{(f)})\mathbf{Y}^{(f)} + \frac{1}{\sigma^2}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega} .
$$

Using this result, the LIML estimator of $\mathbf{\Omega}$ satisfies the following:

$$
\mathbf{\Omega} = \tilde{\mathbf{\Omega}} + \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)'}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{n\sigma^4}\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{\Omega} , \tag{6.32}
$$

where

$$
\tilde{\mathbf{\Omega}} = \frac{1}{n}\mathbf{Y}^{(f)'}(\mathbf{I}_n - \mathbf{P}^{(f)})\mathbf{Y}^{(f)} . \tag{6.33}
$$

This is a fixed-effects estimator for $\mathbf{\Omega}$. In addition, from the relation of (6.32) the following should be satisfied,

$$
\begin{aligned}
\sigma^2 &= \boldsymbol{\beta}'\mathbf{\Omega}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}'\tilde{\mathbf{\Omega}}\boldsymbol{\beta} + \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)'}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{n} .
\end{aligned}
$$

Since (6.29) is minimized for $\boldsymbol{\beta}_2$, consider the determinant of (6.32).

$$|\mathbf{A} + a\mathbf{b}\mathbf{b}'| = |\mathbf{A}|(1 + a\mathbf{b}'\mathbf{A}^{-1}\mathbf{b}) ,$$

From this formula, we obtain

$$|\boldsymbol{\Omega}| = |\tilde{\boldsymbol{\Omega}}| \left(1 + \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{n\sigma^4} \boldsymbol{\beta}'\boldsymbol{\Omega}\tilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Omega}\boldsymbol{\beta} \right) . \tag{6.34}$$

$\boldsymbol{\Omega}$ is still included in the right-hand side. To remove this term we multiply $\boldsymbol{\beta}$ on the right of (6.32),

$$\boldsymbol{\Omega}\boldsymbol{\beta} = \tilde{\boldsymbol{\Omega}}\boldsymbol{\beta} + \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{n\sigma^2} \boldsymbol{\Omega}\boldsymbol{\beta} ,$$

then,

$$\boldsymbol{\Omega}\boldsymbol{\beta} = (1 + \lambda)\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta} , \tag{6.35}$$

i.e., $\boldsymbol{\Omega}\boldsymbol{\beta}$ is solved. From the relation of (6.33),

$$\lambda = \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}(\mathbf{I}_n - \mathbf{P}^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}} . \tag{6.36}$$

Substituting (6.35) into (6.34), we obtain the following by using $\boldsymbol{\beta}'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta} = \sigma^2/(1+\lambda)$,

$$|\boldsymbol{\Omega}| = (1 + \lambda)|\tilde{\boldsymbol{\Omega}}| . \tag{6.37}$$

Since the unknown parameter included in (6.37) is $\boldsymbol{\beta}_2$ this is the concentrated log-likelihood function, and the minimization of (6.29) can be achieved by minimizing λ. Therefore, $\hat{\boldsymbol{\beta}}_2$ minimizing (6.36) is the maximum likelihood estimator.

Finally, we consider the maximum likelihood estimator of the structural coefficient $\boldsymbol{\gamma}_1$. From (6.27), the estimator must satisfy the following:

$$\begin{aligned} \boldsymbol{\gamma}_{1\beta} &= \boldsymbol{\Pi}_{1\cdot}\boldsymbol{\beta} \\ &= (\mathbf{Z}^{(1,f)\prime}\mathbf{Z}^{(1,f)})^{-1}\mathbf{Z}^{(1,f)\prime}(\mathbf{y}^{(1,f)} - \mathbf{Y}^{(2,f)}\boldsymbol{\beta}_2) , \end{aligned} \tag{6.38}$$

where the second equality is from that $\boldsymbol{\Pi}_{2\cdot}\boldsymbol{\beta} = \boldsymbol{\Pi}_{22}\boldsymbol{\beta}_2 - \boldsymbol{\Pi}_{22}\boldsymbol{\beta}_2 = \mathbf{0}$.

Using $\mathbf{G}_n^{(f,f)}$, $\mathbf{H}_n^{(f,f)}$, and $\boldsymbol{\gamma}_{1\beta}$ of (6.27), we rewrite λ of (6.36) as follows:

$$\begin{aligned}
\lambda(\boldsymbol{\beta}_2) &= \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}(\mathbf{P}^{(f)} - \mathbf{P}_1^{(f)})\mathbf{Y}^{(f)}\boldsymbol{\beta}}{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_{1\beta}'\,)\,\mathbf{H}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_{1\beta}\end{pmatrix}} \\
&= \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{Y}^{(f)}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}\mathbf{Z}^{(1,f)}\boldsymbol{\gamma}_{1\beta}}{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_{1\beta}'\,)\,\mathbf{H}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_{1\beta}\end{pmatrix}} \\
&= \frac{\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{Y}^{(f)}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{Y}^{(f)\prime}\mathbf{Z}^{(1,f)}\boldsymbol{\gamma}_{1\beta} + \boldsymbol{\gamma}_{1\beta}'\mathbf{Z}^{(1,f)\prime}\mathbf{Z}^{(1,f)}\boldsymbol{\gamma}_{1\beta}}{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_{1\beta}'\,)\,\mathbf{H}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_{1\beta}\end{pmatrix}} \\
&= \frac{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_{1\beta}'\,)\,\mathbf{G}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_{1\beta}\end{pmatrix}}{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_{1\beta}'\,)\,\mathbf{H}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_{1\beta}\end{pmatrix}}\,,
\end{aligned}$$

where the first equality is from that $(\mathbf{I}_n - \mathbf{P}^{(f)})\mathbf{Z}^{(1,f)} = \mathbf{O}$. The second and third equalities are due to $\mathbf{Z}^{(1,f)}\boldsymbol{\gamma}_{1\beta} = \mathbf{P}_1^{(f)}\mathbf{Y}^{(f)}\boldsymbol{\beta}$ and $\mathbf{P}_1^{(f)}\mathbf{Z}^{(1,f)} = \mathbf{Z}^{(1,f)}$, respectively.

Consider the minimization problem with respect to $\boldsymbol{\beta}_2$ and $\boldsymbol{\gamma}_1$,

$$\lambda(\boldsymbol{\beta}_2,\ \boldsymbol{\gamma}_1) = \frac{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_1'\,)\,\mathbf{G}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_1\end{pmatrix}}{(\,\boldsymbol{\beta}', -\boldsymbol{\gamma}_1'\,)\,\mathbf{H}_n^{(f,f)}\begin{pmatrix}\boldsymbol{\beta}\\ -\boldsymbol{\gamma}_1\end{pmatrix}}\,. \tag{6.39}$$

The first-order condition of $\boldsymbol{\gamma}_1$ is given by

$$-\mathbf{Z}^{(1,f)\prime}\mathbf{Y}^{(f)}\boldsymbol{\beta}_2 + (\mathbf{Z}^{(1,f)\prime}\mathbf{Z}^{(1,f)})\boldsymbol{\gamma}_1 = \mathbf{0}\,.$$

This is the same as (6.38), and thus, the maximum likelihood estimator $(\hat{\boldsymbol{\beta}}_2, \boldsymbol{\gamma}_{1\hat{\beta}})$ can be obtained by this minimization problem. Therefore, (6.39) is the concentrated log-likelihood function for $(\boldsymbol{\beta}_2,\ \boldsymbol{\gamma}_1)$.

[*ii*] We first confirm that Lemma 2.1 holds even for a multivariate model. For $g = 1, \cdots, G$, put

$$\underset{T\times 1}{\Delta\mathbf{v}_{*i}^{(g)}} = \left(\Delta y_{i1}^{(g)},\ \Delta v_{i2}^{(g)},\ ,\cdots,\ \Delta v_{iT}^{(g)}\right)'\,,$$

then, for $\mathbf{L}$ of (3.21) we have

$$\mathbf{L}\Delta\mathbf{v}_{*i}^{(g)} = \xi_i^{(g)}\boldsymbol{\iota} + \mathbf{v}_i^{(g)} \ .$$

Therefore, if the transformation is given by $\mathbf{T} = \mathbf{I}_G \otimes \mathbf{L}$ for $\Delta\mathbf{v}_{*i} = (\Delta\mathbf{v}_{*i}^{(1)'}, \ \cdots, \ \Delta\mathbf{v}_{*i}^{(G)'})'$, then

$$\begin{aligned}
\mathcal{L}_2 &= -\frac{N}{2}\log|\mathbf{T}\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]\mathbf{T}^{'}| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{T}\Delta\mathbf{v}_{*i})^{'}\left(\mathbf{T}\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]\mathbf{T}^{'}\right)^{-1}\mathbf{T}\Delta\mathbf{v}_{*i} \\
&= -N\log|\mathbf{T}| - \frac{N}{2}\log|\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_{*i}^{'}\left(\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]\right)^{-1}\Delta\mathbf{v}_{*i} \ .
\end{aligned}$$

For $\vec{\mathbf{J}}_T^{'}\mathbf{v}_{*i} = \mathbf{v}_i^{*}$,

$$\begin{aligned}
\mathcal{L}_{2\Delta} &= -\frac{N}{2}\log|\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_{*i}^{'}\left(\mathcal{E}\left[\Delta\mathbf{v}_{*i}\Delta\mathbf{v}_{*i}^{'}\right]\right)^{-1}\Delta\mathbf{v}_{*i} \\
&= \frac{N}{2}\log|\boldsymbol{\Omega}_{2\Delta}| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_i^{*'}\boldsymbol{\Omega}_{2\Delta}^{-1}\Delta\mathbf{v}_i^{*} \ .
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Omega}_{2\Delta} &= \begin{pmatrix} \boldsymbol{\Omega}_1 & -\boldsymbol{\Omega} & \mathbf{O} & \cdots & \mathbf{O} \\ -\boldsymbol{\Omega} & 2\boldsymbol{\Omega} & -\boldsymbol{\Omega} & \cdots & \mathbf{O} \\ \mathbf{O} & -\boldsymbol{\Omega} & 2\boldsymbol{\Omega} & \cdots & \mathbf{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & \cdots & -\boldsymbol{\Omega} & 2\boldsymbol{\Omega} \end{pmatrix} , \\
\boldsymbol{\Omega}_1 &= \mathcal{E}[\Delta\mathbf{y}_{i1}\Delta\mathbf{y}_{i1}^{'}] \\
&= (\mathbf{I}_2 - \boldsymbol{\Pi}_\theta^{'})\boldsymbol{\Gamma}_0 + \boldsymbol{\Gamma}_0(\mathbf{I}_2 - \boldsymbol{\Pi}_\theta) \ ,
\end{aligned}$$

and

$$\begin{aligned}
\underset{GT\times 1}{\Delta\mathbf{v}_i^{*}} &= \begin{bmatrix} \Delta\mathbf{y}_{i1} \\ \Delta\mathbf{v}_{it} \end{bmatrix} \\
&= \begin{bmatrix} \Delta\mathbf{y}_{i1} \\ \Delta\mathbf{y}_{it} - \boldsymbol{\Pi}_\theta^{'}\Delta\mathbf{y}_{it-1} \end{bmatrix} , \\
\boldsymbol{\Pi}_\theta^{'} &= \begin{pmatrix} \boldsymbol{\gamma}_1^{'} + \boldsymbol{\beta}_2^{'}\boldsymbol{\Pi}_{12}^{'} & \boldsymbol{\beta}_2^{'}\boldsymbol{\Pi}_{22}^{'} \\ \boldsymbol{\Pi}_{12}^{'} & \boldsymbol{\Pi}_{22}^{'} \end{pmatrix} \\
&= \boldsymbol{\Pi}^{'} \ \ (say,) \ .
\end{aligned}$$

Moreover, for $\mathcal{L}_2$,

$$\boldsymbol{\Omega}_\xi = (\mathbf{I}_2 - \boldsymbol{\Pi}_\theta^{'})\boldsymbol{\Gamma}_0(\mathbf{I}_2 - \boldsymbol{\Pi}_\theta) \ .$$

Therefore, the parameters of $\mathcal{L}_2$ have a one-to-one correspondence with those of $\mathcal{L}_{2\Delta}$:

$$\{\boldsymbol{\theta}_1,\ \boldsymbol{\Pi}_\theta',\ \boldsymbol{\Omega},\ \boldsymbol{\Omega}_\xi\}\ \rightleftarrows\ \{\boldsymbol{\theta}_1,\ \boldsymbol{\Pi}_\theta',\ \boldsymbol{\Omega},\ \boldsymbol{\Omega}_1\}\ .$$

Then, it is sufficient to consider the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\text{TL}}$ in $\mathcal{L}_{2\Delta}$.

Next, we decompose $\mathcal{L}_{2\Delta}$ into the pseudo log-likelihood $\mathcal{L}_{2\cdot 0}$ and $\mathcal{R}_0$, which is the remaining term related to initial values:

$$\mathcal{L}_{2\Delta} = \mathcal{L}_{2\cdot 0} + \mathcal{R}_0\ .$$

For $\boldsymbol{\Omega}_{2\Delta}^{-1}$, we use another expression of (6.18) (cf. Rao (1965)),

$$(\mathbf{C}-\mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}'(\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1}\ ,$$

then,

$$\begin{aligned}\boldsymbol{\Omega}_{2\Delta}^{-1} &= \begin{pmatrix} \boldsymbol{\Omega}_{2\Delta}^{11} & \boldsymbol{\Omega}_{2\Delta}^{21'} \\ \boldsymbol{\Omega}_{2\Delta}^{21} & \boldsymbol{\Omega}_{\Delta}^{-1} + \boldsymbol{\Omega}_{2\Delta}^{21}(\boldsymbol{\Omega}_{2\Delta}^{11})^{-1}\boldsymbol{\Omega}_{2\Delta}^{21'} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{O} & \mathbf{O}' \\ \mathbf{O} & \boldsymbol{\Omega}_{\Delta}^{-1} \end{pmatrix} + \boldsymbol{\Omega}_0\ \ (say,)\ .\end{aligned}$$

Therefore, we obtain

$$\mathcal{R}_0 = \frac{N}{2}\log|\boldsymbol{\Omega}_{2\Delta}^{11}| - \frac{1}{2}\sum_{i=1}^{N}\Delta\mathbf{v}_i^{*'}\boldsymbol{\Omega}_0\Delta\mathbf{v}_i^{*}\ .$$

Consider the order of $\mathcal{R}_0/n$, where $n = NT$. The following submatrix consists of the 1st to G-th rows of $\boldsymbol{\Omega}_0$,

$$\underset{G\times T}{\begin{pmatrix} \boldsymbol{\Omega}_{2\Delta}^{11} & \boldsymbol{\Omega}_{2\Delta}^{21'} \end{pmatrix}}\ ,$$

and let $\boldsymbol{\Omega}_{0t}$ be the $(1,\ t)$ block matrix of the above submatrix. Then, Binder et al. (2005) show that

$$\begin{aligned}\underset{G\times G}{\boldsymbol{\Omega}_{0t}} &= (T+1-t)\,[\,T\boldsymbol{\Omega}_1 - (T-1)\boldsymbol{\Omega}\,]^{-1} \\ &= \frac{T+1-t}{T}\boldsymbol{\Omega}_{0T}\ \ (say,)\ ,\end{aligned}$$

i.e., $\mathbf{\Omega}_{0t} = O((T-t)/T)$. For the leading term of $\mathcal{R}_0/n$, put

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{s,t=2}^{T} \Delta\mathbf{v}_{is}^{'}\mathbf{\Omega}_{2\Delta}^{21}(\mathbf{\Omega}_{2\Delta}^{11})^{-1}\mathbf{\Omega}_{2\Delta}^{21'}\Delta\mathbf{v}_{it} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{s,t=2}^{T} \Delta\mathbf{v}_{is}^{'}\mathbf{\Omega}_{0s}^{'}(\mathbf{\Omega}_{01})^{-1}\mathbf{\Omega}_{0t}\Delta\mathbf{v}_{it} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{s,t=2}^{T} \Delta\mathbf{y}_{is}^{'}\mathbf{\Phi}_{1st}\Delta\mathbf{y}_{it} + \Delta\mathbf{y}_{is-1}^{'}\mathbf{\Phi}_{2st}\Delta\mathbf{y}_{it} + \Delta\mathbf{y}_{is-1}^{'}\mathbf{\Phi}_{3st}\Delta\mathbf{y}_{it-1}\ ,
\end{aligned}
\tag{6.40}
$$

where

$$
\begin{aligned}
\mathbf{\Phi}_{1st} &= \frac{(T+1-s)(T+1-t)}{T^2}\mathbf{\Omega}_{0T}^{'}(\mathbf{\Omega}_{01})^{-1}\mathbf{\Omega}_{0T}\ , \\
\mathbf{\Phi}_{2st} &= -2\mathbf{\Pi}_{\theta}\mathbf{\Phi}_{1st}\ , \\
\mathbf{\Phi}_{3st} &= \mathbf{\Pi}_{\theta}\mathbf{\Phi}_{1st}\mathbf{\Pi}_{\theta}^{'}\ .
\end{aligned}
\tag{6.41}
$$

Regarding the first term of (6.40), when N is fixed, it follows that

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{s,t=2}^{T} \mathcal{E}\left[|\Delta\mathbf{y}_{is}^{'}\mathbf{\Phi}_{1st}\Delta\mathbf{y}_{it}|\right] = O\left(\frac{(\log T)^2}{T}\right)\ . \tag{6.42}
$$

Therefore, this term converges in 1th mean to zero. Since the third term of (6.40) is the same order, we have that

$$
\frac{1}{n}\mathcal{R}_0 \ \xrightarrow{p}\ 0\ .
$$

Therefore,

$$
\frac{1}{n}\mathcal{L}_{2\Delta} \ \xrightarrow{p}\ \frac{1}{n}\mathcal{L}_{2\cdot 0}\ .
$$

Thus, the maximization point $\hat{\boldsymbol{\theta}}_{\text{TL}}$ of $\mathcal{L}_{2\Delta}$ converges in probability to that of $\mathcal{L}_{2\cdot 0}$. From the results of Theorems 2.9 and 2.10, the limit of the maximization point is $\boldsymbol{\theta}_1$.

Finally, we show that the asymptotic distributions are the same. The log-likelihood function has the following parameters:

$$
\mathcal{L}_{2\Delta}(\boldsymbol{\phi}) = \mathcal{L}_{2\cdot 0}(\boldsymbol{\phi}_2) + \mathcal{R}_0(\boldsymbol{\phi})\ ,
$$

where

$$
\boldsymbol{\phi}_2 = \left(\boldsymbol{\theta}_1^{'},\ \mathrm{vec}(\mathbf{\Pi}_{12}),\ \mathrm{vec}(\mathbf{\Pi}_{22}),\ \mathrm{vec}(\mathbf{\Omega})^{'}\right)^{'}\ ,\quad \boldsymbol{\phi} = \left(\boldsymbol{\phi}_2^{'},\ \mathrm{vec}(\mathbf{\Omega}_1)^{'}\right)^{'}\ .
$$

Note that $\mathcal{L}_{2\cdot 0}$ depends only on $\boldsymbol{\phi}_2$. The maximum likelihood estimator $\hat{\boldsymbol{\phi}}$ satisfies the following:

$$\mathbf{0} = \mathbf{s}_2(\hat{\boldsymbol{\phi}}_2) + \mathbf{s}_0(\hat{\boldsymbol{\phi}}) \ ,$$

where

$$\mathbf{s}_2 = \frac{\partial \mathcal{L}_{2\cdot 0}}{\partial \boldsymbol{\phi}_2} \ , \quad \mathbf{s}_0 = \frac{\partial \mathcal{R}_0}{\partial \boldsymbol{\phi}} \ .$$

From the Taylor series for only $\mathbf{s}_2$,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2) &= -\mathbf{H}_{\phi_2}^{-1}\frac{1}{\sqrt{n}}\mathbf{s}_2(\boldsymbol{\phi}_2) + \mathbf{H}_{\phi_2}^{-1}\frac{1}{\sqrt{n}}\mathbf{s}_0(\hat{\boldsymbol{\phi}}) + o_p(1) \ , \\ \mathbf{H}_{\phi_2} &= \operatorname*{plim}_{T\to\infty} \frac{1}{n}\frac{\partial \mathbf{s}_2}{\partial \boldsymbol{\phi}_2'} \ . \end{aligned}$$

Then, the leading term of the second term $(1/\sqrt{n})\mathbf{s}_0(\hat{\boldsymbol{\phi}})$ is given by

$$\begin{aligned} &\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s,t=2}^{T} \Delta\mathbf{y}_{is}'\frac{\partial\hat{\boldsymbol{\Phi}}_{1st}}{\partial\phi_k}\Delta\mathbf{y}_{it} + \Delta\mathbf{y}_{is-1}'\frac{\partial\hat{\boldsymbol{\Phi}}_{2st}}{\partial\phi_k}\Delta\mathbf{y}_{it} + \Delta\mathbf{y}_{is-1}'\frac{\partial\hat{\boldsymbol{\Phi}}_{3st}}{\partial\phi_k}\Delta\mathbf{y}_{it-1} \\ = \ & O_p\left(\frac{(\log T)^2}{\sqrt{T}}\right) \ , \end{aligned}$$

where $\partial\hat{\boldsymbol{\Phi}}_{1st}/\partial\phi_k$ stands for the derivative of (6.41) for each element of $\boldsymbol{\phi}$ evaluated at $\hat{\boldsymbol{\phi}}$. Then,

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2) &= -\mathbf{H}_{\phi_2}^{-1}\frac{1}{\sqrt{n}}\mathbf{s}_2(\boldsymbol{\phi}_2) + o_p(1) \\ &= \sqrt{n}(\breve{\boldsymbol{\phi}}_2 - \boldsymbol{\phi}_2) \ . \end{aligned}$$

Therefore, they have the same asymptotic distribution.

When N is fixed, the distribution of $\sqrt{n}(\breve{\boldsymbol{\theta}}_{\text{PL}} - \boldsymbol{\theta}_1)$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{DL}} - \boldsymbol{\theta}_1)$ are also asymptotically equivalent from Theorems 2.9 and 2.10. □

Proof of Theorem 2.10 : [i] The sampling error is given by

$$\begin{aligned} \sqrt{NT}(\breve{\boldsymbol{\theta}}_{\text{PL}} - \boldsymbol{\theta}) &= \left(\frac{1}{NT}\mathbf{X}^{(f)'}\mathbf{P}^{(f)}\mathbf{X}^{(f)} - \frac{\breve{\lambda}}{NT}\mathbf{X}^{(f)'}(\mathbf{I} - \mathbf{P}^{(f)})\mathbf{X}^{(f)'}\right)^{-1} \\ &\times \left(\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)'}\mathbf{P}^{(f)}\mathbf{u}^{(f)} - \frac{\sqrt{NT}\breve{\lambda}}{NT}\mathbf{X}^{(f)'}(\mathbf{I} - \mathbf{P}^{(f)})\mathbf{u}^{(f)}\right) \ , \end{aligned} \quad (6.43)$$

where $\breve{\lambda} = \min \mathcal{VR}_{2\cdot 0}$. From the smilar arguments of Theorem 2.9, $\sqrt{NT}\breve{\lambda}$ converges in probability to zero. Therefore,

$$\sqrt{NT}(\breve{\boldsymbol{\theta}}_{\text{PL}} - \boldsymbol{\theta}) = \left(\frac{1}{NT}\mathbf{X}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{X}^{(f)}\right)^{-1} \frac{1}{\sqrt{NT}}\mathbf{X}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{u}^{(f)} + o_p(1) \ . \quad (6.44)$$

Similar to the proof of Theorem 2.9, the first term follows that

$$\begin{aligned} \breve{\boldsymbol{\Phi}} &= \frac{1}{NT}\mathbf{X}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{X}^{(f)} \\ &\xrightarrow{p} \boldsymbol{\Pi}_{\text{I}}^{\prime}\left(\mathbf{J}^{\prime}\boldsymbol{\Gamma}_0\mathbf{J}\right)\left(\mathbf{J}^{\prime}\boldsymbol{\Gamma}_0\mathbf{J}\right)^{-1}\left(\mathbf{J}^{\prime}\boldsymbol{\Gamma}_0\mathbf{J}\right)\boldsymbol{\Pi}_{\text{I}} \\ &= \boldsymbol{\Phi} \ . \end{aligned}$$

The second term of (6.44) becomes

$$\frac{1}{\sqrt{NT}}\mathbf{X}^{(f)\prime}\mathbf{P}^{(f)}\mathbf{u}^{(f)} = \frac{1}{NT}\mathbf{X}^{(f)\prime}\mathbf{Z}^{(f)}\left(\frac{1}{NT}\mathbf{Z}^{(f)\prime}\mathbf{Z}^{(f)}\right)^{-1}\frac{1}{\sqrt{NT}}\mathbf{Z}^{(f)\prime}\mathbf{u}^{(f)} \ .$$

For this last term,

$$\begin{aligned} \frac{1}{\sqrt{NT}}\mathbf{e}_k^{\prime}\mathbf{Z}^{(f)\prime}\mathbf{u}^{(f)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{w}_{i(-1)}^{[k]\prime}\mathbf{Q}_T\mathbf{u}_i \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}w_{it-1}^{[k]}u_{it} - \sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{w}_{i(-1)}^{[k]}\bar{u}_i \ , \end{aligned} \quad (6.45)$$

where $\mathbf{w}_{i(-1)}^{[k]} = (w_{i0}^{[k]}, ..., w_{iT-1}^{[k]})'$, $\mathbf{u}_i = (u_{i1}, ..., u_{iT})'$, and

$$\bar{w}_{i(-1)}^{[k]} = \frac{1}{T}\sum_{t=1}^{T}w_{it-1}^{[k]} \ , \qquad \bar{u}_i = \frac{1}{T}\sum_{t=1}^{T}u_{it} \ .$$

The first term of (6.45) converges in distribution to $\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi})$ when $T \to \infty$, so that the first term of (6.44) converges to $\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1})$. The expectation of the second term of (6.45) is given by

$$\mathcal{E}\left[\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{w}_{i(-1)}^{[k]}\bar{u}_i\right] = \sqrt{\frac{N}{T^3}}\mathbf{e}_k^{\prime}\mathcal{E}\left[\mathbf{J}^{\prime}\mathbf{W}_{i(-1)}^{\prime}\boldsymbol{\iota}\boldsymbol{\iota}^{\prime}\mathbf{u}_i\right] \ , \quad (6.46)$$

where $\mathbf{W}_{i(-1)}^{\prime} = (\mathbf{w}_{i0}, \cdots, \mathbf{w}_{iT-1})$ is the $K^* \times T$ matrix. From the result of Akashi and Kunitomo (2015), we have that

$$\begin{aligned} \mathcal{E}\left[\mathbf{W}_{i(-1)}^{\prime}\boldsymbol{\iota}\boldsymbol{\iota}^{\prime}\mathbf{u}_i\right] &= \sum_{h=1}^{T-1}\sum_{j=0}^{T-1-h}(\boldsymbol{\Pi}^{*\prime})^j\mathcal{E}\left[\mathbf{v}_{it}^{*}u_{it}\right] \\ &= T\left(\mathbf{I}_{K^*} - \boldsymbol{\Pi}^{*\prime}\right)^{-1}\mathcal{E}\left[\mathbf{v}_{it}^{*}u_{it}\right] + O(1) \ . \end{aligned}$$

For its variance,

$$\mathcal{V}ar\left[\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{w}_{i(-1)}^{[k]}\bar{u}_i\right] = \frac{1}{T}\mathcal{V}ar\left[\left(\sqrt{T}\bar{w}_{i(-1)}^{[k]}\right)\left(\sqrt{T}\bar{u}_i\right)\right]$$
$$= O\left(\frac{1}{T}\right) .$$

Therefore, this term converges in 2th mean to the expectation of (6.46) when $T \to \infty$.

$$-\left(\frac{1}{NT}\mathbf{X}^{(f)'}\mathbf{P}^{(f)}\mathbf{X}^{(f)}\right)^{-1}\mathbf{X}^{(f)'}\mathbf{Z}^{(f)}\left(\mathbf{Z}^{(f)'}\mathbf{Z}^{(f)}\right)^{-1}\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{w}_{i(-1)}^{[k]}\bar{u}_i ,$$

From the above results, this second term of (6.45) converges in probability to the following:

$$\begin{aligned}\mathbf{b}_d &= -\sqrt{d}\boldsymbol{\rho}^* \\ &= -\sqrt{d}\boldsymbol{\Phi}^{-1}\boldsymbol{\Pi}_{\mathrm{I}}'\mathbf{J}'\left(\mathbf{I}_{K^*}-\boldsymbol{\Pi}^{*'}\right)^{-1}\boldsymbol{\Omega}^*\mathbf{J}_1\boldsymbol{\beta} ,\end{aligned} \tag{6.47}$$

under $N/T \to 0 \leq d < \infty$, where regarding the representation of (6.47) it holds that

$$\breve{\boldsymbol{\Pi}}_{\mathrm{I}} = \left(\mathbf{Z}^{(f)'}\mathbf{Z}^{(f)}\right)^{-1}\mathbf{X}^{(f)'}\mathbf{Z}^{(f)} \xrightarrow{p} \boldsymbol{\Pi}_{\mathrm{I}} .$$

In addtion, we notice that

$$\begin{aligned}\mathcal{E}\left[\mathbf{v}_{it}^*u_{it}\right] &= \mathcal{E}\left[\mathbf{v}_{it}^*\mathbf{v}_{it}'\right]\boldsymbol{\beta} \\ &= \boldsymbol{\Omega}^*\begin{pmatrix}\mathbf{I}_G \\ \mathbf{O}\end{pmatrix}\boldsymbol{\beta} ,\end{aligned}$$

since $\mathbf{v}_{it}$ is defined to be included as the first G elements of $\mathbf{v}_{it}^*$. Therefore, from Slutsky's theorem, (6.43) converges in distribution to

$$\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{\Phi}^{-1}) + \mathbf{b}_d .$$

□

[*ii*] For the sampling error of the corrected estimator, we have

$$\sqrt{NT}(\breve{\breve{\boldsymbol{\theta}}}_{\mathrm{PL}} - \boldsymbol{\theta}) = \sqrt{NT}(\breve{\boldsymbol{\theta}}_{\mathrm{PL}} - \boldsymbol{\theta}) + \sqrt{\frac{N}{T}}\breve{\boldsymbol{\rho}}^* + o_p(1) .$$

