真のデータに線形制約がある際の観測誤差の修正-国民経済計算の場合-（未定稿）

千木良 弘朗　　山本 拓

東北大学大学院経済学研究科　（一財）統計研究会

研究集会「経済統計・政府統計の理論と応用」

2017年2月2日 東京大学

1 はじめに

“観測誤差と線形制約を伴う真のデータ” とは?

例. 国民経済計算における供給表と使用表

供給表:

商品＼産業	製造業	サービス業	総計
製造品	900	100	1000
サービス	300	500	800
総計	1200	600	

使用表:

商品＼産業	製造業	サービス業	消費	総計
製造品	150	50	750	950
サービス	250	100	250	600
賃金	400	250		
営業余剰	350	100		
総計	1150	500		

⇒ 、 、 、 は等しくなければならない

⇒ 真のデータでは等しいが、観測誤差のために等しくなくなっている

真のデータの推定

供給表:

産業 / 商品	製造業	サービス業	総計
製造品	880	90	970
サービス	260	460	720
総計	1140	550	

使用表:

産業 / 商品	製造業	サービス業	消費	総計
製造品	160	40	770	970
サービス	270	150	300	720
賃金	410	280		
営業余剰	300	80		
総計	1140	550		

⇒ 、 、 、 を等しくするよう観測誤差を修正する

⇒ 修正法は国によって違う。アメリカとオランダは統計学の理論に基づいたストーン法（Stone，Champernowne and Meade (1942)）を採用

⇒ 本稿では、ストーン法も部分的に使うが、全く新しいアプローチを導入する

発表の構成

2. ストーン法と先行研究の概観

3. 新しいアプローチ

4. モンテカルロ実験

5. まとめ

3 ストーン法と先行研究の概観

モデル

$$\underset{(N\times 1)}{X} = X^* + u,\ u \sim (0, \Omega_u)$$

$$\underset{(R\times N)}{A'} X^* = h$$

$$\begin{cases} X\text{: 観測されたデータ} \\ X^*\text{: 真のデータ} \\ u\text{: 観測誤差} \\ A、h\text{: } X^* \text{が満たすべき線形制約} \end{cases}$$

例. 国民経済計算における供給表と使用表

供給表:

商品 ＼ 産業	製造業	サービス業	総計
製造品	x_1	x_2	
サービス	x_3	x_4	
総計			

使用表:

商品 ＼ 産業	製造業	サービス業	消費	総計
製造品	x_5	x_6	x_7	
サービス	x_8	x_9	x_{10}	
賃金	x_{11}	x_{12}		
営業余剰	x_{13}	x_{14}		
総計				

→ $A'X^* = h$ は、

$$\begin{bmatrix} 1 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1^* \\ \vdots \\ x_{14}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

ストーン法

$$\min(X - X^*)'\Omega_u^{-1}(X - X^*),\ \text{s.t.} A'X^* = h$$

$$\rightarrow \boxed{\widehat{X}^* = X - \Omega_u A(A'\Omega_u A)^{-1}(A'X - h)}$$

別の解釈:
簡単化のため$h = 0$とすると制約条件は$A'X^* = 0$で

$$\begin{aligned}\widehat{X}^* &= X - \Omega_u A(A'\Omega_u A)^{-1}A'X \\ &= (I - \tilde{A}(A'\tilde{A})^{-1}A')X \qquad (\tilde{A} = \Omega_u A) \\ &= A_\perp(\tilde{A}'_\perp A_\perp)^{-1}\tilde{A}'_\perp X \qquad (\text{“}\perp\text{”は直行補空間})\end{aligned}$$

→ X^*が$A_\perp$に属している（フルランクNではなくランク落ちして$N - R$）ため、Xも$A_\perp$に属するよう変換しているのがストーン法

Ω_uの設定について

ストーン法はΩ_uが解らないと実行不能

→ 統計理論に基づいた解決法はそう多くなく、ある程度恣意的に決められることが多い

		業種						
		農林水産業	鉱業	製造業	建　設　業	電気・ガス・水道業	卸売・小売業	金融・保険業
商品	農林水産業	1,981.2	1.5	8,412.4	250.1	1.7	1,605.1	0.0
	鉱業	0.3	8.4	8,506.8	1,037.0	2,336.4	3.5	0.0
	製造業	3,510.3	281.2	138,525.2	28,727.9	1,669.4	5,702.9	1,517.6
	建　設　業	85.5	9.7	1,410.0	208.1	1,131.2	554.8	160.1
	電気・ガス・水道業	106.8	43.3	6,522.2	500.4	1,335.1	1,128.4	220.5
	卸売・小売業	4.0	0.0	0.0	0.0	0.0	672.0	0.0
	金融・保険業	137.3	42.1	1,419.1	480.4	224.5	1,723.5	1,150.0

野木森 稔,「加重最小二乗法を利用したバランシング・モデル-SUTバランスシステム開発に向けた一考察」, 季刊国民経済計算, 147, 69–89, 2012年
より抜粋

○ 統計理論に基づいた解決法

Weale (1992) 等のアイデア:

$$X_t = X_t^* + u_t,\ u_t \sim (0, \Omega_u),\ t = 1, \cdots, T$$

$$\begin{cases} X_t、X_t^*、u_t \text{は定常} \\ Cov(X_t^*, u_t) = 0 \end{cases}$$

とすると、簡単化のため $E(X_t^*) = 0$ として

$$\begin{aligned} \boxed{\widehat{A'\Omega_u}} &= \frac{1}{T}\sum_{t=1}^{T} A'X_tX_t' \\ &= \frac{1}{T}\sum_{t=1}^{T} A'u_t(X_t^* + u_t)' \xrightarrow{p} \boxed{A'\Omega_u} \end{aligned}$$

⇒ Ω_u ではなく $A'\Omega_u$ の一致推定でストーン法を実行可能にしている

先行研究の問題点と本稿での新しいアプローチ

問題点:

$$\begin{cases} \widehat{A'\Omega_u} \text{は} N << T \text{でないとパフォーマンスが悪い} \\ X_t^* \text{に定常性を仮定するのは非現実的} \\ \text{ストーン法では} u_t \text{を一致推定できない} \\ \Omega_u \text{は一致推定できない} \end{cases}$$

⇒ 新しいアプローチ:

$$\begin{cases} \text{大きな} N \text{に対して良好なパフォーマンス} \\ X_t^* \text{に定常性を仮定しない} \\ u_t \text{（の一部）を一致推定する} \\ \Omega_u \text{を一致推定する} \end{cases}$$

3 新しいアプローチ

モデル

観測誤差にファクターモデルを適用する

$$\boldsymbol{X}_t = \boldsymbol{X}_t^* + \boldsymbol{u}_t = \boldsymbol{X}_t^* + \underset{(N\times q)}{\boldsymbol{C}}\underset{(q\times 1)}{\boldsymbol{F}_t} + \boldsymbol{\varepsilon}_t,\ t = 1, \cdots, T$$

$$\begin{cases} \boldsymbol{C}\text{はファクター負荷} \\ \boldsymbol{F}_t\text{はファクター} \\ q\text{はファクター数} \end{cases}$$

⇒ $\boldsymbol{C}\boldsymbol{F}_t$を一致推定して観測誤差を修正する。$\boldsymbol{\varepsilon}_t$についてはストーン法で修正

（簡単化のため、以下では$h = 0$とする）

モデルの推定の直感的な概観

$$\boldsymbol{X}_t = \boldsymbol{X}_t^* + \boldsymbol{C}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim (0, \Omega),\ \Omega = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix}$$

⇓

$$\boldsymbol{A}_N'\boldsymbol{X}_t = \boldsymbol{A}_N'\boldsymbol{C}\boldsymbol{F}_t + \boldsymbol{A}_N'\boldsymbol{\varepsilon}_t,\ \boldsymbol{A}_N = \boldsymbol{A}\underset{(R\times R)}{\Upsilon_N^{-1}},\ \Upsilon_N = \begin{bmatrix} n_1 & & 0 \\ & \ddots & \\ 0 & & n_R \end{bmatrix},$$

n_rは$\boldsymbol{A}$の第r列のnon-zero要素の数

⇓

この式にBai (2003)の方法を適用して$\boldsymbol{A}_N'\widehat{\boldsymbol{C}\boldsymbol{H}'^{-1}}$と$\widehat{\boldsymbol{H}'\boldsymbol{F}_t}$を得る

$\Downarrow$

$Cov(X_t^*, \varepsilon_t) = 0$、$Cov(F_t, \varepsilon_t) = 0$を仮定すると

$$\frac{1}{T}\sum_{t=1}^{T} \Upsilon_N (A'_N X_t - A'_N \widehat{CH'}^{-1} \widehat{H'F_t}) X'_t \approx \frac{1}{T}\sum_{t=1}^{T} \widehat{A'\varepsilon_t} X'_t \approx \frac{1}{T}\sum_{t=1}^{T} \widehat{A'\varepsilon_t}\varepsilon'_t \approx A'\Omega$$

$\Downarrow$

この式を積率条件としてGMMで$\hat{\sigma}_1^2, \cdots, \hat{\sigma}_N^2$を得る

$\Downarrow$

$$\frac{1}{T}\sum_{t=1}^{T} A'_N X_t X'_t \approx A'_N C \Omega_F C' + A'_N \Omega \qquad (Cov(X_t^*, F_t) = 0\text{を追加})$$
$$= A'_N C H'^{-1} H' \Omega_F H H^{-1} C' + A'_N \Omega$$

において、

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} A'_N X_t X'_t}_{\text{計算可能}} = \underbrace{A'_N C H'^{-1} H' \Omega_F H}_{A'_N\widehat{CH'}^{-1}\text{と}\widehat{H'F_t}\text{で推定可能}} \underbrace{H^{-1}C'}_{\text{未知パラメーター}} + \underbrace{A'_N \Omega}_{\hat{\sigma}_1^2,\cdots,\hat{\sigma}_N^2\text{で推定可能}}$$

$\Downarrow$

この式から$H^{-1}C'$を逆算して$\widehat{H^{-1}C'}$を得る

仮定:

1. $E(F_t) = 0$、$F_t, \varepsilon_t \sim I(0)$

2. Aの第r列と第s列で共通する項目の数c_{rs}が十分小さい
（全てのrについて$\sum_{s=1}^{R} c_{rs}$が$R \to \infty$としても発散しない）

3. $q < R$

4. Aの全行には少なくとも1つnon-zero要素がある

5. $X_t^* \sim I(0)$

6. $N, R, T \to \infty$ with $\frac{N^K}{\sqrt{T}R^{1-W}} \to 0$
($\sum_{r=1}^{R} n_R = O(N^K)$, $1 \leq K \leq 2$、
$n_1, \cdots, n_R$の内で$O(1)$なものの個数を$O(R^W)$, $0 \leq W \leq 1$)

$\widehat{CH'}^{-1}\widehat{H'F}_t \xrightarrow{p} CF_t$ （観測誤差（の一部）の一致推定）
$\hat{\sigma}_1^2, \cdots, \hat{\sigma}_N^2 \xrightarrow{p} \sigma_1^2, \cdots, \sigma_N^2$ （ストーン法の改善）

実際のprocedure

ファクター-ストーン法:

$\widetilde{X}_t^* = X_t - \widehat{CH'}^{-1}\widehat{H'F}_t$に、$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_1^2 & & 0 \\ & \ddots & \\ 0 & & \hat{\sigma}_N^2 \end{bmatrix}$を使ってストーン法を適用
$\to \widehat{X}_t^* = \widetilde{X}_t^* - \hat{\Omega}A(A'\hat{\Omega}A)^{-1}A'\widetilde{X}_t^*$

$X_t^* \sim I(0)$を緩めるには?

