Hi-Stat

Discussion Paper Series

No.245

On the Effect of Nonstationary Initial Conditions in Dynamic Panel Data Models

Kazuhiko Hayakawa

March 2008

Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences

A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
http://hi-stat.ier.hit-u.ac.jp/

On the Effect of Nonstationary Initial Conditions in Dynamic Panel Data Models*

Kazuhiko Hayakawa†
Department of Economics, Hitotsubashi University
JSPS Research Fellow

First Draft: January 2006
This version: March 2008

Abstract

In this paper, we consider dynamic panel data models with possibly nonstationary initial conditions. We derive the asymptotic properties of the GMM estimators with various kinds of instruments when both N and T are large, where N and T denote the dimensions of the cross section and time series. We find that when initial conditions are nonstationary and the degree of heterogeneity, which is measured by the variance ratio of individual effects to the disturbances, is large, the biases and variances of the GMM estimators become small. We demonstrate that this is because the correlation between the lagged dependent variable and instruments gets larger due to the unremoved individual effects. This implies that the instruments become strong when initial conditions are nonstationary and the degree of heterogeneity is large. For the purpose of comparison, we also derive the asymptotic properties of the within groups and the LIML estimators. Numerical studies are conducted to assess the properties of these estimators.

Keywords: Dynamic panel data models, many instruments, generalized method of moments estimator, nonstationary initial conditions, degree of heterogeneity.

JEL classification: C13, C23.

*This paper is a revised version of chapter three of my Ph.D dissertation submitted to Hitotsubashi University and previously circulated under the title "Efficient GMM Estimation of Dynamic Panel Data Models Where Large Heterogeneity May Be Present."

†I am deeply grateful to Taku Yamamoto, Satoru Kanoh, Katsuto Tanaka, Eiji Kurozumi, Ryo Okui, and seminar participants at Hitotsubashi University for helpful comments. I also acknowledge the financial support from the JSPS Fellowship. All remaining errors are mine. Corresponding author: Kazuhiko Hayakawa, E-mail : em031112@ yahoo.co.jp (Remove the space after @.)

1 Introduction

In cross-sectional data models, since the famous work of Angrist and Krueger (1991), the "many instruments" and "weak instruments" problems of the two-stage least squares (2SLS) estimator, which is a special case of the generalized method of moments (GMM) estimator, have been intensively discussed.[1] However, while there are many studies on the many/weak instruments problem in the context of cross-sectional data models, not much research has been conducted in the case of a dynamic panel data model even though this type of model faces the same problems.[2]

For the many instruments problem in dynamic panel data models, it is well known that one of the important features of dynamic panel data models is that the number of available instruments increases as T, the dimension of the time series, expands. Hence, when we use large T panel data, a large number of instruments become available. In such cases, it is suspected that the properties of estimators obtained under large N and fixed T asymptotics cannot explain the finite sample behavior well, creating the need to assess the estimators under large N and large T asymptotics.[3]

There are several studies that provide theoretical discussion of the asymptotic properties of estimators under large N and large T asymptotics.[4] One such study is that by Alvarez and Arellano (2003). They derived the asymptotic properties of the within groups (WG), the GMM, the limited information maximum likelihood (LIML) analog, the non-optimal first-difference GMM, and the random effect pseudo maximum likelihood (RML) estimators and showed that the WG, GMM, LIML, and RML estimators have a different order of asymptotic bias. Another study is that by Bun and Kiviet (2006), who derived the orders of the finite sample bias of several GMM estimators with various kinds of instruments.[5] Yet, another paper that discusses the many instruments problem in dynamic panel data models under large N and large T asymptotics is that by Okui (2005b). Based on Donald and Newey (2001) and Okui (2005a), this work develops a procedure to select the instruments in order to minimize the mean squared error (MSE) of the GMM estimator and improves the accuracy of inference. Finally, although not related to the many instruments problem, Hahn and Kuersteiner's (2002) study provides an important contribution. They derived the asymptotic distribution of the maximum likelihood or the WG estimator and proposed a bias-corrected WG estimator which corrects the bias of $O(T^{-1})$ without efficiency loss.