Then, it is sufficient to show that

$$\begin{aligned}\breve{\boldsymbol{\rho}}^* &= \breve{\boldsymbol{\Phi}}^{-1}\breve{\boldsymbol{\Pi}}_{\mathrm{I}}'\mathbf{J}'\left(\mathbf{I}_{K^*}-\breve{\boldsymbol{\Pi}}^{*'}\right)^{-1}\breve{\boldsymbol{\Omega}}^*\mathbf{J}_1\breve{\boldsymbol{\beta}} \\ &\xrightarrow{p} \boldsymbol{\rho}^* ,\end{aligned}$$

where $\breve{\boldsymbol{\beta}} = (1, -\breve{\boldsymbol{\beta}}_{\text{2PL}}^{'})'$ is the consistent estimator for $\boldsymbol{\beta}$. Regarding $\breve{\boldsymbol{\Pi}}^*$ and $\breve{\boldsymbol{\Omega}}^*$, it is necessary to estimate the companion reduced form separately:

$$\begin{aligned}\underset{K^*\times K^*}{\breve{\boldsymbol{\Pi}}^*} &= \left(\mathbf{Z}_{-1}^{*(f)'}\mathbf{Z}_{-1}^{*(f)}\right)^{-1}\mathbf{Z}_{-1}^{*(f)'}\mathbf{Z}^{*(f)} \quad \xrightarrow{p} \quad \boldsymbol{\Pi}^* ,\\ \underset{K^*\times K^*}{\breve{\boldsymbol{\Omega}}^*} &= \mathbf{Z}^{*(f)'}\left(\mathbf{I}_{K^*} - \mathbf{P}^{*(f)}\right)\mathbf{Z}^{*(f)} \quad \xrightarrow{p} \quad \boldsymbol{\Omega}^* .\end{aligned}$$

where

$$\begin{aligned}\underset{K^*\times N(T-1)}{\mathbf{Z}_{-1}^{*(f)'}} &= \left(\mathbf{z}_{10}^{*(f)}, \cdots, \mathbf{z}_{1T-2}^{*(f)}, \cdots, \mathbf{z}_{N0}^{*(f)}, \cdots, \mathbf{z}_{NT-2}^{*(f)}\right) ,\\ \underset{K^*\times N(T-1)}{\mathbf{Z}^{*(f)'}} &= \left(\mathbf{z}_{11}^{*(f)}, \cdots, \mathbf{z}_{1T-1}^{*(f)}, \cdots, \mathbf{z}_{N1}^{*(f)}, \cdots, \mathbf{z}_{NT-1}^{*(f)}\right) ,\\ \mathbf{P}^{*(f)} &= \mathbf{Z}_{-1}^{*(f)'}\left(\mathbf{Z}_{-1}^{*(f)'}\mathbf{Z}_{-1}^{*(f)}\right)^{-1}\mathbf{Z}_{-1}^{*(f)'} .\end{aligned}$$

Thus, we obtain the desired result. From the relation of (3.11), for $k = 1, \cdots, K^*$, the estimation equation of the companion reduced form becomes

$$\mathbf{e}_k^{'}\mathbf{z}_{it}^{*(f)} = \mathbf{e}_k^{'}\boldsymbol{\Pi}^{*'}\mathbf{z}_{it-1}^{*(f)} + \mathbf{e}_k^{'}\mathbf{v}_{it}^{*(f)} .$$

However, when the identiy $\mathbf{e}_k^{'}\boldsymbol{\Pi}^{*'} = \mathbf{e}_\ell^{'}$ ($\ell = 1, \cdots, K^*$) or $\mathbf{Z}^{*(f)}\mathbf{e}_k = \mathbf{Z}_{-1}^{*(f)}\mathbf{e}_\ell$ holds, there exists no error in estimation:

$$\begin{aligned}\breve{\boldsymbol{\Pi}}^*\mathbf{e}_k &= \left(\mathbf{Z}_{-1}^{*(f)'}\mathbf{Z}_{-1}^{*(f)}\right)^{-1}\left(\mathbf{Z}_{-1}^{*(f)'}\mathbf{Z}_{-1}^{*(f)}\right)\mathbf{e}_\ell\\ &= \mathbf{e}_\ell .\end{aligned}$$

□

Proof of Theorem 2.11 : Following Hahn (2002), we consider the second formulation of (3.24) and use the relations that $\alpha_i + u_{it} = \boldsymbol{\beta}'(\boldsymbol{\pi}_i + \mathbf{v}_{it})$ and

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{22}\boldsymbol{\beta}_2 & \boldsymbol{\Pi}_{22} \end{pmatrix} = \underset{K\times(1+G_2)}{(\boldsymbol{\pi}_1,\ \boldsymbol{\Pi}_2)} .$$

Then, the log-likelihood function is proportional to the following under the limited information method,

$$\mathcal{L}(\boldsymbol{\phi}) = \frac{NT}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{it}'\boldsymbol{\Psi}\mathbf{v}_{it} ,$$

where $\boldsymbol{\Psi} = \boldsymbol{\Omega}^{-1}$,

$$\boldsymbol{\phi}' = (\text{vec}(\boldsymbol{\Psi})', \boldsymbol{\theta}_1', \text{vec}(\boldsymbol{\Pi}_2')', \pi_1^{(1)}, \boldsymbol{\pi}_1^{(2)'}, \cdots, \pi_N^{(1)}, \boldsymbol{\pi}_N^{(2)'}) ,$$

and the error term of the reduced form is as follows:

$$\mathbf{v}'_{it} = (y_{it}^{(1)} - \boldsymbol{\beta}'_2\boldsymbol{\Pi}'_2\mathbf{z}_{it-1} - \boldsymbol{\gamma}_1\mathbf{z}_{it-1}^{(1)} - \pi_i^{(1)}, \mathbf{y}_{it}^{(2)'} - \mathbf{z}'_{it-1}\boldsymbol{\Pi}_2 - \boldsymbol{\pi}_i^{(2)'}) \ .$$

From the discussions of the technical lemmas of Hahn and Kuersteiner (2000), the likelihood ratio process of the VAR model is asymptotically shift normal even if it includes the incidental parameters. Under assumption (A4), for the localized parameter $\boldsymbol{\phi} + \boldsymbol{\delta}/\sqrt{NT}$ around the true value $\boldsymbol{\phi}$, we have

$$\mathcal{L}(\boldsymbol{\phi} + \frac{\boldsymbol{\delta}}{\sqrt{NT}}) - \mathcal{L}(\boldsymbol{\phi}) = \boldsymbol{\Delta}_n^{'}\boldsymbol{\delta} - \frac{1}{2}\mathcal{E}\left[(\boldsymbol{\Delta}_n^{'}\boldsymbol{\delta})^2\right] + o_p(1) \ ,$$

where

$$\boldsymbol{\Delta}_n^{'}\boldsymbol{\delta} \ \ \xrightarrow{d} \ \ \mathcal{N}(0, \lim_{N,\,T\to\infty}\mathcal{E}[(\boldsymbol{\Delta}'\boldsymbol{\delta})^2]) \ ,$$

and the elements are as follows,

$$\boldsymbol{\Delta}_n^{'} = (\boldsymbol{\Delta}'_{\Psi}, \boldsymbol{\Delta}'_{\beta}, \boldsymbol{\Delta}'_{\gamma}, \boldsymbol{\Delta}'_{\Pi}, \boldsymbol{\Delta}'_{\pi_1}, \cdots, \boldsymbol{\Delta}'_{\pi_N}) \ .$$

If the error term follows a normal distribution, then

$$\begin{aligned}
\boldsymbol{\Delta}'_{\Psi} &= \frac{\sqrt{NT}}{2}\mathrm{vec}(\boldsymbol{\Omega})' - \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{v}_{it}\otimes\mathbf{v}_{it})' \ , \\
\boldsymbol{\Delta}'_{\beta} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}'_{it}\boldsymbol{\Psi}\left(\begin{bmatrix} (\boldsymbol{\Pi}'_2\mathbf{w}_{it-1})' \\ \mathbf{0}\end{bmatrix} + \begin{bmatrix} (\boldsymbol{\Pi}'_2\mathbf{J}'\boldsymbol{\mu}_i)' \\ \mathbf{0}\end{bmatrix}\right) \\
&= \boldsymbol{\Delta}'_{w\beta} + \boldsymbol{\Delta}'_{\mu\beta} \ \ (say,) \ , \\
\boldsymbol{\Delta}'_{\gamma} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}'_{it}\boldsymbol{\Psi}\left(\begin{bmatrix} \mathbf{w}_{it-1}^{(1)'} \\ \mathbf{0}\end{bmatrix} + \begin{bmatrix} (\mathbf{J}'_{K_1}\boldsymbol{\mu}_i)' \\ \mathbf{0}\end{bmatrix}\right) \\
&= \boldsymbol{\Delta}'_{w\gamma} + \boldsymbol{\Delta}'_{\mu\gamma} \ \ (say,) \ , \\
\boldsymbol{\Delta}'_{\Pi_2} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}'_{it}\boldsymbol{\Psi}\left(\begin{bmatrix} \beta_2\mathbf{w}'_{it-1} \\ \mathbf{w}'_{it-1} \\ \mathbf{0}\end{bmatrix}, \cdots, \begin{bmatrix} \beta_{G_2}\mathbf{w}'_{it-1} \\ \mathbf{0} \\ \mathbf{w}'_{it-1}\end{bmatrix}\right) \\
&+ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}'_{it}\boldsymbol{\Psi}\left(\begin{bmatrix} \beta_2(\mathbf{J}'\boldsymbol{\mu}_i)' \\ (\mathbf{J}'\boldsymbol{\mu}_i)' \\ \mathbf{0}\end{bmatrix}, \cdots, \begin{bmatrix} \beta_{G_2}(\mathbf{J}'\boldsymbol{\mu}_i)' \\ \mathbf{0} \\ (\mathbf{J}'\boldsymbol{\mu}_i)'\end{bmatrix}\right) \\
&= \boldsymbol{\Delta}'_{w\Pi} + \boldsymbol{\Delta}'_{\mu\Pi} \ \ (say,) \ , \\
\boldsymbol{\Delta}'_{\pi_i} &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbf{v}'_{it}\boldsymbol{\Psi}\mathbf{I}_G \ , \quad (i = 1, \cdots, N) \ ,
\end{aligned}$$

where $\mathbf{J}'_{K_1}$ is a matrix such that $\mathbf{w}_{it-1}^{(1)} = \mathbf{J}'_{K_1}\mathbf{w}_{it-1}$. From (3.12), the term including the individual effect $\boldsymbol{\mu}_i$ is expressed separately. To derive the lower bound for the structural parameter $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_2', \boldsymbol{\gamma}_1')'$ of the first structural equation, it is easier to derive the lower bound for $\boldsymbol{\theta}_1$ and $\boldsymbol{\Pi}_2$ first. The lower bound of the regular estimators is obtained by the following minimization problem. For each $j = 1, \cdots, (G_2 + K_1 + G_2 K)$,

$$\min \mathcal{E}\left[(\mathbf{e}_j'\tilde{\boldsymbol{\Delta}}_1)^2\right] = \mathcal{E}\left[(\mathbf{e}_j'\boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_\Psi'\boldsymbol{\delta}_\Psi^{[j]} - \boldsymbol{\Delta}_\pi'\boldsymbol{\delta}_{\pi_i}^{[j]})^2\right] .$$

Minimizing with respect to $\boldsymbol{\delta}_\Psi^{[j]}$ and $\boldsymbol{\delta}_{\pi_i}^{[j]}$, the lower bound can be evaluated by the inverse matrix of $\mathcal{E}[\tilde{\boldsymbol{\Delta}}_1\tilde{\boldsymbol{\Delta}}_1']$, where

$$\boldsymbol{\Delta}_1' = (\boldsymbol{\Delta}_\beta', \boldsymbol{\Delta}_\gamma', \boldsymbol{\Delta}_{\Pi_2}') , \quad \boldsymbol{\Delta}_\pi' = (\boldsymbol{\Delta}_{\pi_1}', \cdots, \boldsymbol{\Delta}_{\pi_N}') .$$

Therefore, the optimal solution as the linear projection is as follows:

$$\begin{pmatrix} \boldsymbol{\delta}_\Psi^{[j]} \\ \boldsymbol{\delta}_{\pi_i}^{[j]} \end{pmatrix} = \left(\mathcal{E}\begin{bmatrix} \boldsymbol{\Delta}_\Psi\boldsymbol{\Delta}_\Psi' & \boldsymbol{\Delta}_\Psi\boldsymbol{\Delta}_\pi' \\ \boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\Psi' & \boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\pi' \end{bmatrix}\right)^{-1} \mathcal{E}\left[\begin{pmatrix} \boldsymbol{\Delta}_\Psi \\ \boldsymbol{\Delta}_\pi \end{pmatrix}\boldsymbol{\Delta}_1'\mathbf{e}_j\right]$$
$$= \left(\mathcal{E}\begin{bmatrix} \boldsymbol{\Delta}_\Psi\boldsymbol{\Delta}_\Psi' & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\pi' \end{bmatrix}\right)^{-1} \mathcal{E}\left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_1'\mathbf{e}_j \end{pmatrix}\right] ,$$

where the second equality is from that a third-order moment of normal distribution is zero. Therefore, for each j the optimum solution becomes $\boldsymbol{\delta}_\Psi^{[j]} = \mathbf{0}$. For $\boldsymbol{\delta}_{\pi_i}^{[j]}$,

$$\boldsymbol{\delta}_{\pi_i}^{[j]} = (\mathcal{E}[\boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\pi'])^{-1}\mathcal{E}[\boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_1'\mathbf{e}_j]$$

is the optimal solution. Therefore, the variance-covariance matrix of the residuals is given by

$$\begin{aligned} \mathcal{E}[\tilde{\boldsymbol{\Delta}}_1\tilde{\boldsymbol{\Delta}}_1'] &= \mathcal{E}[\boldsymbol{\Delta}_1\boldsymbol{\Delta}_1'] - \mathcal{E}[\boldsymbol{\Delta}_1\boldsymbol{\Delta}_\pi'](\mathcal{E}[\boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\pi'])^{-1}\mathcal{E}[\boldsymbol{\Delta}_\pi'\boldsymbol{\Delta}_1] \\ &= \mathcal{E}[\boldsymbol{\Delta}_{1\cdot w}\boldsymbol{\Delta}_{1\cdot w}' + \boldsymbol{\Delta}_\mu\boldsymbol{\Delta}_\mu'] - \mathcal{E}[\boldsymbol{\Delta}_1\boldsymbol{\Delta}_\pi'](\mathcal{E}[\boldsymbol{\Delta}_\pi\boldsymbol{\Delta}_\pi'])^{-1}\mathcal{E}[\boldsymbol{\Delta}_\pi'\boldsymbol{\Delta}_1] , \end{aligned} \tag{6.48}$$

where

$$\boldsymbol{\Delta}_{1\cdot w}' = (\boldsymbol{\Delta}_{w\beta}', \boldsymbol{\Delta}_{w\gamma}', \boldsymbol{\Delta}_{w\Pi}') , \quad \boldsymbol{\Delta}_\mu' = (\boldsymbol{\Delta}_{\mu\beta}', \boldsymbol{\Delta}_{\mu\gamma}', \boldsymbol{\Delta}_{\mu\Pi}') .$$

In the first term of the second equality of (6.48), we use the following:

$$\mathcal{E}[\boldsymbol{\Delta}_{1\cdot w}\boldsymbol{\Delta}'_\mu] = \mathbf{O} , \tag{6.49}$$

this is because that

$$\tilde{\boldsymbol{\Delta}}_1 = \boldsymbol{\Delta}_{1\cdot w} + \boldsymbol{\Delta}_\mu , \quad \mathcal{E}[\boldsymbol{\Psi}\mathbf{v}_{it}\mathbf{v}_{it}'\boldsymbol{\Psi}(\mathbf{w}_{it-1}, \mathbf{0})'] = \mathbf{O} .$$

For the second term of the second equality of (6.48), using $\mathbf{\Delta}_{\mu} = (1/\sqrt{NT})\sum_i \sum_t \mathbf{\Delta}_{\mu_i}$, we have that

$$
\begin{aligned}
& \mathcal{E}[\mathbf{\Delta}_1 \mathbf{\Delta}'_{\pi}](\mathcal{E}[\mathbf{\Delta}_{\pi}\mathbf{\Delta}'_{\pi}])^{-1}\mathcal{E}[\mathbf{\Delta}'_{\pi}\mathbf{\Delta}_1] \\
= \; & \mathcal{E}[\mathbf{\Delta}_{\mu} \mathbf{\Delta}'_{\pi}](\mathcal{E}[\mathbf{\Delta}_{\pi}\mathbf{\Delta}'_{\pi}])^{-1}\mathcal{E}[\mathbf{\Delta}'_{\pi}\mathbf{\Delta}_{\mu}] \\
= \; & \frac{1}{NT}\left(T\mathbf{\Delta}_{\mu_1}\mathbf{\Psi}, \cdots, T\mathbf{\Delta}_{\mu_N}\mathbf{\Psi}\right) \begin{pmatrix} N\mathbf{\Omega} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & N\mathbf{\Omega} & \mathbf{O} & \cdots \\ \vdots & \vdots & \ddots & \mathbf{O} \\ \mathbf{O} & \cdots & \mathbf{O} & N\mathbf{\Omega} \end{pmatrix} \begin{pmatrix} \frac{T}{NT}\mathbf{\Psi}\mathbf{\Delta}'_{\mu_1} \\ \vdots \\ \frac{T}{NT}\mathbf{\Psi}\mathbf{\Delta}'_{\mu_N} \end{pmatrix} \\
= \; & \frac{1}{N}\sum_{i=1}^{N} \mathbf{\Delta}_{\mu_i}\mathbf{\Psi}\mathbf{\Delta}'_{\mu_i} \\
= \; & \mathcal{E}[\mathbf{\Delta}_{\mu}\mathbf{\Delta}'_{\mu}] \; .
\end{aligned}
$$

Therefore, we obtain the lower bound of asymptotic efficiency for $\boldsymbol{\theta}_1$ and $\mathbf{\Pi}_2$ as follows:

$$
\begin{aligned}
\mathbf{V}_{\theta\Pi_2} &= \lim_{N,\,T\to\infty}\left(\mathcal{E}[\tilde{\mathbf{\Delta}}_1\tilde{\mathbf{\Delta}}'_1]\right)^{-1} \\
&= \lim_{N,\,T\to\infty}\left(\mathcal{E}[\mathbf{\Delta}_{1\cdot w}\mathbf{\Delta}'_{1\cdot w}]\right)^{-1} \; .
\end{aligned}
$$

Thus, this does not depend on the individual effects.

The lower bound for the structural parameter $\boldsymbol{\theta}_1$ is the upper left $(G_2 + K_1) \times (G_2 + K_1)$ matrix of $\mathbf{V}_{\theta\Pi_2}$. It is difficult to directly evaluate this submatrix when $G_2 \geq 2$, so that we consider the concentrated log-likelihood function of $\boldsymbol{\theta}_1$. If $\mathbf{w}_{it}$ could be observed without individual effects, the log-likelihood function would be given by

$$
\mathcal{L}^* = \frac{NT}{2}\log|\mathbf{\Psi}^{-1}| - \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\epsilon}'_{it}\mathbf{\Psi}\boldsymbol{\epsilon}_{it} \; ,
$$

where

$$
\begin{aligned}
\boldsymbol{\epsilon}'_{it} &= (w_{it}^{(1)} - \boldsymbol{\beta}'_2\mathbf{\Pi}'_{22}\mathbf{w}_{it-1}^{(2)} - \boldsymbol{\beta}'_2\mathbf{\Pi}'_{12}\mathbf{w}_{it-1}^{(1)} - \boldsymbol{\gamma}'_1\mathbf{w}_{it-1}^{(1)}, \mathbf{w}_{it}^{(2)'} - \mathbf{w}'_{it-1}\mathbf{\Pi}_2) \quad (6.50) \\
&= (w_{it}^{(1)} - \boldsymbol{\beta}'_2\mathbf{\Pi}'_{22}\mathbf{w}_{it-1}^{(2)} - \boldsymbol{\pi}'_{11}\mathbf{w}_{it-1}^{(1)}, \mathbf{w}_{it}^{(2)'} - \mathbf{w}'_{it-1}\mathbf{\Pi}_2) \; . \quad (6.51)
\end{aligned}
$$

The formulation of (6.50) is the parametarization of Hahn (2002), and under the log-likelihood function $\mathcal{L}^*$, the lower bound is also the same as $\mathbf{V}_{\theta\Pi_2}$. Although (6.51) is the parametarization of Lemma 2.2, the LIML estimators of the structural parameter $\boldsymbol{\theta}_1$ obtained by (6.50) and (6.50) are numerically equal under the constraint $\boldsymbol{\gamma}_1 = \boldsymbol{\pi}_{11} - \mathbf{\Pi}_{12}\boldsymbol{\beta}_2$. Therefore, the concentrated log-likelihood function is also the same. From Lemma 2.2, the function is given by

$$
\lambda^*(\boldsymbol{\theta}_1) = \frac{\boldsymbol{\theta}'\mathbf{G}^*\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{H}^*\boldsymbol{\theta}} \; ,
$$

where

$$\mathbf{G}^* = \begin{pmatrix} \mathbf{Y}^{*'} \\ \mathbf{Z}^{(1,*)'} \end{pmatrix} \mathbf{P}^* \left(\mathbf{Y}^*, \mathbf{Z}^{(1,*)}\right) \tag{6.52}$$

is the $(1+G_2+K_1)\times(1+G_2+K_1)$ matrix and $\mathbf{P}^* = \mathbf{Z}^*(\mathbf{Z}^{*'}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*'}$. The matrices that construct (6.52) are $\mathbf{Y}^{*'} = (\mathbf{Y}_i^{*'})$, $\mathbf{Z}_1^{*'} = (\mathbf{Z}_i^{(1)*'})$, and $\mathbf{Z}^{*'} = (\mathbf{Z}_i^{*'})$, which are $(1+G_2)\times NT$, $K_1 \times NT$, and $K \times NT$ matrices, respectively:

$$\begin{aligned} \mathbf{Y}_i^{*'} &= \left(\mathbf{J}_1^{'}\mathbf{w}_{i1}, \cdots, \mathbf{J}_1^{'}\mathbf{w}_{iT}\right), \\ \mathbf{Z}_i^{(1,*)'} &= \left(\mathbf{J}_{K_1}^{'}\mathbf{w}_{i0}, \cdots, \mathbf{J}_{K_1}^{'}\mathbf{w}_{iT-1}\right), \\ \mathbf{Z}_i^{*'} &= \left(\mathbf{J}^{'}\mathbf{w}_{i0}, \cdots, \mathbf{J}^{'}\mathbf{w}_{iT-1}\right). \end{aligned}$$

$\mathbf{H}^*$ is also defined in the same way as (6.52).

From the concentrated log-likelihood function $\lambda^*(\boldsymbol{\theta}_1)$, the asymptotic variance matrix of $\sqrt{NT}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_1)$ is derived as $\sigma^2\mathbf{\Phi}^{-1}$ under the condtion that $(1/n)\mathrm{tr}(\mathbf{P}^*) = K/n \to 0$. Therefore, we obtain the relation that

$$\mathbf{J}_{G_2+K_1}^{'}\mathbf{V}_{\theta\Pi_2}\mathbf{J}_{G_2+K_1} = \sigma^2\mathbf{\Phi}^{-1} ,$$

where $\mathbf{J}_{G_2+K_1}^{'} = (\mathbf{I}_{G_2+K_1},\ \mathbf{O})$. □

Proof of Theorem 2.12 : Following Anderson et al.(2010), we denote a consistent estimator as follows:

$$\hat{\beta}_g = \phi_g\left(\frac{1}{n}\mathbf{G}_n^{(b,f)}\right) , \quad g = 2, \cdots, 1+G_2 .$$

For any $\boldsymbol{\beta}_2$ and $\mathbf{\Phi}$, the following identities holds in the probability limit because of its consistency,

$$\begin{aligned} \beta_g &= \phi_g(\mathbf{\Theta}) \\ &= \phi_g\left(\begin{bmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{bmatrix} \mathbf{\Phi}\,[\,\boldsymbol{\beta}_2,\ \mathbf{I}_{G_2}\,]\right) , \quad g = 2, \cdots, 1+G_2 , \end{aligned} \tag{6.53}$$

where $\mathbf{\Phi} = (\rho_{g,\ell})$ $(g, \ell = 2, \cdots, 1+G_2)$ is the $G_2 \times G_2$ matrix, which is defined by the VAR process in dynamic panel models. Since $\mathbf{\Omega}^* = \mathbf{\Omega}$ in the case of VAR(1) model, it holds that

$$\mathbf{\Phi} = \mathbf{\Pi}_2^{'}\left(\sum_{s=0}^{\infty}(\mathbf{\Pi}^{'})^s\mathbf{\Omega}\mathbf{\Pi}^s\right)\mathbf{\Pi}_2 .$$

Although we would like to partially differentiate the identity of (6.53) with respect to β_g and $\rho_{g,\ell}$, $\mathbf{\Gamma}_0 = \sum_{s=0}^{\infty}(\mathbf{\Pi}^{'})^s\mathbf{\Omega}\mathbf{\Pi}^s$ is the function of $\mathbf{\Pi}$, and $\mathbf{\Pi} = \mathbf{\Pi}(\boldsymbol{\beta}_2)$ is also the function of the structural parameter; i.e., $\mathbf{\Gamma}_0$ depends on $\boldsymbol{\beta}_2$. Then, the partial differentiation can be defined by the following lemma.

Lemma 2.4 : [i] For g, put $\beta_g + h$ and take

$$\begin{aligned}\mathrm{vec}(\mathbf{\Omega}_h) &= (\mathbf{I}_{G^2} - \mathbf{\Pi}_h^{'} \otimes \mathbf{\Pi}_h^{'})\mathrm{vec}\,(\mathbf{\Gamma}_0 + \mathbf{D}_{1h})\ , \\ \mathbf{D}_{1h} &= \mathbf{\Pi}_{2h}(\mathbf{\Pi}_{2h}^{'}\mathbf{\Pi}_{2h})^{-1}(\mathbf{\Pi}_2^{'}\mathbf{\Gamma}_0\mathbf{\Pi}_2 - \mathbf{\Pi}_{2h}^{'}\mathbf{\Gamma}_0\mathbf{\Pi}_{2h})(\mathbf{\Pi}_{2h}^{'}\mathbf{\Pi}_{2h})^{-1}\mathbf{\Pi}_{2h}^{'}\ .\end{aligned} \quad (6.54)$$

Then, there exists an expression equivalent to the following as $h \to 0$,

$$\frac{\partial \phi_g}{\partial \beta_\ell} = \delta_\ell^{(g)}\ ,$$

where $\delta_g^{(g)} = 1$ and $\delta_\ell^{(g)} = 0$ $(g \neq \ell)$.

[ii] Take $\mathbf{\Omega}_h$ as follows:

$$\mathrm{vec}(\mathbf{\Omega}_h) = (\mathbf{I}_{G^2} - \mathbf{\Pi}^{'} \otimes \mathbf{\Pi}^{'})\mathrm{vec}\left(\mathbf{\Gamma}_0 + \mathbf{\Pi}_2(\mathbf{\Pi}_2^{'}\mathbf{\Pi}_2)^{-1}\mathbf{D}_{2h}(\mathbf{\Pi}_2^{'}\mathbf{\Pi}_2)^{-1}\mathbf{\Pi}_2^{'}\right)\ ,$$

where the (ℓ, m) and (m, ℓ) elements of $\mathbf{D}_{2h}$ are h, and the other elements are zero.

Then, for $\ell, m = 2, \cdots, 1 + G_2$, there exists an expression equivalent to the following as $h \to 0$,

$$\frac{\partial \phi_g}{\partial \rho_{\ell m}} = 0\ .$$

Proof : $\mathbf{\Pi}_{2h}^{'}$ and $\mathbf{\Pi}_h^{'}$ in (6.54) denote the coefficients of the reduced form corresponding to $\beta_g + h$. $\mathrm{rank}(\mathbf{\Pi}_{2h}^{'}) = G_2$ because of consistency, so that $(\mathbf{\Pi}_{2h}^{'}\mathbf{\Pi}_{2h})$ is nonsingular. For $\beta_g + h$, $\mathbf{\Gamma}_h$ and $\mathbf{\Phi}_h$ are expressed as follows:

$$\begin{aligned}\mathbf{\Phi}_h &= \mathbf{\Pi}_{2h}^{'}\mathbf{\Gamma}_h\mathbf{\Pi}_{2h}\ , \\ \mathrm{vec}(\mathbf{\Gamma}_h) &= (\mathbf{I}_{G^2} - \mathbf{\Pi}_h^{'} \otimes \mathbf{\Pi}_h^{'})^{-1}\mathrm{vec}(\mathbf{\Omega})\ ,\end{aligned}$$

where $\mathbf{\Gamma}_h$ stands for the variance-covariance matrix of the VAR process and the second equality is from that

$$\mathrm{vec}(\mathbf{ADC}) = (\mathbf{C}^{'} \otimes \mathbf{A})\mathrm{vec}(\mathbf{D})\ .$$

Since $\mathbf{\Omega} = \mathbf{B}^{-1}\mathbf{\Sigma}(\mathbf{B}^{-1})^{'}$, $\mathbf{\Omega}$ is the free parameter for $\boldsymbol{\beta}_2$, where $\mathbf{B}$ is the coefficient matrix of (4.11) and $\mathbf{\Sigma} = \mathcal{E}[\mathbf{u}_{it}\mathbf{u}_{it}^{'}]$.

Therefore, we can take $\mathrm{vec}(\mathbf{\Omega}) = \mathrm{vec}(\mathbf{\Omega}_h)$. Then,

$$\begin{aligned}\mathrm{vec}(\mathbf{\Gamma}_{hh}) &= (\mathbf{I}_{G^2} - \mathbf{\Pi}_h^{'} \otimes \mathbf{\Pi}_h^{'})^{-1}\mathrm{vec}(\mathbf{\Omega}_h) \\ &= \mathrm{vec}\,(\mathbf{\Gamma}_0 + \mathbf{D}_{1h})\ .\end{aligned}$$

$\boldsymbol{\Gamma}_{hh}$ is the symmetric matrix, and it follows that $\lim_{h\to 0} |\boldsymbol{\Gamma}_0 + \mathbf{D}_{1h}| = |\boldsymbol{\Gamma}_0| > 0$, where $\boldsymbol{\Gamma}_0$ is positive definite because $\boldsymbol{\Omega} > \mathbf{O}$. Therefore, considering all leading principal minors, we can take a positive definite matrix in a neighbourhood of $\boldsymbol{\Gamma}_0$. Similarly,

$$\boldsymbol{\Omega}_h = \boldsymbol{\Gamma}_{hh} - \boldsymbol{\Pi}_h^{'}\boldsymbol{\Gamma}_{hh}\boldsymbol{\Pi}_h$$

is also the symmetric matrix such that $\lim_{h\to 0} |\boldsymbol{\Omega}_h| = |\boldsymbol{\Omega}| > 0$ and belongs to a parameter space of variace-covariance matrices. Then, the corresponding $\boldsymbol{\Phi}_{hh}$ is given by

$$\begin{aligned}
\boldsymbol{\Phi}_{hh} &= \boldsymbol{\Pi}_{2h}^{'}\boldsymbol{\Gamma}_{hh}\boldsymbol{\Pi}_{2h} \\
&= \boldsymbol{\Pi}_{2h}^{'}\left(\boldsymbol{\Gamma}_0 + \mathbf{D}_{1h}\right)\boldsymbol{\Pi}_{2h} \\
&= \boldsymbol{\Pi}_2^{'}\boldsymbol{\Gamma}_0\boldsymbol{\Pi}_2 \\
&= \boldsymbol{\Phi}\ .
\end{aligned}$$

Thus, $\boldsymbol{\Phi}_{hh}$ becomes invariant. Consider the difference in the following identities:

$$\begin{aligned}
& \phi_g(\boldsymbol{\beta}_2 + \mathbf{e}_g h, \boldsymbol{\Phi}_{hh}) - \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi}) = (\beta_g + h) - \beta_g \\
\Rightarrow\quad & \phi_g(\boldsymbol{\beta}_2 + \mathbf{e}_g h, \boldsymbol{\Phi}) - \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi}) = h\ . \qquad (6.55)
\end{aligned}$$

Therefore, the limit of (6.55) divided by h is the same as the partial derivative with respect to β_g. For $\ell \neq g$, it also holds that $\partial\phi_g/\partial\beta_\ell = 0$.