Nelson and Plosser (1982): GNPは$I(1)$
$\rightarrow X_t^* \sim I(1)$とする

$\Rightarrow$ $Cov(X_t^*, \varepsilon_t) = 0$・$Cov(X_t^*, F_t) = 0$を仮定しても $\frac{1}{T}\sum_{t=1}^{T} X_t^* \varepsilon_t \approx 0$・$\frac{1}{T}\sum_{t=1}^{T} X_t^* F_t \approx 0$とならず問題

$\Rightarrow$ $\Delta X_t = \Delta X_t^* + C\Delta F_t + \Delta\varepsilon_t$だと厳しい仮定が必要

Perron (1989): GNPは構造変化を伴うトレンド定常
Diebold and Rudebusch (1989): GNPは長期記憶過程

$\Rightarrow$ エルゴード性が成立する範囲までX_t^*に構造変化付きトレンドや長期記憶過程を入れて一般化する

4 モンテカルロ実験

DGPと実験の設定

$$X_t = X_t^* + u_t = X_t^* + CF_t + \varepsilon_t$$

$$\begin{cases} X_t^* \sim \text{トレンド定常VAR}(1) & \\ F_t \sim \text{期待値0の定常VAR}(1) & (C\Omega_F C'/Var(X_t^*) \approx 5\%) \\ \varepsilon_t \sim \text{期待値0の定常VAR}(1) & (\Omega/Var(X_t^*) \approx 5\%) \\ X_t^*、F_t、\varepsilon_t\text{は全て正規乱数で互いに独立} & \end{cases}$$

$\Rightarrow Var(X_t)$の内、$Var(X_t^*)$が約90%でΩ_uが約10%

$Var(X_t^*)$、$C\Omega_F C'$、Ω全てにおいて最小の分散と最大の分散には約500〜1000倍の差をつけ、不均一性を表現

- X_t^*のVAR(1)パラメーターと定数・トレンド項は乱数で生成して固定
- F_tとε_tのVAR(1)パラメーターは乱数で生成して固定
- Cは乱数で生成して固定
- Aは$-1, 0, 1, 2$しかとらないsparse行列として乱数で生成して固定
- $q = 2$で既知
- $N = 25, 100, 400$、$R = N/4, \sqrt{N}$、$T = 25, 50, 800$
- 繰り返し回数1000回

各修正法の具体的な計算

○ 原始的なストーン法（Stone1）

$\Omega_u = I_N$として$\widehat{X_t^*} = X_t - A(A'A)^{-1}A'X_t$

○ Weale (1992)の方法（Weale）

demean、detrendしたX_t^dより$\widehat{A'\Omega_u} = \frac{1}{T}\sum_{t=1}^{T} A'X_tX_t^{d\prime}$として$\widehat{X_t^*} = X_t - \widehat{\Omega_u A}(\widehat{A'\Omega_u A})^{-1}A'X_t$

○ ファクター-ストーン法（F-Stone）

先述の通りだが、

$$\begin{cases} \frac{1}{T}\sum_{t=1}^{T} \Upsilon_N(A_N'X_t - A_N'\widehat{CH'^{-1}}\widehat{H'F_t})X_t^{d\prime} \approx A'\Omega \\ \frac{1}{T}\sum_{t=1}^{T} A_N'X_tX_t^{d\prime} \approx A_N'CH'^{-1}H'\Omega_F HH^{-1}C' + A_N'\Omega \end{cases}$$

のようにX_t^dで計算する。また、$\hat{\sigma}_i^2 \leq 0$だった時は$\hat{\sigma}_i^2 = 0.01 \times \frac{1}{T}\sum_{t=1}^{T}(x_{it}^d)^2$とする

推定精度の尺度

MSE$= E\{(X_t^* - \widehat{X_t^*})'(X_t^* - \widehat{X_t^*})\}$だと、

$\begin{cases} \text{変動の大きい} x_{it}^* - \hat{x}_{it}^* \text{だけにdominateされてしまい不公平} \\ N \text{や} R \text{を変えるとDGPのパラメーターの値の変化を反映して} \\ \text{しまい、純粋なsample sizeの効果を捉えられない} \end{cases}$

⇒ 基準化MSE（sMSE）：

$$\text{sMSE}= \frac{1}{N}E\left\{(X_t^* - \widehat{X_t^*})' \begin{bmatrix} Var(x_1^*) & & 0 \\ & \cdots & \\ 0 & & Var(x_N^*) \end{bmatrix}^{-1} (X_t^* - \widehat{X_t^*})\right\}$$

$$= \frac{1}{N}E\left\{\sum_{i=1}^{N} \frac{(x_{it}^* - \hat{x}_{it}^*)^2}{Var(x_i^*)}\right\}$$

⇒ $\begin{cases} x_{it}^* = \mu_i + \delta_i t + z_{it},\ E(z_{it}) = 0,\ z_{it} \sim I(0) \\ \hat{x}_{it}^* = \mu_i + \delta_i t \quad \text{（確定項だけ捉えている“最低限”の推定量）} \end{cases}$

とすると、$E\left\{\frac{(x_{it}^* - \hat{x}_{it}^*)^2}{Var(x_i^*)}\right\} = E\left\{\frac{z_{it}^2}{Var(x_i^*)}\right\} = 1$

⇒ $\begin{cases} \text{sMSE} < 1 \cdots \text{最低限の規準はクリア} \\ \text{sMSE} \geq 1 \cdots \text{推定精度に問題あり} \end{cases}$

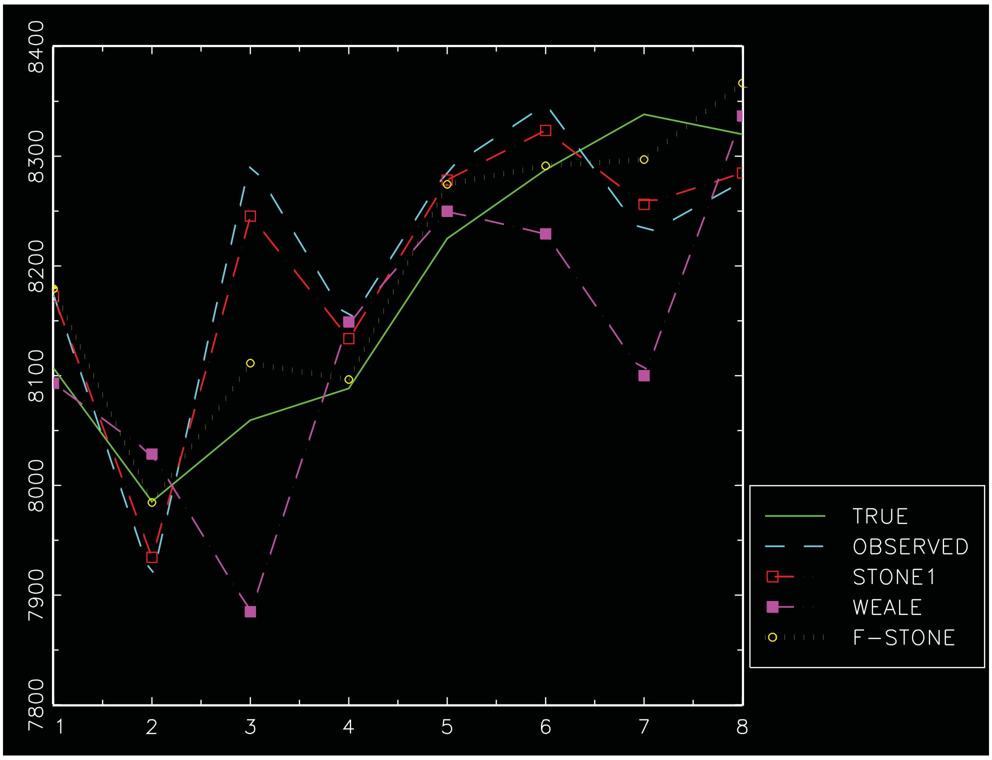
（実際はsMSEが$t = 1, \cdots, T$について得られるので$\frac{1}{T}\sum_{t=1}^{T} \text{sMSE}_t$を計算）

sMSE

		Stone1			Weale			F-Stone		
		T			T			T		
N	R	25	50	800	25	50	800	25	50	800
25	5	0.137	0.136	0.136	0.294	0.178	0.062	0.144	0.109	0.064
	7	0.152	0.152	0.151	0.398	0.222	0.060	0.143	0.106	0.057
100	10	0.138	0.139	0.139	0.599	0.307	0.068	0.144	0.110	0.059
	25	0.192	0.192	0.193	.	0.632	0.081	0.147	0.106	0.052
400	20	0.116	0.115	0.115	2.005	0.524	0.081	0.135	0.105	0.053
	100	0.137	0.138	0.138	.	.	0.195	0.139	0.104	0.045

実験データの時系列プロット

$N = 100$、$R = 25$、$T = 50$



5 まとめ

本稿の貢献

観測誤差を伴う真のデータの推定について、

- 観測誤差（の一部）の一致推定による直接的な推定
- （それ以外の観測誤差に関する）修正精度の改善

を提案した

推定精度

sMSE という尺度では、概ね **F-Stone>Stone1>Weale**

今後の課題

$X_t^* \sim I(0)$ という仮定をどこまで緩められるか?

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

Kazuhiko Hayakawa
Department of Economics,
Hiroshima University

Meng Qi
Department of Economics,
Hiroshima University

Jörg Breitung
Center of Econometrics and Statistics
University of Cologne

January 16, 2017

Abstract

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

*The authors are grateful to Graham Elliott, Jan Kiviet, Koichi Maekawa, Peter Phillips, Hiroshi Yamada, Jun Yu and the participants of SETA 2015 at Hitotsubashi University, CFE 2015 at University of London, ESEM at Geneva, 2016 Japan-Korea Allied Conference in Econometrics at Hitotsubashi University and seminars at Singapore Management University and Nanyang Technological University for helpful comments. This research is supported by KAKENHI(25780153, 15H01943, 16H03606).

1 Introduction

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

More recently, an alternative IV estimator has been proposed in the literature, where a forward demeaning (detrending) is applied to the model while backward demeaning (detrending) is applied to the instruments. We call that IV estimator the double filter IV (DFIV) estimator since unobserved heterogeneity is filtered out forward and backward. Perhaps, the first paper that considers the DFIV estimator is Moon and Phillips (2000) where a near integrated autoregressive panel data model is studied. However, they did not provide theoretical discussion on the DFIV estimator. Hayakawa (2009) considers the DFIV estimator in a stationary panel autoregressive model and derives the asymptotic properties when both N and T are large. A novel feature of the DFIV estimator is that it has the same asymptotic distribution as the bias-corrected FE estimator when T and N are large. Since the DFIV estimator simply uses variables derived from past means as instruments, as opposed to the commonly used level variables, it is quite easy to use in practice[1]. From the theoretical point of view, the DFIV estimator has addressed the trade-off problem of using many instruments. Although many instruments are required to improve efficiency, the DFIV estimator becomes efficient despite the same number of instruments as the parameters is used. Hence, the DFIV estimator becomes efficient with the minimal number of instruments. This property has the advantage that it does not cause a large finite sample bias induced by using many instruments. Thereby, the trade-off between bias and efficiency of the generalized method of moments (GMM) estimator is addressed: both the bias and variance of the DFIV estimator become small simultaneously.

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

The DFIV estimator is also extended to a panel VAR model(Hayakawa, 2016), dynamic panel simultaneous equation model (Akashi and Kunitomo, 2015) and an infinite order panel autoregressive model (Lee, Okui and Shintani, 2016). However, while there are several nice features as above, unfortunately, the asymptotic equivalence between the DFIV and bias-corrected FE estimators are only proved in the context of (V)AR models, which are somewhat restrictive in practice. One of the purposes of this paper is to demonstrate that this equivalence result holds for more general case with additional regressors. Specifically, we demonstrate that the asymptotic distributions of the DFIV and bias-corrected fixed effect estimator with large N and T are identical for linear panel data models including dynamic models as well as static panel data models with weakly exogenous regressors. Moreover, we demonstrate that this equivalence result holds even when the errors are heteroskedastic and heterogeneous time trends are included in the model, which are not allowed in Hayakawa (2009, 2016). We also investigate the efficiency property of the DFIV and related GMM estimators when T is small and N is large. We conduct Monte Carlo simulation to investigate the finite sample behavior of estimators. Consequently, we find that the DFIV/DFGMM estimators tend to outperform the fixed effects estimator and IV/GMM estimators using instruments in levels.

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

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

2 Model and estimators

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

2.1 Fixed effects model

Consider a panel data model with fixed effects, given by

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

where $\mathbf{w}_{it}$ and $\boldsymbol{\delta}$ are $k \times 1$ vectors. We assume that the error term v_{it} is serially and cross-sectionally uncorrelated. The fixed effect η_i may be correlated with the regressor $\mathbf{w}_{it}$. Also, we assume that the regressor is weakly exogenous in the sense that $E(\mathbf{w}_{it}v_{is}) = \mathbf{0}$ for $t \le s$ and $E(\mathbf{w}_{it}v_{is}) \neq \mathbf{0}$ for $t > s$.