With regard to the weak instruments problem in dynamic panel data models, it is well known that the first difference GMM estimator of Holtz-Eakin, Newey and Rosen

[1] For recent studies, see Andrews and Stock (2006) and the papers cited therein.

[2] An analysis of the many instruments problem in the context of static panel data models with predetermined variables is provided by Ziliak (1997).

[3] Recent studies on dynamic panel data estimators under large N and fixed T asymptotics include Hsiao (2003), Arellano (2003a), and Baltagi (2005).

[4] As for simulation studies examining the finite sample properties of several dynamic panel estimators when both N and T are large, an example is Judson and Owen (1999).

[5] In fact, they consider an asymptotic expansion.

(1988) and Arellano and Bond (1991) suffers from the weak instruments problem when persistency is strong and/or the degree of heterogeneity is large (e.g. Blundell and Bond 1998; Blundell, Bond and Windmeijer 2000). Blundell and Bond (1998) therefore propose to use the system GMM estimator of Arellano and Bover (1995) which does not suffer from the weak instruments problem even when persistency is strong and is more efficient than the first difference GMM estimator. Because of these desirable properties, it is a common strategy in empirical studies to use the system GMM estimator to avoid the weak instruments problem and improve efficiency.[6]

However, we should note that all the results mentioned above are derived under the assumption of (mean) stationary initial conditions, which may not hold in practice.[7] Although it is well known that the initial conditions do not matter for long time series, this is not the case for panel data, the time series dimension of which is usually short. Hence, the treatment of initial conditions is an important issue in dynamic panel data models. Recent papers that discuss the initial condition problem are Arellano (2003) and Kiviet (2007).[8] Arellano (2003) provides a comprehensive discussion of the initial conditions problem, and one of the issues raised that is relevant to this paper is the discussion of the asymptotic bias of the inconsistent instrumental variables estimator derived from the invalid moment conditions.[9] Kiviet (2007), on the other hand, conducts a large scale Monte Carlo simulation for the GMM estimators with various kinds of weighting matrices when initial conditions are nonstationary.

This paper attempts to contribute to the literature on initial conditions in dynamic panel data models. Specifically, we relax the assumption used in Alvarez and Arellano (2003) to allow for nonstationary initial conditions, and derive the asymptotic properties of the GMM estimators using various kinds of instruments. Although this extension seems to be trivial, it is shown that there are significant differences in the properties of estimators. Indeed, we show that the strength of instruments is closely related to the assumption of initial conditions, and that the first difference GMM estimator does not always suffer from the weak instruments problem even when persistency is strong. In fact, we show that, in some cases, the instruments becomes strong. In particular, we investigate the relationship between the initial conditions and the degree of heterogeneity that is measured by the variance ratio of individual effects to the disturbances. Our focus is the effect of large heterogeneity on the performance of estimators, an issue that, as highlighted by Kiviet (2007), has been hardly discussed in the literature.[10]

We find that if the initial conditions are stationary, the GMM estimators with instruments in levels have large bias and variability when the degree of heterogeneity is

[6]Note that Bun and Windmeijer (2007) show that the system GMM estimator suffers from the weak instruments problem when the degree of heterogeneity is large.

[7]As is well known, the system GMM estimator is not consistent when initial conditions are nonstationary since the moment conditions are invalid.

[8]The first to consider initial conditions in dynamic panel data models is Anderson and Hsiao (1982).

[9]See Sections 6.4 and 6.5 in Arellano (2003).

[10]Some exceptions are Bun and Kiviet (2006) and Hayakawa (2007a), which theoretically discuss the relationship between the finite sample bias of several estimators and the degree of heterogeneity.

large, i.e., when the variance ratio of the individual effects to the disturbances is large. This result is consistent with the literature (see Bun and Kiviet (2006) and Hayakawa (2007)). However, in the case of nonstationary initial conditions, we show that the GMM estimators with instruments in levels have small bias and variability when the degree of heterogeneity is large. We demonstrate that when initial conditions are nonstationary, an additional correlation between the lagged dependent variable and instruments appears. Especially, we find that, when the degree of heterogeneity is large, the instruments become strong. Also, we find that, when the degree of heterogeneity is not so large, instruments may be weak, depending on initial conditions. For the GMM estimators with instruments in first difference or backward orthogonal deviation (BOD), they are not affected by the degree of heterogeneity when initial conditions are stationary, while their performance may improve when initial conditions are nonstationary and the degree of heterogeneity is large.