Next, we consider the case of [ii]. Take $\mathrm{vec}(\boldsymbol{\Omega}) = \mathrm{vec}(\boldsymbol{\Omega}_h)$, then,

$$\mathrm{vec}(\boldsymbol{\Gamma}_h) = \mathrm{vec}\left(\boldsymbol{\Gamma}_0 + \boldsymbol{\Pi}_2(\boldsymbol{\Pi}_2^{'}\boldsymbol{\Pi}_2)^{-1}\boldsymbol{\Pi}_2^{'}\mathbf{D}_{2h}\boldsymbol{\Pi}_2(\boldsymbol{\Pi}_2^{'}\boldsymbol{\Pi}_2)^{-1}\boldsymbol{\Pi}_2^{'}\right)\ .$$

Therefore,

$$\begin{aligned}
\boldsymbol{\Phi}_h &= \boldsymbol{\Pi}_2^{'}\left(\boldsymbol{\Gamma}_0 + \boldsymbol{\Pi}_2(\boldsymbol{\Pi}_2^{'}\boldsymbol{\Pi}_2)^{-1}\mathbf{D}_{2h}(\boldsymbol{\Pi}_2^{'}\boldsymbol{\Pi}_2)^{-1}\boldsymbol{\Pi}_2^{'}\right)\boldsymbol{\Pi}_2 \\
&= \boldsymbol{\Phi} + \mathbf{D}_{2h}\ .
\end{aligned}$$

For fixed $\boldsymbol{\beta}_2$, the difference in the identities becomes

$$\begin{aligned}
& \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi}_h) - \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi}) = \beta_g - \beta_g \\
\Rightarrow\quad & \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi} + \mathbf{D}_{2h}) - \phi_g(\boldsymbol{\beta}_2, \boldsymbol{\Phi}) = 0\ . \qquad (6.56)
\end{aligned}$$

The limit of (6.56) divided by h is the same as the partial derivative with respect to $\rho_{g,\ell} = \rho_{\ell,g}$. □

We prepare the lemma for the order of $(1/n)\mathbf{G}_n^{(b,f)}$.

Lemma 2.5 : Suppose (A1), (A2), and (A4′). For $T \to \infty$ and any N,

$$\mathbf{G}_1^{(b)} = \sqrt{n}\left(\frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{Y}^{(f)} - \mathbf{G}_{10}^{(b)}\right) = O_p(1) \ ,$$

where

$$\mathbf{G}_{10}^{(b)} \ = \ \mathcal{E}\left[\frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{Y}^{(f)}\right] = \frac{1}{T-1}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{z}_{it}^{(b)}\mathbf{y}_{it}^{(f)\prime}] \ .$$

Proof : We show that the variance of each (k,g) element is $O(1)$.

$$\begin{aligned}
& \mathcal{V}ar\left[\mathbf{e}_k^{\prime}\mathbf{G}_1^{(b)}\mathbf{e}_g\right] \\
= \ & \mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathbf{e}_g^{\prime}\mathbf{y}_{it}^{(f)}\mathbf{z}_{it}^{(b)\prime}\mathbf{e}_k\right] \\
= \ & \mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} w_{it}^{[g]}w_{it-1}^{[k]} - (1-c_t)w_{it}^{[g]}w_{it-1}^{[k]} - c_t\bar{w}_{it-1,0}^{[k]}w_{it}^{[g]} - w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]} + \bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]}\right] \ ,
\end{aligned} \tag{6.57}$$

where the first equality is from the assumption of $i.i.d.$ for $i = 1, \cdots, N$, and the second equality is due to (6.23). It is sufficient to show that the variances of the first to fifth terms are bounded.

First, the variance of the first term is $O(1)$ from the result of Akashi and Kunitomo (2012). For the fifth term,

$$\begin{aligned}
\mathcal{E}\left[(\bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]})^2\right] \ &= \ \mathcal{E}\left[\frac{1}{t}\frac{1}{T-t}\left(\frac{1}{\sqrt{t}}\sum_{h=1}^{t} w_{ih-1}^{[k]}\right)^2\left(\frac{1}{\sqrt{T-t}}\sum_{h=t}^{T} w_{ih+1}^{[k]}\right)^2\right] \\
&= \ O\left(\frac{1}{t}\frac{1}{T-t}\right) \ ,
\end{aligned}$$

then,

$$\begin{aligned}
\mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\bar{w}_{it-1,0}^{[k]}\tilde{w}_{it,T}^{[g]}\right] \ &\leq \ \frac{1}{T}\sum_{t=1}^{T-1}O(\frac{1}{\sqrt{t}\sqrt{(T-t)}})\sum_{s=1}^{T-1}O(\frac{1}{\sqrt{s}\sqrt{(T-s)}}) \\
&\leq \ \frac{O((\sqrt{T})^2)}{T} \\
&= \ O(1) \ ,
\end{aligned}$$

where the first and second inequalities are due to the CS inequality. For the fourth term of (6.57), we obtain that

$$
\begin{aligned}
& \mathcal{V}ar\left[w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]}\right] \\
= \;& \mathcal{E}\left[(w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]})^2\right] - \mathcal{E}[w_{it-1}^{[k]2}]\mathcal{E}[\tilde{w}_{it,T}^{[g]2}] + \mathcal{E}[w_{it-1}^{[k]2}]\mathcal{E}[\tilde{w}_{it,T}^{[g]2}] - \left(\mathcal{E}[w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]}]\right)^2 \\
\leq \;& \sqrt{\mathcal{V}ar[w_{it-1}^{[k]2}]}\sqrt{\mathcal{V}ar[\tilde{w}_{it,T}^{[g]2}]} + \mathcal{E}[w_{it-1}^{[k]2}]\mathcal{E}[\tilde{w}_{it,T}^{[g]2}] \\
= \;& O\left(\frac{1}{t}\right) + O\left(\frac{1}{t}\right) .
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathcal{V}ar\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} w_{it-1}^{[k]}\tilde{w}_{it,T}^{[g]}\right] &\leq \frac{1}{T}\sum_{t=1}^{T-1} O(\frac{1}{\sqrt{t}})\sum_{s=1}^{T-1} O(\frac{1}{\sqrt{s}}) \\
&= \frac{O((\sqrt{T})^2)}{T} .
\end{aligned}
$$

Similarly, the variances of the second and third terms of (6.57) are evaluated as $O(1)$. Since $\mathcal{E}[\mathbf{G}_1^{(b)}] = \mathbf{O}$ by definition, we obtain the desired result. □

Return to the proof of theorem. We represent the partial derivatives by the following $(1+G_2)\times(1+G_2)$ partitioned matrix. For $\mathbf{\Theta} = (\theta_{hj})$,

$$
\begin{aligned}
\mathbf{T}^{(g)} &= \left(\frac{\partial \phi_g}{\partial \theta_{hj}}\right) \qquad (g = 2,\cdots,G \; ; \; h,j = 1,\cdots,G) \\
&= \begin{pmatrix} \tau_{11}^{(g)} & \boldsymbol{\tau}_2^{(g)\prime} \\ \boldsymbol{\tau}_2^{(g)} & \mathbf{T}_{22}^{(g)} \end{pmatrix} .
\end{aligned}
$$

The partial derivatives with respect to β_g and $\rho_{\ell m}$ become

$$
\begin{aligned}
\mathrm{tr}\left(\mathbf{T}^{(g)}\frac{\partial \mathbf{\Theta}}{\partial \beta_j}\right) &= \delta_j^{(g)} , \\
\mathrm{tr}\left(\mathbf{T}^{(g)}\frac{\partial \mathbf{\Theta}}{\partial \rho_{hj}}\right) &= 0 ,
\end{aligned}
$$

then, the following conditions are obtained by the result of Anderson et al.(2010),

$$
\begin{aligned}
2\tau_{11}^{(g)}\mathbf{\Phi}\boldsymbol{\beta} + 2\mathbf{\Phi}\boldsymbol{\tau}_2^{(g)} &= \mathbf{e}_g , \\
\tau_{11}^{(g)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2' + \boldsymbol{\tau}_2^{(g)}\boldsymbol{\beta}_2' + \boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(g)\prime} + \mathbf{T}_{22}^{(g)} &= \mathbf{O} . \qquad (6.58)
\end{aligned}
$$

Consider a linear approximation to ϕ_g in the following. Put

$$\begin{aligned}
\mathbf{S}^{(b)} &= \sqrt{n}\left(\frac{1}{n}\mathbf{G}_n^{(b,f)} - \mathbf{G}_{10}^{(b)'}(\mathbf{G}_{20}^{(b)})^{-1}\mathbf{G}_{10}^{(b)}\right) \\
&= \begin{pmatrix} s_{11}^{(g)} & \mathbf{s}_2^{(b)'} \\ \mathbf{s}_2^{(b)} & \mathbf{S}_{22}^{(b)} \end{pmatrix} , \\
\mathbf{G}_2^{(b)} &= \sqrt{n}\left(\frac{1}{n}\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)} - \mathbf{G}_{20}^{(b)}\right) ,
\end{aligned}$$

where

$$\mathbf{G}_{20}^{(b)} = \mathcal{E}\left[\frac{1}{n}\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)}\right] = \frac{1}{T-1}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{z}_{it}^{(b)}\mathbf{z}_{it}^{(b)'}] .$$

Smilar to Lemma 2.5, we have that $\mathbf{G}_2^{(b)} = O_p(1)$. Then, using the Taylor series for $\mathbf{S}^{(b)}$ around $(\mathbf{G}_{10}^{(b)}, \mathbf{G}_{20}^{(b)})$,

$$\mathbf{S}^{(b)} = O_p(1) . \tag{6.59}$$

From the mean-value theorem, we have that

$$\begin{aligned}
&\sqrt{n}\left(\phi_g\left(\frac{1}{n}\mathbf{G}_n^{(b,f)}\right) - \phi_g\left(\mathbf{G}_{10}^{(b)'}(\mathbf{G}_{20}^{(b)})^{-1}\mathbf{G}_{10}^{(b)}\right)\right) \\
&= \sqrt{n}\left(\phi_g\left(\frac{1}{n}\mathbf{G}_n^{(b,f)}\right) - \beta_g\right) \\
&= \tau_{11*}^{(g)}s_{11}^{(b)} + 2\boldsymbol{\tau}_{2*}^{(g)'}\mathbf{s}_2^{(b)} + \mathrm{tr}\left(\mathbf{T}_{22*}^{(g)}\mathbf{S}_{22}^{(b)}\right) \\
&= \tau_{11}^{(g)}s_{11}^{(b)} + 2\boldsymbol{\tau}_{2}^{(g)'}\mathbf{s}_2^{(b)} + \mathrm{tr}\left(\mathbf{T}_{22}^{(g)}\mathbf{S}_{22}^{(b)}\right) + r_n^{(g)} ,
\end{aligned}$$

where the first equality is from [iii] of assumption (A4), and in the secon equality, $(\tau_{11*}^{(g)}, \boldsymbol{\tau}_{2*}^{(g)'}, \mathrm{vec}(\mathbf{T}_{22*}^{(g)})')$ denote the derivatives evaluated at some mean-values. For the remaining term, when $T \to \infty$,

$$\begin{aligned}
r_n^{(g)} &= (\tau_{11*}^{(g)} - \tau_{11}^{(g)})s_{11}^{(b)} + 2(\boldsymbol{\tau}_{2*}^{(g)} - \boldsymbol{\tau}_2^{(g)})'\mathbf{s}_2^{(b)} + \mathrm{tr}\left((\mathbf{T}_{22*}^{(g)} - \mathbf{T}_{22}^{(g)})\mathbf{S}_{22}^{(b)}\right) \\
&= o_p(1) \times O_p(1) ,
\end{aligned}$$

this is based on [i], [ii] of (A4), and (6.59). Using (6.58), we obtain the following expression:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) = \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}' + (\mathbf{0},\ \boldsymbol{\Phi}^{-1})\right]\mathbf{S}^{(b)}\boldsymbol{\beta} + o_p(1) ,$$

where

$$\boldsymbol{\tau}_{11} = \begin{pmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{pmatrix} ,$$

The asymptotic distribution is as follows:

$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2\right) &= (\boldsymbol{\beta}'\mathbf{S}^{(b)}\boldsymbol{\beta})\boldsymbol{\tau}_{11}+(\mathbf{0},\ \boldsymbol{\Phi}^{-1})\mathbf{S}^{(b)}\boldsymbol{\beta}+o_p(1)\\ &= \boldsymbol{\Phi}^{-1}(\mathbf{0},\ \mathbf{I}_{G_2})\boldsymbol{\Pi}'\frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}+o_p(1)\\ &\stackrel{d}{\to} \mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{\Phi}^{-1})\ ,\end{aligned}$$

where the second equality is from that $\mathbf{G}_{10}^{(b)}\boldsymbol{\beta}=(1/(T-1))\sum_t\mathcal{E}[\mathbf{z}_{it-1}^{(b)}u_{it}^{(f)}]=\mathbf{0}$ and that under $\mathrm{rank}(\mathbf{P}^{(b)})=K<\infty$,

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{S}^{(b)}\boldsymbol{\beta} &= \left(\frac{1}{\sqrt{n}}\mathbf{u}^{(f)'}\mathbf{Z}^{(b)}\right)\left(\frac{1}{n}\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)}\right)^{-1}\left(\frac{1}{n}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right)\\ &\stackrel{p}{\to} 0\ .\end{aligned}$$

The asymptotic normality is due to Theorem 2.9. □

Proof of Theorem 2.13 : Suppose that $|\boldsymbol{\Omega}_\xi|>0$. For $t\ge 1$,

$$\begin{aligned}\begin{pmatrix} y_{it}^{(1,\ell)}\\ y_{it}^{(2,\ell)}\end{pmatrix} &= \boldsymbol{\Pi}'\begin{pmatrix} y_{it-1}^{(1,\ell)}\\ y_{it-1}^{(2,\ell)}\end{pmatrix}+\boldsymbol{\xi}_i+\mathbf{v}_{it}\\ &= (\mathbf{I}_2-\boldsymbol{\Pi}')^{-1}(\mathbf{I}_2-\boldsymbol{\Pi}'^{t})\boldsymbol{\xi}_i+\sum_{h=0}^{t-1}\boldsymbol{\Pi}'^{h}\mathbf{v}_{it-h}\ ,\end{aligned}$$

since $y_{i0}^{(g,\ell)}=0$ $(g=1,\ 2)$. Then, for $t\ge 1$,

$$\mathcal{E}\left[\begin{pmatrix} y_{it-1}^{(1,\ell)}\\ y_{it-1}^{(2,\ell)}\end{pmatrix}\boldsymbol{\xi}_i'\right]=(\mathbf{I}_2-\boldsymbol{\Pi}')^{-1}(\mathbf{I}_2-\boldsymbol{\Pi}'^{t-1})\boldsymbol{\xi}_i\boldsymbol{\xi}_i'\ ,$$

this is because that $\boldsymbol{\xi}_i$ is also a constant by the following:

$$\boldsymbol{\xi}_i=-(\mathbf{I}_2-\boldsymbol{\Pi}')\mathbf{w}_{i0}\ .$$

For $t\ge 2$,

$$\mathcal{E}\left[\begin{pmatrix} y_{it-1}^{(1,\ell)}\\ y_{it-1}^{(2,\ell)}\end{pmatrix}\mathbf{v}_{it-h}'\right]=\boldsymbol{\Pi}'^{h-1}\boldsymbol{\Omega}\ ,\ \ (1\le h\le t-1)\ ,$$

otherwise, it is $\mathbf{O}$.

$$\begin{aligned}&\mathcal{E}\left[\begin{pmatrix} \mathbf{y}_{i,-1}^{(1,\ell)'}\\ \mathbf{y}_{i,-1}^{(2,\ell)'}\end{pmatrix}\mathbf{J}_T\boldsymbol{\xi}_i'\right]\\ =\ & (\mathbf{I}_2-\boldsymbol{\Pi}')^{-1}\left((T-1)\mathbf{I}_2-\boldsymbol{\Pi}(\mathbf{I}_2-\boldsymbol{\Pi}')^{-1}(\mathbf{I}_2-\boldsymbol{\Pi}'^{T-2})-\boldsymbol{\Pi}'^{T-1}\right)\boldsymbol{\xi}_i\boldsymbol{\xi}_i'\\ =\ & (\mathbf{I}_2-\boldsymbol{\Pi}')^{-2}\left((T-1)\mathbf{I}_2-T\boldsymbol{\Pi}'+\boldsymbol{\Pi}'^{T}\right)\boldsymbol{\xi}_i\boldsymbol{\xi}_i'\ .\end{aligned}$$

After some calculation, we have that

$$
\begin{aligned}
& \mathcal{E}\left[\begin{pmatrix} \mathbf{y}_{i,-1}^{(1,\ell)\prime} \\ \mathbf{y}_{i,-1}^{(2,\ell)\prime} \end{pmatrix} \mathbf{J}_T(\mathbf{v}_i^{(1)},\ \mathbf{v}_i^{(2)})\right] \\
= \; & \frac{1}{T}(\mathbf{I}_2 - \mathbf{\Pi}')^{-1}\left((T-1)\mathbf{I}_2 - \mathbf{\Pi}(\mathbf{I}_2-\mathbf{\Pi}')^{-1}(\mathbf{I}_2 - \mathbf{\Pi}'^{T-2}) - \mathbf{\Pi}'^{T-1}\right)\mathbf{\Omega}\ .
\end{aligned}
$$

The following relation exists by the definition of $\mathbf{Q}_T$:

$$
\mathcal{E}\left[\begin{pmatrix} \mathbf{y}_{i,-1}^{(1,\ell)\prime} \\ \mathbf{y}_{i,-1}^{(2,\ell)\prime} \end{pmatrix} \mathbf{Q}_T(\mathbf{v}_i^{(1)},\ \mathbf{v}_i^{(2)})\right] = -\mathcal{E}\left[\begin{pmatrix} \mathbf{y}_{i,-1}^{(1,\ell)\prime} \\ \mathbf{y}_{i,-1}^{(2,\ell)\prime} \end{pmatrix} \mathbf{J}_T(\mathbf{v}_i^{(1)},\ \mathbf{v}_i^{(2)})\right]\ ,
$$

Put

$$
\frac{1}{T}(\mathbf{I}_2 - \mathbf{\Pi}')^{-2}\left((T-1)\mathbf{I}_2 - T\mathbf{\Pi}' + \mathbf{\Pi}'^{T}\right) = \begin{pmatrix} \pi_T^{(11)} & \pi_T^{(12)} \\ \pi_T^{(21)} & \pi_T^{(22)} \end{pmatrix}\ ,
$$

then, each element is $O(1)$.

Using the result of Hsiao and Zhou (2015), the score function $s_{11,i}$ is evaluated as follows:

$$
\frac{1}{N}\sum_{i=1}^{N}\mathcal{E}\left[\omega^{11}\mathbf{y}_{i,-1}^{(1,\ell)\prime}\mathbf{Q}_T\mathbf{v}_i^{(1)} + \omega^{12}\mathbf{y}_{i,-1}^{(1,\ell)\prime}\mathbf{Q}_T\mathbf{v}_i^{(2)}\right] = -\frac{N}{N}\pi_T^{(11)}\ .
$$

We evaluate $\mathbf{\Omega}_\xi$ at $\bar{\mathbf{\Omega}}_N$, then,

$$
\mathbf{\Psi}_T = \mathbf{\Omega} + T\bar{\mathbf{\Omega}}_N\ .
$$

Therefore,

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N}\mathcal{E}\left[\psi_T^{11}\mathbf{y}_{i,-1}^{(1,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \psi_T^{12}\mathbf{y}_{i,-1}^{(1,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)}\right] \\
= \; & \frac{1}{N}\sum_{i=1}^{N}\pi_T^{(11)}\left[\psi_T^{11}(\omega_{11} + T\xi_i^{(1)2}) + \psi_T^{12}(\omega_{12} + T\xi_i^{(1)}\xi_i^{(2)})\right] \\
& + \pi_T^{(12)}\left[\psi_T^{11}(\omega_{12} + T\xi_i^{(1)}\xi_i^{(2)}) + \psi_T^{12}(\omega_{22} + T\xi_i^{(2)2})\right] \\
= \; & \pi_T^{(11)}\left[\psi_T^{11}\left(\omega_{11} + T\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(1)2}\right) + \psi_T^{12}\left(\omega_{21} + T\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(2)}\xi_i^{(1)}\right)\right] \\
& + \pi_T^{(12)}\left[\psi_T^{11}\left(\omega_{12} + T\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(1)}\xi_i^{(2)}\right) + \psi_T^{12}\left(\omega_{22} + T\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(2)2}\right)\right] \\
= \; & \pi_T^{(11)}\ ,
\end{aligned}
$$

since

$$\psi_T^{(11)}\psi_{T,11} + \psi_T^{(12)}\psi_{T,21} = 1 \ , \quad \psi_T^{(11)}\psi_{T,12} + \psi_T^{(12)}\psi_{T,22} = 0 \ ,$$

because $\mathbf{\Psi}_T\mathbf{\Psi}_T^{-1} = \mathbf{I}_2$. Therefore, for any N and T, we obtain

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{E}[s_{11,i}] = 0 \ ,$$

and thus, this is also 0 under $N,\ T \to \infty$. The same is true for $(s_{12,i},\ s_{22,i})$. For the score function of $\mathbf{\Omega}_\xi$, we use the following:

$$\begin{aligned}
-\frac{N}{2}\log|\mathbf{\Omega}_{\xi v}| &= -\frac{N(T-1)}{2}\log|\mathbf{\Omega}| - \frac{N}{2}\log|\mathbf{\Psi}_T| \ , \\
\frac{\partial \mathbf{\Psi}_T^{-1}}{\partial \omega_{\xi,11}} &= -\frac{T\psi_{T,22}}{|\mathbf{\Psi}_T|}\mathbf{\Psi}_T^{-1} + \frac{1}{|\mathbf{\Psi}_T|}\begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix} \\
&= \begin{pmatrix} \ddot{\psi}_T^{11} & \ddot{\psi}_T^{12} \\ \ddot{\psi}_T^{21} & \ddot{\psi}_T^{22} \end{pmatrix} \ (say,) \ .
\end{aligned} \tag{6.60}$$

We evaluate $\mathbf{\Omega}_\xi$ at $\bar{\mathbf{\Omega}}_N$,

$$\begin{aligned}
&\frac{1}{N}\mathcal{E}\left[\frac{\partial \mathcal{L}_2}{\partial \omega_{\xi,11}}\right] \\
= \ & -\frac{1}{2}\frac{T\psi_{T,22}}{|\mathbf{\Psi}_T|} - \frac{1}{2N}\sum_{i=1}^{N}\mathcal{E}\left[\ddot{\psi}_T^{11}\mathbf{v}_i^{(1,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \ddot{\psi}_T^{12}\mathbf{v}_i^{(1,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)}\right] \\
& -\frac{1}{2N}\sum_{i=1}^{N}\mathcal{E}\left[\ddot{\psi}_T^{21}\mathbf{v}_i^{(2,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(1,\ell)} + \ddot{\psi}_T^{22}\mathbf{v}_i^{(2,\ell)\prime}\mathbf{J}_T\mathbf{v}_i^{(2,\ell)}\right] \\
= \ & -\frac{1}{2}\frac{T\psi_{T,22}}{|\mathbf{\Psi}_T|} - \frac{1}{2}\left(\ddot{\psi}_T^{11}\psi_{T,11} + \ddot{\psi}_T^{12}\psi_{T,12} + \ddot{\psi}_T^{21}\psi_{T,21} + +\ddot{\psi}_T^{22}\psi_{T,22}\right) \\
= \ & -\frac{1}{2}\frac{T\psi_{T,22}}{|\mathbf{\Psi}_T|} - \frac{1}{2}\left[\left(-\frac{T\psi_{T,22}}{|\mathbf{\Psi}_T|}\right)\left(\psi_T^{11}\psi_{T,11} + \psi_T^{12}\psi_{T,12} + \psi_T^{21}\psi_{T,21} + \psi_T^{22}\psi_{T,22}\right) - 1\right] \\
= \ & 0 \ ,
\end{aligned}$$

since

$$\left(\psi_T^{11}\psi_{T,11} + \psi_T^{12}\psi_{T,12}\right) + \left(\psi_T^{21}\psi_{T,21} + \psi_T^{22}\psi_{T,22}\right) = 2 \ ,$$

which is due to $\mathbf{\Psi}_T\mathbf{\Psi}_T^{-1} = \mathbf{I}_2$. Similarly, the expectations of score functions for other elements of $\mathbf{\Omega}_\xi$ and $\mathbf{\Omega}$ are zero. Therefore, the estimators are consistent from the assumptions. Then,

$$\hat{\mathbf{\Omega}}_\xi - \left(\bar{\mathbf{\Omega}}_\xi + o(1)\right) = o_p(1) \ ,$$

since $\hat{\boldsymbol{\Omega}}_\xi - \bar{\boldsymbol{\Omega}}_N = o_p(1)$. That is, $\bar{\boldsymbol{\Omega}}_\xi$ is the limit of $\hat{\boldsymbol{\Omega}}_\xi$.

We consider consistency under $|\boldsymbol{\Omega}_\xi| = 0$.

$$|\boldsymbol{\Psi}_T| = |\boldsymbol{\Omega}| + T\left[(\omega_{\xi,22}\omega_{11} + \omega_{\xi,11}\omega_{22}) - (\omega_{\xi,21}\omega_{12} + \omega_{\xi,12}\omega_{21})\right] ,$$

where $T^2|\boldsymbol{\Omega}_\xi| = 0$ does not appear. For some $\boldsymbol{\iota}_\xi$ it holds that $\boldsymbol{\Omega}_\xi = \boldsymbol{\iota}_\xi\boldsymbol{\iota}_\xi'$ by the assumption $|\boldsymbol{\Omega}_\xi| \geq 0$, and thus,

$$|\boldsymbol{\Psi}_T| = |\boldsymbol{\Omega}|\left(1 + T\boldsymbol{\iota}_\xi'\boldsymbol{\Omega}^{-1}\boldsymbol{\iota}_\xi\right) > 0 .$$

Therefore,

$$\frac{\partial\boldsymbol{\Psi}_T^{-1}}{\partial\omega_{\xi,11}} = -\frac{T\omega_{22}}{|\boldsymbol{\Psi}_T|}\boldsymbol{\Psi}_T^{-1} + \frac{1}{|\boldsymbol{\Psi}_T|}\begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix} ,$$

where each element becomes $O(1)$. This is the same as (6.60), and thus, the consistency is obtained.

In the following, $|\boldsymbol{\Omega}_\xi| > 0$ is assumed again. Then, for the Hessian, we have the same result as that of Theorem 2.7. Under the assumption $||\boldsymbol{\xi}_i|| < \infty$, (6.11) holds. In addition, for $\mathbf{w}_{it} = (w_{it}^{(1)},\ w_{it}^{(2)})'$, we have the following state-space representation:

$$\begin{aligned} \mathcal{E}\left[\mathbf{w}_{it}\right] &= \boldsymbol{\Pi}'^t\mathbf{w}_{i0} , \\ \mathcal{E}\left[\mathbf{w}_{it}\mathbf{w}_{it}'\right] &= \sum_{h=0}^{t-1}\boldsymbol{\Pi}'^h\boldsymbol{\Omega}\boldsymbol{\Pi}^h + \boldsymbol{\Pi}'^t\left(\mathbf{w}_{i0}\mathbf{w}_{i0}'\right)\boldsymbol{\Pi}^t . \end{aligned}$$

Using these relations, $\mathbf{H}_{\phi\phi}$ becomes the same under $T \to \infty$. Therefore,

$$\sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi}) = -\mathbf{H}_{\phi\phi}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{s}_i^{(\ell)} + o_p(1) .$$

Consider the asymptotic normality. Although for each i, $\mathcal{E}[\mathbf{s}_i^{(\ell)}]$ is not zero, the sum becomes 0:

$$\begin{aligned} \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{s}_i^{(\ell)} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{s}_i^{(\ell)} - \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathcal{E}\left[\mathbf{s}_i^{(\ell)}\right] \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^N\left(\mathbf{s}_i^{(\ell)} - \mathcal{E}\left[\mathbf{s}_i^{(\ell)}\right]\right) . \end{aligned}$$

For the variance-covariance matrix,

$$\frac{1}{NT}\sum_{i=1}^N\left(\mathcal{E}\left[\mathbf{s}_i^{(\ell)}\mathbf{s}_i^{(\ell)'}\right] - \boldsymbol{\mu}_i^{(\ell)}\boldsymbol{\mu}_i^{(\ell)'}\right) \xrightarrow{p} -\mathbf{H}_{\phi\phi} , \tag{6.61}$$

where $\boldsymbol{\mu}_i^{(\ell)} = \mathcal{E}[\mathbf{s}_i^{(\ell)}]$. This is because that the first term converges to $\mathbf{G}_\phi$ from the proof of Theorem 2.7. Regarding the second term of (6.61), using the fact that $\boldsymbol{\mu}_i^{(\ell)}$ is $O(1)$ by $\boldsymbol{\Psi}_T^{-1} = O(1/T)$ of (6.12), we have

$$\frac{1}{T}\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\mu}_i^{(\ell)}\boldsymbol{\mu}_i^{(\ell)'} = O\left(\frac{1}{T}\right) \ .$$

For non-zero vector $\mathbf{a}$, there exists the fourth moment of

$$\frac{1}{\sqrt{T}}\mathbf{a}^{'}\left(\mathcal{E}\left[\mathbf{s}_i^{(\ell)}\mathbf{s}_i^{(\ell)'}\right] - \boldsymbol{\mu}_i^{(\ell)}\boldsymbol{\mu}_i^{(\ell)'}\right) \ . \tag{6.62}$$

Thus, the generalized Lindeberg-Feller condition holds. Therefore, the asymptotic variance-covariance matrix of $\sqrt{NT}(\hat{\boldsymbol{\phi}}_{\text{TL}} - \boldsymbol{\phi})$ is $-\mathbf{H}_{\phi\phi}^{-1}$, so we obtain the desired result. □

Proof of Theorem 2.14 : $\mathbf{G}_n^{(f,b)}$ is decomposed as follows:

$$\underset{3\times 3}{\mathbf{G}_n^{(f,b)}} = \mathbf{G}_{n1}^{(f,b)} + \mathbf{G}_{n2}^{(f,b)} + \mathbf{G}_{n2}^{(f,b)'} + \mathbf{G}_{n3}^{(f,b)} \ , \tag{6.63}$$

where

$$\begin{aligned}
\mathbf{G}_{n1}^{(f,b)} &= \boldsymbol{\Theta}_{\text{I}}^{'}\boldsymbol{\Pi}_{\text{I}n}^{'}\mathbf{Z}^{(f)'}\mathbf{P}^{(b)}\mathbf{Z}^{(f)}\boldsymbol{\Pi}_{\text{I}n}\boldsymbol{\Theta}_{\text{I}} \ ,\\
\mathbf{G}_{n2}^{(f,b)} &= \boldsymbol{\Theta}_{\text{I}}^{'}\boldsymbol{\Pi}_{\text{I}n}^{'}\mathbf{Z}^{(f)'}\mathbf{P}^{(b)}\left(\mathbf{V}^{(f)},\ \mathbf{0}\right) ,\\
\mathbf{G}_{n3}^{(f,b)} &= \left(\mathbf{V}^{(f)},\ \mathbf{0}\right)^{'}\mathbf{P}^{(b)}\left(\mathbf{V}^{(f)},\ \mathbf{0}\right) \ ,\\
\underset{2\times(2+K_{2n})}{\boldsymbol{\Pi}_{\text{I}n}^{'}} &= \begin{pmatrix} 0 & \pi_{22} & \boldsymbol{\pi}_{2n}^{'} \\ 1 & 0 & \mathbf{0}^{'} \end{pmatrix}\\
&= \left(\underset{2\times 2}{\boldsymbol{\Pi}_{1\text{I}}^{'}},\ \boldsymbol{\Pi}_{1n}^{'}\right)\ (say,) \ ,\\
\underset{1\times K_{2n}}{\boldsymbol{\pi}_{2n}^{'}} &= \left(\frac{\pi_2}{\sqrt{K_{2n}}},\ \cdots,\ \frac{\pi_2}{\sqrt{K_{2n}}}\right) \ .
\end{aligned}$$

We prepare the notations as follows:

$$\begin{aligned}
\underset{K_n\times n}{\mathbf{Z}^{(f)'}} &= \left(\mathbf{Z}_1^{(f)'},\ \cdots,\ \mathbf{Z}_{T-1}^{(f)'}\right) \ ,\\
\underset{K_n\times n}{\mathbf{Z}^{(b)'}} &= \left(\mathbf{Z}_1^{(b)'},\ \cdots,\ \mathbf{Z}_{T-1}^{(b)'}\right) \ ,
\end{aligned}$$

where $K_n = 2 + K_{2n}$. From the definition of the forward filter,

$$\begin{aligned}
\underset{K_n\times N}{\mathbf{Z}_t^{(f)'}} &= \tilde{\mathbf{W}}_t^{'} - \tilde{\mathbf{V}}_{tT}^{'}\\
&= \left(\tilde{\mathbf{W}}_{1,t}^{'} - \tilde{\mathbf{W}}_{2,t}^{'}\right) - \tilde{\mathbf{V}}_{tT}^{'}\ \ (say,) \ ,
\end{aligned}$$

where

$$
\begin{aligned}
\underset{K_n\times N}{\tilde{\mathbf{W}}'_{1,t}} &= f_t\mathbf{W}'_{t-1}\ ,\quad \mathbf{W}'_{t-1}=(\mathbf{w}_{it-1})\ ,\\
\tilde{\mathbf{W}}'_{2,t} &= \frac{f_t}{T-t}\sum_{h=1}^{T-t}\left(\mathbf{\Pi}_n^{*'}\right)^h\mathbf{W}'_{t-1}\ ,\\
\underset{K_n\times K_n}{\mathbf{\Pi}_n^{*'}} &= \begin{pmatrix}\mathbf{\Pi}_1 & \mathbf{\Pi}'_{2n}\\ \mathbf{O} & \pi_{33}\mathbf{I}_{K_{2n}}\end{pmatrix}\ ,\\
\mathbf{\Pi}_1 &= \begin{pmatrix}\pi_{11} & \beta_2\pi_{22}\\ 0 & \pi_{22}\end{pmatrix}\ ,\\
\mathbf{\Pi}'_{2n} &= \begin{pmatrix}\beta_2\boldsymbol{\pi}'_{2n}\\ \boldsymbol{\pi}'_{2n}\end{pmatrix}\ ,\\
\underset{K_n\times N}{\tilde{\mathbf{V}}'_{tT}} &= \frac{f_t}{T-t}\sum_{h=1}^{T-t}\mathbf{\Phi}_h\mathbf{V}^{*'}_{T-h}\ ,\quad \mathbf{V}^{*'}_t=(\mathbf{v}^*_{it})\ ,\\
\mathbf{\Phi}_h &= \sum_{s=0}^{h-1}(\mathbf{\Pi}_n^{*'})^s\ .
\end{aligned}
$$

For large K_n, we present the following lemma.