This model includes several models as special cases:

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

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

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

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

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

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

In a matrix form, the model (1) can be written as

$$\mathbf{y}_i = \mathbf{W}_i \boldsymbol{\delta} + \eta_i \boldsymbol{\iota}_T + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{2}$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{iT})'$, $\mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT})'$, $\boldsymbol{\iota}_T = (1, ..., 1)'$ and $\mathbf{v}_i = (v_{i1}, ..., v_{iT})'$. Define the following matrix that can be used to remove fixed effects:

$$\underset{(T_1 \times T)}{\mathbf{F}^{\iota}_T} = \text{diag}\left(c^{\iota}_1, c^{\iota}_2, ..., c^{\iota}_{T_1}\right) \left[\begin{array}{c|ccccc|c} 1 & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} \\ \hline 0 & 1 & \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} \\ 0 & 0 & 1 & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & \frac{-1}{T-3} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & 0 & \cdots & \cdots & 0 & 1 & -1 \end{array}\right] = \begin{bmatrix} F^{\iota}_{11} & \mathbf{F}^{\iota}_{12} & F^{\iota}_{13} \\ \mathbf{0}_{T_2 \times 1} & \mathbf{F}^{\iota}_{22} & \mathbf{F}^{\iota}_{23} \end{bmatrix}$$

$$= \{f^{\iota}_{st}\} = \begin{cases} 0 & \text{if } s > t \\ c^{\iota}_t = 1 + O\left(\frac{1}{T-t}\right) & \text{if } s = t \\ \frac{-c^{\iota}_s}{T-s} = O\left(\frac{1}{T-s}\right) & \text{if } s < t \end{cases} \tag{3}$$

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

Multiplying (2) by $\mathbf{F}^{\iota}_T$, the model to be estimated becomes

$$\dot{\mathbf{y}}^{\iota}_i = \dot{\mathbf{W}}^{\iota}_i \boldsymbol{\delta} + \dot{\mathbf{v}}^{\iota}_i, \qquad (i = 1, ..., N) \tag{4}$$

where $\dot{\mathbf{y}}^{\iota}_i = \mathbf{F}^{\iota}_T \mathbf{y}_i = \left(\dot{y}^{\iota}_{i1}, ..., \dot{y}^{\iota}_{iT_1}\right)'$, $\dot{\mathbf{W}}^{\iota}_i = \mathbf{F}^{\iota}_T \mathbf{W}_i = \left(\dot{\mathbf{w}}^{\iota}_{i1}, ..., \dot{\mathbf{w}}^{\iota}_{iT_1}\right)'$ and $\dot{\mathbf{v}}^{\iota}_i = \mathbf{F}^{\iota}_T \mathbf{v}_i = \left(\dot{v}^{\iota}_{i1}, ..., \dot{v}^{\iota}_{iT_1}\right)'$ with $\dot{y}^{\iota}_{it} = c^{\iota}_t \left[y_{it} - (y_{i,t+1} + \cdots + y_{iT})/(T-t)\right]$, $\dot{\mathbf{w}}^{\iota}_{it} = c^{\iota}_t \left[\mathbf{w}_{it} - (\mathbf{w}_{i,t+1} + \cdots + \mathbf{w}_{iT})/(T-t)\right]$ and $\dot{v}^{\iota}_{it} = c^{\iota}_t \left[v_{it} - (v_{i,t+1} + \cdots + v_{iT})/(T-t)\right]$ for $t = 1, ..., T_1$. Note that the fixed effects η_i is removed by taking a deviation from future means. The tth row of (4) can be written as

$$\dot{y}^{\iota}_{it} = \dot{\mathbf{w}}^{\iota\prime}_{it} \boldsymbol{\delta} + \dot{v}^{\iota}_{it}, \qquad (t = 1, ..., T_1;\ i = 1, ..., N) \tag{5}$$

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

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

$$\underset{(T_1 \times T)}{\mathbf{B}^{\iota}_T} = \text{diag}\left(c^{\iota}_{T_1}, \cdots, c^{\iota}_2, c^{\iota}_1\right) \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ \frac{-1}{2} & \frac{-1}{2} & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ \frac{-1}{T-3} & \frac{-1}{T-3} & \cdots & \frac{-1}{T-3} & 1 & 0 & 0 \\ \frac{-1}{T-2} & \frac{-1}{T-2} & \cdots & \frac{-1}{T-2} & \frac{-1}{T-2} & 1 & 0 \\ \frac{-1}{T-1} & \frac{-1}{T-1} & \cdots & \frac{-1}{T-1} & \frac{-1}{T-1} & \frac{-1}{T-1} & 1 \end{bmatrix} = \left\{b^{\iota}_{st}\right\}. \tag{6}$$

$\mathbf{B}^{\iota}_T$ is obtained by rotating $\mathbf{F}^{\iota}_T$. The mathematical relationship between $\mathbf{F}^{\iota}_T$ and $\mathbf{B}^{\iota}_T$ is given in (70) in the appendix. Using this, we define an instrumental variable $\ddot{\mathbf{W}}^{\iota}_i = \mathbf{B}^{\iota}_T \mathbf{W}_i =$

$(\ddot{\mathbf{w}}_{i2}^{\iota}, ..., \ddot{\mathbf{w}}_{iT}^{\iota})'$ where[2]

$$\ddot{\mathbf{w}}_{it}^{\iota} = c_{T-t+1}^{\iota}\left[\mathbf{w}_{it} - \frac{\mathbf{w}_{i,t-1} + \cdots + \mathbf{w}_{i1}}{t-1}\right], \qquad (i = 1, ..., N;\ t = 2, ..., T). \tag{7}$$

Note that the first period is lost due to the difference property of the transformation matrix (6). The transformation that induces $\ddot{\mathbf{w}}_{it}^{\iota}$ is called the backward orthogonal deviation (BOD) transformation as opposed to FOD transformation.

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

$$\widehat{\boldsymbol{\delta}}_{IV}^{\iota} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T_1} \ddot{\mathbf{w}}_{it}^{\iota}\dot{\mathbf{w}}_{it}^{\iota\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T_1} \ddot{\mathbf{w}}_{it}^{\iota}\dot{y}_{it}^{\iota}\right). \tag{8}$$

We call $\widehat{\boldsymbol{\delta}}_{IV}^{\iota}$ the *double filter instrumental variable* (DFIV) estimator since it is based on forward and backward filtering.

How are the moment conditions derived?

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

$$(y_{it} - \bar{y}_i) = (\mathbf{w}_{it} - \bar{\mathbf{w}}_i)'\boldsymbol{\delta} + (v_{it} - \bar{v}_i), \qquad (i = 1, ..., N;\ t = 1, ..., T) \tag{9}$$

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

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

Hence, the covariance between the regressors and error term in (9) becomes

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

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

[2] Compared with the form given in Hayakawa (2009), the coefficient c_{T-t+1} is slightly different. However, this is inconsequential and does not affect the main result of this paper. Also, although it is possible to remove c_{T-t+1}^{ι} from $\ddot{\mathbf{w}}_{it}^{\iota}$, we keep it to simplify the mathematical relashionship between $\mathbf{F}_{T}^{\iota}$ and $\mathbf{B}_{T}^{\iota}$(see (70)).

2.2 Trend model

Next, we consider a panel data model with usual fixed effects and heterogeneous time trends, given by[3]

$$y_{it} = \mathbf{w}_{it}'\boldsymbol{\delta} + \eta_i + \lambda_i t + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T). \tag{10}$$

In this model, since both η_i and λ_i can be correlated with $\mathbf{w}_{it}$, the FE estimator augmented with time trend is inconsistent when T is small. In a matrix form, this model can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \eta_i\boldsymbol{\iota}_T + \lambda_i\boldsymbol{\tau}_T + \mathbf{v}_i, \qquad (i = 1, ..., N) \tag{11}$$

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

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

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

Multiplying (12) to (11), we have the following transformed model

$$\dot{\mathbf{y}}_i^{\tau} = \dot{\mathbf{W}}_i^{\tau}\boldsymbol{\delta} + \dot{\mathbf{v}}_i^{\tau}, \qquad (i = 1, ..., N) \tag{14}$$

where $\dot{\mathbf{y}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{y}_i = (\dot{y}_{i1}^{\tau}, ..., \dot{y}_{iT_2}^{\tau})'$, $\dot{\mathbf{W}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{W}_i = (\dot{\mathbf{w}}_{i1}^{\tau}, ..., \dot{\mathbf{w}}_{iT_2}^{\tau})'$, and $\dot{\mathbf{v}}_i^{\tau} = \mathbf{F}_T^{\tau}\mathbf{v}_i = (\dot{v}_{i1}^{\tau}, ..., \dot{v}_{iT_2}^{\tau})'$. The tth row of (14) can be written as

$$\dot{y}_{it}^{\tau} = \dot{\mathbf{w}}_{it}^{\tau\prime}\boldsymbol{\delta} + \dot{v}_{it}^{\tau}, \qquad (t = 1, ..., T_2; i = 1, ..., N) \tag{15}$$

where

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

and f_{st}^{τ} is defined in (13).

[3]Panel data models with heterogeneous time trends are studied by, say, Wansbeek and Knaap (1999), and Phillips and Sul (2007) etc..

Next, to introduce an instrumental variables, we define

$$\begin{aligned}\underset{(T_2\times T)}{\mathbf{B}_T^{\tau}} &= \mathbf{B}_T^{\tau 1}\begin{bmatrix} \frac{2(-2+3)}{2\cdot 1} & \frac{2(-2)}{2\cdot 1} & 1 & 0 & 0 & \cdots & \cdots & 0 \\ \frac{2(-4+6)}{3\cdot 2} & \frac{2(-4+3)}{3\cdot 2} & \frac{2(-4)}{3\cdot 2} & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \frac{2(-2T_4+3T_4)}{T_3T_4} & \frac{2(-2T_4+3T_5)}{T_3T_4} & \cdots & \cdots & \frac{2(-2T_4)}{T_3T_4} & 1 & \ddots & \vdots \\ \frac{2(-2T_3+3T_3)}{T_2T_3} & \frac{2(-2T_3+3T_4)}{T_2T_3} & \cdots & \cdots & \frac{2(-2T_3+3)}{T_2T_3} & \frac{2(-2T_3)}{T_2T_3} & 1 & 0 \\ \frac{2(-2T_2+3T_2)}{T_1T_2} & \frac{2(-2T_2+3T_3)}{T_1T_2} & \cdots & \cdots & \frac{2(-2T_2+6)}{T_1T_2} & \frac{2(-2T_2+3)}{T_1T_2} & \frac{2(-2T_2)}{T_1T_2} & 1 \end{bmatrix} \\ &= \{b_{st}^{\tau}\} \end{aligned} \tag{16}$$

where $\mathbf{B}_T^{\tau 1} = \text{diag}\left(c_{T_2}^{\tau}, \cdots c_2^{\tau}, c_1^{\tau}\right)$. Note that $\mathbf{B}_T^{\tau}$ can be obtained by rotating $\mathbf{F}_T^{\tau}$ (see (70) in the appendix). Using this, we define an instrumental variable $\ddot{\mathbf{W}}_i^{\tau} = \mathbf{B}_T^{\tau}\mathbf{W}_i = (\mathbf{z}_{i3}^{\tau}, ..., \mathbf{z}_{iT}^{\tau})'$ where its tth row is given by

$$\ddot{\mathbf{w}}_{it}^{\tau} = b_{t-2,t}^{\tau}\mathbf{w}_{it} + b_{t-2,t-1}^{\tau}\mathbf{w}_{i,t-1} + \cdots + b_{t-2,1}^{\tau}\mathbf{w}_{i1}, \qquad (i = 1, ..., N; t = 3, ..., T) \tag{17}$$

with b_{st}^{τ} being defined in (16). Note that the first two periods are lost due to the difference property of the transformation matrix $\mathbf{B}_T^{\tau}$.

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

$$\widehat{\boldsymbol{\delta}}_{IV}^{\tau} = \left(\sum_{i=1}^{N}\sum_{t=3}^{T_2}\ddot{\mathbf{w}}_{it}^{\tau}\dot{\mathbf{w}}_{it}^{\tau\prime}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=3}^{T_2}\ddot{\mathbf{w}}_{it}^{\tau}\dot{y}_{it}^{\tau}\right). \tag{18}$$

This is the DFIV estimator for the model with fixed effects and time trend given by (10).

2.3 Unified model

To derive the asymptotic properties of the proposed IV estimators (8) and (18) in the next section, we formulate the above two models (1) and (11) in a unified framework. For this, let us define a variable d such that $d = 1$ corresponds to the FE model while $d = 2$ corresponds to the trend model. Also, let us define $\mathbf{C}_T$ and $\mathbf{F}_T$ such that $\mathbf{C}_T = \boldsymbol{\iota}_T$ and $\mathbf{F}_T = \mathbf{F}_T^{\iota}$ for the FE model, and $\mathbf{C}_T = (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T)$ and $\mathbf{F}_T = \mathbf{F}_T^{\tau}$ for the trend model. Thereby, the case $(d, T_d, \mathbf{F}_T, \mathbf{C}_T) = (1, T_1, \mathbf{F}_T^{\iota}, \boldsymbol{\iota}_T)$ corresponds to the FE model while $(d, T_d, \mathbf{F}_T, \mathbf{C}_T) = (2, T_2, \mathbf{F}_T^{\tau}, (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T))$ corresponds to the trend model. Note that $\mathbf{F}_T$ has the properties such that $\mathbf{F}_T\mathbf{C}_T = \mathbf{0}$, $\mathbf{F}_T\mathbf{F}_T' = \mathbf{I}_{T_d}$ and

$$\mathbf{F}_T'\mathbf{F}_T = \mathbf{Q}_T = \mathbf{I}_T - \mathbf{C}_T\left(\mathbf{C}_T'\mathbf{C}_T\right)^{-1}\mathbf{C}_T' = \mathbf{I}_T - \mathbf{R}_T, \tag{19}$$

$$\mathbf{R}_T = \begin{cases} \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' & \text{FE model} \\ \frac{2(2T+1)}{T(T-1)}\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \frac{12}{T(T-1)(T+1)}\boldsymbol{\tau}_T\boldsymbol{\tau}_T' - \frac{6}{T(T-1)}\left(\boldsymbol{\iota}_T\boldsymbol{\tau}_T' + \boldsymbol{\tau}_T\boldsymbol{\iota}_T'\right) & \text{trend model} \end{cases}. \tag{20}$$

Using these, the models (1) and (11) can be written as

$$\mathbf{y}_i = \mathbf{W}_i\boldsymbol{\delta} + \mathbf{C}_T\boldsymbol{\eta}_i + \mathbf{v}_i, \qquad (i = 1, ..., N). \tag{21}$$

Multiplying (21) by $\mathbf{F}_T$ we obtain the following transformed model:

$$\mathbf{y}_i^* = \mathbf{W}_i^*\boldsymbol{\delta} + \mathbf{v}_i^* \tag{22}$$

where $\mathbf{y}_i^* = \mathbf{F}_T\mathbf{y}_i = (y_{i1}^*, ..., y_{iT_d}^*)'$. $\mathbf{W}_i^* = \mathbf{F}_T\mathbf{W}_i = (\mathbf{w}_{i1}^*, ..., \mathbf{w}_{iT_d}^*)'$ and $\mathbf{v}_i^* = \mathbf{F}_T\mathbf{v}_i = (v_{i1}^*, ..., v_{iT_d}^*)'$. The tth row of (22) can be written as

$$y_{it}^* = \mathbf{w}_{it}^{*\prime}\boldsymbol{\delta} + v_{it}^*, \qquad (i = 1, ..., N; t = 1, ..., T_d) \tag{23}$$

Note that the models (5) and (15) are the special cases of (23).

Similarly, let $\mathbf{w}_{it}^{**}$, $(i = 1, ..., N;\ t = d+1, ..., T)$ denote $\ddot{\mathbf{w}}_{it}^{\iota}$ for FE model given by (7), and $\ddot{\mathbf{w}}_{it}^{\tau}$ for the trend model given by (17), respectively. Since $E\left(\sum_{t=d+1}^{T_d} \mathbf{w}_{it}^{**} v_{it}^*\right) = \mathbf{0}$, we have the following instrumental variable estimator

$$\widehat{\boldsymbol{\delta}}_{IV}^B = \left(\sum_{i=1}^N \sum_{t=d+1}^{T_d} \mathbf{w}_{it}^{**}\mathbf{w}_{it}^{*\prime}\right)^{-1} \left(\sum_{i=1}^N \sum_{t=d+1}^{T_d} \mathbf{w}_{it}^{**} y_{it}^*\right). \tag{24}$$

This is the DFIV estimator for the unified model given by (23). Note that the previous two estimators (8) and (18) are the special case of (24).

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

$$\check{\mathbf{y}}_i^* = \check{\mathbf{W}}_i^{*\prime}\boldsymbol{\delta} + \check{\mathbf{v}}_i^*, \qquad (i = 1, ..., N) \tag{25}$$

where $\check{\mathbf{y}}_i^* = \left(y_{i,d+1}^*, ..., y_{iT_d}^*\right)'$, $\check{\mathbf{W}}_i^* = \left(\mathbf{w}_{i,d+1}^*, ..., \mathbf{w}_{iT_d}^*\right)'$ and $\check{\mathbf{v}}_i^* = \left(v_{i,d+1}^*, ..., v_{iT_d}^*\right)'$. For this model, we consider the moment conditions given by $E(\mathbf{Z}_i^{B\prime}\check{\mathbf{v}}_i^*) = \mathbf{0}$ where $\mathbf{Z}_i^B = \text{diag}(\mathbf{z}_{i,d_0+1}^{B\prime}, ..., \mathbf{z}_{iT_d}^{B\prime})$, $\mathbf{z}_{it}^B = (\mathbf{w}_{i,t-\ell+1}^{**\prime}, ..., \mathbf{w}_{it}^{**\prime})'$, $(1 \le \ell \le t-d)$ and $\mathbf{v}_i^* = (v_{i,d_0+1}^*, ..., v_{iT_d}^*)'$. The corresponding one-step GMM estimator is given by[4]

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}_{GMM}^B &= \left[\left(\sum_{i=1}^N \check{\mathbf{W}}_i^{*\prime}\mathbf{Z}_i^B\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\mathbf{Z}_i^B\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\check{\mathbf{W}}_i^*\right)\right]^{-1} \\
&\quad \times \left[\left(\sum_{i=1}^N \check{\mathbf{W}}_i^{*\prime}\mathbf{Z}_i^B\right)\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\mathbf{Z}_i^B\right)^{-1}\left(\sum_{i=1}^N \mathbf{Z}_i^{B\prime}\check{\mathbf{y}}_i^*\right)\right] \\
&= \left[\sum_{t=d_0+1}^{T_d} \underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^B\left(\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{Z}}_t^B\right)^{-1}\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{W}}_t^*\right]^{-1}\left[\sum_{t=d_0+1}^{T_d} \underline{\mathbf{W}}_t^{*\prime}\underline{\mathbf{Z}}_t^B\left(\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{Z}}_t^B\right)^{-1}\underline{\mathbf{Z}}_t^{B\prime}\underline{\mathbf{y}}_t^*\right]
\end{aligned} \tag{26}$$

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

In order to compare the IV estimator with the FE estimator in the next section, we further reformulate (25) in terms of $\mathbf{y}_i$, $\mathbf{W}_i$ and $\mathbf{v}_i$. For this, let us define $\mathbf{L}_T = (\mathbf{0}_{T_d \times d}, \mathbf{I}_{T_d})$. Then, by noting that $\mathbf{L}_T\mathbf{y}_i = (y_{i,d+1}, ..., y_{iT})'$ and $\mathbf{L}_T\mathbf{W}_i = (\mathbf{w}_{i,d+1}, ..., \mathbf{w}_{iT})'$, the model (25) can be written as

$$\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{y}_i = \mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i\boldsymbol{\delta} + \mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i, \qquad (i = 1, ..., N). \tag{27}$$

[4]The two-step GMM estimator is not considered in this paper since it requires a computation of large dimensional weighting matrix. Indeed, if the number of moment conditions $m = \sum_{t=d_0+1}^{T_d} m_t$ exceeds the sample size where m_t denotes the number of instruments used at period t, the optimal weighting matrix cannot be computed. However, the one-step GMM estimator (26) can be computed as long as $m_t < N$ for all t even when $m > N$.

Similarly, by using $\mathbf{K}_T = (\mathbf{I}_{T_d}, \mathbf{0}_{T_d \times d})$ and $\mathbf{K}_T \mathbf{W}_i = (\mathbf{w}_{i1}, ..., \mathbf{w}_{iT_d})'$, we have $\mathbf{W}_i^{**} = (\mathbf{w}_{i,d+1}^{**}, ..., \mathbf{w}_{iT_d}^{**})' = \mathbf{B}_{T_d} \mathbf{K}_T \mathbf{W}_i$ where $\mathbf{B}_T$ denotes $\mathbf{B}_T^{\iota}$ for the FE model and $\mathbf{B}_T^{\tau}$ for trend model. Using these, the moment conditions $E\left(\sum_{t=d+1}^{T_d} \mathbf{w}_{it}^{**} v_{it}^*\right) = \mathbf{0}$ can be written as $E(\mathbf{W}_i' \mathbf{K}_T' \mathbf{B}_{T_d}' \mathbf{F}_{T_d} \mathbf{L}_T \mathbf{v}_i) = \mathbf{0}$, and the IV estimator (24) can be written as

$$\widehat{\boldsymbol{\delta}}_{IV}^B = \left(\sum_{i=1}^N \mathbf{W}_i^{**\prime} \mathbf{W}_i^*\right)^{-1} \left(\sum_{i=1}^N \mathbf{W}_i^{**\prime} \mathbf{y}_i^*\right) = \left[\sum_{i=1}^N \mathbf{W}_i' \mathbf{K}_T' \mathbf{B}_{T_d}' \mathbf{F}_{T_d} \mathbf{L}_T \mathbf{W}_i\right]^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{K}_T' \mathbf{B}_{T_d}' \mathbf{F}_{T_d} \mathbf{L}_T \mathbf{y}_i. \quad (28)$$

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

$$\widehat{\boldsymbol{\delta}}_{FE} = \left[\sum_{i=1}^N \mathbf{W}_i' \mathbf{Q}_T \mathbf{W}_i\right]^{-1} \sum_{i=1}^N \mathbf{W}_i' \mathbf{Q}_T \mathbf{y}_i \quad (29)$$

where $\mathbf{Q}_T$ is defined in (19).

Also, for later use, we define IV and GMM estimators using instrument in levels. The IV estimator for model (23) using instruments $\mathbf{w}_{it}$ is given by

$$\widehat{\boldsymbol{\delta}}_{IV}^L = \left(\sum_{i=1}^N \sum_{t=1}^{T_d} \mathbf{w}_{it} \mathbf{w}_{it}^{*\prime}\right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^{T_d} \mathbf{w}_{it} y_{it}^*\right). \quad (30)$$

The GMM estimator based on the moment condition $E(\mathbf{Z}_i^{L\prime} \mathbf{v}_i^*) = \mathbf{0}$ where $\mathbf{Z}_i^L = \text{diag}(\mathbf{z}_{i1}^{L\prime}, ..., \mathbf{z}_{iT_d}^{L\prime})$, $\mathbf{z}_{it}^L = (\mathbf{w}_{i,t-\ell+1}', ..., \mathbf{w}_{it}')'$, $(1 \le \ell \le t)$ is given by

$$\widehat{\boldsymbol{\delta}}_{GMM}^L = \left[\sum_{t=1}^{T_d} \underline{\mathbf{W}}_t^{*\prime} \underline{\mathbf{Z}}_t^L \left(\underline{\mathbf{Z}}_t^{L\prime} \underline{\mathbf{Z}}_t^L\right)^{-1} \underline{\mathbf{Z}}_t^{L\prime} \underline{\mathbf{W}}_t^*\right]^{-1} \left[\sum_{t=1}^{T_d} \underline{\mathbf{W}}_t^{*\prime} \underline{\mathbf{Z}}_t^L \left(\underline{\mathbf{Z}}_t^{L\prime} \underline{\mathbf{Z}}_t^L\right)^{-1} \underline{\mathbf{Z}}_t^{L\prime} \underline{\mathbf{y}}_t^*\right] \quad (31)$$

where $\underline{\mathbf{Z}}_t^L = (\mathbf{z}_{1t}^L, ..., \mathbf{z}_{Nt}^L)'$.

3 Asymptotic properties

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

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

3.1 Fixed T and large N case

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

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad (t = 1, ..., T;\ i = 1, ..., N). \quad (32)$$

We make the following assumption

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

$$y_{i0} = \frac{\eta_i}{1-\alpha} + e_{i0} \tag{33}$$

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

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

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

The model after the FOD transformation is given by

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

Let $\mathbf{z}_{it}^{L(\ell)} = (y_{i,t-\ell}, ..., y_{i,t-1})'$ and $\mathbf{z}_{it}^{B(\ell)} = (y_{i,t-\ell}^{**}, ..., y_{i,t-1}^{**})'$ be an $m_t^L \times 1$ and $m_t^B \times 1$ vectors of instruments, respectively, where ℓ denotes the maximum length of instruments used in each period and $y_{i,t-k}^{**} = c_{T-t+k+1}^{\iota}\left[y_{i,t-k} - (y_{i,t-k-1} + \cdots + y_{i0})/(t-k)\right], (1 \leq k \leq \ell)$. Note that $m_t^L, m_t^B = t$ for $t < \ell$ and $m_t^L, m_t^B = \ell$ for $t \geq \ell$.

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

$$\widehat{\alpha}_{GMM}^{L(\ell)} = \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{L(\ell)} \mathbf{y}_{t-1}^*\right)^{-1} \left(\sum_{t=1}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{L(\ell)} \mathbf{y}_t^*\right), \tag{34}$$

$$\widehat{\alpha}_{GMM}^{B(\ell)} = \left(\sum_{t=2}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{B(\ell)} \mathbf{y}_{t-1}^*\right)^{-1} \left(\sum_{t=2}^{T-1} \mathbf{y}_{t-1}^{*\prime} \mathbf{P}_t^{B(\ell)} \mathbf{y}_t^*\right) \tag{35}$$

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

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

$$Avar\left(\widehat{\alpha}_{GMM}^{L(\ell)}\right) = \sigma_v^2 \left[\sum_{t=1}^{T-1} E\left(y_{i,t-1}^* \mathbf{z}_{it}^{L(\ell)\prime}\right) \left[E\left(\mathbf{z}_{it}^{L(\ell)} \mathbf{z}_{it}^{L(\ell)\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^{L(\ell)} y_{i,t-1}^*\right)\right]^{-1}, \tag{36}$$

$$Avar\left(\widehat{\alpha}_{GMM}^{B(\ell)}\right) = \sigma_v^2 \left[\sum_{t=1}^{T-1} E\left(y_{i,t-1}^* \mathbf{z}_{it}^{B(\ell)\prime}\right) \left[E\left(\mathbf{z}_{it}^{B(\ell)} \mathbf{z}_{it}^{B(\ell)\prime}\right)\right]^{-1} E\left(\mathbf{z}_{it}^{B(\ell)} y_{i,t-1}^*\right)\right]^{-1}. \tag{37}$$