Lemma 2.6 : Suppose the assumption (A1′) and (A2). For $N,\ T,\ K_{2n}\to\infty$,

$$
\begin{aligned}
&[i] && \frac{1}{n}\mathbf{\Pi}'_{\mathrm{In}}\tilde{\mathbf{W}}'_1\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{In}} \xrightarrow{p} \underset{2\times 2}{\mathbf{\Phi}^*}\ ,\\
&[ii] && \frac{1}{n}\mathbf{\Pi}'_{\mathrm{In}}\tilde{\mathbf{W}}'_2\tilde{\mathbf{W}}_2\mathbf{\Pi}_{\mathrm{In}} \xrightarrow{p} \underset{2\times 2}{\mathbf{O}}\ ,\\
&[iii] && \frac{1}{n}\mathbf{\Pi}'_{\mathrm{In}}\tilde{\mathbf{W}}'_0\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{In}} \xrightarrow{p} \mathbf{O}\ ,\\
&[iv] && \frac{1}{n}\mathbf{\Pi}'_{\mathrm{In}}\tilde{\mathbf{V}}'_T\tilde{\mathbf{V}}_T\mathbf{\Pi}_{\mathrm{In}} \xrightarrow{p} \mathbf{O}\ ,\\
&[v] && \frac{1}{n}\mathbf{V}^{(f)'}\mathbf{P}^{(b)}\mathbf{V}^{(f)} \xrightarrow{p} \underset{2\times 2}{\mathbf{O}}\ .
\end{aligned}
$$

Proof : [i] $\tilde{\mathbf{W}}'_1=(\tilde{\mathbf{W}}'_{1,t})$ is the $K_n\times n$ matrix and the sum of periods is given by

$$
\begin{aligned}
\frac{1}{n}\mathbf{\Pi}'_{\mathrm{In}}\tilde{\mathbf{W}}'_1\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{In}} &= \frac{1}{n}\sum_{t=1}^{T-1}f_t^2\mathbf{\Pi}'_{\mathrm{In}}\mathbf{W}'_{t-1}\mathbf{W}_{t-1}\mathbf{\Pi}_{\mathrm{In}}\\
&= \frac{1}{n}\sum_{t=1}^{T-1}\sum_{i=1}^{N}f_t^2\mathbf{\Pi}'_{\mathrm{In}}\mathbf{w}_{it-1}\mathbf{w}'_{it-1}\mathbf{\Pi}_{\mathrm{In}}\ .
\end{aligned}
$$

We first show that the following is covariance stationary $(t = 1, \cdots, T-1)$ even when $K_{2n} \to \infty$,

$$\underset{2\times 1}{\mathbf{w}_{it,n}} = \underset{2\times K_n}{\mathbf{\Pi}'_{\mathrm{I}n}} \underset{K_n\times 1}{\mathbf{w}_{it-1}} \ .$$

For all t, it holds that $\mathcal{E}[\mathbf{w}_{it,n}] = \mathbf{0}$.

$$\begin{aligned} \mathcal{E}\left[\mathbf{w}_{i(t+h),n}\mathbf{w}'_{t,n}\right] &= \mathbf{\Pi}'_{\mathrm{I}n}\mathcal{E}\left[\mathbf{w}_{i(t+h-1)}\mathbf{w}'_{it-1}\right]\mathbf{\Pi}_{\mathrm{I}n} \\ &= \mathbf{\Pi}'_{\mathrm{I}n}(\mathbf{\Pi}^{*'}_n)^h\mathbf{\Gamma}_{0n}\mathbf{\Pi}_{\mathrm{I}n} \ , \end{aligned} \tag{6.64}$$

where

$$\begin{aligned} \mathbf{\Gamma}_{0n} &= \mathcal{E}\left[\mathbf{w}_{it-1}\mathbf{w}'_{it-1}\right] \\ &= \begin{pmatrix} \mathbf{\Gamma}_1 & \mathbf{\Gamma}'_{1n} \\ \mathbf{\Gamma}_{1n} & \sigma_3^2\mathbf{I}_{K_{2n}} \end{pmatrix} , \\ \sigma_3^2 &= \frac{\omega_3}{1-\pi_{33}^2} , \\ (\mathbf{\Pi}^{*'}_n)^h &= \begin{pmatrix} \mathbf{\Pi}_1^h & \mathbf{\Phi}_{1h}\mathbf{\Pi}'_{2n} \\ \mathbf{O} & \pi_{33}^h\mathbf{I}_{K_{2n}} \end{pmatrix} , \\ \mathbf{\Phi}_{1h} &= \sum_{s=0}^{h-1}\pi_{33}^s\mathbf{\Pi}_1^{h-1-s} \ . \end{aligned}$$

Therefore,

$$\begin{aligned} \mathcal{E}\left[\mathbf{w}_{i(t+h),n}\mathbf{w}'_{it,n}\right] &= \mathbf{\Pi}'_{1\mathrm{I}}\left(\mathbf{\Pi}_1^h\mathbf{\Gamma}_1 + \mathbf{\Phi}_{1h}\mathbf{\Pi}'_{2n}\mathbf{\Gamma}_{1n}\right)\mathbf{\Pi}_{1\mathrm{I}} + \pi_{33}^h\mathbf{\Pi}'_{1n}\mathbf{\Gamma}_{1n}\mathbf{\Pi}_{1\mathrm{I}} \\ &\quad +\mathbf{\Pi}'_{1\mathrm{I}}\left(\mathbf{\Pi}_1^h\mathbf{\Gamma}'_{1n}\mathbf{\Pi}_{1n} + \sigma_3^2\mathbf{\Phi}_{1h}\mathbf{\Pi}'_{2n}\mathbf{\Pi}_{1n}\right) + \sigma_3^2\pi_{33}^h\mathbf{\Pi}'_{1n}\mathbf{\Pi}_{1n} \ . \end{aligned}$$

We show that $\mathbf{\Pi}'_{1n}\mathbf{\Pi}_{1n}$, $\mathbf{\Pi}'_{2n}\mathbf{\Pi}_{1n}$, $\mathbf{\Pi}'_{1n}\mathbf{\Gamma}_{1n}$, $\mathbf{\Pi}'_{2n}\mathbf{\Gamma}_{1n}$, and $\mathbf{\Gamma}_1$ converge to a constant under $K_{2n} \to \infty$. Since

$$\boldsymbol{\pi}'_{2n}\boldsymbol{\pi}_{2n} = \frac{1}{K_{2n}}\sum_{k=1}^{K_{2n}}\pi_2^2 = \pi_2^2 \ ,$$

the following do not depend on K_{2n},

$$\mathbf{\Pi}'_{1n}\mathbf{\Pi}_{1n} = \begin{pmatrix} \pi_2^2 & 0 \\ 0 & 0 \end{pmatrix} , \quad \mathbf{\Pi}'_{2n}\mathbf{\Pi}_{1n} = \begin{pmatrix} \beta_2\pi_2^2 & 0 \\ \pi_2^2 & 0 \end{pmatrix} .$$

We derive the $2 \times K_{2n}$ matrix $\mathbf{\Gamma}_{1n}^{'}$. For $k = 3, \cdots, 2 + K_{2n}$,

$$\begin{aligned}
\mathcal{E}\left[w_{it}^{(2)} w_{it-1}^{(k)}\right] &= \frac{\pi_2}{\sqrt{K_{2n}}} \sum_{s=0}^{\infty} \pi_{22}^{s} \mathcal{E}\left[w_{it}^{(k)} w_{it-s}^{(k)}\right] \\
&= \frac{1}{\sqrt{K_{2n}}} \frac{\pi_2 \sigma_3^2}{1 - \pi_{22}\pi_{33}} \\
&= \frac{\gamma_{12}}{\sqrt{K_{2n}}} \ (say,) \ ,
\end{aligned}$$

where the first equality is from that $\mathcal{E}[w_{it-1}^{(2)} w_{it-1}^{(k)}] = 0$ and $\mathcal{E}[w_{it-1}^{(k)} v_t^{(2)}] = 0$. Similarly, $\mathcal{E}[w_{it}^{(1)} w_{it-1}^{(k)}]$ is derived. Then,

$$\mathbf{\Gamma}_{1n}^{'} = \begin{pmatrix} \frac{1}{\sqrt{K_{2n}}} \frac{\beta_2 \gamma_{12}}{1-\pi_{11}\pi_{33}} & \cdots & \frac{1}{\sqrt{K_{2n}}} \frac{\beta_2 \gamma_{12}}{1-\pi_{11}\pi_{33}} \\ \frac{\gamma_{12}}{\sqrt{K_{2n}}} & \cdots & \frac{\gamma_{12}}{\sqrt{K_{2n}}} \end{pmatrix} .$$

Thus, $\mathbf{\Pi}_{1n}^{'} \mathbf{\Gamma}_{1n}$ and $\mathbf{\Pi}_{2n}^{'} \mathbf{\Gamma}_{1n}$ do not depend on K_{2n}. From the independence between the K_{2n} variables,

$$\begin{aligned}
\mathcal{V}ar\left[w_{it}^{(2)}\right] &= \mathcal{V}ar\left[\sum_{s=0}^{\infty} \pi_{22}^{s} \frac{\pi_2}{\sqrt{K_{2n}}} \sum_{k=3}^{2+K_{2n}} w_{i(t-1-s)}^{(k)}\right] + \mathcal{V}ar\left[\sum_{s=0}^{\infty} \pi_{22}^{s} v_{t-s}^{(2)}\right] \\
&= \pi_2^2 \mathcal{V}ar\left[\sum_{s=0}^{\infty} \pi_{22}^{s} w_{i(t-1-s)}^{(3)}\right] + \frac{\omega_{22}}{1 - \pi_{22}^2} .
\end{aligned}$$

Similarly,

$$\mathbf{\Gamma}_1 = \mathcal{E}\left[\begin{pmatrix} w_{it}^{(1)} \\ w_{it}^{(2)} \end{pmatrix} \left(w_{it}^{(1)}, \ w_{it}^{(2)}\right)\right]$$

does not depend on K_{2n}. From the above, the elements of autocovariance matrix (6.64) are finite and depend only on the difference h. Let

$$r(h)_g = \mathbf{e}_g^{'} \mathcal{E}\left[\mathbf{w}_{i(t+h),n} \mathbf{w}_{it,n}^{'}\right] \mathbf{e}_g$$

be the autocovariance, where $\mathbf{e}_g^{'} = (0, \cdots, 1, \cdots, 0)$ whose g-th element is only unity ($g = 1, 2$). Under $h \to \infty$, $\mathbf{\Pi}_1^h$, $\mathbf{\Phi}_{1h}$, and π_3^h converge to zero, so that

$$\sum_{h=0}^{\infty} [r(h)_g]^2 < \infty .$$

When $\mathbf{w}_{it,n}$ follows a normal distribution, the Lindgren et al. (2014, Ch. 2) show that

$$\frac{1}{T} \sum_{t=1}^{T-1} \mathbf{w}_{it,n} \mathbf{w}_{it,n}^{'} \xrightarrow[T \to \infty]{p} \mathcal{E}\left[\mathbf{w}_{it,n} \mathbf{w}_{it,n}^{'}\right] = \mathbf{\Phi}^{*} ,$$

where

$$\mathbf{\Phi}^* = \mathbf{\Pi}_{1\mathrm{I}}^{'}\mathbf{\Gamma}_1\mathbf{\Pi}_{1\mathrm{I}} + \mathbf{\Pi}_{1n}^{'}\mathbf{\Gamma}_{1n}\mathbf{\Pi}_{1\mathrm{I}} + \mathbf{\Pi}_{1\mathrm{I}}^{'}\mathbf{\Gamma}_{1n}^{'}\mathbf{\Pi}_{1n} + \sigma_3^2\mathbf{\Pi}_{1n}^{'}\mathbf{\Pi}_{1n} \ .$$

From the $i.i.d.$ assumption,

$$\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_1^{'}\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{I}n} \quad \stackrel{p}{\longrightarrow} \quad \mathbf{\Phi}^* - \frac{1}{n}\sum_{t=1}^{T-1}\sum_{i=1}^{N}(1-f_t^2)\mathbf{w}_{it,n}\mathbf{w}_{it,n}^{'} \ .$$

For the second term,

$$\mathcal{E}\left[\frac{1}{n}\sum_{t=1}^{T-1}\sum_{i=1}^{N}|(1-f_t^2)\mathbf{e}_g^{'}\mathbf{w}_{it,n}\mathbf{w}_{it,n}^{'}\mathbf{e}_g|\right] = O\left(\frac{\log T}{T}\right) \ ,$$

since $(1-f_t^2) = O(1/T-t)$. Therefore, this term converges in probability to zero.

[ii] $\tilde{\mathbf{W}}_2^{'} = (\tilde{\mathbf{W}}_{2,t}^{'})$ is the $K_n \times n$ matrix and it follows that

$$\mathcal{E}\left[\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_2^{'}\tilde{\mathbf{W}}_2\mathbf{\Pi}_{\mathrm{I}n}\right] = \frac{1}{T}\sum_{t=1}^{T-1}\frac{f_t^2}{(T-t)^2}\sum_{h,s=1}^{T-t}\mathbf{\Pi}_{\mathrm{I}n}^{'}\left(\mathbf{\Pi}_n^{*'}\right)^h\mathbf{\Gamma}_{0n}\left(\mathbf{\Pi}_n^*\right)^s\mathbf{\Pi}_{\mathrm{I}n} \ .$$

Similar to the result of [i], for any $(h,\ s)$, $\mathbf{\Pi}_{\mathrm{I}n}^{'}\left(\mathbf{\Pi}_n^{*'}\right)^h\mathbf{\Gamma}_{0n}\left(\mathbf{\Pi}_n^*\right)^s\mathbf{\Pi}_{\mathrm{I}n}$ does not depend on K_{2n}. In addtion, $|\pi_{11}|$, $|\pi_{22}|$, and $|\pi_{33}|$ are less than 1, so that for any t,

$$\mathbf{e}_g^{'}\sum_{h,s=1}^{T-t}\mathbf{\Pi}_{\mathrm{I}n}^{'}\left(\mathbf{\Pi}_n^{*'}\right)^h\mathbf{\Gamma}_{0n}\left(\mathbf{\Pi}_n^*\right)^s\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g = O(1) \ .$$

Then,

$$\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_2^{'}\tilde{\mathbf{W}}_2\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right] = O\left(\frac{1}{T}\right) \ ,$$

since $\sum_t O(t^{-2}) = O(1)$.

[iii] $\tilde{\mathbf{W}}_0^{'} = (\tilde{\mathbf{w}}_{it-1,0})$ is the $K_n \times n$ matrix whose elements consist of (6.69).

$$\begin{aligned}\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_0^{'}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right] &= \frac{1}{T}\sum_{t=1}^{T-1}\frac{f_t^2}{t}\mathcal{E}\left[(\sqrt{t}\mathbf{e}_g^{'}\bar{\mathbf{w}}_{it,n})^2\right] \\ &= O\left(\frac{\log T}{T}\right) \ ,\end{aligned}$$

where $f_t^2 < 1$. Define

$$\underset{2\times 1}{\bar{\mathbf{w}}_{it,n}} = \frac{1}{t}\sum_{h=0}^{t-1}\mathbf{w}_{ih,n} \ ,$$

it follows that $\sqrt{t}\mathbf{e}_g^{'}\bar{\mathbf{w}}_{it,n} = O_p(1)$. Therefore, we obtain the desired result.

[iv] $\tilde{\mathbf{V}}_T^{'} = (\tilde{\mathbf{V}}_{tT}^{'})$ is the $K_n \times n$ matrix. Using the fact that $\mathcal{E}[\mathbf{v}_{it}^{*}\mathbf{v}_{is}^{*'}] = \mathbf{O}$ $(s \neq t)$,

$$\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{\Pi}_{\text{In}}^{'}\tilde{\mathbf{V}}_T^{'}\tilde{\mathbf{V}}_T\mathbf{\Pi}_{\text{In}}\mathbf{e}_g\right] = \frac{1}{T}\sum_{t=1}^{T-1}\frac{f_t^2}{(T-t)^2}\sum_{h=1}^{T-t}\mathbf{e}_g^{'}\mathbf{\Pi}_{\text{In}}^{'}\mathbf{\Phi}_h\mathbf{\Omega}_n^{*}\mathbf{\Phi}_h^{'}\mathbf{\Pi}_{\text{In}}\mathbf{e}_g \ .$$

From the following relation,

$$\mathbf{\Omega}_n^{*\frac{1}{2}} = \begin{pmatrix} \mathbf{\Omega}^{\frac{1}{2}} & \mathbf{O}^{'} \\ \mathbf{O} & \omega_3^{\frac{1}{2}}\mathbf{I}_{K_{2n}} \end{pmatrix} ,$$

we obtain the expressions that

$$\underset{2\times K_n}{\mathbf{\Pi}_{\text{In}}^{'}\mathbf{\Phi}_h\mathbf{\Omega}_n^{*\frac{1}{2}}} = \left(\sum_{s=0}^{h-1}\mathbf{\Pi}_{1\text{I}}^{'}\mathbf{\Pi}_1^{s}\mathbf{\Omega}^{\frac{1}{2}}, \quad \sum_{s=0}^{h-1}\mathbf{\Pi}_{1\text{I}}^{'}\mathbf{\Phi}_{1s}\mathbf{\Pi}_{2n}^{'} + \pi_{33}^{s}\omega_3^{\frac{1}{2}}\mathbf{\Pi}_{1n}^{'}\right) .$$

$\mathbf{\Pi}_{2n}^{'}\mathbf{\Pi}_{2n}$ does not depend on K_{2n}, so that for all h, $\mathbf{e}_g^{'}\mathbf{\Pi}_{\text{In}}^{'}\mathbf{\Phi}_h\mathbf{\Omega}_n^{*}\mathbf{\Phi}_h^{'}\mathbf{\Pi}_{\text{In}}\mathbf{e}_g = O(1)$. That is,

$$\sum_{h=1}^{T-t}\mathbf{e}_g^{'}\mathbf{\Pi}_{\text{In}}^{'}\mathbf{\Phi}_h\mathbf{\Omega}_n^{*}\mathbf{\Phi}_h^{'}\mathbf{\Pi}_{\text{In}}\mathbf{e}_g = O(T-t) \ .$$

Therefore,

$$\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{\Pi}_{\text{In}}^{'}\tilde{\mathbf{V}}_T^{'}\tilde{\mathbf{V}}_T\mathbf{\Pi}_{\text{In}}\mathbf{e}_g\right] = O\left(\frac{\log T}{T}\right) \ .$$

[v] $\mathbf{V}^{(f)'} = (\mathbf{v}_{it}^{(f)})$ is the $2 \times n$ matrix. Let $\mathbf{P}_2^{(b)} = \mathbf{Z}_2^{(b)}\left(\mathbf{Z}_2^{(b)'}\mathbf{Z}_2^{(b)}\right)^{-1}\mathbf{Z}_2^{(b)'}$ be the projection matrix consisting of the K_{2n} strongly exogenous variables, where $\mathbf{Z}_1^{(b)'} = (\mathbf{z}_{it-1}^{(1,b)},\ \mathbf{z}_{it-1}^{(2,b)})$ is the $G_* \times n$ instrumental variable matrix consisting of the $G_* = 2$ lagged endogenous variables. We use the decomposition for the projection matrix (cf. Amemiya, 1985),

$$\mathbf{P}^{(b)} = \mathbf{P}_2^{(b)} + \bar{\mathbf{P}}_1^{(b)} \ . \tag{6.65}$$

Then,

$$\frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\mathbf{P}^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g = \frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\mathbf{P}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g + \frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\bar{\mathbf{P}}_1^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g \ , \tag{6.66}$$

where

$$
\begin{aligned}
\bar{\mathbf{P}}_1^{(b)} &= \mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\left(\mathbf{Z}_1^{(b)'}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\right)^{-1}\mathbf{Z}_1^{(b)'}\mathbf{Q}_2^{(b)} \ , \\
\mathbf{Q}_2^{(b)} &= \mathbf{I}_n - \mathbf{P}_2^{(b)} \ .
\end{aligned}
$$

For the first term of (6.66), using the fact that $\mathcal{E}[\mathbf{v}_{is}^{(f)}\mathbf{v}_{it}^{(f)'}] = \mathbf{\Omega}$ $(s = t)$ or $\mathbf{O}$ $(s \neq t)$,

$$
\begin{aligned}
\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\mathbf{P}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g\right] &= \mathcal{E}\left[\frac{1}{n}\mathrm{tr}\left(\mathbf{P}_2^{(b)}\mathbf{e}_g^{'}\mathbf{\Omega}\mathbf{e}_g\mathbf{I}_n\right)\right] \\
&= \frac{K_{2n}}{n}\mathbf{e}_g^{'}\mathbf{\Omega}\mathbf{e}_g \\
&\to 0 \ .
\end{aligned}
$$

The first equality is from that the K_{2n} variables are strongly exogenous:

$$
\mathcal{E}\left[\mathbf{V}^{(f)}\mathbf{e}_g\mathbf{e}_g^{'}\mathbf{V}^{(f)'}|\mathbf{Z}_2^{(b)}\right] = \mathcal{E}\left[\mathbf{V}^{(f)}\mathbf{e}_g\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\right] \ .
$$

For the second term of (6.66), we consider the usual normalization because $G_* < \infty$:

$$
\frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\bar{\mathbf{P}}_1^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g = \underbrace{\frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}}_{1\times 2}\underbrace{\left(\frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\right)^{-1}}_{2\times 2}\frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{Q}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g \ .
$$

For the third term,

$$
\begin{aligned}
\frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{Q}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g &= \frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{V}^{(f)}\mathbf{e}_g + \frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{P}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g \qquad (6.67) \\
&\xrightarrow{p} \mathbf{0} \ ,
\end{aligned}
$$

since the first term of (6.67) converges to zero by (6.77), and the second term becomes the following by the CS inequality,

$$
\left|\frac{1}{n}\mathbf{e}_h^{'}\mathbf{Z}_1^{(b)'}\mathbf{P}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g\right| \leq \left(\frac{1}{n}\mathbf{e}_h^{'}\mathbf{Z}_1^{(b)'}\mathbf{Z}_1^{(b)}\mathbf{e}_h\right)^{\frac{1}{2}}\left(\frac{1}{n}\mathbf{e}_g^{'}\mathbf{V}^{(f)'}\mathbf{P}_2^{(b)}\mathbf{V}^{(f)}\mathbf{e}_g\right)^{\frac{1}{2}} \ .
$$

Since $K_{2n}/n \to 0$, the lemma is verified. □

We return to the proof of theorem. $\mathbf{Z}^{(f)'}$ is the $K_n \times n$ matrix as follows:

$$
\begin{aligned}
\mathbf{Z}^{(f)'} &= \tilde{\mathbf{W}}' - \tilde{\mathbf{V}}_T^{'} \\
&= \left(\tilde{\mathbf{W}}_0^{'}, \ \cdots, \ \tilde{\mathbf{W}}_{T-1}^{'}\right) - \left(\tilde{\mathbf{V}}_{1T}^{'}, \ \cdots, \ \tilde{\mathbf{V}}_{(T-1)T}^{'}\right) \ .
\end{aligned}
$$

Then, the first term of (6.63) is decomposed as follows:

$$\begin{aligned}\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\mathbf{Z}^{(f)}\mathbf{P}^{(b)}\mathbf{Z}^{(f)'}\mathbf{\Pi}_{\mathrm{I}n} &= \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}'\mathbf{P}^{(b)}\tilde{\mathbf{W}}\mathbf{\Pi}_{\mathrm{I}n} - \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}'\mathbf{P}^{(b)}\tilde{\mathbf{V}}_T\mathbf{\Pi}_{\mathrm{I}n} \\ &\quad -\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{V}}_T'\mathbf{P}^{(b)}\tilde{\mathbf{W}}\mathbf{\Pi}_{\mathrm{I}n} + \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{V}}_T'\mathbf{P}^{(b)}\tilde{\mathbf{V}}_T\mathbf{\Pi}_{\mathrm{I}n} \ .\end{aligned} \tag{6.68}$$

Moreover, this first term is decomposed as follows:

$$\begin{aligned}\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}'\mathbf{P}^{(b)}\tilde{\mathbf{W}}\mathbf{\Pi}_{\mathrm{I}n} &= \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_1'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{I}n} - \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_1'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_2\mathbf{\Pi}_{\mathrm{I}n} \\ &\quad -\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_2'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{I}n} + \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_2'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_2\mathbf{\Pi}_{\mathrm{I}n} \\ &= \mathbf{G}_{n11} - \mathbf{G}_{n12} - \mathbf{G}_{n12} + \mathbf{G}_{n13} \ \ (say,) \ .\end{aligned}$$

For the backward filter,

$$\begin{aligned} f_t\mathbf{w}_{it-1} &= \mathbf{z}_{it-1}^{(b)} + \frac{f_t}{t}(\mathbf{w}_{it-2} + \cdots + \mathbf{w}_{i,-1}) \ , \\ &= \mathbf{z}_{it-1}^{(b)} + \tilde{\mathbf{w}}_{it-1,0} \ \ (say,). \end{aligned} \tag{6.69}$$

Using this expression, the $K_n \times n$ matrix is given by

$$\tilde{\mathbf{W}}_1' = \mathbf{Z}^{(b)'} + \tilde{\mathbf{W}}_0' \ , \tag{6.70}$$

then,

$$\begin{aligned}\mathbf{G}_{n11} &= \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)}\mathbf{\Pi}_{\mathrm{I}n} + \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\mathbf{Z}^{(b)'}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n} \\ &\quad +\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\mathbf{Z}^{(b)}\mathbf{\Pi}_{\mathrm{I}n} + \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n} \ .\end{aligned} \tag{6.71}$$

Since the maximum eigenvalue of $\mathbf{P}^{(b)}$ is unity, so that this fourth term is evaluated as follows by [*iv*] of Lemma 2.7,

$$\begin{aligned}\frac{1}{n}\mathbf{e}_g'\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\mathbf{P}^{(b)}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g &\leq \frac{1}{n}\mathbf{e}_g'\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g \\ &\stackrel{p}{\to} 0 \ .\end{aligned}$$

Similarly, the first term of (6.71) becomes

$$\begin{aligned}\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)}\mathbf{\Pi}_{\mathrm{I}n} &= \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_1'\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{I}n} - \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_1'\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n} \\ &\quad -\frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\tilde{\mathbf{W}}_1\mathbf{\Pi}_{\mathrm{I}n} + \frac{1}{n}\mathbf{\Pi}_{\mathrm{I}n}'\tilde{\mathbf{W}}_0'\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n} \\ &\stackrel{p}{\to} \mathbf{\Phi}^* \ ,\end{aligned} \tag{6.72}$$

because of (6.70), [i], and [ii] of Lemma 2.7. Therefore, we obtain

$$\frac{1}{n}\mathbf{G}_{n11} \ \stackrel{p}{\longrightarrow}\ \mathbf{\Phi}^* \ .$$

Since $(1/n)\mathbf{G}_{n13}$ converges to $\mathbf{O}$ by [iv] of Lemma 2.7, and $\mathbf{P}^{(b)}$ is idempotent, we have

$$\left(\mathbf{e}_g^{'}\frac{1}{n}\mathbf{G}_{n12}\mathbf{e}_h\right)^2 \leq \left(\mathbf{e}_g^{'}\frac{1}{n}\mathbf{G}_{n11}\mathbf{e}_g\right)\left(\mathbf{e}_h^{'}\frac{1}{n}\mathbf{G}_{n13}\mathbf{e}_h\right) \ .$$

Thus, the first term of (6.63) also converges in probability to $\mathbf{\Phi}^*$. Since the fourth term of (6.68) can be ignored by [iv] of Lemma 2.7, it holds that

$$\frac{1}{n}\mathbf{G}_{n1}^{(f,b)} \ \stackrel{p}{\longrightarrow}\ \mathbf{\Phi}^* \ .$$