Note that the asymptotic variance $Avar\left(\widehat{\alpha}_{GMM}^{L(\ell)}\right)$ is a function of α and the variance ratio $r = \sigma_\eta^2/\sigma_v^2$ whereas $Avar\left(\widehat{\alpha}_{GMM}^{B(\ell)}\right)$ is a function of α only since a covariance-stationarity is imposed in Assumption 1 (see Theorem 1). Figure 1 shows the asymptotic variances of $\widehat{\alpha}_{GMM}^{L(\ell)}$ and $\widehat{\alpha}_{GMM}^{B(\ell)}$ based on (36) and (37) with various lag length of instruments for the cases with $\alpha = 0.3, 0.6, 0.9$, $r = 0.2, 1, 5$ and $T = 10$. For the detail of computation of (36) and (37), see appendix. From the figure, it is found that when $r = 0.2$, the asymptotic variances of the two GMM estimators are very similar regardless of lag length of instruments ℓ. This implies that many lags are not required to improve efficiency. Also it is found that $\widehat{\alpha}_{GMM}^{L}$ tends to be more efficient than $\widehat{\alpha}_{GMM}^{B}$. This is because one estimation period is lost in $\widehat{\alpha}_{GMM}^{B}$ compared to $\widehat{\alpha}_{GMM}^{L}$. However, when $r = 1$ and $r = 5$, the result dramatically changes. From the figure, it is found that $\widehat{\alpha}_{GMM}^{B}$ is little affected by lag length ℓ and hence, we do not need to use many lags; in view of the figure, one lagged instrument is sufficient to obtain nearly efficient estimator. However, this is not the case for $\widehat{\alpha}_{GMM}^{L}$. When $r = 1$ and $r = 5$, the lag length of instruments substantially affects the asymptotic variance. When r is large, $\widehat{\alpha}_{GMM}^{L}$ with one or two lagged instruments are far less efficient than $\widehat{\alpha}_{GMM}^{B}$ despite the same number of instruments is used. But as the lag length increases, efficiency improve. What is striking is that efficiency gain when lag length ℓ is increased from one to two or two from three is substantial. If lag length is more than four, the asymptotic variances of $\widehat{\alpha}_{GMM}^{L}$ and $\widehat{\alpha}_{GMM}^{B}$ are very similar. This implies that although many instruments leads to efficiency gain, in the current case, three or four lags are sufficient to obtain reasonably efficient $\widehat{\alpha}_{GMM}^{L}$. In other words, using higher order lags does not contribute to efficiency gain so much. Thus, this result supports the use of a few lagged variable as instruments. Based on this results, in the following, we mainly consider $\widehat{\alpha}_{GMM}^{L}$ with $\ell = 3$ and $\widehat{\alpha}_{GMM}^{B}$ with $\ell = 1$. From the figure, we also find an interesting relationship between the variance ratio r and instruments lag length ℓ. For example, consider the case with $(\alpha, r) = (0.6, 0.2)$ and $(\alpha, r) = (0.6, 5)$. Comparing these two cases, we find that $\widehat{\alpha}_{GMM}^{L}$ becomes substantially less efficient when ℓ is small if r is increased from 0.2 to 5, while efficiency loss is not so evident when ℓ is large. For instance, the variance of $\widehat{\alpha}_{GMM}^{L}$ with $r = 5$ is about 11.7 times larger than that with $r = 0.2$ when $\ell = 1$. However, when $\ell = 5$, the variance of $\widehat{\alpha}_{GMM}^{L}$ with $r = 5$ is only 1.8 times larger than that with $r = 0.2$. This indicates that the variance inflation of $\widehat{\alpha}_{GMM}^{L}$ caused by a large r is more evident when the instruments lag length is small compared with a large ℓ; in other words, the extent of reactivity of the variance of $\widehat{\alpha}_{GMM}^{L}$ to a large r can be reduced by using many instruments, i.e., a large ℓ.

Next, we investigate the effect of time length T on the efficiency of IV and GMM estimators with level and BOD instruments. In addition to the GMM estimators (34) and (35), we consider IV estimators given by

$$\begin{aligned}
\widehat{\alpha}_{IV}^{L} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} y_{i,t-1} y_{i,t-1}^{*}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T-1} y_{i,t-1} y_{it}^{*}\right), \\
\widehat{\alpha}_{IV}^{B} &= \left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} y_{i,t-1}^{**} y_{i,t-1}^{*}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=2}^{T-1} y_{i,t}^{**} y_{it}^{*}\right).
\end{aligned}$$

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

$$Avar\left(\widehat{\alpha}_{GMM}^{L(t)}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1} \psi_t^2 \left(1 - \frac{r(\frac{1+\alpha}{1-\alpha})}{1 + r\{(\frac{1+\alpha}{1-\alpha}) + (t-1)\}}\right)\right]^{-1}, \tag{38}$$

$$Avar\left(\widehat{\alpha}_{GMM}^{L(3)}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1} \psi_t^2 \left(1 - \frac{r(\frac{1+\alpha}{1-\alpha})}{1 + r\{(\frac{1+\alpha}{1-\alpha}) + m_t^L\}}\right)\right]^{-1}, \tag{39}$$

$$Avar\left(\widehat{\alpha}_{GMM}^{L(1)}\right) = (1-\alpha^2)\left[\sum_{t=1}^{T-1}\psi_t^2\left(\frac{1}{1+r(\frac{1+\alpha}{1-\alpha})}\right)\right]^{-1}, \quad (40)$$

$$Avar\left(\widehat{\alpha}_{IV}^{L}\right) = (1-\alpha^2)\left[1+r\left(\frac{1+\alpha}{1-\alpha}\right)\right]\left[\sum_{t=1}^{T-1}\psi_t\right]^{-2}, \quad (41)$$

$$Avar\left(\widehat{\alpha}_{GMM}^{B(1)}\right) = (1-\alpha^2)\left[\sum_{t=2}^{T-1}\psi_t^2\left(1-\frac{\alpha\phi_{t-1}}{t-1}\right)^2 A_t^{-1}\right]^{-1}, \quad (42)$$

$$Avar\left(\widehat{\alpha}_{IV}^{B}\right) = (1-\alpha^2)\left(\sum_{t=2}^{T-1}c_{T-t+1}^{\iota 2}A_t\right)\left[\sum_{t=2}^{T-1}\psi_t c_{T-t+1}^{\iota}\left(1-\frac{\alpha\phi_{t-1}}{t-1}\right)^2\right]^{-2} \quad (43)$$

where

$$\psi_t = c_t^{\iota}\left[1-\frac{\alpha\phi_{T-t}}{T-t}\right], \quad (44)$$

$$\phi_j = \frac{1-\alpha^j}{1-\alpha} = 1+\alpha+\cdots+\alpha^{j-1}, \quad (45)$$

$$A_t = \left[1-\frac{2\alpha\phi_{t-1}}{t-1}+\frac{1}{(t-1)^2}\left\{\frac{(t-1)(1+\alpha)}{1-\alpha}-\frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}\right]. \quad (46)$$

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

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

3.2 Large T and large N case

In order to derive the asymptotic property for large N and large T, we first make the following assumptions.

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

$$E(v_{it}|\mathbf{w}_{it}, ..., \mathbf{w}_{i1}, \eta_i) = 0. \quad (47)$$

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

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

$$\mathbf{w}_{it} = \begin{cases} \boldsymbol{\mu}_i + \boldsymbol{\xi}_{it} & \text{FE model} \\ \boldsymbol{\mu}_i + \boldsymbol{\kappa}_i t + \boldsymbol{\xi}_{it} & \text{trend model} \end{cases} \tag{48}$$

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d} \boldsymbol{\xi}_{it}\boldsymbol{\xi}'_{it} \xrightarrow{p} \boldsymbol{\Gamma}_0, \tag{49}$$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d} \boldsymbol{\xi}_{it}v_{it} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}) \tag{50}$$

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

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

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

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

$$(a) \quad \frac{1}{NT}\sum_{i=1}^{N} \mathbf{W}'_i\mathbf{K}'_T\mathbf{B}'_{T_d}\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N} \mathbf{W}'_i\mathbf{Q}_T\mathbf{W}_i + O_p\left(\frac{\log T}{T}\right), \tag{51}$$

$$(b) \quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{W}'_i\mathbf{Q}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d} \boldsymbol{\xi}_{it}v_{it} - O_p\left(\sqrt{\frac{N}{T}}\right), \tag{52}$$

$$(c) \quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{W}'_i\mathbf{K}'_T\mathbf{B}'_{T_d}\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d} \boldsymbol{\xi}_{it}v_{it} + o_p(1). \tag{53}$$

Lemma 1(a) is a result for the denominator of the FE and DFIV estimators, and Lemma 1(b) and (c) are the results fore the numerator of the FE and DFIV estimators, respectively. Lemma 1(a) indicates that the denominators of $\widehat{\boldsymbol{\delta}}_{FE}$ and $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ are asymptotically equivalent when T is large. Also, comparing (b) and (c), we find that the second term of the right-hand side makes a significant difference in FE and DFIV estimators. When N/T converges to a non-zero constant, the second term in (b) becomes $O_p(1)$, and because of this, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{FE}$ is not centered around the true value as shown in Theorem 2 below. This bias is due the incidental parameter problem. Contrary to the FE estimator, the second term of (c) vanishes asymptotically. Hence, as shown in Theorem 4 below, the asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ is centered around the true value.

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

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

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

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

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

$$\widehat{\boldsymbol{\delta}}_{BCFE} = \widehat{\boldsymbol{\delta}}_{FE} - \frac{1}{T}\widehat{\boldsymbol{\Gamma}}_0^{-1}\widehat{\bar{\mathbf{h}}} \tag{54}$$

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

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

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

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

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

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

This theorem implies that $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ and $\widehat{\boldsymbol{\delta}}_{BCFE}$ have the same asymptotic distribution when both N and T are large. This is because $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ does not suffer from the incidental parameter problem nor many instruments problem.

For the GMM estimator, since it is quite involved to derive the large T and large N asymptotic properties for the general model (21), we instead consider a simple AR(1) model given by (32). The asymptotic distributions of IV/GMM estimators for the AR(1) model (32) under large N and large T asymptotics is given in the following theorem.

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

$$(a)\quad \sqrt{NT}\left[\widehat{\alpha}_{FE} - \left(\alpha - \frac{1}{T}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \qquad \text{if } T^3/N \to 0,$$

$$(b)\quad \sqrt{NT}\left[\widehat{\alpha}_{GMM}^{L(t)} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2),$$
$$\text{if } (\log T)^2/N \to \infty \text{ and } T/N \to c, (0 \le c \le 1),$$

$$(c)\quad \sqrt{NT}\left(\widehat{\alpha}_{GMM}^{L(1)} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\left(1 + \frac{r(1+\alpha)}{1-\alpha}\right)\right), \qquad \text{if } T/N \to c, (0 \le c \le 1),$$

$$(d) \quad \sqrt{NT}\left(\widehat{\alpha}_{IV}^{L}-\alpha\right) \xrightarrow{d} \mathcal{N}\left(0,(1-\alpha^2)\left(1+\frac{r(1+\alpha)}{1-\alpha}\right)\right),$$

$$(e) \quad \sqrt{NT}\left(\widehat{\alpha}_{GMM}^{B(1)}-\alpha\right) \xrightarrow{d} \mathcal{N}(0,1-\alpha^2), \qquad \text{if } T/N \to c, (0 \le c \le 1),$$

$$(f) \quad \sqrt{NT}\left(\widehat{\alpha}_{IV}^{B}-\alpha\right) \xrightarrow{d} \mathcal{N}(0,1-\alpha^2).$$

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

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

In the AR(1) model given by (32), Hahn and Kuersteiner (2002) demonstrate that $\mathcal{N}(0,1-\alpha^2)$ is the minimal asymptotic distribution under the normality assumption on v_{it}. Hence, the bias-corrected FE estimator given by $\widehat{\alpha}_{BCFE}=[(T+1)/T]\widehat{\alpha}_{FE}+(1/T)$, the GMM estimators using all instruments in levels and using one lagged new instrument, and the IV estimator using new instruments are asymptotically efficient. Given this, it is conjectured that $\widehat{\boldsymbol{\delta}}_{BCFE}$ and $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ are asymptotically efficient under certain conditions. However, a formal discussion is beyond the scope of the present paper and left as a future topic.

4 Monte Carlo simulation

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

4.1 Design

The data are generated as

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

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

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

or

$$\mathbf{p}_{it}=\mathbf{\Phi}\mathbf{p}_{i,t-1}+\mathbf{c}_\eta\eta_i+\varphi\mathbf{c}_\lambda\lambda_i t+\boldsymbol{\varepsilon}_{it} \tag{55}$$

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

$$\boldsymbol{\Phi} = \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix}. \tag{56}$$

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

$$\begin{aligned} \mathbf{p}_{it} &= \mathbf{a}_i + \varphi \mathbf{b}_i t + \boldsymbol{\zeta}_{it}, \\ \boldsymbol{\zeta}_{it} &= \boldsymbol{\Phi}\boldsymbol{\zeta}_{i,t-1} + \boldsymbol{\varepsilon}_{it} \end{aligned} \tag{57}$$

where

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

Data for y_{it} and x_{it} are generated from (57). For the sample size, we consider $T = 10, 25, 50, 100$ and $N = 50, 100, 250$. For parameter values, we consider $\alpha = 0.4, 0.8$, $\beta = 1$, $\rho = 0.5$, $\theta = -0.2$, $\tau_\eta = 0.5$, $\tau_\lambda = 0.5$. v_{it}, e_{it}, η_i and λ_i are independently generated as $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$, $e_{it} \sim \mathcal{N}(0, \sigma_e^2)$, $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$ and $\lambda_i \sim \mathcal{N}(0, \sigma_\lambda^2)$ with $\sigma_v^2 = 1$ $\sigma_e^2 = 0.16$, $\sigma_\eta^2 = 1, 5$ and $\sigma_\lambda^2 = 1$[6]. We report the median bias, interquartile range (IQR), median absolute error (MAE) and empirical size with 5% significance level based on 2,000 replications.

4.2 Estimators to be compared

We consider seven estimators. The first is the FE estimator $\widehat{\boldsymbol{\delta}}_{FE}$ given in (29).[7] The second is the IV estimator $\widehat{\boldsymbol{\delta}}_{IV}^{L}$ given in (30) where instruments in levels are used. The third and fourth are the GMM estimator $\widehat{\boldsymbol{\delta}}_{GMM}^{L}$ given in (31) where instruments in levels are used. For the choice of lag length of instruments, we consider $\ell = 1$ and 3. The corresponding GMM estimators are denoted as "LEV1" and "LEV3", respectively. The fifth estimator is the DFIV estimator $\widehat{\boldsymbol{\delta}}_{IV}^{B}$ defined in (24). The last two estimators are the GMM estimator $\widehat{\boldsymbol{\delta}}_{GMM}^{B}$ defined in (26) where backward filtered instruments are used. For the choice of lag length of instruments, we consider $\ell = 1$ and 3. The corresponding GMM estimators are denoted as "BOD1" and "BOD3", respectively. For the computation of standard errors, we use those obtained under large N and fixed T since they are more accurate than those obtained under large N and large T(see Hayakawa, 2015).

4.3 Results

Simulation results are provided in Tables 1-4. We first consider the model with fixed effects only. From Tables 1 and 2, we find that the FE estimator for α is severely biased when $T = 10$. However, as T gets larger, the bias becomes smaller as expected since the FE estimator is consistent when T is large. However, in terms of accuracy of inference, the sizes are severely

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

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

distorted even when T is large, say, $T = 100$. This is because the asymptotic distribution of the FE estimator is not centered around the true value due to the incidental parameter problem. Also, note that increase in N does not reduce the bias since the bias of FE estimator does not depend on N. With regard to the FE estimator of β, the performance is better than those of α. However, it still shows some bias and size distortions. This result implies that the FE estimator does not work even when T is large. Also, note that a widely acceptable bias-correction method is not available since the regressor is weakly exogenous[8]. With regard to the IV and GMM estimators, in terms of MAE, the IV estimators using instruments in levels perform worst among the four estimators mainly due to the large dispersions. With regard to the remaining three estimators, they perform very similarly in terms of MAE when $T = 10$. However, as T gets larger, IV and GMM estimators using new instruments outperform the GMM estimator using instruments in levels. With regard to the choice of IV or GMM estimators using new instruments, it is observed that the GMM estimator tends to have slightly smaller MAEs than IV estimator. In terms of accuracy of inference, IV and GMM estimators using new instruments have almost correct empirical sizes in all cases while the GMM estimator using instruments in levels have large size distortions especially when $T = 10$ and $\alpha = 0.8$. We find that the efficiency of GMM estimator using instruments in levels depends substantially on the lag length ℓ. Comparing the IQRs with $\ell = 1$ and 3, the reduction of dispersion with $\ell = 3$ is substantial though it induces many instruments. Contrary to IV/GMM with instruments in levels, the effects of ℓ on the performance of the GMM with new instruments are minor and the IQRs are relatively smaller than those of GMM with instruments in levels. This result is consistent with the theoretical implication that using new instruments leads to efficient estimation. Considering overall performance, we may conclude that the IV or GMM estimator using new instruments with $\ell = 1$ tend to perform best in many cases.

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

5 Conclusion

In this paper, we have proposed a new instrumental variable estimator for panel data models including static and dynamic models with weakly exogenous variables and with fixed effects and/or heterogeneous time trends. We showed that the new IV estimator called the DFIV

[8]If the regressors are strictly exogenous, bias-corrected FE estimators such as Bun and Carree (2005) can be used.

estimator is consistent and has the same asymptotic distribution as the bias-corrected fixed effects estimator, which is sometimes infeasible, when both N and T are large. This implies that the DFIV estimator is as efficient as the fixed effects estimator. Monte Carlo simulation results revealed that the DFIV and DFGMM estimators tend to perform better than the conventional IV/GMM estimators using instruments in levels in almost all cases.

Appendix

Derivation of $\mathbf{F}_T$

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

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

Multiplying $\mathbf{D}_T$ by (1) and noting that $\mathbf{D}_T \boldsymbol{\iota}_T = \mathbf{0}$, we have

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

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

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

where $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2}$ is the *upper* triangular Cholesky factorization of $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1}$ with[9]

$$\underset{(T_1 \times T_1)}{\mathbf{D}_T \mathbf{D}_T'} = \begin{bmatrix} 2 & -1 & 0 & & & 0 \\ -1 & 2 & -1 & & & \\ 0 & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & 0 \\ & & & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 2 \end{bmatrix}.$$

To compute $\left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1/2}$, we need to derive the inverse matrix $\mathbf{H}^{\iota} = \left(\mathbf{D}_T \mathbf{D}_T'\right)^{-1} = \{h_{st}^{\iota}\}$.[10] Using the results by El-Mikkawy and Karawia (2006), we have

$$h_{st}^{\iota} = \begin{cases} \frac{T_1}{T_1+1} & \text{if } s = t = 1 \text{ or } s = t = T_1 \\ \frac{s(T_1-s+1)}{T_1+1} & \text{if } s = t < T_1 \\ \frac{s(T_1-t+1)}{T_1+1} & \text{if } s < t \\ \frac{t(T_1-s+1)}{T_1+1} & \text{if } s > t \end{cases}$$

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

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

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

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

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

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

ℓ_{ij} can be solved sequentially as follows:

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

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

Next, we consider a model with individual effects and heterogeneous time trends given by (11). To remove both η_i and λ_i from the model, we take second differences. In terms of a model in matrix, this corresponds to multiplying by $\mathbf{D}_{T_1}\mathbf{D}_T, (T_2 \times T)$, we have

$$\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{y}_i = \mathbf{D}_{T_1}\mathbf{D}_T\mathbf{W}_i\boldsymbol{\delta} + \mathbf{D}_{T_1}\mathbf{D}_T\mathbf{v}_i.$$

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

$$\mathbf{F}_T^\tau = \left(\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T_1}'\right)^{-1/2}\mathbf{D}_{T_1}\mathbf{D}_T,$$

where $\left(\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T_1}'\right)^{-1/2}$ is the *upper* triangular Cholesky factorization of $\left(\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T_1}'\right)^{-1}$. To compute $\mathbf{F}_T^\tau$, we need to derive the inverse matrix $\mathbf{H}^\tau = \left(\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T_1}'\right)^{-1} = \{h_{st}^\tau\}$ with[11]

$$\underset{(T_2\times T_2)}{\mathbf{D}_{T_1}\mathbf{D}_T\mathbf{D}_T'\mathbf{D}_{T_1}'} = \begin{bmatrix} 6 & -4 & 1 & & & 0 \\ -4 & 6 & -4 & & & \\ 1 & -4 & 6 & -4 & & \\ & & \ddots & \ddots & \ddots & 1 \\ & & & -4 & 6 & -4 \\ 0 & & & 1 & -4 & 6 \end{bmatrix}.$$

Using the results by Dow (2003), we have[12]

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

where

$$a_{t0} = -(3 + 2t + T_2)d_t/c, \qquad a_{t1} = 3t(1 + T_2)d_t/c, \qquad a_{t2} = (3 + 5t + T_2 + 3tT_2)d_t/c,$$

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

[12] See also Chen (2013) for an alternative expression.

$$\begin{aligned}
b_{t0} &= (5-2t+3T_2)e_t/c, \qquad b_{t1} = -3(1+T_2)(4-t+2T_2)e_t/c,\\
b_{t2} &= (1+5t+12T_2+3tT_2+12T_2^2+3T_2^3)e_t/c, \qquad b_{t3} = (1-t)e_t/6,\\
d_t &= (T_2-t+1)(T_2-t+2), \qquad e_t = t(t+1), \qquad c = 6(T_2+1)(T_2+2)(T_2+3).
\end{aligned}$$

Using these results and applying the algorithm of Cholesky factorization introduced above where $\mathbf{A} = \mathbf{H}^{\tau}$, after a lengthy calculation, we obtain the explicit expression of $\mathbf{F}_T^{\tau}$ as in (12).

Derivation of asymptotic variances (36) and (37)

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

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

Using this, we have

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

Hence, (36) and (37) can be written as

$$Avar\left(\widehat{\alpha}_{GMM}^L\right) = \sigma_v^2 \left[\sum_{t=1}^{T-1} \mathbf{f}_t' \mathbf{V}_T \mathbf{L}_{(t-\ell+1:t)}' \left[\mathbf{L}_{(t-\ell+1:t)} \mathbf{V}_T \mathbf{L}_{(t-\ell+1:t)}'\right]^{-1} \mathbf{L}_{(t-\ell+1:t)} \mathbf{V}_T \mathbf{f}_t\right]^{-1} \quad (58)$$

$$Avar\left(\widehat{\alpha}_{GMM}^B\right) = \sigma_v^2 \left[\sum_{t=2}^{T-1} \mathbf{f}_t' \mathbf{V}_T \mathbf{B}_{(t-\ell+1:t)}' \left[\mathbf{B}_{(t-\ell+1:t)} \mathbf{V}_T \mathbf{B}_{(t-\ell+1:t)}'\right]^{-1} \mathbf{B}_{(t-\ell+1:t)} \mathbf{V}_T \mathbf{f}_t\right]^{-1} \quad (59)$$

Figure 1 is described based on (58) and (59) numerically without deriving the explicit form of expectations such as $E\left(\mathbf{z}_{it}^{L(\ell)} y_{i,t-1}^*\right)$.

Proof of Theorem 1

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

$$y_{i,t-1} = \frac{\eta_i}{1-\alpha} + \xi_{i,t-1} \quad (60)$$

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

$$y_{i,t-1}^* = \psi_t \xi_{i,t-1} - c_t \left(\frac{\phi_{T-t} v_{it} + \cdots + \phi_1 v_{i,T-1}}{T-t}\right). \quad (61)$$

where ϕ is defined in (45). Using (60) and (61), and under Assumption 1, we have

$$E(\mathbf{z}_{it}^L y_{i,t-1}^*) = \psi_t \left(\frac{\sigma_v^2}{1-\alpha^2}\right) (\alpha^{\ell-1}, \cdots, 1)', \quad (62)$$

$$\left[E\left(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime}\right)\right]^{-1} = \mathbf{V}_\ell^{-1} = \frac{1}{\sigma_v^2} \left[(\sqrt{\lambda} \boldsymbol{\iota}_\ell)(\sqrt{\lambda} \boldsymbol{\iota}_\ell)' + \boldsymbol{\Phi}_\ell\right]^{-1} \quad (63)$$

where $\lambda = \sigma_\mu^2/\sigma_v^2 = r/(1-\alpha)^2$, $\boldsymbol{\iota}_\ell$ is an ℓ dimensional column vector of ones, and $\mathbf{V}_\ell$ is the upper-left $\ell \times \ell$ matrix of $\mathbf{V}_T$. The explicit expression of (63) is obtained as follows. By using the Sherman-Morrison-Woodbury inversion formula

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

and the decomposition of $\mathbf{V}_\ell^{-1}$ [13]

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

where

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

we obtain

$$\left[E(\mathbf{z}_{it}^L \mathbf{z}_{it}^{L\prime})\right]^{-1} = \sigma_v^{-2}\left[\mathbf{C}'\mathbf{C} - \frac{\lambda}{1+\lambda\boldsymbol{\iota}_\ell'\mathbf{C}'\mathbf{C}\boldsymbol{\iota}_\ell}\mathbf{C}'\mathbf{C}\boldsymbol{\iota}_\ell\boldsymbol{\iota}_\ell'\mathbf{C}'\mathbf{C}\right]. \quad (64)$$

By substituting (62) and (64) into (36), we obtain (38). The results (39) and (40) are obtained from (38). Next, we consider $Var(\widehat{\alpha}_{GMM}^{B(1)})$. First, note that $y_{i,t-1}^{**}$ can be written as

$$y_{i,t-1}^{**} = c_{T-t+1}^{\iota}\left[\xi_{i,t-1} - \frac{\xi_{i,t-2} + \cdots + \xi_{i0}}{t-1}\right]. \quad (65)$$

Then, using (61) and (65), we obtain

$$E\left(y_{i,t-1}^{**} y_{i,t-1}^{*}\right) = \left(\frac{\sigma_v^2}{1-\alpha^2}\right)\psi_t c_{T-t+1}^{\iota}\left(1 - \frac{\phi_{t-1}}{t-1}\right). \quad (66)$$

Also, from (60), we obtain

$$\begin{aligned} E[(y_{i,t-1}^{**})^2] &= c_{T-t+1}^{\iota 2} E\left[\xi_{i,t-1} - \frac{1}{t-1}(\xi_{i,0} + \cdots + w_{i,t-2})\right]^2 \\ &= c_{T-t+1}^{\iota 2}\left[\frac{\sigma_v^2}{1-\alpha^2}\left(1 - \frac{2\alpha\phi_{t-1}}{t-1}\right) + \frac{1}{(t-1)^2}E(\xi_{i0} + \cdots + \xi_{i,t-1})^2\right]. \quad (67) \end{aligned}$$

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

$$E(\xi_{i0} + \cdots + \xi_{i,t-1})^2 = \frac{\sigma_v^2}{1-\alpha^2}\left[\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right]. \quad (68)$$

By substituting this into (67), we get

$$E[(y_{i,t-1}^{**})^2] = \left(\frac{\sigma_v^2}{1-\alpha^2}\right) c_{T-t+1}^{\iota 2} A_t \quad (69)$$

where A_t is defined in (46). The result (42) is obtained by substituting (66) and (69) into (36).