For the fourth term of (6.63), it follows that

$$\frac{1}{n}\mathbf{G}_{n3}^{(f,b)} \ \stackrel{p}{\longrightarrow}\ \mathbf{O} \ ,$$

by [v] of Lemma 2.7. Therefore, we obtain

$$\frac{1}{n}\mathbf{G}_{n}^{(f,b)} \ \stackrel{p}{\longrightarrow}\ \underset{3\times 3}{\mathbf{G}_0} = \mathbf{\Theta}_{\mathrm{I}}^{'}\mathbf{\Phi}^*\mathbf{\Theta}_{\mathrm{I}} \ .$$

Regarding $(1/n)\mathbf{H}_n^{(f,b)}$,

$$\frac{1}{n}\mathbf{H}_n^{(f,b)} = \frac{1}{n}\begin{pmatrix} \mathbf{y}^{(1,f)'} \\ \mathbf{X}^{(f)'} \end{pmatrix}\left(\mathbf{y}^{(1,f)},\ \mathbf{X}^{(f)}\right) - \frac{1}{n}\mathbf{G}_n^{(f,b)} \ . \tag{6.73}$$

Thus, evaluating the following is sufficient:

$$\begin{pmatrix} \mathbf{y}^{(1,f)'} \\ \mathbf{X}^{(f)'} \end{pmatrix}\left(\mathbf{y}^{(1,f)},\ \mathbf{X}^{(f)}\right) = \mathbf{H}_{n1}^{(f)} + \mathbf{H}_{n2}^{(f)} + \mathbf{H}_{n2}^{(f)'} + \mathbf{H}_{n3}^{(f)} \ ,$$

where

$$\begin{aligned} \mathbf{H}_{n1}^{(f)} &= \mathbf{\Theta}_{\mathrm{I}}^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\mathbf{Z}^{(f)'}\mathbf{Z}^{(f)}\mathbf{\Pi}_{\mathrm{I}n}\mathbf{\Theta}_{\mathrm{I}} \ , \\ \mathbf{H}_{n2}^{(f)} &= \mathbf{\Theta}_{\mathrm{I}}^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\mathbf{Z}^{(f)'}\left(\mathbf{V}^{(f)},\ \mathbf{0}\right) \ , \\ \mathbf{H}_{n3}^{(f)} &= \left(\mathbf{V}^{(f)},\ \mathbf{0}\right)^{'}\left(\mathbf{V}^{(f)},\ \mathbf{0}\right) \ . \end{aligned}$$

$(1/n)\mathbf{H}_{n1}^{(f)}$ converges in probability to $\mathbf{G}_0$ by using the arugumens of $(1/n)\mathbf{G}_{n1}^{(f)}$. $(1/n)\mathbf{H}_{n2}^{(f)}$ converge to $\mathbf{O}$. For $(1/n)\mathbf{H}_{n3}^{(f)}$,

$$\frac{1}{n}\mathbf{H}_{n3}^{(f,b)} \ \stackrel{p}{\longrightarrow}\ \underset{3\times 3}{\mathbf{H}_0} = \begin{pmatrix} \mathbf{\Omega} & \mathbf{0}^{'} \\ \mathbf{0} & 0 \end{pmatrix} \ .$$

Therefore,

$$\frac{1}{n}\mathbf{H}_n^{(f,b)} \xrightarrow{p} \mathbf{H}_0 \ .$$

Following Akashi and Kunitomo (2015), we derive $\mathbf{b}_1 = \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ under many instruments variavles, where $\hat{\boldsymbol{\theta}} = (1, \ -\hat{\boldsymbol{\theta}}_{\text{DL}})$. Define

$$\begin{aligned}
\lambda_{1n} &= \sqrt{n}\lambda \ , \\
\mathbf{G}_{1n} &= \sqrt{n}\left(\frac{1}{n}\mathbf{G}_n^{(f,b)} - \mathbf{G}_0\right) \ , \\
\mathbf{H}_{1n} &= \sqrt{n}\left(\frac{1}{n}\mathbf{H}_n^{(f,b)} - \mathbf{H}_0\right) \ .
\end{aligned}$$

For $\mathbf{G}_{1n}$, it holds that $(1/\sqrt{n})\mathbf{G}_{n2}^{(f,b)} = O_p(1)$ by the argument of (6.77) and $(1/\sqrt{n})\mathbf{G}_{n3}^{(f,b)} = O_p(1)$ by (6.80). For $\mathbf{G}_{n1}^{(f,b)}$, the cross terms that appear from (6.68) to (6.72) are $O_p(\log T/T)$. For instance, the cross term of (6.72) becomes

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{n}\mathbf{\Pi}_{\text{I}n}^{'}\tilde{\mathbf{W}}_1^{'}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\text{I}n}\right] &= \frac{1}{T}\sum_{t=1}^{T-1}\frac{f_t^2}{t}\mathcal{E}\left[\mathbf{w}_{it,n}\sum_{h=0}^{t-1}\mathbf{w}_{ih,n}^{'}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T-1}O\left(\frac{1}{t}\right) \\
&= O\left(\frac{\log T}{T}\right) \ .
\end{aligned}$$

Using the result of Lemma 2.6,

$$\begin{aligned}
\mathbf{G}_{1n} &= O_p(1) + O_p(\sqrt{NT}\frac{\log T}{T}) \\
&= O_p(1) \ ,
\end{aligned}$$

because $\sqrt{NT}/T = O(T^{-1/4})$ by the assumption. Then, we have that $\mathbf{H}_{1n} = O_p(1)$. Since λ and $\hat{\boldsymbol{\theta}}_{\text{DL}}$ are the continuous functions of $\mathbf{G}_n^{(f,b)}$ and $\mathbf{H}_n^{(f,b)}$, $\lambda_{1n} = O_p(1)$ and $\mathbf{b}_1 = O_p(1)$. Substituting these into (3.48) and using $\mathbf{G}_0\boldsymbol{\theta} = \mathbf{0}$ under $K_n/n \to 0$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left[\mathbf{G}_{1n} - \lambda_{1n}\mathbf{H}_0\right]\boldsymbol{\theta} + \frac{1}{\sqrt{n}}\mathbf{G}_0\mathbf{b}_1 \qquad (6.74) \\
&= \frac{\lambda_{1n}}{\sqrt{n}}\frac{1}{\sqrt{n}}\mathbf{H}_{1n}\boldsymbol{\theta} + \frac{\lambda_{1n}}{\sqrt{n}}\left(\mathbf{H}_0 + \frac{1}{\sqrt{n}}\mathbf{H}_{1n}\right)\frac{1}{\sqrt{n}}\mathbf{b}_1 \\
&= o_p\left(\frac{1}{\sqrt{n}}\right) \ .
\end{aligned}$$

By multiplying $\boldsymbol{\theta}^{'}$ on the left of (6.74),

$$\lambda_{1n} = \frac{\boldsymbol{\theta}^{'}\mathbf{G}_{1n}\boldsymbol{\theta}}{\boldsymbol{\theta}^{'}\mathbf{H}_0\boldsymbol{\theta}} + o_p(1) \ .$$

In addtion, multiply by $(\mathbf{0},\ \mathbf{I}_{G_2+K_1})$ on the left of (6.74) and substitute λ_{1n}, then,

$$\boldsymbol{\Phi}^{*}\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{DL}} - \boldsymbol{\theta}_1) = (\mathbf{0},\ \mathbf{I}_2)\left(\mathbf{I}_3 - \frac{1}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\left[\begin{array}{c}\boldsymbol{\Omega}\boldsymbol{\beta}\\ 0\end{array}\right]\boldsymbol{\theta}^{'}\right)\mathbf{G}_{1n}\boldsymbol{\theta} + o_p(1) \ . \quad (6.75)$$

From the relation that $\boldsymbol{\Theta}_{\text{I}}\boldsymbol{\theta} = \mathbf{0}$,

$$\mathbf{G}_{1n}\boldsymbol{\theta} = \frac{1}{\sqrt{n}}\boldsymbol{\Theta}_{\text{I}}^{'}\boldsymbol{\Pi}_{\text{In}}^{'}\mathbf{Z}^{(f)'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} + \frac{1}{\sqrt{n}}\left(\mathbf{V}^{(f)},\ \mathbf{0}\right)^{'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} \ .$$

Therefore, we obtain

$$\boldsymbol{\Phi}^{*}\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{DL}} - \boldsymbol{\theta}_1) = \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{\text{In}}^{'}\mathbf{Z}^{(f)'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} + \frac{1}{\sqrt{n}}\left(\mathbf{u}^{(\perp,f)},\ \mathbf{0}\right)^{'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} + o_p(1) \ , \quad (6.76)$$

where

$$\underset{n\times 1}{\mathbf{u}^{(f)}} = (u_{it}^{(f)}) \ ,$$
$$\underset{1\times n}{\mathbf{u}^{(\perp,f)'}} = (0,1)\left(\mathbf{I}_2 - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}\right)\mathbf{V}^{(f)'} \ .$$

In the following, we evaluate the effects of the forward filter. From the relations that

$$u_{it}^{(f)} = \frac{1}{f_t}\left(u_{it} - \bar{u}_{it,T}\right) \ ,$$

and $b_t = f_t$, we use the following expression:

$$\underset{K_n\times 1}{\mathbf{z}_{it}^{(f)}} = \mathbf{z}_{it}^{(b)} + \left(\tilde{\mathbf{w}}_{it-1,0} - \tilde{\mathbf{w}}_{it-1,T}\right) \ .$$

Then, the first term of (6.76) becomes the followings by $\mathbf{P}^{(b)}\mathbf{Z}^{(b)} = \mathbf{Z}^{(b)}$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\boldsymbol{\Pi}_{\text{In}}^{'}\mathbf{Z}^{(f)'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\boldsymbol{\Pi}_{\text{In}}^{'}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)} + \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\boldsymbol{\Pi}_{\text{In}}^{'}\left(\tilde{\mathbf{W}}_0 - \tilde{\mathbf{W}}_T\right)^{'}\mathbf{P}^{(b)}\mathbf{u}^{(f)} \\ &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\boldsymbol{\Pi}_{\text{In}}^{'}\mathbf{W}^{'}\mathbf{u} - r_{11g} - r_{12g} + r_{13g} - r_{14g} \ , \quad (6.77)\end{aligned}$$

where

$$\begin{aligned}
r_{11g} &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\bar{\mathbf{W}}_0^{'}\mathbf{u} \ , \\
r_{12g} &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\mathbf{W}^{'}\bar{\mathbf{u}}_T - \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\bar{\mathbf{W}}_0^{'}\bar{\mathbf{u}}_T \ , \\
r_{13g} &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\left(\tilde{\mathbf{W}}_0 - \tilde{\mathbf{W}}_T\right)^{'}\mathbf{P}^{(b)}\mathbf{u}_{\sim} \ , \\
r_{14g} &= \frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\left(\tilde{\mathbf{W}}_0 - \tilde{\mathbf{W}}_T\right)^{'}\mathbf{P}^{(b)}\bar{\mathbf{u}}_{T\sim} \ ,
\end{aligned}$$

and the notations are as follows:

$$\begin{aligned}
\underset{n\times 1}{\mathbf{u}} &= (u_{it}) \ , \\
\mathbf{u}_{\sim} &= (u_{it\sim}) \ , \ \ u_{it\sim} = \frac{u_{it}}{f_t} \ , \\
\bar{\mathbf{u}}_T &= (\bar{u}_{it,T}) \ , \ \ \bar{u}_{it,T} = \frac{1}{T-t+1}(u_{it} + \cdots + u_{iT}) \ , \\
\bar{\mathbf{u}}_{T\sim} &= \left(\frac{\bar{u}_{it,T}}{f_t}\right) \ , \\
\underset{K_n\times n}{\bar{\mathbf{W}}_0} &= (\bar{\mathbf{w}}_{it-1,0}) \ , \ \ \bar{\mathbf{w}}_{it-1,0} = \frac{1}{t}(\mathbf{w}_{it-2} + \cdots + \mathbf{w}_{i,-1}) \ , \\
\tilde{\mathbf{W}}_0 &= (\tilde{\mathbf{w}}_{it-1,0}) \ , \\
\tilde{\mathbf{W}}_T &= (\tilde{\mathbf{w}}_{it-1,T}) \ , \ \ \tilde{\mathbf{w}}_{it-1,T} = \frac{f_t}{T-t}(\mathbf{w}_{it} + \cdots + \mathbf{w}_{iT-1}) \ .
\end{aligned}$$

(***)We show that the terms from r_{11g} to r_{14g} are asymptotically negligible. For r_{11g}, using $\mathcal{E}[r_{11g}] = 0$ and the i.i.d. assumption,

$$\begin{aligned}
\mathcal{V}ar[r_{11g}] &= \frac{1}{T}\sum_{t=1}^{T-1}\mathcal{V}ar\left[\mathbf{e}_g^{'}\bar{\mathbf{w}}_{it,n}u_{it}\right] \\
&= O\left(\frac{\log T}{T}\right) \ ,
\end{aligned}$$

where the first equality is from that $\bar{\mathbf{w}}_{is,n}u_{is}$ and $\bar{\mathbf{w}}_{it,n}u_{it}$ are uncorrelated ($s \neq t$).

Regarding the first term of r_{12g}, similar to the result of Akashi and Kunitomo (2015),

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{n}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\mathbf{W}^{'}\bar{\mathbf{u}}_T\bar{\mathbf{u}}_T^{'}\mathbf{W}\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right] &= \frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathcal{E}\left[\mathbf{e}_g^{'}\mathbf{w}_{it,n}\bar{u}_{it,T}\bar{u}_{is,T}\mathbf{w}_{is,n}^{'}\mathbf{e}_g\right] \\
&= O\left(\frac{\log T}{T}\right) \ .
\end{aligned}$$

The second term of r_{12g} is the following by the CS inequality,

$$\begin{aligned}
\mathcal{E}\left[|\frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\bar{\mathbf{W}}_0^{'}\bar{\mathbf{u}}_T|\right] &\leq \sqrt{\frac{N}{T}}\sum_{t=1}^{T-1}\left(\frac{1}{t}\mathcal{E}\left[(\sqrt{t}\mathbf{e}_g^{'}\bar{\mathbf{w}}_{it,n})^2\right]\right)^{\frac{1}{2}}\left(\frac{1}{t}\mathcal{E}\left[(\sqrt{t}\bar{u}_{it,T})^2\right]\right)^{\frac{1}{2}} \\
&= \sqrt{\frac{N}{T}}O(\log T) \\
&= O\left(\frac{\log T}{T^{\frac{1}{4}}}\right), \qquad (6.78)
\end{aligned}$$

where the last equation is due to assumption (A1$'$).

For the first term of r_{13g}, using the CS inequality,

$$\begin{aligned}
|\frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_0^{'}\mathbf{P}^{(b)}\mathbf{u}_{\sim}| &\leq \left(\frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\tilde{\mathbf{W}}_0^{'}\tilde{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right)^{\frac{1}{2}}\left(\frac{1}{\sqrt{n}}\mathbf{u}_{\sim}^{'}\mathbf{P}^{(b)}\mathbf{u}_{\sim}\right)^{\frac{1}{2}} \\
&\leq \left(\frac{f_{T-1}^{-2}}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\bar{\mathbf{W}}_0^{'}\bar{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right)^{\frac{1}{2}}\left(\frac{1}{\sqrt{n}}\mathbf{u}_{\sim}^{'}\mathbf{P}^{(b)}\mathbf{u}_{\sim}\right)^{\frac{1}{2}}, \qquad (6.79)
\end{aligned}$$

where the second inequality is from that $1/2 = f_{T-1}^2 \leq f_t^2$. Regarding this first term, similar to (6.78),

$$\mathcal{E}\left[\frac{1}{\sqrt{n}}\mathbf{e}_g^{'}\mathbf{\Pi}_{\mathrm{I}n}^{'}\bar{\mathbf{W}}_0^{'}\bar{\mathbf{W}}_0\mathbf{\Pi}_{\mathrm{I}n}\mathbf{e}_g\right] = O\left(\frac{\log T}{T^{\frac{1}{4}}}\right).$$

For the second term of (6.79), it is sufficient to show that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbf{u}_{\sim}^{'}\mathbf{P}^{(b)}\mathbf{u}_{\sim} &= \frac{1}{\sqrt{n}}\mathbf{u}_{\sim}^{'}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim} + \frac{1}{\sqrt{n}}\mathbf{u}_{\sim}^{'}\bar{\mathbf{P}}_1^{(b)}\mathbf{u}_{\sim} \\
&= O_p(1) + o_p(1). \qquad (6.80)
\end{aligned}$$

Let p_{hj} be the element of $\mathbf{P}_2^{(b)}$. For $h,\ j,\ \ell,\ m = 1,\ \cdots,\ n = N(T-1)$, we have the following expression:

$$\mathcal{E}\left[\frac{1}{n}\left(\mathbf{u}_{\sim}^{'}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim}\right)^2|\mathbf{Z}_2^{(b)}\right] = \frac{1}{n}\sum_{h,\,j,\,\ell,\,m}^{n} p_{hj}p_{\ell m}\mathcal{E}\left[u_{h\sim}u_{j\sim}u_{\ell\sim}u_{m\sim}\right],$$

where $u_{h\sim}$ are mutually independent given $\mathbf{Z}_2^{(b)}$. In the case of $\{h = j = \ell = m\}$, it is evaluated as $O(K_{2n}/n)$ since $p_{hh} \leq 1$. In the case of $\{h = \ell \neq j = m\}$ or $\{h = m \neq j = \ell\}$, it is also evaluated as $O(K_{2n}/n)$ by the fact that $\mathbf{P}_2^{(b)} = \mathbf{P}_2^{(b)2} = \mathbf{P}_2^{(b)'}$. Thus, the leading term becomes the case when $\{h = j \neq \ell = m\}$:

$$\begin{aligned}
\frac{1}{n}\sum_{h}^{n}\sum_{\ell}^{n} p_{hh}p_{\ell\ell}\mathcal{E}\left[u_{h\sim}^2\right]\mathcal{E}\left[u_{\ell\sim}^2\right] &\leq \frac{f_{T-1}^{-4}\sigma^4}{n}\sum_{h}^{n}p_{hh}\sum_{\ell}^{n}p_{\ell\ell} \\
&= O\left(\frac{(\mathrm{tr}(\mathbf{P}_2^{(b)}))^2}{n}\right), \qquad (6.81)
\end{aligned}$$

where the inequality is from that $p_{hh},\ p_{\ell\ell} \geq 0$. Since $K_{2n}^2/n = O(1)$ by the assumption, the order of the first term in (6.80) is verified. For the second term of (6.80),

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathbf{u}_{\sim}'\bar{\mathbf{P}}_1^{(b)}\mathbf{u}_{\sim} &= \frac{1}{n}\mathbf{u}_{\sim}'\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\left(\frac{1}{n}\mathbf{Z}_1^{(b)\prime}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(b)\prime}\mathbf{Q}_2^{(b)}\mathbf{u}_{\sim} \\ &= o_p(1)\ , \qquad (6.82)\end{aligned}$$

since

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(b)\prime}\mathbf{Q}_2^{(b)}\mathbf{u}_{\sim} &= \frac{1}{\sqrt{n}}\mathbf{Z}_1^{(b)\prime}\mathbf{u}_{\sim} - \frac{1}{\sqrt{n}}\mathbf{Z}_1^{(b)\prime}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim} \\ &= O_p(1)\ ,\end{aligned}$$

this reason is that $1/2 \leq f_t^2 < 1$ and Lemma 2.7 below. Therefore, the first term of (6.82) is $O_p(1/\sqrt{n})$.

Lemma 2.7 : Suppose the assumptions (A1$'$) and (A2). For $N,\ T,\ K_n \to \infty$,

$$\frac{1}{\sqrt{n}}\mathbf{W}_1'\mathbf{P}_2^{(b)}\mathbf{u} = O_p(1)\ .$$

where $\mathbf{W}_1' = (\mathbf{w}_{-1}^{(1)},\ \mathbf{w}_{-1}^{(2)})'$ is the $2 \times n$ matrix consisting of the lagged endogenous variables.

Proof : We first consider the second element $w_{it-1}^{(2)}$ in the following reduced form:

$$\begin{bmatrix} w_{it-1}^{(1)} \\ w_{it-1}^{(2)} \end{bmatrix} = \mathbf{\Pi}_1 \begin{bmatrix} w_{it-2}^{(1)} \\ w_{it-2}^{(2)} \end{bmatrix} + \mathbf{\Pi}_{2n}'\mathbf{w}_{it-1}^{(n)} + \mathbf{v}_{it}\ , \qquad (6.83)$$

where $(\mathbf{w}_{-1}^{(1)},\ \mathbf{w}_{-1}^{(2)})' = (w_{it-1}^{(1)},\ w_{it-1}^{(2)}\ )'$ do not include individual effects and $\mathbf{w}_{i(t-1)}^{(n)\prime} = (w_{it-1}^{(3)},\ \cdots,\ w_{it-1}^{(K_n)})$. Using the moving average representation, we have

$$\begin{aligned}w_{it-1}^{(2)} &= \sum_{s=0}^{t-2}\pi_{22}^s v_{i(t-s-1)}^{(2)} + \sum_{s=0}^{t-2}\pi_{22}^s\boldsymbol{\pi}_{2n}'\mathbf{w}_{i(t-s-1)}^{(n)} + \pi_{22}^{t-1}w_{i0}^{(2)} \\ &= \bar{v}_{it-1}^{(2)} + \bar{w}_{it-1,n}^{(2)} + w_{it-1,0}^{(2)}\ \ (say,)\ .\end{aligned}$$

In the $n \times 1$ vector representation,

$$\mathbf{w}_{-1}^{(2)} = \bar{\mathbf{v}}^{(2)} + \bar{\mathbf{w}}_n^{(2)} + \mathbf{w}_0^{(2)}\ , \qquad (6.84)$$

where $\bar{\mathbf{v}}^{(2)} = (\bar{v}_{it-1}^{(2)})$, $\bar{\mathbf{w}}_n^{(2)} = (\bar{w}_{it-1,n}^{(2)})$, and $\mathbf{w}_0^{(2)} = (\bar{w}_{it-1,0}^{(2)})$.

For the first term of (6.84), we decompose $\bar{\mathbf{v}}^{(2)}$ into the sum of the vectors $(t = 2, \cdots, T-1)$ as follows:

$$\underset{n\times 1}{\bar{\mathbf{v}}^{(2)}} = \sum_{s=2}^{T-1} \pi_{22}^{T-1-s} \mathbf{v}_{[s]}^{(2)} ,$$

where $\mathbf{v}_{[s]}^{(2)} = (\mathbf{v}_{1[s]}^{(2)\prime}, \cdots, \mathbf{v}_{N[s]}^{(2)\prime})'$ and the construction of the $(T-1)\times 1$ vector $\mathbf{v}_{i[s]}^{(2)}$ is as follows:

$$\mathbf{v}_{i[2]}^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ v_{i1}^{(2)} \end{bmatrix} , \quad \mathbf{v}_{i[3]}^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ v_{i1}^{(2)} \\ v_{i2}^{(2)} \end{bmatrix} , \quad \cdots , \quad \mathbf{v}_{i[T-2]}^{(2)} = \begin{bmatrix} 0 \\ 0 \\ v_{i1}^{(2)} \\ \vdots \\ v_{iT-4}^{(2)} \\ v_{iT-3}^{(2)} \end{bmatrix} , \quad \mathbf{v}_{i[T-1]}^{(2)} = \begin{bmatrix} 0 \\ v_{i1}^{(2)} \\ v_{i2}^{(2)} \\ \vdots \\ v_{iT-3}^{(2)} \\ v_{iT-2}^{(2)} \end{bmatrix} .$$

Note that the elements in each low are mutually independent. Therefore, the elemnet of $\mathbf{v}_{[s]}^{(2)}$ are also mutually independent. Meanwhile, $\mathbf{u}_i$ is given by

$$\underset{(T-1)\times 1}{\mathbf{u}_i} = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT-2} \\ u_{iT-1} \end{bmatrix} ,$$

where $\mathcal{E}[v_{it}^{(2)} u_{it}] \neq 0$. From the decomposition,

$$\begin{aligned} \frac{1}{\sqrt{n}} |\bar{\mathbf{v}}^{(2)\prime} \mathbf{P}_2^{(b)} \mathbf{u}| &\leq \frac{1}{\sqrt{n}} \sum_{t=2}^{T-1} |\pi_{22}|^{T-1-t} \left(\mathbf{v}_{[t]}^{(2)\prime} \mathbf{P}_2^{(b)} \mathbf{v}_{[t]}^{(2)} \right)^{\frac{1}{2}} \left(\mathbf{u}' \mathbf{P}_2^{(b)} \mathbf{u} \right)^{\frac{1}{2}} \\ &= \left(\frac{1}{\sqrt{n}} \mathbf{u}' \mathbf{P}_2^{(b)} \mathbf{u} \right)^{\frac{1}{2}} \sum_{t=2}^{T-1} |\pi_{22}|^{T-1-t} \left(\frac{1}{\sqrt{n}} \mathbf{v}_{[t]}^{(2)\prime} \mathbf{P}_2^{(b)} \mathbf{v}_{[t]}^{(2)} \right)^{\frac{1}{2}} \\ &= O_p(1) , \end{aligned}$$

because that the first term is $O_p(1)$ by (6.80), and for the second term it is shown that

$$\mathcal{E} \left[\left(\frac{1}{\sqrt{n}} \mathbf{v}_{[t]}^{(2)\prime} \mathbf{P}_2^{(b)} \mathbf{v}_{[t]}^{(2)} \right)^2 \right] = O\left(\frac{K_{2n}^2}{n} \right) .$$

Then, we obtain the result since $|\pi_{22}| < 1$.

The second term of (6.84) is also $O_p(1)$ since

$$\begin{aligned}
\mathcal{E}\left[\mathcal{E}\left[\left(\frac{1}{\sqrt{n}}\bar{\mathbf{w}}_n^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u}\right)^2 |\mathbf{Z}_2^{(b)}\right]\right] &= \frac{1}{n}\mathcal{E}\left[\bar{\mathbf{w}}_n^{(2)\prime}\mathbf{P}_2^{(b)}\left(\sigma^2\mathbf{I}_n\right)\mathbf{P}_2^{(b)}\bar{\mathbf{w}}_n^{(2)}\right] \\
&\leq \frac{\sigma^2}{n}\mathcal{E}\left[\bar{\mathbf{w}}_n^{(2)\prime}\bar{\mathbf{w}}_n^{(2)}\right] \\
&= O(1)\ .
\end{aligned}$$

Similarly, the third term of (6.84) is as follows given the initial value:

$$\mathcal{E}\left[\mathcal{E}\left[\left(\frac{1}{\sqrt{n}}\mathbf{w}_0^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u}\right)^2 |\mathbf{Z}_2^{(b)},\ \mathbf{w}_0^{(2)}\right]\right] = O(1)\ .$$

The orders for the three terms in (6.84) are $O_p(1)$. Therefore,

$$\frac{1}{\sqrt{n}}\mathbf{w}_{-1}^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u} = O_p(1)\ . \tag{6.85}$$

Similarly, the order of the first element $w_{it-1}^{(1)}$ of (6.83) becomes $O_p(1)$ by using the moving average representation. Thus, we obtain the desired result.

□

We return to the proof of theorem. The second term of r_{13g} is also $o_p(1)$ under the similar arguments. Regarding r_{14g}, using the similar arguments as used for the second term of r_{12g}, we obtain

$$r_{14g} = O_p\left(\frac{\log T}{T^{\frac{1}{4}}}\right)\ ,$$

The first column of the second term in (6.76) is asymptotically negligible. From the decomposition of (6.65),

$$\frac{1}{\sqrt{n}}\mathbf{u}^{(\perp,f)\prime}\mathbf{P}^{(b)}\mathbf{u}^{(f)} = \frac{1}{\sqrt{n}}\mathbf{u}^{(\perp,f)\prime}\mathbf{P}_2^{(b)}\mathbf{u}^{(f)} + \frac{1}{\sqrt{n}}\mathbf{u}^{(\perp,f)\prime}\bar{\mathbf{P}}_1^{(b)}\mathbf{u}^{(f)}\ .$$

Then, the second term becomes as follows:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbf{u}^{(\perp,f)\prime}\bar{\mathbf{P}}_1^{(b)}\mathbf{u}^{(f)} &= \frac{1}{n}\mathbf{u}^{(\perp,f)\prime}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\left(\frac{1}{n}\mathbf{Z}_1^{(b)\prime}\mathbf{Q}_2^{(b)}\mathbf{Z}_1^{(b)}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(b)\prime}\mathbf{Q}_2^{(b)}\mathbf{u}^{(f)} \\
&= o_p(1)\ .
\end{aligned}$$

We evaluate the effects of the forward filter in the second term of (6.76).

$$\frac{1}{\sqrt{n}}\mathbf{u}^{(\perp,f)\prime}\mathbf{P}_2^{(b)}\mathbf{u}^{(f)} = \frac{1}{\sqrt{n}}\tilde{\mathbf{u}}^{\perp\prime}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim} - r_{21} - r_{22} + r_{23}\ , \tag{6.86}$$

where

$$
\begin{aligned}
r_{21} &= \frac{1}{\sqrt{n}} \tilde{\mathbf{u}}^{\perp\prime} \mathbf{P}_2^{(b)} \bar{\mathbf{u}}_{T\sim} , \\
r_{22} &= \frac{1}{\sqrt{n}} \tilde{\mathbf{u}}_T^{\perp\prime} \mathbf{P}_2^{(b)} \mathbf{u}_{\sim} , \\
r_{23} &= \frac{1}{\sqrt{n}} \tilde{\mathbf{u}}_T^{\perp\prime} \mathbf{P}_2^{(b)} \bar{\mathbf{u}}_{T\sim} ,
\end{aligned}
$$

and the notations are as follows:

$$
\begin{aligned}
\underset{n\times 1}{\tilde{\mathbf{u}}^{\perp}} &= \left(f_t u_{it}^{\perp} \right) , \quad u_{it}^{\perp} = (0,1) \left(\mathbf{I}_2 - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}} \right) \mathbf{v}_{it} , \\
\tilde{\mathbf{u}}_{T\sim}^{\perp} &= \left(\tilde{u}_{it,T}^{\perp} \right) , \quad \tilde{u}_{it,T}^{\perp} = \frac{f_t}{T-t} (u_{it+1}^{\perp} + \cdots + u_{iT}^{\perp}) .
\end{aligned}
$$

Similar to r_{13} and r_{14}, the terms from r_{21} to r_{23} are $O(\log T / T^{\frac{1}{4}})$. Furthermore, the first term of (6.86) disappears under $K_{2n}/n \to 0$. This is because that the leading term of (6.81) becomes zero due to

$$
\mathcal{E} \left[u_{it} u_{it}^{\perp} \right] = 0 .
$$

Then,

$$
\mathcal{E} \left[\left(\frac{1}{\sqrt{n}} \tilde{\mathbf{u}}^{\perp\prime} \mathbf{P}_2^{(b)} \mathbf{u}_{\sim} \right)^2 \right] = O\left(\frac{K_{2n}}{n} \right) .
$$

Therefore, we obtain that

$$
\begin{aligned}
\boldsymbol{\Phi}^* \sqrt{n} (\hat{\boldsymbol{\theta}}_{\mathrm{DL}} - \boldsymbol{\theta}_1) &= \frac{1}{\sqrt{n}} \boldsymbol{\Pi}_{In}' \mathbf{W}' \mathbf{u} + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{w}_{it,n} u_{it} \right) + o_p(1) .
\end{aligned}
$$

The 2×1 vector $(1/\sqrt{N}) \sum_i \mathbf{w}_{it,n} u_{it}$ $(t = 1, \cdots, T-1)$ is the martingale difference sequence for any N, K_{2n}. Therefore, from the martingale central limit theorem and [*i*] of Lemma 2.6, it holds that

$$
\boldsymbol{\Phi}^* \sqrt{n} (\hat{\boldsymbol{\theta}}_{\mathrm{DL}} - \boldsymbol{\theta}_1) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Phi}^*) . \tag{6.87}
$$

Thus, [ii] of Theorem 2.14 is verified.

Finally, we consider the sampling error of the GMM estimator:

$$
\boldsymbol{\Phi}^* \sqrt{n} (\hat{\boldsymbol{\theta}}_{\mathrm{DG}} - \boldsymbol{\theta}_1) = \frac{1}{\sqrt{n}} \boldsymbol{\Pi}_{In}' \mathbf{W}' \mathbf{u} + \frac{1}{\sqrt{n}} \left(\tilde{\mathbf{v}}^{(2)}, \ \mathbf{0} \right)' \mathbf{P}_2^{(b)} \mathbf{u}_{\sim} + o_p(1) , \tag{6.88}
$$

where

$$\underset{n\times 1}{\tilde{\mathbf{v}}^{(2)}} = \left(\tilde{v}_{it}^{(2)}\right) , \quad \tilde{v}_{it}^{(2)} = f_t v_{it}^{(2)} .$$

For the first low of the second term in (6.88),

$$\begin{aligned}
\mathcal{E}\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{v}}^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim}\right] &= \frac{1}{\sqrt{n}}\sum_{h=1}^{n} p_{hh}\mathcal{E}\left[\tilde{v}_h^{(2)}u_{h\sim}\right] \\
&= \left(\frac{K_{2n}^2}{n}\right)^{\frac{1}{2}}\mathcal{E}\left[v_{it}^{(2)}u_{it}\right] ,
\end{aligned}$$

where the second equality is form that $f_t/f_t = 1$ for all $h = h$ $(h = 1, \cdots, N(T-1))$. For its second moment,

$$\mathcal{E}\left[\left(\frac{1}{\sqrt{n}}\tilde{\mathbf{v}}^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim}\right)^2\right] = \frac{K_{2n}^2}{n}\left(\mathcal{E}\left[v_{it}^{(2)}u_{it}\right]\right)^2 + O\left(\frac{K_{2n}}{n}\right) .$$

From the variance formula,

$$\mathcal{V}ar\left[\frac{1}{\sqrt{n}}\tilde{\mathbf{v}}^{(2)\prime}\mathbf{P}_2^{(b)}\mathbf{u}_{\sim}\right] = O\left(\frac{K_{2n}}{n}\right) .$$

this term converges in probability to its expectation under $K_{2n}^2/n \to d_2$. Thus, [i] of Theorem 2.14 is verified. □

Proof of Theorem 3.1 : The denominator of λ becomes the following by the proof of Theorem 2.9:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\hat{\boldsymbol{\theta}}'\mathbf{H}_n^{(f,b)}\hat{\boldsymbol{\theta}} \\
&\xrightarrow{p} \left(\boldsymbol{\beta}', -\boldsymbol{\gamma}_1'\right)\begin{pmatrix}\boldsymbol{\Omega} & \mathbf{O} \\ \mathbf{O} & \mathbf{O}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta} \\ -\boldsymbol{\gamma}_1\end{pmatrix} \\
&= \sigma^2 ,
\end{aligned}$$

where $\hat{\boldsymbol{\theta}} = (1, \ -\hat{\boldsymbol{\theta}}_{\text{DL}}')'$. Following Hayashi (2000, Ch. 3), we prepare the notations as follows:

$$\begin{aligned}
\mathbf{m}_y &= \frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{y}^{(1,f)} , \\
\mathbf{M}_x &= \frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{X}^{(f)} , \\
\mathbf{M}_z &= \frac{\hat{\sigma}^2}{n}\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)} .
\end{aligned}$$

The orthogonal condition on a sample is given by

$$\begin{aligned}\mathbf{m}_y - \mathbf{M}_x\hat{\boldsymbol{\theta}}_{\text{DG}} &= \mathbf{H}_{xz}\left(\mathbf{m}_y - \mathbf{M}_x\boldsymbol{\theta}_1\right) \\ &= \frac{1}{n}\mathbf{H}_{xz}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)} \ ,\end{aligned}$$

where

$$\begin{aligned}\mathbf{H}_{xz} &= \mathbf{I}_K - \mathbf{M}_x(\mathbf{M}_x^{'}\mathbf{M}_z^{-1}\mathbf{M}_x)^{-1}\mathbf{M}_x^{'}\mathbf{M}_z^{-1} \ , \\ \hat{\boldsymbol{\theta}}_{\text{DG}} &= (\mathbf{M}_x^{'}\mathbf{M}_z^{-1}\mathbf{M}_x)^{-1}\mathbf{M}_x^{'}\mathbf{M}_z^{-1}\mathbf{m}_y \ .\end{aligned}$$

Then,

$$\begin{aligned}n\lambda &= n\frac{\frac{1}{n}\hat{\boldsymbol{\theta}}^{'}\mathbf{G}_n^{(f,b)}\hat{\boldsymbol{\theta}}}{\hat{\sigma}^2} \\ &= n\left(\mathbf{m}_y - \mathbf{M}_x\hat{\boldsymbol{\theta}}_{\text{DL}}\right)^{'}\mathbf{M}_z^{-1}\left(\mathbf{m}_y - \mathbf{M}_x\hat{\boldsymbol{\theta}}_{\text{DL}}\right) \\ &= \left(\mathbf{M}_x\sqrt{n}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_{\text{DL}}) + \frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right)^{'}\mathbf{M}_z^{-1}\left(\mathbf{M}_x\sqrt{n}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_{\text{DL}}) + \frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right) \\ &= n\left(\mathbf{m}_y - \mathbf{M}_x\hat{\boldsymbol{\theta}}_{\text{DG}}\right)^{'}\mathbf{M}_z^{-1}\left(\mathbf{m}_y - \mathbf{M}_x\hat{\boldsymbol{\theta}}_{\text{DG}}\right) + o_p(1) \\ &= \left(\frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right)^{'}\mathbf{H}_{xz}^{'}\mathbf{M}_z^{-1}\mathbf{H}_{xz}\left(\frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) \ ,\end{aligned}$$

where the fourth equality is from that $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{DL}} - \boldsymbol{\theta}_1) = \sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{DG}} - \boldsymbol{\theta}_1) + o_p(1)$ due to Corollary 2.1. Since $\mathbf{M}_z^{-1}$ is positive definite, expressed as $\mathbf{M}_z^{-1} = \mathbf{L}_z^{'}\mathbf{L}_z$,

$$\begin{aligned}\mathbf{H}_{xz}^{'}\mathbf{M}_z^{-1}\mathbf{H}_{xz} &= \mathbf{H}_{xz}^{'}\mathbf{L}_z^{'}\left[\mathbf{L}_z - \mathbf{L}_z\mathbf{M}_x((\mathbf{L}_z\mathbf{M}_x)^{'}\mathbf{L}_z\mathbf{M}_x)^{-1}(\mathbf{L}_z\mathbf{M}_x)^{'}\mathbf{L}_z\right] \\ &= \mathbf{H}_{xz}^{'}\mathbf{L}_z^{'}\left[\mathbf{I}_K - \mathbf{L}_z\mathbf{M}_x((\mathbf{L}_z\mathbf{M}_x)^{'}\mathbf{L}_z\mathbf{M}_x)^{-1}(\mathbf{L}_z\mathbf{M}_x)^{'}\right]\mathbf{L}_z \\ &= \mathbf{H}_{xz}^{'}\mathbf{L}_z^{'}\mathbf{Q}_{xz}\mathbf{L}_z \ \ (say,) \\ &= \mathbf{L}_z^{'}\mathbf{Q}_{xz}\mathbf{L}_z \ .\end{aligned}$$

Therefore,

$$n\lambda = \left(\frac{1}{\sqrt{n}}\mathbf{L}_z\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right)^{'}\mathbf{Q}_{xz}\left(\frac{1}{\sqrt{n}}\mathbf{L}_z\mathbf{Z}^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) \ .$$

Under null hypothesis, using the result of Theorem 2.9 and that $(\mathbf{L}_z^{'}\mathbf{L}_z)^{-1} \xrightarrow{p} \sigma^2\mathbf{J}^{'}\mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}^{'}]\mathbf{J}$,

$$\frac{1}{\sqrt{n}}\mathbf{L}_z\mathbf{Z}^{(b)'}\mathbf{u}^{(f)} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_K) \ , \tag{6.89}$$

where $\mathbf{Q}_{xz}$ is idempotent because $\text{rank}(\mathbf{Q}_{xz}) = \text{tr}(\mathbf{Q}_{xz}) = K-(G_2+K_1)$. Applying Cochran's theorem, we obtain

$$n\lambda^{(f,b)} \xrightarrow{d} \chi^2_{K_2-G_2} .$$

□

Proof of Theorem 3.2 : Similar to the proof of Theorem 3.1, we prepare the notations as follows:

$$\begin{aligned} \mathbf{M}_{1x} &= \frac{1}{n}\mathbf{Z}_1^{(b)'}\mathbf{X}^{(f)} , \\ \mathbf{M}_{1z} &= \frac{\bar{\sigma}^2}{n}\mathbf{Z}_1^{(b)'}\mathbf{Z}_1^{(b)} , \\ \mathbf{H}_{1xz} &= \mathbf{I}_{K+G_{21}} - \mathbf{M}_{1x}(\mathbf{M}_{1x}'\mathbf{M}_{1z}^{-1}\mathbf{M}_{1x})^{-1}\mathbf{M}_{1x}'\mathbf{M}_{1z}^{-1} , \end{aligned}$$

where $\mathbf{Z}_1^{(b)} = (\mathbf{Y}^{(21,b)}, \mathbf{Z}^{(b)})$ and $\mathbf{Y}^{(21,b)'} = (\mathbf{y}_{it}^{(21)})$ are the $n \times G_{21}$ matrices. Assuming that $\mathbf{y}_{it}^{(21)}$ is also exogenous and applying Theorem 3.1,

$$\bar{\sigma}^2 \xrightarrow{p} \sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta} ,$$

where $\boldsymbol{\Omega}$ becomes the $(1+G_{22}) \times (1+G_{22})$ matrix. Moreover,

$$\begin{aligned} n\lambda_1 &= \left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right)'\mathbf{Q}_{1xz}\left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) \\ &\xrightarrow{d} \chi^2_{K_2-G_{22}} , \end{aligned}$$

where $\mathbf{L}_{1z}'\mathbf{L}_{1z} = \mathbf{M}_{1z}^{-1}$ and

$$\mathbf{Q}_{1xz} = \mathbf{I}_{K+G_{21}} - \mathbf{L}_{1z}\mathbf{M}_{1x}((\mathbf{L}_{1z}\mathbf{M}_{1x})'\mathbf{L}_{1z}\mathbf{M}_{1x})^{-1}(\mathbf{L}_{1z}\mathbf{M}_{1x})' .$$

We define the $K \times (G_{21}+K)$ matrix $\mathbf{J}_{21}'$ such that

$$\mathbf{Z}^{(b)'} = \mathbf{J}_{21}'\mathbf{Z}_1^{(b)'} .$$

Assuming that $\mathbf{y}_{it}^{(21)}$ is endogenous and applying Theorem 3.1,

$$\begin{aligned} n\lambda &= \left(\frac{1}{\sqrt{n}}\mathbf{L}_z\mathbf{J}_{21}'\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right)'\mathbf{Q}_{xz}\left(\frac{1}{\sqrt{n}}\mathbf{L}_z\mathbf{J}_{21}'\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) \\ &= \left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right)'\mathbf{P}_{xz}\left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) , \end{aligned}$$

where

$$\mathbf{P}_{xz} = (\mathbf{L}_z\mathbf{J}_{21}'\mathbf{L}_{1z}^{-1})'\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}'\mathbf{L}_{1z}^{-1} .$$

Therefore,

$$n\lambda_1 - n\lambda = \left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right)^{'}(\mathbf{Q}_{xz}-\mathbf{P}_{xz})\left(\frac{1}{\sqrt{n}}\mathbf{L}_{1z}\mathbf{Z}_1^{(b)'}\mathbf{u}^{(f)}\right) + o_p(1) .$$

Finally, we show that $(\mathbf{Q}_{xz}-\mathbf{P}_{xz})$ is an idempotent matrix.

$$\begin{aligned}
\mathbf{P}_{xz}^2 &= (\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1})^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1}(\mathbf{L}_{1z}^{-1})^{'}\mathbf{J}_{21}\mathbf{L}_z^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= (\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1})^{'}\mathbf{Q}_{xz}\mathbf{L}_z(\mathbf{J}_{21}^{'}\mathbf{M}_{1z}\mathbf{J}_{21})\mathbf{L}_z^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= (\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1})^{'}\mathbf{Q}_{xz}^2\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= \mathbf{P}_{xz} ,
\end{aligned}$$

where the third equality is from that $(\mathbf{L}_z^{'}\mathbf{L}_z)^{-1} = \mathbf{M}_z = \mathbf{J}_{21}^{'}\mathbf{M}_{1z}\mathbf{J}_{21}$, and $\bar{\sigma}^2$ is supposed to be the denominator of λ. In addition,

$$\begin{aligned}
(\mathbf{L}_{1z}\mathbf{M}_{1x})^{'}\mathbf{P}_{xz} &= (\mathbf{L}_{1z}\mathbf{M}_{1x})^{'}(\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1})^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= \mathbf{M}_{1x}^{'}\mathbf{L}_{1z}^{'}(\mathbf{L}_{1z}^{'})^{-1}\mathbf{J}_{21}\mathbf{L}_z^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= (\mathbf{L}_z\mathbf{M}_x)^{'}\mathbf{Q}_{xz}\mathbf{L}_z\mathbf{J}_{21}^{'}\mathbf{L}_{1z}^{-1} \\
&= \mathbf{O} .
\end{aligned}$$

Thus, $\mathbf{Q}_{xz}\mathbf{P}_{xz} = \mathbf{P}_{xz}$. From the above,

$$(\mathbf{Q}_{xz}-\mathbf{P}_{xz})^2 = \mathbf{Q}_{xz}-\mathbf{P}_{xz}$$

it is the idempotent matrix. The degree of freedom becomes

$$\begin{aligned}
\mathrm{tr}(\mathbf{Q}_{xz}-\mathbf{P}_{xz}) &= [(K+G_{21})-(G_2+K_1)] - [K-(G_2+K_1)] \\
&= G_{21} .
\end{aligned}$$

Therefore, we obtain the desired result. □

Proof of Theorem 3.3 : [i] We first consider the case when $N < \infty$ ($c_1 > 0$). From the derivation of Akashi and Kunitomo (2015),

$$\sqrt{n}(\tilde{\lambda} - c_1) = \frac{\boldsymbol{\theta}^{'}[\mathbf{G}_1 - \sqrt{c_1 c_{1*}}\mathbf{H}_1]\boldsymbol{\theta}}{\boldsymbol{\theta}^{'}\mathbf{H}_0\boldsymbol{\theta}} + o_p(1) ,$$

and

$$
\begin{aligned}
& [\mathbf{G}_1 - \sqrt{c_1 c_{1*}} \mathbf{H}_1] \begin{bmatrix} 1 \\ -\boldsymbol{\theta}_1 \end{bmatrix} \qquad (6.90) \\
= \quad & \frac{1}{\sqrt{n}} \boldsymbol{\Theta}_{\mathrm{I}}^{'} \boldsymbol{\Pi}_{\mathrm{I}}^{'} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'} \mathbf{P}_t^{(b)} \mathbf{u}_t^{(f)} + \frac{1}{\sqrt{n}} \left[\sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{pmatrix} \mathbf{P}_t^{(b)} \mathbf{u}_t^{(f)} - r_n \begin{pmatrix} \boldsymbol{\Omega\beta} \\ \mathbf{0} \end{pmatrix} \right] \\
& - \frac{\sqrt{c_1 c_{1*}}}{\sqrt{q_n}} \boldsymbol{\Theta}_{\mathrm{I}}^{'} \boldsymbol{\Pi}'_{\mathrm{I}} \sum_{t=1}^{T-1} \mathbf{Z}_{t-1}^{(f)'} [\mathbf{I}_N - \mathbf{P}_t^{(b)}] \mathbf{u}_t^{(f)} \\
& - \frac{\sqrt{c_1 c_{1*}}}{\sqrt{q_n}} \left[\sum_{t=1}^{T-1} \begin{pmatrix} \mathbf{V}_t^{(f)'} \\ \mathbf{O} \end{pmatrix} [\mathbf{I}_N - \mathbf{P}_t^{(b)}] \mathbf{u}_t^{(f)} - q_n \begin{pmatrix} \boldsymbol{\Omega\beta} \\ \mathbf{0} \end{pmatrix} \right] ,
\end{aligned}
$$

where $q_n = n - r_n$, $r_n = K(T-1)$ and

$$
\begin{aligned}
\mathbf{G}_1 &= \sqrt{n} \left(\frac{1}{n} \mathbf{G}^{(f,b)} - \mathbf{G}_0 \right) , \\
\mathbf{H}_1 &= \sqrt{n} \left(\frac{1}{n} \mathbf{H}^{(f,b)} - \mathbf{H}_0 \right) .
\end{aligned}
$$

Multiplying $\boldsymbol{\theta}$ on the left of (6.90), we have that $\boldsymbol{\theta}' \boldsymbol{\Theta}_{\mathrm{I}}^{'} = \mathbf{0}$ and $\sqrt{c_1 c_{1*}} / \sqrt{q_n} = c_{1*} / \sqrt{n}$. Then,

$$
\begin{aligned}
\sqrt{n}(\tilde{\lambda} - c_1) &= \frac{1}{\sqrt{n}\sigma^2} \sum_{t=1}^{T-1} \mathbf{u}_t^{(f)'} [\mathbf{P}_t^{(b)} - c_{1*}(\mathbf{I}_N - \mathbf{P}_t^{(b)})] \mathbf{u}_t^{(f)} + o_p(1) \\
&= \frac{1}{\sqrt{n}\sigma^2} \sum_{t=1}^{T-1} \mathbf{u}_t^{(f)'} \frac{1}{1-c_1} (\mathbf{P}_t^{(b)} - c_1 \mathbf{I}_N) \mathbf{u}_t^{(f)} + o_p(1) \\
&= \frac{1}{\sqrt{n}\sigma^2} \sum_{t=1}^{T-1} \mathbf{u}_t^{(f)'} \mathbf{N}_t^{(b)} \mathbf{u}_t^{(f)} + o_p(1) \\
&= \frac{1}{\sqrt{T}\sigma^2} \sum_{t=1}^{T-1} \frac{1}{\sqrt{N}} \mathbf{u}_t^{'} \mathbf{N}_t^{(b)} \mathbf{u}_t + o_p(1) ,
\end{aligned}
$$

where the third equality is from that

$$
\mathbf{N}_t^{(b)} = \frac{1}{1-c_1} (\mathbf{P}_t^{(b)} - c_1 \mathbf{I}_N) , \qquad (6.91)
$$

and at the fourth equality the effect of the forward filter is asymptotically negligible even in $N < \infty$ by the result of Akashi and Kunitomo (2015). $(1/\sqrt{N}) \mathbf{u}_t \mathbf{N}_t^{(b)} \mathbf{u}_t$ is the martingale difference sequence, since $c_1 = K/N$, so that

$$
\begin{aligned}
\mathcal{E}_{t-1}[\frac{1}{\sqrt{N}} \mathbf{u}_t^{'} \mathbf{N}_t^{(b)} \mathbf{u}_t] &= \frac{\sigma^2}{\sqrt{N}} \mathrm{tr}(\mathbf{N}_t^{(b)}) \\
&= \frac{\sigma^2}{\sqrt{N}} \frac{1}{1-c_1} (K - \frac{K}{N} N) \\
&= 0 .
\end{aligned}
$$

The conditional variance of each t is given by

$$
\begin{aligned}
\mathcal{E}_{t-1}[(\mathbf{u}_t'\mathbf{N}_t^{(b)}\mathbf{u}_t)^2] &= \sum_i\sum_j\sum_k\sum_\ell n_{ij}^{(t)}n_{k\ell}^{(t)}\mathcal{E}_t[u_{it}u_{jt}u_{kt}u_{\ell t}] \\
&= \mathcal{E}[u_{it}^4]\mathbf{n}_t'\mathbf{n}_t + \sum_i\sum_{k\neq i} n_{ii}^{(t)}n_{kk}^{(t)}(\mathcal{E}[u_{it}^2])^2 \\
&\quad + \sum_i\sum_{j\neq i} n_{ij}^{(t)}n_{ij}^{(t)}\mathcal{E}[u_{it}^2]\mathcal{E}[u_{it}^2] + \sum_i\sum_{j\neq i} n_{ij}^{(t)}n_{ji}^{(t)}(\mathcal{E}[u_{it}^2])^2 \\
&= (\mathcal{E}[u_{it}^4]-3\sigma^4)\mathbf{n}_t'\mathbf{n}_t + \sigma^4[\mathrm{tr}(\mathbf{N}_t^{(b)})]^2 + 2\sigma^4\mathrm{tr}([\mathbf{N}_t^{(b)}]^2) \\
&= (\mathcal{E}[u_{it}^4]-3\sigma^4)\mathbf{n}_t'\mathbf{n}_t + 2\sigma^4\mathrm{tr}([\mathbf{N}_t^{(b)}]^2)\ ,
\end{aligned}
$$

where $\mathbf{n}_t$ denotes the $N\times 1$ vector, which consists of the diagonal elements $n_{ii}^{(t)}$ of $\mathbf{N}_t^{(b)}$. Applying the martingale central limit theorem, for $T\to\infty$,

$$
\sqrt{n}(\tilde{\lambda}-c_1) \ \xrightarrow{d}\ \mathcal{N}(0,\sigma_\lambda^2)\ ,
$$

where

$$
\begin{aligned}
\sigma_\lambda^2 &= \lim_{T\to\infty}\frac{1}{n\sigma^4}\sum_{t=1}^{T-1} 2\sigma^4\mathrm{tr}([\mathbf{N}_t^{(b)}]^2) + (\mathcal{E}[u_{it}^4]-3\sigma^4)\mathcal{E}[\mathbf{n}_t'\mathbf{n}_t] \\
&= 2c_{1*} + \frac{\mathcal{E}[u_{it}^4]-3\sigma^4}{\sigma^4(1-c_1)^2}\left(\lim_{T\to\infty}\frac{1}{n}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{p}_t'\mathbf{p}_t]-c_1^2\right)\ ,
\end{aligned}
\tag{6.92}
$$

and $\mathbf{p}_t$ is the $N\times 1$ vector consisting of the elements $p_{ii}^{(t)}$ of $\mathbf{P}_t^{(b)}$. This is because that $[\mathbf{N}_t^{(b)}]^2 = \mathbf{P}_t^{(b)} + c_{1*}^2(\mathbf{I}_N - \mathbf{P}_t^{(b)})$ and

$$
\frac{1}{n}\sum_{t=1}^{T-1}\mathrm{tr}([\mathbf{N}_t^{(b)}]^2) = \frac{r_n}{n} + \frac{q_n}{n}c_{1*}^2 \ \longrightarrow\ c_{1*}\ .
$$

From $n_{ii}^{(t)} = (p_{ii}^{(t)}-c_1)/(1-c_1)$,

$$
\frac{1}{n}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{n}_t'\mathbf{n}_t] \ \longrightarrow\ \frac{1}{(1-c_1)^2}\left(\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{p}_t'\mathbf{p}_t]-2c_1^2+c_1^2\right)\ .
$$

However, the second term of (6.92) disappears since $\mathcal{E}[u_{it}^4]=3\sigma^4$ due to the assumption of normality. □

[*ii*] We consider the case when $N\to\infty$ ($c_1=0$). From the result of Akashi and Kunitomo (2015), we have the following regardless of c_1,

$$
\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(\frac{1}{\sqrt{n}})\ ,
$$

where $(1, -\tilde{\boldsymbol{\theta}}_{\text{LI}}^{'})' = \tilde{\boldsymbol{\theta}}$ and the first element of $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is zero by definition. Then,

$$n\tilde{\lambda} = \frac{1}{\tilde{\sigma}^2}\left(\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta} - 2\boldsymbol{\theta}'\mathbf{G}^{(f,b)}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{G}^{(f,b)}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})\right) , \qquad (6.93)$$

where $\tilde{\boldsymbol{\theta}}^{'}(1/q_n)\mathbf{H}^{(f)}\tilde{\boldsymbol{\theta}} = \tilde{\sigma}^2$. For the third term, normalized by $\sqrt{T}$,

$$\frac{1}{\sqrt{T}} \times \sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\left(\frac{1}{n}\mathbf{G}^{(f,b)}\right)\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_p(\frac{1}{\sqrt{T}}) .$$

Similarly, for the second term of (6.93),

$$\begin{aligned} & \frac{1}{\sqrt{T}} \times 2\boldsymbol{\theta}'\left(\sqrt{n}\mathbf{G}_0 + \mathbf{G}_1\right)\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ = \quad & O_p(\frac{1}{\sqrt{T}}) + O_p(\sqrt{\frac{N}{T}}\frac{\log T}{\sqrt{T}}) , \end{aligned}$$

since $\boldsymbol{\theta}'\mathbf{G}_0 = \mathbf{0}$ under $c_1 = 0$, and $\mathbf{G}_1 = O_p(1) + O_p(\sqrt{N/T}\log T)$. Therefore, in the following quantity, the second and third terms of (6.93) can be asymptotically ignored,

$$\begin{aligned} \frac{n\tilde{\lambda} - KT}{\sqrt{2KT}} &= \frac{1}{\bar{\sigma}^2}\frac{\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta}}{\sqrt{2KT}} - \frac{KT}{\sqrt{2KT}} + o_p(1) \\ &= \frac{1}{\tilde{\sigma}^2}\frac{\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta} - KT\sigma^2 + KT\sigma^2}{\sqrt{2KT}} - \frac{KT}{\sqrt{2KT}} + o_p(1) \\ &= \frac{\sigma^2}{\tilde{\sigma}^2}\frac{\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta} - KT\sigma^2}{\sqrt{2KT\sigma^4}} + \frac{1}{\tilde{\sigma}^2}(\sigma^2 - \tilde{\sigma}^2)\frac{KT}{\sqrt{2KT}} + o_p(1) \\ &= \frac{\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta} - KT\sigma^2}{\sqrt{2KT\sigma^4}} + O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{\log T}{\sqrt{T}}) + o_p(1) , \qquad (6.94) \end{aligned}$$

where the last equality is from that $(\tilde{\sigma}^2 - \sigma^2) = O_p(1/\sqrt{n}) + O_p(\log T/T)$. Regarding the first term of (6.94),

$$\begin{aligned} \frac{\boldsymbol{\theta}'\mathbf{G}^{(f,b)}\boldsymbol{\theta} - KT\sigma^2}{\sqrt{2KT\sigma^4}} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\frac{\mathbf{u}_t^{(f)'}\mathbf{P}_t^{(b)}\mathbf{u}_t^{(f)} - K\sigma^2}{\sqrt{2K\sigma^4}} + o_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\frac{\mathbf{u}_t'\mathbf{P}_t^{(b)}\mathbf{u}_t - K\sigma^2}{\sqrt{2K\sigma^4}} + o_p(1) . \end{aligned}$$

This is the martingale difference sequence since $\mathcal{E}_{t-1}[\mathbf{u}_t'\mathbf{P}_t^{(b)}\mathbf{u}_t - K\sigma^2] = 0$. The conditional variance of each t is given by

$$\begin{aligned} & \mathcal{E}_{t-1}[(\mathbf{u}_t'\mathbf{P}_t^{(b)}\mathbf{u}_t - K\sigma^2)^2] \\ = \quad & (\mathcal{E}[u_{it}^4] - 3\sigma^4)\mathbf{p}_t'\mathbf{p}_t + \sigma^4[\text{tr}(\mathbf{P}_t^{(b)})]^2 + 2\sigma^4\text{tr}([\mathbf{P}_t^{(b)}]^2) - 2\text{tr}(\mathbf{P}_t^{(b)})K\sigma^4 + K^2\sigma^4 \\ = \quad & (\mathcal{E}[u_{it}^4] - 3\sigma^4)\mathbf{p}_t'\mathbf{p}_t + 2K\sigma^4 . \end{aligned}$$

Similar to Akashi and Kunitomo (2015), we obtain that $\mathcal{E}[(\mathbf{u}_t'\mathbf{P}_t^{(b)}\mathbf{u}_t - K\sigma^2)^4] < \Delta$. Applying the martingale central limit theorem, for $T \to \infty$,

$$\frac{n\tilde{\lambda} - KT}{\sqrt{2KT}} \xrightarrow{d} \mathcal{N}(0, \sigma_\lambda^2) \; ,$$

where

$$\sigma_\lambda^2 \;=\; 1 + \frac{\mathcal{E}[u_{it}^4] - 3\sigma^4}{2K\sigma^4}\left(\lim_{N,T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}\mathcal{E}[\mathbf{p}_t'\mathbf{p}_t]\right) \; .$$

However, if $N \to \infty$, then

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{N} p_{ii}^2 \;&=\; \frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{N}\frac{1}{N^2}\mathbf{z}_{it-1}^{(b)\prime}\left(\frac{1}{N}\mathbf{Z}_t^{(b)\prime}\mathbf{Z}_t^{(b)}\right)^{-1}\mathbf{z}_{it-1}^{(b)} \\
&=\; O_p(\frac{NT}{N^2T}) \\
&=\; o_p(1) \; .
\end{aligned}$$

Since this is a bounded random variable with $0 \le (1/T)\sum_t\sum_i p_{ii}^2 \le (1/T)\sum_t \mathrm{tr}(\mathbf{P}_t^{(b)}) \le K$, the convergence in probability also means the convergence in 1th mean to zero (cf. Sen and Singer, 1993). Thus,

$$\lim_{N,T\to\infty}\mathcal{E}\left[\frac{1}{T}\sum_{t=1}^{T-1}\mathbf{p}_t'\mathbf{p}_t\right] = 0 \; ,$$

or $\sigma_\lambda^2 = 1$.

Finally, for the adjusted degree of freedom $d_T = 2KT - (G_2 + K_1)$, it holds that $d_T/2KT \to 1$ and $(G_2 + K_1)/d_T \to 0$. Therefore, t_0 converges in distribution to the standard normal distribution. □

Proof of Theorem 3.4 : We present the following two lemmas.