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

Next, we derive the asymptotic variances of $\widehat{\alpha}_{IV}^{L}$ and $\widehat{\alpha}_{IV}^{B}$. Using (60) and (61) and the fact that v_{it}^{*} is serially uncorrelated, we obtain

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

Using these, we obtain the asymptotic variance of $\widehat{\alpha}_{IV}^{L}$ as in (41). The asymptotic variance of $\widehat{\alpha}_{IV}^{B}$ can be derived similarly. Using (66) and (69), and under Assumption 1, we obtain

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

From these, the asymptotic variance of $\widehat{\alpha}_{IV}^{L}$ is obtained as (43).

Proof of Lemma 1

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

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

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

$$\mathbf{B}_T=\mathcal{I}_{T_d}\mathbf{F}_T\mathcal{I}_T \tag{70}$$

where

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

and $\mathcal{I}_T^2=\mathcal{I}_T'\mathcal{I}_T=\mathbf{I}_T$. Furthermore, using (48), $\mathbf{W}_i$ can be written as

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

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

Proof of (a): Note the following decomposition:

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i+\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T-\mathbf{Q}_T\right)\mathbf{W}_i. \tag{71}$$

Using $\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{C}_T = \mathbf{B}_{T_d}\mathbf{K}_T\mathbf{C}_T = \mathbf{Q}_T\mathbf{C}_T = \mathbf{0}$ and (19), the second term of (71) can be further decomposed as

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{Q}_T\right)\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Xi}_i'\left(\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{I}_T\right)\mathbf{\Xi}_i + \frac{1}{NT}\sum_{i=1}^{N}\mathbf{\Xi}_i'\mathbf{R}_T\mathbf{\Xi}_i. \tag{72}$$

To consider the first term of right-hand side of (72), we derive the explicit form of $\mathbf{A}_T = \mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T - \mathbf{I}_T$. Using (70), $\mathbf{F}_{T_d} = \mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'$ and

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

we have

$$\begin{aligned}
\mathbf{A}_T &= \mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'\mathbf{L}_T - \mathbf{I}_T = \mathbf{K}_T'\mathcal{I}_{T_d}\mathbf{F}_{T_d}'\mathcal{I}_{T_{2d}}\mathbf{L}_{T_d}\mathbf{F}_T\mathbf{L}_T'\mathbf{L}_T - \mathbf{I}_T \\
&= \left(\mathbf{K}_T'\mathcal{I}_{T_d}\mathbf{L}_T\right)\mathbf{F}_T'\left(\mathbf{L}_{T_d}'\mathcal{I}_{T_{2d}}\mathbf{L}_{T_d}\right)\mathbf{F}_T\left(\mathbf{L}_T'\mathbf{L}_T\right) - \mathbf{I}_T \\
&= \begin{bmatrix} -\mathbf{I}_d & \mathcal{I}_d\mathbf{F}_{23}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} & \mathcal{I}_d\mathbf{F}_{23}'\mathcal{I}_{T_{2d}}\mathbf{F}_{23} \\ \mathbf{0}_{T_{2d}\times d} & \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} - \mathbf{I}_{T_{2d}} & \mathcal{I}_{T_{2d}}\mathbf{F}_{22}'\mathcal{I}_{T_{2d}}\mathbf{F}_{23} \\ \mathbf{0}_{d\times d} & \mathbf{0}_{d\times T_{2d}} & -\mathbf{I}_d \end{bmatrix} \\
&= \{\mathbf{A}_{ij}\}, (i,j = 1,2,3).
\end{aligned} \tag{73}$$

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

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

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

we have

$$\begin{aligned}
\underset{(d\times T_{2d})}{\mathbf{A}_{12}} &= \mathcal{I}_d\mathbf{F}_{23}'\mathcal{I}_{T_{2d}}\mathbf{F}_{22} \\
&= \begin{bmatrix} f_{T_d,T}f_{d+1,d+1} & f_{T_d-1,T}f_{d+2,d+2} + f_{T_d,T}f_{d+1,d+2} & \cdots & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T}f_{T_d-\ell+1,T_d} \\ f_{T_d,T_d+1}f_{d+1,d+1} & f_{T_d-1,T_d+1}f_{d+2,d+2} + f_{T_d,T_d+1}f_{d+1,d+2} & \cdots & \sum_{\ell=1}^{T_{2d}} f_{d+\ell,T_d+1}f_{T_d-\ell+1,T_d} \end{bmatrix} \\
&= \left\{a_{12}^{jk}\right\} = \sum_{\ell=1}^{k} f_{T_d-\ell+1,T-j+1}f_{d+\ell,d+k} = \sum_{\ell=1}^{k-1} f_{T_d-\ell+1,T-j+1}f_{d+\ell,d+k} + f_{T_d-k+1,T-j+1}f_{d+k,d+k}
\end{aligned}$$

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

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

$$
\begin{aligned}
\underset{(T_{2d}\times T_{2d})}{\mathbf{A}_{22}} &= \mathcal{I}_{T_{2d}} \mathbf{F}'_{22} \mathcal{I}_{T_{2d}} \mathbf{F}_{22} - \mathbf{I}_{T_{2d}} = \begin{bmatrix} a_{22}^{11} & a_{22}^{12} & \cdots & a_{22}^{1,T_{2d}} \\ 0 & a_{22}^{22} & \cdots & a_{22}^{2,T_{2d}} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{22}^{T_{2d},T_{2d}} \end{bmatrix} \\
&= \left\{a_{22}^{jk}\right\} = \begin{cases} 0 & \text{if } j>k \\ f_{T_d-j+1,T_d-j+1} f_{d+j,d+j} - 1 & \text{if } j=k \\ \sum_{\ell=1}^{k} f_{T_d-\ell+1,T_d-j+1} f_{d+\ell,d+k} & \text{if } j<k \end{cases} \\
&= \begin{cases} \begin{cases} 0 & \text{if } j>k \\ c^{\iota}_{T-j} c^{\iota}_{j+1} - 1 = O\left(\frac{1}{j}\right) + O\left(\frac{1}{(T-j)}\right) + O\left(\frac{1}{j^2} + \frac{1}{(T-j)^2}\right) & \text{if } j=k \\ \sum_{\ell=1}^{k-j+1} \frac{c^{\iota}_{\ell+1} c^{\iota}_{T-\ell}}{(T-\ell-1)\ell} = O\left(\frac{\log T}{T}\right) & \text{if } j<k \\ \qquad (j,k=1,...,T_{2d}) & \end{cases} & \text{for FE model} \\ \begin{cases} 0 & \text{if } j>k \\ c^{\tau}_{T-j-1} c^{\tau}_{j+2} - 1 = O\left(\frac{1}{j}\right) + O\left(\frac{1}{(T-j)}\right) & \text{if } j=k \\ \sum_{\ell=1,\ell>j,\ell<k}^{k} \frac{4c^{\tau}_{\ell+2} c^{\tau}_{T-\ell-1}(3j-\ell+3)(2T-3k+\ell-3)}{\ell(\ell+1)(T-\ell-2)(T-\ell-3)} & \\ -\frac{2c^{\tau}_{j+2} c^{\tau}_{T-j-1}(2T+j-3k-3)}{(T-j-2)(T-j-3)} - \frac{2c^{\tau}_{k+2} c^{\tau}_{T-k-1}(3j-k+3)}{k(k+1)} = O\left(\frac{\log T}{T}\right) & \text{if } j<k \\ \qquad (j,k=1,...,T_{2d}) & \end{cases} & \text{for trend model} \end{cases},
\end{aligned}
$$

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

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

We now assess the first term of (72). Using $\boldsymbol{\Xi}_i = (\boldsymbol{\Xi}_{1i}, \boldsymbol{\Xi}_{2i}, \boldsymbol{\Xi}_{3i})'$ where $\boldsymbol{\Xi}_{1i}$ is $d \times k$, $\boldsymbol{\Xi}_{2i}$ is $T_{2d} \times k$, and $\boldsymbol{\Xi}_{3i}$ is $d \times k$, we have

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

where

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

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

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

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

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

$$
\begin{aligned}
E(\mathbf{S}_{2i}) &= \sum_{t=2}^{T_d} a_{12}^{1,t-1} E\left(\boldsymbol{\xi}_{i1} \boldsymbol{\xi}_{it}'\right) = O\left(\frac{\log T}{T}\right) \sum_{t=2}^{T_1} \boldsymbol{\Gamma}_{i,t-1} = O\left(\frac{\log T}{T}\right), \\
E(\mathbf{S}_{3i}) &= a_{13}^{1,1} E\left(\boldsymbol{\xi}_{i1} \boldsymbol{\xi}_{iT}'\right) = O\left(\frac{\log T}{T}\right) \boldsymbol{\Gamma}_{i,T-1} = O\left(\frac{\log T}{T}\right).
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
E(\mathbf{S}_{5i}) &= \sum_{t=2}^{T_d} a_{23}^{t-1,1} E\left(\boldsymbol{\xi}_{it} \boldsymbol{\xi}_{iT}'\right) = \sum_{t=2}^{T_d} O\left(\frac{\log T}{T}\right) \boldsymbol{\Gamma}_{i,T-t} = O\left(\frac{\log T}{T}\right), \\
E(\mathbf{S}_{6i}) &= -E\left(\boldsymbol{\xi}_{iT} \boldsymbol{\xi}_{iT}'\right) = -\boldsymbol{\Gamma}_{i0} = O(1).
\end{aligned}
$$

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

$$
\frac{1}{NT} \sum_{i=1}^{N} \boldsymbol{\Xi}_i' \mathbf{A}_T \boldsymbol{\Xi}_i = O_p\left(\frac{\log T}{T}\right). \tag{74}
$$

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

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

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

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

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

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

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

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

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

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{A}_T\boldsymbol{\Xi}_i = O_p\left(\frac{\log T}{T}\right). \tag{75}$$

Next, we consider the second term of (72). Let us define $\mathbf{H}_i = \boldsymbol{\Xi}_i'\mathbf{R}_T\boldsymbol{\Xi}_i$. Then, for the FE model, using (20), we have

$$\begin{aligned} E(\mathbf{H}_i) &= \frac{1}{T}E(\boldsymbol{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\boldsymbol{\Xi}_i) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{is}'\right) \\ &= \boldsymbol{\Gamma}_{i0} + \frac{1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\left(\boldsymbol{\Gamma}_{i,t-s} + \boldsymbol{\Gamma}_{i,t-s}'\right) = O\left(1\right) \end{aligned} \tag{76}$$

where we used $\|\sum_{t=s+1}^{T}\boldsymbol{\Gamma}_{i,t-s}\| = \|\sum_{l=1}^{T-s}\boldsymbol{\Gamma}_{i,l}\| \le \sum_{l=1}^{T-s}\|\boldsymbol{\Gamma}_{i,l}\| < \sum_{l=1}^{\infty}\|\boldsymbol{\Gamma}_{i,l}\| < \infty$. For the trend model, using (20), we have

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

Using Assumption 3 and (76), we have

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\Xi}_i'\mathbf{R}_T\boldsymbol{\Xi}_i = O_p\left(\frac{1}{T}\right). \tag{77}$$

By combining (74), (75), and (77), we obtain

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

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

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

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

$$\begin{aligned}E(\mathbf{h}_i) &= \frac{-1}{T}E\left[\left(\sum_{t=1}^{T}\boldsymbol{\xi}_{it}\right)\left(\sum_{s=1}^{T}v_{is}\right)\right] = \frac{-1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}E\left(\boldsymbol{\xi}_{it}v_{is}\right) + \frac{-1}{T}\sum_{s,t=1,t\le s}^{T}E\left(\boldsymbol{\xi}_{it}v_{is}\right) \\ &= \frac{-1}{T}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\boldsymbol{\phi}_{i,t-s} = O\left(1\right)\end{aligned} \tag{78}$$

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

$$\mathbf{h}_i = -\frac{2\left(2T+1\right)\boldsymbol{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{v}_i}{T\left(T-1\right)} - \frac{12\boldsymbol{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\tau}_T'\mathbf{v}_i}{T\left(T-1\right)\left(T+1\right)} + \frac{6\left(\boldsymbol{\Xi}_i'\boldsymbol{\iota}_T\boldsymbol{\tau}_T'\mathbf{v}_i + \boldsymbol{\Xi}_i'\boldsymbol{\tau}_T\boldsymbol{\iota}_T'\mathbf{v}_i\right)}{T\left(T-1\right)}$$

$$= -\mathbf{h}_{1i} - \mathbf{h}_{2i} + \mathbf{h}_{3i}.$$

Using Assumption 3 and (78), we have

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

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

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

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

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

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

The first term converges in distribution to $\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ by Assumption 4. To derive the order of the second term, let us define $\mathbf{s}_i = \boldsymbol{\Xi}_i' \mathbf{A}_T \mathbf{v}_i$. Then, using (73), and $\mathbf{v}_i = (\mathbf{v}_{i1}', \mathbf{v}_{i2}', \mathbf{v}_{i3}')'$ where $\mathbf{v}_{i1}$ and $\mathbf{v}_{i3}$ are $d \times 1$ and $\mathbf{v}_{i2}$ is $T_{2d} \times 1$, $\mathbf{s}_i$ can be decomposed as

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

where

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

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

FE model Since $d = 1$ in the FE model, we have

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

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

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

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

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

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

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

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

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

For the covariances, we have

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

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

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

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

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

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

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

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

Trend model

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

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

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

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

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

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

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

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

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

For the covariances of trend model,

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

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

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

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

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

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

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

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

Proof of Theorems 2 and 3

We first provide a proof of Theorem 2. Using (77) and Assumption 4, we have

$$
\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}\boldsymbol{\xi}_{it}' + O_p\left(\frac{1}{T}\right) \xrightarrow{p} \boldsymbol{\Gamma}_0. \tag{79}
$$

Next, we have the following decomposition

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

Hence, using Assumption 4, as $N, T \to \infty$ with $N/T \to \kappa, (0 < \kappa < \infty)$, we obtain

$$
\sqrt{NT}(\widehat{\boldsymbol{\delta}}_{FE} - \boldsymbol{\delta}) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{v}_i
$$

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

$$\overset{d}{\to} \mathcal{N}(\sqrt{\kappa}\boldsymbol{\Gamma}_0^{-1}\bar{\mathbf{h}}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^{-1})$$

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

Proof of Theorem 4

Using Lemma 1(a), (79) and Assumption 4, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{W}_i = \frac{1}{NT}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{Q}_T\mathbf{W}_i + O_p\left(\frac{\log T}{T}\right) \to^p \boldsymbol{\Gamma}_0. \quad (80)$$

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

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{K}_T'\mathbf{B}_{T_d}'\mathbf{F}_{T_d}\mathbf{L}_T\mathbf{v}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=d+1}^{T_d}\boldsymbol{\xi}_{it}v_{it} + o_p(1) \to^d \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}). \quad (81)$$

Combining (80) and (81), we obtain the result.

References

Ahn, S. C. and P. Schmidt (1995) "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5-27.

Akashi, K. and N. Kunitomo (2015) "The limited information maximum likelihood approach to dynamic panel structural equation models," *Annals of the Institute of Statistical Mathematics*, 67, 39–73.

Alvarez, J. and M. Arellano (2003) "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121-1159.

Anderson, T. W. and C. Hsiao (1981) "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598-606.

Arellano, M. (2003) *Panel Data Econometrics*, Oxford: Oxford University Press.

Arellano, M. and S. Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.

Arellano, M. and O. Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29-51.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

Bun, M. J. G. and M. A. Carree (2005) "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business and Economic Statistics*, 23, 200-210.

Chen, Y. (2013) "A New Algorithm for Computing the Determinant and the Inverse of a Pentadiagonal Toeplitz Matrix," *Engineering*, 5, p. 25.

Dow, M. (2003) "Explicit inverses of Toeplitz and associated matrices," *ANZIAM Journal*, 44, E185–E215.

El-Mikkawy, M. and A. Karawia (2006) "Inversion of General Tridiagonal Matrices," *Applied Mathematics Letters*, 19, 712–720.

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*, 70, 1639-1657.

Hayakawa, K. (2006) "Efficient GMM estimation of dynamic panel data models where large heterogeneity may be present."

Hayakawa, K. (2009) "A Simple Efficient Instrumental Variable Estimator in Panel AR(p) Models When Both N and T Are Large," *Econometric Theory*, 25, 873-890.

Hayakawa, K. (2015) "The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T Are Large," *Econometric Theory*, 31, 647-667.

Hayakawa, K. (2016) "Improved GMM Estimation of Panel VAR Models," *Computational Statistics & Data Analysis.*

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988) "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371-1395.

Hsiao, C. and Q. Zhou (2015) "First Difference or Forward Demeaning: Implications for the Method of Moments Estimators." Working paper.

Lee, Y.J., R. Okui, and M. Shintani (2016) "Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes." Working paper.

Moon, H. R. and P. C. B. Phillips (2000) "Estimation of Autoregressive Roots Near Unity Using Panel Data," *Econometric Theory*, 16, 927-997.

Nickell, S. J. (1981) "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417-1426.

Okui, R. (2009) "The Optimal Choice of Moments in Dynamic Panel Data Models," *Journal of Econometrics*, 151, 1-16.

Phillips, P. C. B. and H. R. Moon (1999) "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1111.

Phillips, P. C. B. and D. Sul (2007) "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," *Journal of Econometrics*, 127, 162-188.

Wansbeek, T. J. and T. Knaap (1999) "Estimating a Dynamic Panel Data Model with Heterogeneous Trends," *Annales d'Economie et de Statistique*, 55-56, 331-349.

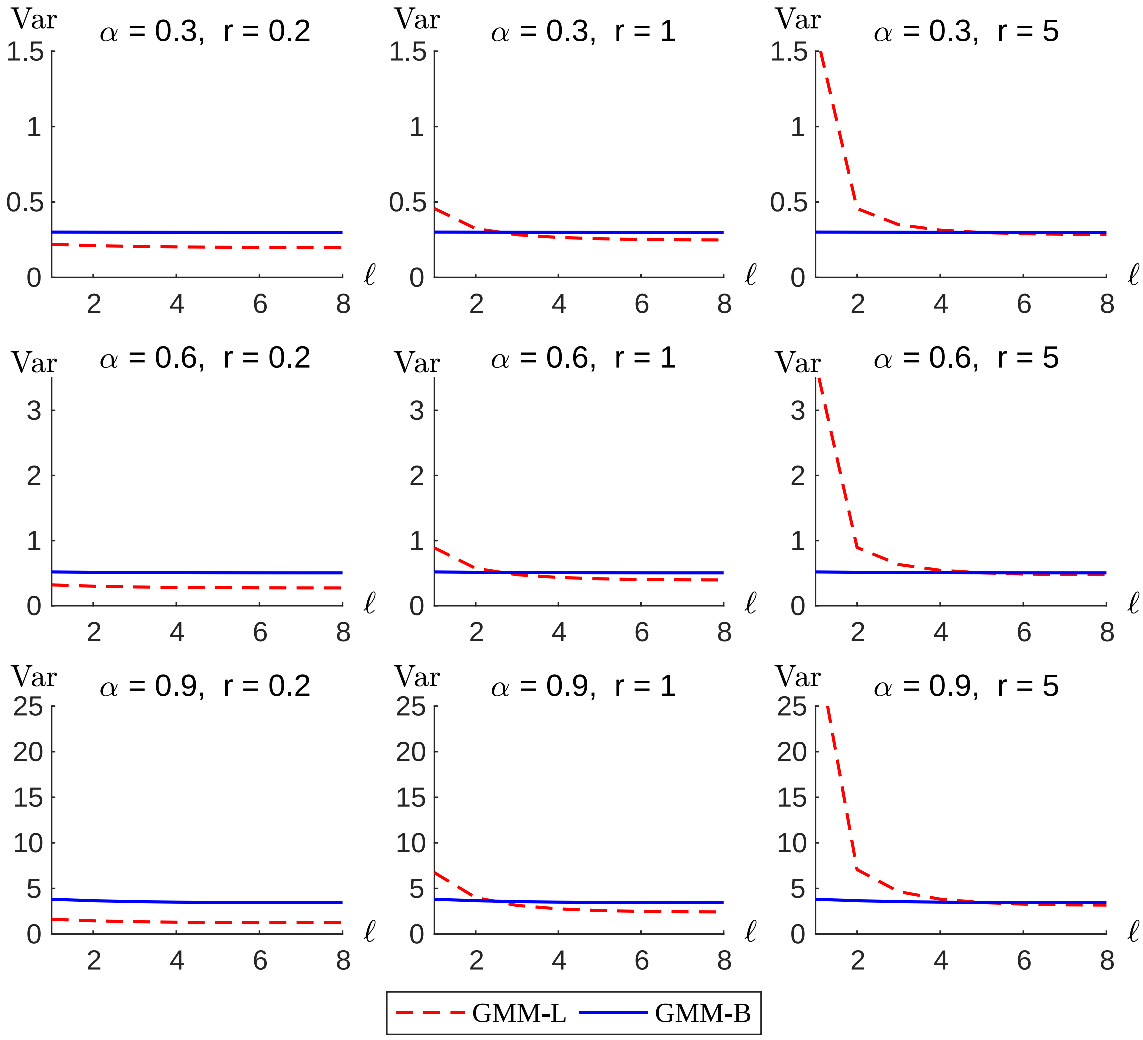


Figure 1: Asymptotic variance of GMM estimators with various instruments lag length ℓ ($T = 10$).

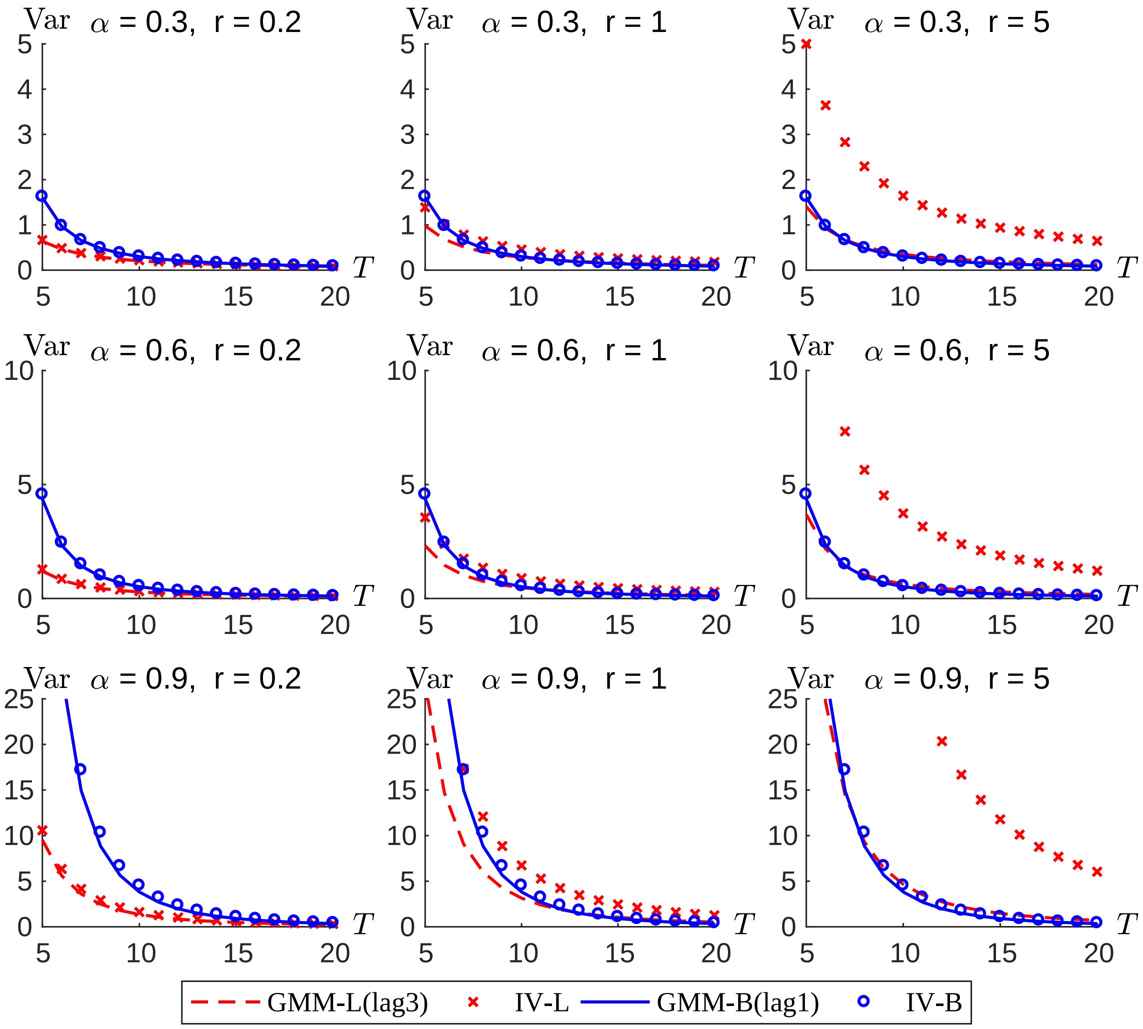


Figure 2: Asymptotic variance of IV/GMM estimators with various T.

Table 1: Fixed effects model: $\alpha = 0.4$, $\beta = 1.0$

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

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

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

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

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

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

Table 1(cont.): Fixed effects model: $\alpha = 0.4$, $\beta = 1.0$

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

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

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

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

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

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

Table 2: Fixed effects model: $\alpha = 0.8,\ \ \beta = 1.0$

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

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

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

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

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

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

Table 2(cont.): Fixed effects model: $\alpha = 0.8$, $\beta = 1.0$

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

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

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

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

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

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

Table 3: Trend model: $\alpha = 0.4$, $\beta = 1.0$,

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

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

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

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

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

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

Table 3(cont.): Trend model: $\alpha = 0.4,\ \ \beta = 1.0,$

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

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

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

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

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

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

Table 4: Trend model: $\alpha = 0.8,\ \ \beta = 1.0,$

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

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

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

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

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

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

Table 4(cont.): Trend model: $\alpha = 0.8,\ \ \beta = 1.0,$

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

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

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

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

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

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