Lemma 3.2 : Suppose that for the g-th reduced form, there exists a missing variable in $\mathbf{z}_{it}^{\{1\}}$. Let $\hat{\omega}_{gg}^{\{1\}}$ be the estimator for the variance of error term and $\hat{\omega}_{gg}$ be based on the true instrumental variable $\mathbf{z}_{it}$. Then,

$$\operatorname*{plim}_{n\to\infty}\,(\hat{\omega}_{gg}^{\{1\}} - \hat{\omega}_{gg}) = \delta_{g1} > 0 \; . \qquad (6.95)$$

Proof : For the true instrumental variable $\mathbf{z}_{it}$, the following estimator is consistent:

$$\begin{aligned}\hat{\omega}_{gg} &= \frac{1}{n}\mathbf{y}^{(g,f)'}\mathbf{Q}_0'\mathbf{Q}_0\mathbf{y}^{(g,f)} \\ &= \frac{1}{n}(\mathbf{Z}^{(f)}\boldsymbol{\pi}_g + \mathbf{v}^{(g,f)})'\mathbf{Q}^{\mathrm{m}\prime}{}_0\mathbf{Q}_0(\mathbf{Z}^{(f)}\boldsymbol{\pi}_g + \mathbf{v}^{(g,f)}) \\ &= \omega_{gg} + o_p(1)\ , \end{aligned} \tag{6.96}$$

where

$$\mathbf{Q}_0 = \mathbf{I}_n - \mathbf{Z}^{(f)}(\mathbf{Z}^{(b)'}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(b)'}$$

is idempotent but asymmetric, and the third equality is from that the results of (6.25), (6.45), and

$$\mathbf{Q}_0\mathbf{Z}^{(f)} = \mathbf{O}\ . \tag{6.97}$$

For any candidate $\mathbf{z}_{it}^{\{1\}}$,

$$\begin{aligned}\hat{\omega}_{gg}^{\{1\}} &= \frac{1}{n}\mathbf{y}^{(g,f)'}\mathbf{Q}_1'\mathbf{Q}_1\mathbf{y}^{(g,f)} \\ &= \frac{1}{n}(\mathbf{Z}^{(f)}\boldsymbol{\pi}_g + \mathbf{v}^{(g,f)})'\mathbf{Q}_1'\mathbf{Q}_1(\mathbf{Z}^{(f)}\boldsymbol{\pi}_g + \mathbf{v}^{(g,f)}) \\ &= \frac{1}{n}\boldsymbol{\pi}_g'\mathbf{Z}^{(f)'}\mathbf{Q}_1'\mathbf{Q}_1\mathbf{Z}^{(f)}\boldsymbol{\pi}_g + \omega_{gg} + o_p(1)\ ,\end{aligned}$$

where

$$\mathbf{Q}_1 = \mathbf{I}_n - \mathbf{Z}_{\{1\}}^{(f)}(\mathbf{Z}_{\{1\}}^{(b)'}\mathbf{Z}_{\{1\}}^{(f)})^{-1}\mathbf{Z}_{\{1\}}^{(b)'}$$

is generated by $\mathbf{z}_{it}^{\{1\}(f)}$ and $\mathbf{z}_{it}^{\{1\}(b)}$, and the third equality is form that $\mathcal{E}[\mathbf{Z}_{\{1\}}^{(b)'}\mathbf{v}^{(g,f)}] = \mathbf{0}$ and

$$\frac{1}{n}\mathbf{Z}^{(f)'}\mathbf{Q}_1'\mathbf{Q}_1\mathbf{v}^{(g,f)} \ \xrightarrow{p}\ \mathbf{0}\ .$$

Therefore, for any $\mathbf{z}_{it}^{\{1\}}$, it holds that

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}(\hat{\omega}_{gg}^{\{1\}} - \hat{\omega}_{gg}) &= \operatorname*{plim}_{n\to\infty}\frac{1}{n}(\mathbf{Q}_1\mathbf{Z}^{(f)}\boldsymbol{\pi}_g)'(\mathbf{Q}_1\mathbf{Z}^{(f)}\boldsymbol{\pi}_g) \\ &= \delta^{(g)} \\ &\geq 0\ ,\end{aligned}$$

in the probability limit.

We show that if $\mathbf{z}_{it}^{\{1\}}$ has a missing variable, then $\delta^{(g)} \neq 0$. Consider the regression of $\boldsymbol{\pi}_g'\mathbf{w}_{it-1}$ on $\mathbf{w}_{it-1}^{\{1\}}$ on the population. For some t,

$$\boldsymbol{\pi}_g'\mathbf{w}_{it-1} = \boldsymbol{\rho}_g'\mathbf{w}_{it-1}^{\{1\}} + \epsilon_{it}\ ,$$

where the $K_{\{1\}} \times 1$ coefficient is given by

$$\boldsymbol{\rho}_g = \left(\mathcal{E}\left[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{\{1\}'} \right] \right)^{-1} \mathcal{E}\left[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{'} \right] \boldsymbol{\pi}_g \ .$$

Then, the expectation of the error term becomes

$$\mathcal{E}[\epsilon_{it}] = \boldsymbol{\pi}_g^{'} \mathcal{E}[\mathbf{w}_{it-1}] - \boldsymbol{\rho}_g^{'} \mathcal{E}[\mathbf{w}_{it-1}^{\{1\}}] = 0 \ .$$

For its variance, using $\mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \epsilon_{it}] = \mathbf{0}$,

$$\begin{aligned}
& \mathcal{V}ar[\epsilon_{it}] \\
= \ & \mathcal{E}[(\boldsymbol{\pi}_g^{'} \mathbf{w}_{it-1} - \boldsymbol{\rho}_g^{'} \mathbf{w}_{it-1}^{\{1\}})^2] \\
= \ & \boldsymbol{\pi}_g^{'} \mathcal{E}[\mathbf{w}_{it-1} \mathbf{w}_{it-1}^{'}] \boldsymbol{\pi}_g - \boldsymbol{\rho}_g^{'} \mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{\{1\}'}] \boldsymbol{\rho}_g \\
= \ & \boldsymbol{\pi}_g^{'} \left(\mathcal{E}[\mathbf{w}_{it-1} \mathbf{w}_{it-1}^{'}] - \mathcal{E}[\mathbf{w}_{it-1} \mathbf{w}_{it-1}^{\{1\}'}] (\mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{\{1\}'}])^{-1} \mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{'}] \right) \boldsymbol{\pi}_g \ .
\end{aligned}$$

Now, we have that

$$\begin{aligned}
& \frac{1}{n} \mathbf{Z}^{(f)'} \mathbf{Q}_1^{'} \mathbf{Q}_1 \mathbf{Z}^{(f)} \\
\xrightarrow{p} \ & \mathcal{E}[\mathbf{w}_{it-1} \mathbf{w}_{it-1}^{'}] - \mathcal{E}[\mathbf{w}_{it-1} \mathbf{w}_{it-1}^{\{1\}'}] (\mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{\{1\}'}])^{-1} \mathcal{E}[\mathbf{w}_{it-1}^{\{1\}} \mathbf{w}_{it-1}^{'}] \\
= \ & \boldsymbol{\Gamma}^{\{1\}} \ \ (say,) \ .
\end{aligned}$$

Therefore,

$$\mathcal{V}ar[\epsilon_{it}] = \delta_{g1} \ .$$

Suppose that $\delta_{g1} = 0$. Then, it holds that $\mathcal{V}ar[\epsilon_{it}] = 0$ or $\epsilon_{it} = 0$, and

$$\boldsymbol{\pi}_g^{'} \mathbf{w}_{it-1} = \boldsymbol{\rho}_g^{'} \mathbf{w}_{it-1}^{\{1\}} \ .$$

Since for g, $\mathbf{z}_{it}^{\{1\}}$ does not contain a variable $w_{it-1}^{[k]}$, its coefficient should be $\pi_g^{[k]} \neq 0$. Dividing both sides by $\pi_g^{[k]}$ and transposing the variables other than $w_{it-1}^{[k]}$ to the right-hand side,

$$\mathbf{w}_{it-1}^{[k]} = \boldsymbol{\rho}_{0g}^{'} \mathbf{w}_{it-1}^{\{01\}} \ ,$$

where $\mathbf{w}_{it-1}^{\{01\}} = \mathbf{w}_{it-1} \cup \mathbf{w}_{it-1}^{\{1\}}$ is the $K^{\{01\}} \times 1$ vector. Expressed as the $n \times 1$ vector and $n \times K^{\{01\}}$ matrix,

$$\mathbf{w}^{[k]} = \mathbf{W}^{\{01\}} \boldsymbol{\rho}_{0g} \ .$$

If $\boldsymbol{\rho}_{0g} = \mathbf{0}$, then $\mathbf{w}^{[k]} = \mathbf{0}$. This contradicts rank$(\mathbf{W}) = K^*$ by assumption (A5). If $\boldsymbol{\rho}_{0g} \neq \mathbf{0}$, then $\mathbf{w}^{[k]}$ becomes a linear combination of other variables. However,

$w_{it-1}^{[k]}$ is not included on the right-hand side, which also contradicts rank$(\mathbf{W}) = K^*$. Therefore, it is verified that $\delta_{g1} > 0$. □

If there exists no missing variable, then $\delta^{(g)} = 0$. Therefore, we examine the higher order in the following lemma.

Lemma 3.3 : Suppose that $\mathbf{z}_{it}^{\{1\}}$ includes the true instrumental variables $\mathbf{z}_{it}$. Provided $d < \infty$, for all $g = 1, \cdots, G$,

$$n(\hat{\omega}_{gg} - \hat{\omega}_{gg}^{\{1\}}) \xrightarrow{d} \omega_{gg}\chi^2_{g,\, K_{\{1\}}-K} + \mathcal{N}(0, \sigma_{g1}^2) \ .$$

Proof : Using $\mathbf{Q}_1\mathbf{Z}^{(f)} = \mathbf{O}$,

$$\begin{aligned}
& n(\hat{\omega}_{gg} - \hat{\omega}_{gg}^{\{1\}}) \\
= \ & \mathbf{v}^{(g,f)\prime}\mathbf{Q}_0^{\prime}\mathbf{Q}_0\mathbf{v}^{(g,f)} - \mathbf{v}^{(g,f)\prime}\mathbf{Q}_1^{\prime}\mathbf{Q}_1\mathbf{v}^{(g,f)} \\
= \ & \mathbf{v}^{(g,f)\prime}(-\mathbf{P}_0^{\prime} - \mathbf{P}_0 + \mathbf{P}_0^{\prime}\mathbf{P}_0 + \mathbf{P}_1^{\prime} + \mathbf{P}_1^{\prime} - \mathbf{P}_1^{\prime}\mathbf{P}_1)\mathbf{v}^{(g,f)} \\
= \ & \mathbf{v}^{(g,f)\prime}(-2\mathbf{P}_0 + \mathbf{P}_0^{\prime}\mathbf{P}_0 + 2\mathbf{P}_1 - \mathbf{P}_1^{\prime}\mathbf{P}_1)\mathbf{v}^{(g,f)} \ ,
\end{aligned} \quad (6.98)$$

where

$$\mathbf{P}_0 = \mathbf{Z}^{(f)}(\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(f)})^{-1}\mathbf{Z}^{(b)\prime} \ , \quad \mathbf{P}_1 = \mathbf{Z}_{\{1\}}^{(f)}(\mathbf{Z}_{\{1\}}^{(b)\prime}\mathbf{Z}_{\{1\}}^{(f)})^{-1}\mathbf{Z}_{\{1\}}^{(b)\prime} \ .$$

However, from the derivation of Theorem 2.9,

$$\begin{aligned}
& \mathbf{v}^{(g,f)\prime}\mathbf{P}_1^{\prime}\mathbf{P}_1\mathbf{v}^{(g,f)} \\
= \ & \frac{1}{\sqrt{n}}\mathbf{v}^{(g,f)\prime}\mathbf{Z}_{\{1\}}^{(b)}\left(\frac{1}{n}\mathbf{Z}_{\{1\}}^{(f)\prime}\mathbf{Z}_{\{1\}}^{(b)}\right)^{-1}\frac{1}{n}\mathbf{Z}_{\{1\}}^{(f)\prime}\mathbf{Z}_{\{1\}}^{(f)}\left(\frac{1}{n}\mathbf{Z}_{\{1\}}^{(b)\prime}\mathbf{Z}_{\{1\}}^{(f)}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_{\{1\}}^{(b)\prime}\mathbf{v}^{(g,f)} \\
= \ & \mathbf{v}^{(g,f)\prime}\mathbf{P}_1^{(b)}\mathbf{v}^{(g,f)} + o_p(1) \ ,
\end{aligned}$$

where

$$\mathbf{P}_1^{(b)} = \mathbf{Z}_{\{1\}}^{(b)}(\mathbf{Z}_{\{1\}}^{(b)\prime}\mathbf{Z}_{\{1\}}^{(b)})^{-1}\mathbf{Z}_{\{1\}}^{(b)\prime} \ .$$

Similarly,

$$\mathbf{v}^{(g,f)\prime}\mathbf{P}_0^{\prime}\mathbf{P}_0\mathbf{v}^{(g,f)} = \mathbf{v}^{(g,f)\prime}\mathbf{P}^{(b)}\mathbf{v}^{(g,f)} + o_p(1) \ .$$

Using the simlar arguments of Theorems 2.9 and 2.11 under $d < \infty$, we have the following relations between the forward and backward filters:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathbf{Z}^{(f)\prime}\mathbf{v}^{(g,f)} &= \frac{1}{\sqrt{n}}\mathbf{Z}^{(b)}\mathbf{v}^{(g,f)} - \mathbf{b}_{g\{0\}} + o_p(1) \ , \\
\frac{1}{\sqrt{n}}\mathbf{Z}_{\{1\}}^{(f)\prime}\mathbf{v}^{(g,f)} &= \frac{1}{\sqrt{n}}\mathbf{Z}_{\{1\}}^{(b)}\mathbf{v}^{(g,f)} - \mathbf{b}_{g\{1\}} + o_p(1) \ ,
\end{aligned}$$

where similar to (6.47), the constant vectors $\mathbf{b}_{g\{0\}}$ and $\mathbf{b}_{g\{1\}}$ are given by

$$\begin{aligned}\mathbf{b}_{g\{0\}} &= -d^{\frac{1}{2}}\mathbf{J}'\left(\mathbf{I}_{G_*} - \mathbf{\Pi}^{*'}\right)^{-1}\mathbf{\Omega}^*\mathbf{e}_g \;, \\ \mathbf{b}_{g\{1\}} &= -d^{\frac{1}{2}}\mathbf{J}'_{K_{\{1\}}}\left(\mathbf{I}_{G_*} - \mathbf{\Pi}^{*'}\right)^{-1}\mathbf{\Omega}^*\mathbf{e}_g \;.\end{aligned}$$

Then,

$$\begin{aligned}& 2\mathbf{v}^{(g,f)'}\mathbf{P}_1\mathbf{v}^{(g,f)} \\ =\;& 2\mathbf{v}^{(g,f)'}\mathbf{P}_1^{(b)}\mathbf{v}^{(g,f)} - 2\mathbf{b}'_{g\{1\}}\left(\frac{1}{n}\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{Z}^{(f)}_{\{1\}}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{v}^{(g,f)} + o_p(1) \\ =\;& 2\mathbf{v}^{(g,f)'}\mathbf{P}_1^{(b)}\mathbf{v}^{(g,f)} - q_{n1} + o_p(1) \;\; (say,) \;,\end{aligned}$$

where the second term q_{n1} converges to a normal distribution if $d > 0$ and converges to zero if $d = 0$. In addtion,

$$2\mathbf{v}^{(g,f)'}\mathbf{P}_0\mathbf{v}^{(g,f)} = 2\mathbf{v}^{(g,f)'}\mathbf{P}^{(b)}\mathbf{v}^{(g,f)} - q_{n0} + o_p(1) \;\; (say,) \;.$$

Therefore, (6.98) is expressed as

$$n(\hat{\omega}_{gg} - \hat{\omega}^{\{1\}}_{gg}) = \mathbf{v}^{(g,f)'}(\mathbf{P}_1^{(b)} - \mathbf{P}^{(b)})\mathbf{v}^{(g,f)} + (q_{n0} - q_{n1}) + o_p(1) \;. \tag{6.99}$$

For the second term,

$$(q_{n0} - q_{n1}) \;\; \xrightarrow{d} \;\; \mathcal{N}(0, \sigma^2_{g1}) \;,$$

where

$$\sigma^2_{g1} = 4\omega_{gg}\left(-\mathbf{b}'_{g\{0\}}, \mathbf{b}'_{g\{1\}}\right)\begin{pmatrix} (\mathbf{J}'_{01}\mathbf{\Gamma}_{\{1\}}\mathbf{J}_{01})^{-1} & (\mathbf{J}'_{01}\mathbf{\Gamma}_{\{1\}}\mathbf{J}_{01})^{-1}\mathbf{J}'_{01} \\ \mathbf{J}_{01}(\mathbf{J}'_{01}\mathbf{\Gamma}_{\{1\}}\mathbf{J}_{01})^{-1} & \mathbf{\Gamma}^{-1}_{\{1\}} \end{pmatrix}\begin{pmatrix} -\mathbf{b}_{g\{0\}} \\ \mathbf{b}_{g\{1\}} \end{pmatrix} \;,$$

$\mathbf{\Gamma}_{\{1\}} = \mathcal{E}[\mathbf{w}^{\{1\}}_{it-1}\mathbf{w}^{\{1\}'}_{it-1}]$, and $\mathbf{J}'_{01}$ is the $K \times K_{\{1\}}$ matrix such that

$$\mathbf{Z}^{(b)} = \mathbf{Z}^{(b)}_{\{1\}}\mathbf{J}_{01} \;.$$

Using $\mathbf{J}_{01}$, we rewrite the first term of (6.99) as follows:

$$\begin{aligned}& \frac{1}{\omega_{gg}}\mathbf{v}^{(g,f)'}(\mathbf{P}_1^{(b)} - \mathbf{P}^{(b)})\mathbf{v}^{(g,f)} \\ =\;& \frac{1}{n\omega_{gg}}\mathbf{v}^{(g,f)'}\mathbf{Z}^{(b)}_{\{1\}}\left(\left(\frac{1}{n}\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{Z}^{(b)}_{\{1\}}\right)^{-1} - \mathbf{J}_{01}\left(\mathbf{J}'_{01}\frac{1}{n}\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{Z}^{(b)}_{\{1\}}\mathbf{J}_{01}\right)^{-1}\mathbf{J}'_{01}\right)\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{v}^{(g,f)} \\ =\;& \frac{1}{n\omega_{gg}}\mathbf{v}^{(g,f)'}\mathbf{Z}^{(b)}_{\{1\}}\mathbf{L}'_1\left(\mathbf{I}_{K_{\{1\}}} - (\mathbf{J}'_{01}\mathbf{L}_1^{-1})'(\mathbf{J}'_{01}\mathbf{L}_1^{-1}(\mathbf{J}'_{01}\mathbf{L}_1^{-1})')^{-1}\mathbf{J}'_{01}\mathbf{L}_1^{-1}\right)\mathbf{L}_1\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{v}^{(g,f)} \\ =\;& \left(\frac{1}{\sqrt{n\omega_{gg}}}\mathbf{L}_1\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{v}^{(g,f)}\right)'\mathbf{Q}^{(b)}_{\{1\}}\left(\frac{1}{\sqrt{n\omega_{gg}}}\mathbf{L}_1\mathbf{Z}^{(b)'}_{\{1\}}\mathbf{v}^{(g,f)}\right) \;\; (say,) \;,\end{aligned}$$

where the second equality is from that $((1/n)\mathbf{Z}^{(b)\prime}_{\{1\}}\mathbf{Z}^{(b)}_{\{1\}})^{-1} = \mathbf{L}_1^{\prime}\mathbf{L}_1$. Similar to (6.89),

$$\frac{1}{\sqrt{n\omega_{gg}}}\mathbf{L}_1\mathbf{Z}^{(b)\prime}_{\{1\}}\mathbf{v}^{(g,f)} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}_{K_{\{1\}}}) .$$

Since $\mathrm{tr}(\mathbf{Q}^{(b)}_{\{1\}}) = K_{\{1\}} - K$, we obtain

$$n(\hat{\omega}_{gg} - \hat{\omega}^{\{1\}}_{gg}) \xrightarrow{d} \omega_{gg}\chi^2_{g,\, K_{\{1\}}-K} .$$

□

In the following, we present the proof of theorem. If for g-th reduced form, $\mathbf{z}^{\{1\}}_{it}$ has at least one missing variable $w^{[k]}_{it-1}$, then there exists some $\delta > 0$ by Lemma 3.2,

$$\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}}) \xrightarrow{p} \delta > 0 .$$

For any ϵ_0 and ϵ_1 such that $\delta > \epsilon_1 > \epsilon_0 > 0$, when $T \to \infty$,

$$\begin{aligned}
& \Pr(\mathrm{PIC}_1 > \mathrm{PIC}_{1,0}) \\
= & \Pr\left(\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}}) + G(K_{\{1\}} - K)\frac{\log n}{n} > 0\right) \\
\geq & \Pr\left(\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}}) \geq \epsilon_1 \,\cap\, |G(K_{\{1\}} - K)\frac{\log n}{n}| \leq \epsilon_0\right) \\
\geq & \Pr\left(\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}}) \geq \epsilon_1\right) + \Pr\left(|G(K_{\{1\}} - K)\frac{\log n}{n}| \leq \epsilon_0\right) - 1 \\
\to & \; 1 .
\end{aligned}$$

This is because that the first term converges to unity, and the second term also converges to unity by $\log n/n \to 0$.

If there exists no missing variable, then

$$K < K_{\{1\}} ,$$

i.e., $\mathbf{z}_{it}$ is included in $\mathbf{z}^{\{1\}}_{it}$. Then,

$$\begin{aligned}
\Pr(\mathrm{PIC}_1 > \mathrm{PIC}_{1,0}) &= \Pr\left(n\left(\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}})\right) + G(K_{\{1\}} - K)\log n > 0\right) \\
&\to 1 .
\end{aligned}$$

This is because that $(K_{\{1\}} - K)\log n \to +\infty$, and the following holds by Lemma 3.3,

$$\begin{aligned}
n\left(\mathrm{tr}(\hat{\mathbf{\Omega}}_{\{1\}}) - \mathrm{tr}(\hat{\mathbf{\Omega}})\right) &= -\sum_{g=1}^{G}\left(\omega_{gg}\chi^2_{g,\, K_{\{1\}}-K} + \mathcal{N}(0, \sigma^2_{g1})\right) + o_p(1) \\
&= O_p(1) .
\end{aligned}$$

$\square$

Proof of Theorem 3.5 : We first consider the case when the true $\mathbf{z}_{it}$ is not included in $\mathbf{z}_{it}^{\{1\}}$. From the proof of Theorem 3.4,

$$\begin{aligned}
\hat{\mathbf{\Omega}}_{\{1\}} - \hat{\mathbf{\Omega}} &= \frac{1}{n}\mathbf{Y}^{(f)\prime}\mathbf{Q}_1^{\prime}\mathbf{Q}_1\mathbf{Y}^{(f)} - \frac{1}{n}\mathbf{Y}^{(f)\prime}\mathbf{Q}_0^{\prime}\mathbf{Q}_0\mathbf{Y}^{(f)} \\
&= \frac{1}{n}\mathbf{\Pi}^{\prime}\mathbf{Z}^{(f)\prime}\mathbf{Q}_1^{\prime}\mathbf{Q}_1\mathbf{Z}^{(f)}\mathbf{\Pi} + o_p(1) \\
&\xrightarrow{p} \mathbf{\Pi}^{\prime}\mathbf{\Gamma}^{\{1\}}\mathbf{\Pi}\ .
\end{aligned}$$

Similar to the argument of Lemma 3.2, it is shown that the $K \times K$ matrix $\mathbf{\Gamma}^{\{1\}}$ is positive definite. From the assumption [ii] of (A5), the $G \times G$ matrix $\mathbf{\Pi}^{\prime}\mathbf{\Gamma}^{\{1\}}\mathbf{\Pi}$ is also positive definite. Therefore, from an inequality for determinant (cf. Abadir, 2005), we have that $|\hat{\mathbf{\Omega}}_{\{1\}}| > |\hat{\mathbf{\Omega}}|$ for sufficiently large T. That is,

$$\log(|\hat{\mathbf{\Omega}}_{\{1\}}|) > \log(|\hat{\mathbf{\Omega}}|)\ .$$

Next, consider the case when $\mathbf{z}_{it}$ is included in $\mathbf{z}_{it}^{\{1\}}$. For $T \to \infty$, it is sufficient to show that

$$\begin{aligned}
n(\mathrm{PIC}_2 - \mathrm{PIC}_{2,0}) &= n\left[\log(|\hat{\mathbf{\Omega}}_{\{1\}}|) - \log(|\hat{\mathbf{\Omega}}|)\right] + G(K_{\{1\}} - K)\log n \\
&\to +\infty\ .
\end{aligned}$$

For the first term, we apply the mean value theorem:

$$\begin{aligned}
n\left[\log(|\hat{\mathbf{\Omega}}_{\{1\}}|) - \log(|\hat{\mathbf{\Omega}}|)\right] &= n\sum_{g,h=1}^{G} f_{gh}(\hat{\omega}_{gh}^{\{1\}} - \hat{\omega}_{gh}) \\
&= O_p(1)\ ,
\end{aligned}$$

where f_{gh} denotes the derivative evaluated between $\hat{\mathbf{\Omega}}_{\{1\}}$ and $\hat{\mathbf{\Omega}}$, and the second equality is from that

$$n(\hat{\omega}_{gh} - \hat{\omega}_{gh}^{\{1\}}) = O_p(1)\ .$$

This is because that from the argument of Lemma 3.3 also holds for the off-diagonal elements $g \neq h$ of $\mathbf{\Omega}$. Thus, we obtain the desired result. $\square$

Proof of Lemma 3.1 : [i] For any (i,t), it is necessary that

$$\begin{aligned}
y_{it}^{(1)} &= \boldsymbol{\beta}_2^{\prime}\mathbf{y}_{it}^{(2)} + \boldsymbol{\gamma}_1^{\prime}\mathbf{z}_{it}^{(1)} + (\alpha_i + u_{it}) \\
&= \boldsymbol{\pi}_{11}^{\prime}\mathbf{z}_{it}^{(1)} + \boldsymbol{\pi}_{21}^{\prime}\mathbf{z}_{it}^{(2)} + (\pi_i^{(1)} + v_{it}^{(1)})\ ,
\end{aligned}$$

since $\boldsymbol{\gamma}_2 = \mathbf{0}$. Substituting the reduced form of $\mathbf{y}_{it}^{(2)}$,

$$(\boldsymbol{\beta}_2'\boldsymbol{\Pi}_{12}' + \boldsymbol{\gamma}_1' - \boldsymbol{\pi}_{11}')\mathbf{z}_{it}^{(1)} + (\boldsymbol{\beta}_2'\boldsymbol{\Pi}_{22}' - \boldsymbol{\pi}_{21}')\mathbf{z}_{it}^{(2)} + (\alpha_i + u_{it}) - \boldsymbol{\beta}'(\boldsymbol{\pi}_i + \mathbf{v}_{it}) = 0 \ .$$

Take the first-difference,

$$\boldsymbol{\delta}_\pi'(\mathbf{w}_{it-1} - \mathbf{w}_{it}) + (u_{it} - u_{it+1}) - \boldsymbol{\beta}'(\mathbf{v}_{it} - \mathbf{v}_{it+1}) = 0 \ , \tag{6.100}$$

where

$$\begin{aligned} \boldsymbol{\delta}_\pi' &= (\boldsymbol{\beta}_2'\boldsymbol{\Pi}_{12}' + \boldsymbol{\gamma}_1' - \boldsymbol{\pi}_{11}',\ \boldsymbol{\beta}_2'\boldsymbol{\Pi}_{22}' - \boldsymbol{\pi}_{21}') \\ &= (-\boldsymbol{\beta}'\boldsymbol{\Pi}_{1\cdot}' + \boldsymbol{\gamma}_1',\ -\boldsymbol{\beta}'\boldsymbol{\Pi}_{2\cdot}') \end{aligned}$$

is the $1 \times (K_1 + K_2)$ vector. In the case when $\mathcal{E}[\boldsymbol{\pi}_i^* \mathbf{v}_{it}^{*'}] = \mathbf{O}$. Multiplying $\mathbf{z}_{it} = \mathbf{J}'(\mathbf{w}_{it-1} + \boldsymbol{\mu}_i)$ on the right and taking the expectation,

$$\boldsymbol{\delta}_\pi'\mathcal{E}\left[(\mathbf{w}_{it-1} - \mathbf{w}_{it})\mathbf{w}_{it-1}'\right] = \mathbf{0}' \ ,$$

or

$$\boldsymbol{\delta}_\pi'\mathbf{J}'(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1)\mathbf{J} = \mathbf{0}' \ , \tag{6.101}$$

where $\boldsymbol{\Gamma}_0 = \mathcal{E}[\mathbf{w}_{it-1}\mathbf{w}_{it-1}]$ and $\boldsymbol{\Gamma}_1 = \mathcal{E}[\mathbf{w}_{it}\mathbf{w}_{it-1}] = \boldsymbol{\Pi}^{*'}\boldsymbol{\Gamma}_0$. Similarly, multiplying $\mathbf{z}_{it-1},\ \mathbf{z}_{it-2}, \cdots$ on the right of (6.100) and taking the expectations,

$$\boldsymbol{\delta}_\pi'\mathbf{J}'(\boldsymbol{\Gamma}_s - \boldsymbol{\Gamma}_{s+1})\mathbf{J} = \mathbf{0}' \ , \quad (s = 1, 2, \cdots) \ , \tag{6.102}$$

where $\boldsymbol{\Gamma}_h = (\boldsymbol{\Pi}^{*'})^h\boldsymbol{\Gamma}_0$. If we add up (6.101) and (6.102), then

$$\boldsymbol{\delta}_\pi'\mathbf{J}'\left[(\boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma}_1) + (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2) + (\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_3) + \cdots\right]\mathbf{J} = \mathbf{0}' \ .$$

From $\boldsymbol{\Gamma}_\infty = \mathbf{O}$, it follows that

$$\boldsymbol{\delta}_\pi'\mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J} = \mathbf{0}' \ . \tag{6.103}$$

Since $\mathbf{J}'\boldsymbol{\Gamma}_0\mathbf{J}$ is nonsingular by $\boldsymbol{\Gamma}_0 > \mathbf{O}$, we obtain

$$\boldsymbol{\delta}_\pi = \mathbf{0} \ .$$

In the case when $\mathcal{E}[\boldsymbol{\pi}_i^* \mathbf{v}_{it}^{*'}] \neq \mathbf{O}$, applying the forward filter,

$$\boldsymbol{\delta}_\pi'\mathbf{z}_{it}^{(f)} + u_{it}^{(f)} + \boldsymbol{\beta}'\mathbf{v}_{it}^{(f)} = 0 \ .$$

Since $\mathbf{z}_{it}^{(b)}$ does not include $\boldsymbol{\pi}_i^*$,

$$\boldsymbol{\delta}_\pi' \sum_{t=1}^{T-1} \mathcal{E}\left[\mathbf{z}_{it}^{(f)}\mathbf{z}_{it}^{(b)'}\right] = \mathbf{0} \ .$$

If divided both sides by T, then the left-hand side converges to $\mathbf{J}'\mathbf{\Gamma}_0\mathbf{J}$ as $T \to \infty$, i.e., it is the same as (6.103). Thus, the necessity is verified.

Suppose that $\mathbf{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}$. From the argument of the necessity, for any $\boldsymbol{\gamma}_2$,

$$-\boldsymbol{\beta}'\mathbf{\Pi}_{2\cdot}' + \boldsymbol{\gamma}_2' = \mathbf{0}' .$$

Then, we have that $\boldsymbol{\gamma}_2 = \mathbf{0}$ since $\mathbf{\Pi}_{2\cdot}\boldsymbol{\beta} = \mathbf{0}$. Thus, the sufficiency is verified.

[ii] When $\boldsymbol{\gamma}_2 = \mathbf{0}$,

$$\begin{aligned} \boldsymbol{\gamma}_1 &= \boldsymbol{\pi}_{11} - \mathbf{\Pi}_{12}\boldsymbol{\beta}_2 , \\ \mathbf{\Pi}_{22}\boldsymbol{\beta}_2 &= \boldsymbol{\pi}_{21} . \end{aligned}$$

Therefore, $\boldsymbol{\gamma}_1$ is uniquely determined given $\boldsymbol{\beta}_2$. If there exists $\boldsymbol{\beta}_2$, then $\text{rank}([\boldsymbol{\pi}_{21},\ \mathbf{\Pi}_{22}]) = \text{rank}(\mathbf{\Pi}_{22})$. Moreover, if $\boldsymbol{\beta}_2$ is unique, then $\text{rank}(\mathbf{\Pi}_{22}) = G_2$.

Next, consider the sufficiency. Multiplying the reduced (3.9) by $\boldsymbol{\beta}$, which satisfies the rank conditions,

$$\begin{aligned} \boldsymbol{\beta}'\mathbf{y}_{it} &= \boldsymbol{\beta}'\mathbf{\Pi}_{1\cdot}\mathbf{z}_{it-1}^{(1)} + \boldsymbol{\beta}'\mathbf{\Pi}_{2\cdot}\mathbf{z}_{it}^{(2)} + \boldsymbol{\beta}'\boldsymbol{\pi}_i + \boldsymbol{\beta}'\mathbf{v}_{it} \\ &= \boldsymbol{\gamma}_1\mathbf{z}_{it}^{(1)} + \alpha_i + u_{it} , \end{aligned}$$

where $\alpha_i = \boldsymbol{\beta}'\boldsymbol{\pi}_i$ and $u_{it} = \boldsymbol{\beta}'\mathbf{v}_{it}$. In other words, the first structural equation with an exclusion restriction is obtained by $\mathbf{\Pi}$; however, $\boldsymbol{\beta}_2$ is uniquely determined by the rank condition. For instance,

$$\boldsymbol{\beta}_2 = \left(\mathbf{\Pi}_{22}'\mathbf{\Pi}_{22}\right)^{-1}\mathbf{\Pi}_{22}'\boldsymbol{\pi}_{21} .$$

Thus, $(\boldsymbol{\beta}_2,\ \boldsymbol{\gamma}_1)'$ is a function of $\mathbf{\Pi}$. □

Proof of Theorem 3.6 : Following Cragg and Donald (1993), we consider the following constrained minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{\pi}} \quad q(\boldsymbol{\pi}) = n(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi})'\tilde{\mathbf{W}}^{-1}(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}) , \\ &\text{s.t.} \quad \text{rank}(\mathbf{\Pi}_{2\cdot}) = G_* , \end{aligned} \tag{6.104}$$

where $\boldsymbol{\pi} = \text{vec}(\mathbf{\Pi})$, $\tilde{\boldsymbol{\pi}} = \text{vec}(\tilde{\mathbf{\Pi}})$, and

$$\begin{aligned} \tilde{\mathbf{\Pi}} &= \left(\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)}\right)^{-1}\mathbf{Z}^{(b)\prime}\mathbf{Y}^{(f)} , \\ \tilde{\mathbf{W}} &= \tilde{\mathbf{\Omega}} \otimes \left(\frac{1}{n}\mathbf{Z}^{(b)\prime}\mathbf{Z}^{(b)}\right)^{-1} , \\ \tilde{\mathbf{\Omega}} &= \mathbf{H}_{n1}^{(f,b)} . \end{aligned}$$

Although $\tilde{\mathbf{\Pi}}$ is slightly different from the instrumental variable estimator of Theorem 1.5, it is also the consistent estimator. Then, the following lemma holds.

Lemma 3.4 : Let $\bar{\mathbf{\Pi}}$ be the solution of (6.104). Under the assumptions of Theorem 3.6,

$$q(\bar{\mathbf{\Pi}}) \quad \xrightarrow{d} \quad \chi^2_{(G-G_*)(K_2-G_*)} \ . \tag{6.105}$$

Provided rank$(\mathbf{\Pi}_{2\cdot}) < G_*$, there exists $\bar{\bar{\mathbf{\Pi}}}$ such that $q(\bar{\mathbf{\Pi}}) < q(\bar{\bar{\mathbf{\Pi}}})$ *a.s.*, and then,

$$q(\bar{\bar{\mathbf{\Pi}}}) \quad \xrightarrow{d} \quad \chi^2_{(G-G_*)(K_2-G_*)} \ . \tag{6.106}$$

Proof : For (6.105), it follows that

$$\sqrt{n}(\tilde{\boldsymbol{\pi}} - \boldsymbol{\pi}) \quad \xrightarrow{d} \quad \mathcal{N}(\mathbf{0},\ \text{plim}\ \tilde{\mathbf{W}}) \ .$$

Therefore, the assumptions of theorem 1 of Cragg and Donald (1993) are satisfied. For (6.106), the conditions of their Theorem 2 are also satisfied by the construction of $\tilde{\mathbf{W}}$. □

We show the case [ii] of Theorem 3.6. Then, the degree of freedom for (6.106) is changed into $(G_2 - G_{2*})(K_2 - G_{2*})$ by replacing $\mathbf{\Pi}_{2\cdot}$ with $\mathbf{\Pi}_{22}$ in the constraint (6.104). It is sufficient to show that the minimum value $q(\bar{\mathbf{\Pi}})$ is numerically equal to the sum of eigenvalues. In the case of [ii], the constraint on rank means that there exist the $L_2 = (G_2 - G_{2*})$ linearly independent vectors $(\boldsymbol{\beta}_{[1]}, \cdots, \boldsymbol{\beta}_{[L_2]}) = \mathbf{B}_*$ $(G_2 \times L_2)$ such that

$$\mathbf{\Pi}_{2\cdot}\mathbf{J}_2\mathbf{B}_* = \mathbf{O} \ , \tag{6.107}$$

where $\mathbf{J}_2^{'} = (\mathbf{0},\ \mathbf{I}_{G_2})$. Note that if $\mathbf{J}_2$ and $\mathbf{B}_*$ are replaced with $\mathbf{I}_G$ and the $G \times (G - G_*)$ matrices, respectively, then the case [i] of Theorem 3.6 can be verified.

We apply the following standardization,

$$\mathbf{B}_*^{'}\mathbf{J}_2^{'}\tilde{\mathbf{\Omega}}\mathbf{J}_2\mathbf{B}_* = \mathbf{I} \ . \tag{6.108}$$

Given $\mathbf{B}_*$, consider the minimization problem of $q(\boldsymbol{\pi})$ under the constraints of (6.107):

$$\begin{aligned} \mathbf{J}_{22}^{'}\mathbf{\Pi}\mathbf{J}_2\mathbf{B}_* = \mathbf{O} \quad &\Leftrightarrow \quad (\mathbf{B}_*^{'}\mathbf{J}_2^{'} \otimes \mathbf{J}_{22}^{'})\text{vec}(\mathbf{\Pi}) = \mathbf{0} \\ &\Leftrightarrow \quad \mathbf{R}\boldsymbol{\pi} = \mathbf{0}\ \ (say,)\ , \end{aligned}$$

where $\mathbf{J}_{22}^{'} = (\mathbf{O},\ \mathbf{I}_{K_2})$. That is,

$$\min_{\boldsymbol{\pi},\,\boldsymbol{\mu}}\ q(\boldsymbol{\pi}) + \boldsymbol{\mu}^{'}\mathbf{R}\boldsymbol{\pi}\ .$$

Solving by Lagrange's method of undetermined multipliers, we obtain

$$\bar{\boldsymbol{\pi}} = \left(\mathbf{I} + \tilde{\mathbf{W}}\mathbf{R}^{'}(\mathbf{R}\tilde{\mathbf{W}}\mathbf{R}^{'})^{-1}\mathbf{R}\right)\tilde{\boldsymbol{\pi}}\ .$$

Substituting this into $q(\boldsymbol{\pi})$,

$$\begin{aligned}
q(\bar{\mathbf{\Pi}}) &= n(\mathbf{R}\tilde{\boldsymbol{\pi}})^{'}(\mathbf{R}\tilde{\mathbf{W}}\mathbf{R}^{'})^{-1}(\mathbf{R}\tilde{\boldsymbol{\pi}}) \\
&= (\mathbf{R}\tilde{\boldsymbol{\pi}})^{'}\left((\mathbf{B}_{*}^{'}\mathbf{J}_{2}^{'}\otimes\mathbf{J}_{22}^{'})(\tilde{\mathbf{\Omega}}\otimes(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1})(\mathbf{B}_{*}^{'}\mathbf{J}_{2}^{'}\otimes\mathbf{J}_{22}^{'})^{'}\right)^{-1}(\mathbf{R}\tilde{\boldsymbol{\pi}}) \\
&= (\mathbf{R}\tilde{\boldsymbol{\pi}})^{'}\left(\mathbf{B}_{*}^{'}\mathbf{J}_{2}^{'}\tilde{\mathbf{\Omega}}\mathbf{B}_{*}\mathbf{J}_{2}\otimes\mathbf{J}_{22}^{'}(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1}\mathbf{J}_{22}\right)^{-1}(\mathbf{R}\tilde{\boldsymbol{\pi}}) \\
&= (\mathbf{R}\tilde{\boldsymbol{\pi}})^{'}\left(\mathbf{I}\otimes(\mathbf{Z}_{2}^{(b)'}\mathbf{Q}_{1}^{(b)}\mathbf{Z}_{2}^{(b)})^{-1}\right)^{-1}(\mathbf{R}\tilde{\boldsymbol{\pi}}) \\
&= \mathrm{vec}(\mathbf{J}_{22}^{'}\tilde{\mathbf{\Pi}}\mathbf{J}_{2}\mathbf{B}_{*})^{'}\left(\mathbf{I}\otimes\mathbf{Z}_{2}^{(b)'}\mathbf{Q}_{1}^{(b)}\mathbf{Z}_{2}^{(b)}\right)\mathrm{vec}(\mathbf{J}_{22}^{'}\tilde{\mathbf{\Pi}}\mathbf{J}_{2}\mathbf{B}_{*})\ ,
\end{aligned}$$

where the fourth equality is based on (6.108) and that for the $(K_1+K_2)\times(K_1+K_2)$ partitioned matrix $(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1}$,

$$\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)} = \begin{pmatrix} \mathbf{Z}_{1}^{(b)'}\mathbf{Z}_{1}^{(b)} & \mathbf{Z}_{1}^{(b)'}\mathbf{Z}_{2}^{(b)} \\ \mathbf{Z}_{2}^{(b)'}\mathbf{Z}_{1}^{(b)} & \mathbf{Z}_{2}^{(b)'}\mathbf{Z}_{2}^{(b)} \end{pmatrix}, \tag{6.109}$$

$$\mathbf{Q}_{1}^{(b)} = \mathbf{I} - \mathbf{Z}_{1}^{(b)}(\mathbf{Z}_{1}^{(b)'}\mathbf{Z}_{1}^{(b)})^{-1}\mathbf{Z}_{1}^{(b)'}\ , \tag{6.110}$$

we apply the formula (6.18) to the $K_2\times K_2$ submatrix in the lower right of (6.109).

Now, the objective function is concentrated for $\mathbf{B}_{*}$. We minimize the function with respect to $\mathbf{B}_{*}$. Put

$$\mathrm{vec}(\mathbf{J}_{22}^{'}\tilde{\mathbf{\Pi}}\mathbf{J}_{2}\mathbf{B}_{*}) = \mathrm{vec}(\tilde{\mathbf{\Pi}}_{2\cdot}\mathbf{J}_{2}\boldsymbol{\beta}_{[1]},\cdots,\ \tilde{\mathbf{\Pi}}_{2\cdot}\mathbf{J}_{2}\boldsymbol{\beta}_{[L_2]})\ ,$$

$\mathbf{F}^{(b)} = \mathbf{J}_{2}^{'}\tilde{\mathbf{\Pi}}_{2\cdot}^{'}(\mathbf{Z}_{2}^{(b)'}\mathbf{Q}_{1}^{(b)}\mathbf{Z}_{2}^{(b)})\tilde{\mathbf{\Pi}}_{2\cdot}\mathbf{J}_{2}$, and $\tilde{\mathbf{\Omega}}_{2} = \mathbf{J}_{2}^{'}\tilde{\mathbf{\Omega}}\mathbf{J}_{2}$. Then,

$$\begin{aligned}
q(\bar{\mathbf{\Pi}}) &= \boldsymbol{\beta}_{[1]}^{'}\mathbf{F}^{(b)}\boldsymbol{\beta}_{[1]} + \cdots + \boldsymbol{\beta}_{[L_2]}^{'}\mathbf{F}^{(b)}\boldsymbol{\beta}_{[L_2]} \\
&= (\tilde{\mathbf{\Omega}}_{2}^{\frac{1}{2}}\boldsymbol{\beta}_{[1]})^{'}(\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}})(\tilde{\mathbf{\Omega}}_{2}^{\frac{1}{2}}\boldsymbol{\beta}_{[1]}) + \cdots + (\tilde{\mathbf{\Omega}}_{2}^{\frac{1}{2}}\boldsymbol{\beta}_{[L_2]})^{'}(\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}})(\tilde{\mathbf{\Omega}}_{2}^{\frac{1}{2}}\boldsymbol{\beta}_{[L_2]}) \\
&\geq \bar{\mathbf{c}}_{[1]}^{'}(\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}})\bar{\mathbf{c}}_{[1]} + \cdots + \bar{\mathbf{c}}_{[L_2]}^{'}(\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_{2}^{-\frac{1}{2}})\bar{\mathbf{c}}_{[L_2]} \\
&= \bar{\boldsymbol{\beta}}_{[1]}^{'}\mathbf{F}^{(b)}\bar{\boldsymbol{\beta}}_{[1]} + \cdots + \bar{\boldsymbol{\beta}}_{[L_2]}^{'}\mathbf{F}^{(b)}\bar{\boldsymbol{\beta}}_{[L_2]} \\
&= \lambda_{21} + \cdots + \lambda_{2L_2}\ ,
\end{aligned} \tag{6.111}$$

where the third inequality is from that for the orthogonal vectors $(\tilde{\mathbf{\Omega}}_2^{\frac{1}{2}}\boldsymbol{\beta}_{[1]}, \cdots, \tilde{\mathbf{\Omega}}_2^{\frac{1}{2}}\boldsymbol{\beta}_{[L_2]})$, the sum of the quadratic forms is minimized by the eigenvectors (cf. Amemiya, 1985). In fact, for the following characteristic equation,

$$\left(\tilde{\mathbf{\Omega}}_2^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_2^{-\frac{1}{2}} - \lambda\mathbf{I}\right)\bar{\mathbf{c}} = \mathbf{0} , \tag{6.112}$$

even if the eigenvalues overlap, the orthogonal eigenvector $\bar{\mathbf{c}}$ exists since the corresponding matrix is symmetric (cf. Abadir and Magunus, 2005). If $\bar{\boldsymbol{\beta}} = \tilde{\mathbf{\Omega}}_2^{-\frac{1}{2}}\bar{\mathbf{c}}$, then the third inequality and standardization (6.108) are satisfied. The fifth equality of (6.111) is from that the minimum value is represented by the sum of the eigenvalues from the smaller of (6.112); however, these eigenvalues are equivalent to the generalized eigenvalues of $\mathbf{F}^{(b)}(= \mathbf{J}_2'\mathbf{G}_{n1}^{(f,b)}\mathbf{J}_2)$, since

$$\mid \tilde{\mathbf{\Omega}}_2^{-\frac{1}{2}'}\mathbf{F}^{(b)}\tilde{\mathbf{\Omega}}_2^{-\frac{1}{2}} - \lambda\mathbf{I} \mid = 0 \ \Leftrightarrow \ \mid \mathbf{F}^{(b)} - \lambda\tilde{\mathbf{\Omega}}_2 \mid = 0 \ .$$

Finally, we show that $\mathbf{F}^{(b)}$ and $\mathbf{J}_2'\mathbf{G}_{n1}^{(f,b)}\mathbf{J}_2$ are numerically equal. For (6.109), using the inverse of the partitioned matrix,

$$\begin{aligned} \mathbf{J}_{22}'(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1} &= \mathbf{J}_{22}'\begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{12}' & \mathbf{M}_{22} \end{pmatrix}^{-1} \\ &= \mathbf{J}_{22}'\begin{pmatrix} \mathbf{S}_{11}^{-1} & -\mathbf{S}_{11}^{-1}\mathbf{M}_{12}\mathbf{M}_{22}^{-1} \\ -\mathbf{M}_{22}^{-1}\mathbf{M}_{12}'\mathbf{S}_{11}^{-1} & \mathbf{S}_{22}^{-1} \end{pmatrix} \ (say,) \ . \end{aligned}$$

Then,

$$\begin{aligned} &\mathbf{F}^{(b)} \\ =& \left(\mathbf{J}_{22}'(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1}\mathbf{Z}^{(b)'}\mathbf{Y}^{(f)}\mathbf{J}_2\right)'\left(\mathbf{Z}_2^{(b)'}\mathbf{Q}_1^{(b)}\mathbf{Z}_2^{(b)}\right)\mathbf{J}_{22}'(\mathbf{Z}^{(b)'}\mathbf{Z}^{(b)})^{-1}\mathbf{Z}^{(b)'}\mathbf{Y}^{(f)}\mathbf{J}_2 \\ =& \mathbf{J}_2'\mathbf{Y}^{(f)'}\bar{\mathbf{S}}'\left(\mathbf{Z}_2^{(b)'}\mathbf{Q}_1^{(b)}\mathbf{Z}_2^{(b)}\right)\bar{\mathbf{S}}\mathbf{Y}^{(f)}\mathbf{J}_2 \ , \end{aligned} \tag{6.113}$$

where

$$\bar{\mathbf{S}} = \mathbf{S}_{22}^{-1}\mathbf{Z}_2^{(b)'} - \mathbf{M}_{22}^{-1}\mathbf{M}_{12}'\mathbf{S}_{11}^{-1}\mathbf{Z}_1^{(b)'} \ .$$

For the central term of (6.113), we have that

$$\begin{aligned} \bar{\mathbf{S}}'\left(\mathbf{Z}_2^{(b)'}\mathbf{Q}_1^{(b)}\mathbf{Z}_2^{(b)}\right)\bar{\mathbf{S}} &= \bar{\mathbf{S}}'\mathbf{S}_{22}\left(\mathbf{S}_{22}^{-1}\mathbf{Z}_2^{(b)'} - \mathbf{M}_{22}^{-1}\mathbf{M}_{12}'\mathbf{S}_{11}^{-1}\mathbf{Z}_1^{(b)'}\right) \\ &= \bar{\mathbf{S}}'\mathbf{S}_{22}\left(\mathbf{S}_{22}^{-1}\mathbf{Z}_2^{(b)'} - \mathbf{S}_{22}^{-1}\mathbf{M}_{12}'\mathbf{M}_{11}^{-1}\mathbf{Z}_1^{(b)'}\right) \\ &= \mathbf{Q}_1^{(b)}\left(\mathbf{Z}_2^{(b)'}\mathbf{S}_{22}^{-1}\mathbf{Z}_2^{(b)}\right)\mathbf{Q}_1^{(b)} \ , \end{aligned}$$

where the second equation is due to the fact that

$$\mathbf{M}_{22}^{-1}\mathbf{M}_{12}^{'}\mathbf{S}_{11}^{-1} = \mathbf{S}_{22}^{-1}\mathbf{M}_{12}^{'}\mathbf{M}_{11}^{-1} \ .$$

Therefore,

$$\begin{aligned}
\mathbf{F}^{(b)} &= \mathbf{J}_2^{'}\mathbf{Y}^{(f)'}\left(\mathbf{Q}_1^{(b)}\mathbf{Z}_2^{(b)'}\left(\mathbf{Z}_2^{(b)'}\mathbf{Q}_1^{(b)}\mathbf{Z}_2^{(b)}\right)^{-1}\mathbf{Z}_2^{(b)}\mathbf{Q}_1^{(b)}\right)\mathbf{Y}^{(f)}\mathbf{J}_2 \\
&= \mathbf{J}_2^{'}\mathbf{Y}^{(f)'}\left(\mathbf{P}^{(b)} - \mathbf{P}_1^{(b)}\right)\mathbf{Y}^{(f)}\mathbf{J}_2 \\
&= \mathbf{J}_2^{'}\mathbf{G}_{n1}^{(f,b)}\mathbf{J}_2 \ ,
\end{aligned}$$

where the first equality is an expression of the LIML estimator defined by Goldberger (1964), and the second equality is due to the partitioned matrix of projection matrix (cf. Amemiya, 1985). □

References

[1] Abadir, K. M. and J. R. Magnus (2005), *Matrix Algebra*, Cambridge University Press.

[2] Acemoglu, D., S. Johnson, J. Robinson and P. Yared (2008), "Income and Democracy," *American Economic Review*, Vol. 98, 808-842.

[3] Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," *Proceedings of the 2nd International Symposium on Information Theory*, edited by N. Petrov and F. Csàdki, 267-281.

[4] Akashi, K. and N. Kunitomo (2012), "Some Properties of the LIML Estimator in a Dynamic Panel Structural Equation," *Journal of Econometrics*, Vol. 166, 167-183.

[5] Akashi, K. and Kunitomo, N. (2015), "The Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equation Models," *Annals of the Institute of Statistical Mathematics*, Vol. 67, 39-73.

[6] Alonso-Borrego, C. and Arellano, M. (1999), "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business and Economic Statistics*, Vol. 17, 36-49.

[7] Alvarez, J. and M. Arellano (2003), "The Time Series and Cross Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, Vol. 71, 1121-1159.

[8] Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press.

[9] Amisano, G and C. Giannini (1996), *Topics in Structural VAR Econometrics*, Springer.

[10] Anatolyev, S. (2013), "Instrumental Variables Estimation and Inference in the Presence of Many Exogenous Regressors," *Econometrics Journal*, Vol. 16, 27-72.

[11] Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc.

[12] Anderson, T. W. (1978), "Repeated Measurements on Autoregressive Processes," *Journal of the American Statistical Association*, Vol. 73, 371-378.

[13] Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysisi*, Third Edition, John Wiley & Sons, Inc.

[14] Anderson, T. W. (2005), "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics*, Vol. 127, 1-16.

[15] Anderson, T. W. (2010), "The LIML Estimator Has Finite Moments!," *Journal of Econometrics*, Vol. 157, 359-361.

[16] Anderson, T. W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, Vol.50, 1009-1027.

[17] Anderson, T. W. and N. Kunitomo (1992), "Tests of over-identification and Predeterminedness in Simultaneous Equation Models," *Journal of Econometrics*, Vol.54, 49-78.

[18] Anderson, T. W. and N. Kunitomo (1994), "Asymptotic Robustness of Tests of over-identification and Predeterminedness," *Journal of Econometrics*, Vol.62, 383-414.

[19] Anderson, T. W., N. Kunitomo and Y. Matsushita (2010), "On the Asymptotic Optimality of the LIML Estimator with Possibly Many Instruments," *Journal of Econometrics*, Vol.157, 191-204.

[20] Anderson, T. W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, Vol. 76, 598-606.

[21] Anderson, T. W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models with Panel Data," *Journal of Econometrics*, Vol. 18, 47-82.

[22] Anderson, T. W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[23] Anderson, T. W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[24] Andrews, D. W. and B. Lu (2001), "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 101, 123-164.

[25] Angrist, J. D. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, Vol. 106, 979-1014.

[26] Arellano, M. (2003a), *Panel Data Econometrics*, Oxford University Press.

[27] Arellano, M. (2003b), "Modelling Optimal Instrumental Variables for Dynamic Panel Data Models," CEMFI Working Paper No. 0310.

[28] Arellano, M. and S. R. Bond (1991), "Some Tests of Specification for Panel Data : Monte Carlo Evidence and an Application to Employment Equation," *Review of Economics Studies*, Vol. 58, 277-297.

[29] Arellano, M. and O. Bover (1995), "Another Look at the Instrumental Variable Estimation of Error- Components Models," *Journal of Econometrics*, Vol. 68, 29-51.

[30] Barro, R. J. and X. Sala-i-Martin (1995), *Economic Growth*, McGraw Hill.

[31] Bhargava, A. (1991), "Identification and Panel Data Models with Endogenous Regressors ," *The Review of Economic Studies*, Vol. 58, 129-140.

[32] Bhargava, A. and J. D. Sargan (1983), "Estimating Dynamic Random Effects Models From Panel Data Covering Short Time Periods ," *Econometrica*, Vol. 51, 1635-1659.

[33] Binder, M., C. Hsiao and M. H. Pesaran (2005), "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration ," *Econometric Theory*, Vol. 21, 795-837.

[34] Blundell, R. and S. Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models ," *Journal of Econometrics*, Vol. 87, 115-143.

[35] Blundell, R. and S. Bond (2000), "GMM Estimation with Persistent Panel Data : An Application to Production Function," *Econometric Review*, Vol. 19, 321-340.

[36] Bun, M. J. G. and F. Windmeijer (2010), "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models," *Econometrics Journal*, Vol. 13, 95-126.

[37] Chao. J. C., N. R. Swanson, J. A. Hausman, W. K. Newey and T. Woutersen (2012), "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments," *Econometric Theory*, Vol. 28, 42-86.

[38] Cragg, J. G. and S. G. Donald (1993), "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, Vol. 9, 222-240.

[39] Davis, R. A., P. Zang and T. Zheng (2016), "Sparse Vector Autoregressive Modeling," *Journal of Computational and Graphical Statistics*, Vol. 25, 1077-1096.

[40] Doornik, J. A. (2006), *Introduction to Ox: An Object-Oriented Matrix Language.*

[41] Goldberger, A. S. (1964), *Econometric Theory*, John Wiley & Sons, Inc.

[42] Grassetti, L. (2011), "A Note on Transformed Likelihood Approach in Linear Dynamic Panel Models," *Statistical Methods and Applications*, Vol. 20, 221-240.

[43] Hahn, J. (2002), "Optimal Inference with Many Instruments," *Econometric Theory*, Vol. 18, 140-168.

[44] Hahn, J. and G. Kuersteiner (2000), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large," Massachusetts Institute of Technology Department of Economics Working Paper Series 01-17, Massachusetts Institute of Technology.

[45] Hahn, J. and G. Kuersteiner (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large," *Econometrica*, Vol. 70, 1639-1657.

[46] Han, C. and P. C. B. Phillips (2013), "First Difference Maximum Likelihood and Dynamic Panel Estimation," *Journal of Econometrics*, Vol. 175, 35-45.

[47] Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica*, Vol. 46, 1251-1271.

[48] Hayakawa, K. (2006a), "The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T are Large," Hi-Stat Discussion Paper No.129, Hitotsubashi University.

[49] Hayakawa, K. (2006b), "Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present," Hi-Stat Discussion Paper No.130, Hitotsubashi University.

[50] Hayakawa, K. (2008), "On the Effect of Nonstationary Initial Conditions in Dynamic Panel Data Models," Hi-Stat Discussion Paper No.245, Hitotsubashi University.

[51] Hayakawa, K. (2009), "A Simple Efficient Instrumental Variable Estimator in Panel AR(p) Models When Both N and T are Large," *Econometric Theory*, Vol. 25, 873-890.

[52] Hayakawa, K. (2014), "Alternative Over-identifying Restriction Test in GMM with Grouped Moment Conditions," mimeo, Hiroshima University.

[53] Hayakawa, K. and M. H. Pesaran (2015), "Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models with Cross-Sectional Heteroskedasticity," *Journal of Econometrics*, Vol. 188, 111-134.

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

[55] Holtz-Eakin, D., W. Newey and H. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, Vol. 56, 1371-1395.

[56] Hsiao, C. (1983), "Identification," in *Handbook of Econometrics*, Vol. I, edited by Z. Griliches and M. Intriligator, 223-283, North-Holland.

[57] Hsiao, C. (2014), *Analysis of Panel Data*, 3rd Edition, Cambridge University Press.

[58] Hsiao, C., M. H. Pesaran and K. Tahmiscioglu (2002), "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Series," *Journal of Econometrics*, Vol. 109, 107-150.

[59] Hsiao, C. and J. Zhang (2015), "IV, GMM or Likelihood Approach to Estimate Dynamic Panel Models When Either N or T or Both are Large," *Journal of Econometrics*, Vol. 187, 312-322.

[60] Hsiao, C. and Q. Zhou (2015), "Statistical Inference for Panel Dynamic Simultaneous Equations Models," *Journal of Econometrics*, Vol. 189, 383-396.

[61] Hsiao, C. and Q. Zhou (2016), "Asymptotic Distribution of Quasi-Maximum Likelihood Estimation of Dynamic Panels Using Long Difference Transformation When Both N and T Are Large," *Statistical Methods and Applications*, forthcoming.

[62] Huang, Y. and M. G. Quibria (2013), "The global partnership for inclusive growth," WIDER Working Paper No. 59, United Nations University.

[63] Koopmans, T. C. and W. C. Hood (1953), "The Estimation of Simultaneous Linear Economic Relationships," *Studies in Econometric Method*, edited by W. C. Hood and T. C. Koopmans, 112-199.

[64] Kunitomo, N. (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, Vol. 75, 693-700.

[65] Kunitomo, N. (2012), "An optimal modification of the LIML estimation for many instruments and persistent heteroscedasticity," *Annals of Institute of Statistical Mathematics*, Vol. 64, 881-910.

[66] Kunitomo, N. and K. Akashi (2010), "An Aysmptotically Optimal Modification of the Panel LIML Estimation for Individual Heteroscedasticity," CIRJE Discussion Paper No.780, University of Tokyo.

[67] Kruiniger, H. (2013), "Quasi ML Estimation of the Panel AR(1) Model with Arbitrary Initial Conditions," *Journal of Econometrics*, Vol. 173, 175-188.

[68] Lee, Y. and R. Okui (2012), "Hahn-Hausman test as a specification test," *Journal of Econometrics*, Vol. 167, 133-139.

[69] Lindgren, G., H. Rootzen and M. Sandsten (2013), *Stationary Stochastic Processes for Scientists and Engineers* , CRC Press.

[70] Moral-Benito, E. (2013), "Likelihood-based Estimation of Dynamic Panels with Predetermined Regressors," *Journal of Business and Economic Statistics*, Vol. 31, 451-472.

[71] Morimune, K. (1984), *Estimation and Test of Economic Model*, Kyoritsu Shuppan Co., Ltd. (in Japanese)

[72] Morimune, K. and T. Sawa (1980), "Decision Rules for the Choice of Structural Equations," *Journal of Econometrics*, Vol. 14, 329-347.

[73] Neyman, J. and E. L. Scott (1948), "Consistent Estimation from Partially Consistent Observations," *Econometrica*, Vol. 16, 1-32.

[74] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica*, Vol. 49, 1417-1426.

[75] Phillips, P. C. B. and C. Han (2014), "The True Limit Distributions of the Anderson-Hsiao IV Estimators in Panel Autoregression," Cowles Foundation Discussion Paper No. 1963, Yale Universtiy.

[76] Phillips, P. C. B. and H. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, Vol. 67, 1057-1111.

[77] Rao, C. R. (1965), *Linear Statistical Inference and its Applications*, John Wiley & Sons, Inc.

[78] Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, Vol. 6, 461-464.

[79] Sen, P. K. and J. M. Singer (1993), *Large Sample Methods in Statistics*, Chaoman & Hall.

[80] So, B. S. and D. W. Shin (1999), "Recursive Mean Adjustment in Time Series Inferences," *Statistics and Probability Letters*, Vol. 43, 65-73.

[81] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models ," *Econometrica*, Vol. 50, 1-25.

[82] White, H. (2001), *Asymptotic Theory for Econometricians*, Revised Edition, Princeton University Press.

[83] Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

[84] Zeller, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias ," *Journal of the American Statistical Association*, Vol. 57, 348-368.