For the purpose of comparison, we also derive the asymptotic properties of the WG estimator and the LIML analog estimators with the same kinds of instruments as the GMM estimators. As a result, we find that although the WG estimator is not affected by the degree of heterogeneity in the case of stationary initial conditions, a large degree of heterogeneity helps the WG estimator to have a small bias. This feature causes a problem in the bias-corrected WG estimator of Hahn and Kuersteiner (2002). Since, as will be described, the bias corrected WG estimator has the correction term $1/T$, it is upwardly biased by construction when the initial conditions are nonstationary and the degree of heterogeneity is large. With regards to the LIML estimators, similar results are found as in the case of the GMM estimators. With stationary initial conditions, the performance of the LIML estimators with instruments in levels are negatively affected if the degree of heterogeneity is large. However, with nonstationary initial conditions, the bias and variability of the LIML estimators with instruments in levels become quite small when the degree of heterogeneity is large. We also find that the GMM and LIML estimators with instruments in backward orthogonal deviation have the same asymptotic properties.

The remainder of this paper is organized as follows. Section 2 introduces the model and the assumptions and defines the GMM estimator. Section 3 investigates the effect of the degree of heterogeneity and initial conditions on the GMM estimators with instruments in levels. Section 4 considers the removal of the individual effects from the instruments and derives the asymptotic properties of the GMM estimators. Section 5 derives the asymptotic properties of the WG and LIML estimators for the purpose of comparison. Section 6 then reports Monte Carlo simulation results to assess the theoretical implications. Finally, Section 7 provides some concluding remarks.

Note that throughout the paper, T_* denotes $T-1$ or $T-2$ when the range of a summation in estimators is $t = 1, ..., T-1$ or $t = 2, ..., T-1$, respectively. All the proofs of theorems are included in the appendix.

2 The model, the assumptions, and the GMM estimator

We consider an AR(1) panel data model given by

$$y_{it} = \alpha y_{i,t-1} + \eta_i + v_{it}, \qquad i = 1, ..., N \quad and \quad t = 1, ..., T \tag{1}$$

where α is the parameter of interest with $|\alpha| < 1$ and v_{it} has mean zero given $\eta_i, y_{i0}, ..., y_{i,t-1}$.[11] We assume that y_{i0} is observable. By letting $x_{it} = y_{i,t-1}$, $y_i = (y_{i,1}, ..., y_{i,T})'$, $x_i = (x_{i,1}, ..., x_{i,T})'$, $\iota_T = (1, ..., 1)'$ and $v_i = (v_{i,1}, ..., v_{i,T})$, (1) can be expressed in a vector form as follows:

$$y_i = \alpha x_i + \eta_i \iota_T + v_i. \tag{2}$$

To define the GMM estimator, let us define the forward orthogonal deviation (FOD) transformation matrix F as follows:

$$F = diag\left[\sqrt{\frac{T-1}{T}}, ..., \sqrt{\frac{1}{2}}\right] \begin{bmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}. \tag{3}$$

F is a $(T-1) \times T$ matrix such that $F'F = Q_T$ and $FF' = I_{T-1}$, where $Q_T = I_T - \iota_T \iota_T'/T$.

Premultiplying F in (2), we have

$$y_i^* = \alpha x_i^* + v_i^*, \tag{4}$$

where $y_i^* = Fy_i$, $x_i^* = Fx_i$, and $v_i^* = Fv_i$. The t-th element of v_i^* is given by

$$v_{it}^* = c_t \left[v_{i,t} - \frac{1}{T-t}(v_{i,t+1} + \cdots + v_{i,T}) \right], \qquad t = 1, ..., T-1 \tag{5}$$

where $c_t^2 = (T-t)/(T-t+1)$.

Next, we define the GMM estimator. Let z_{it} be a generic instruments vector that is orthogonal to v_{it}^*. Then, the GMM estimator can be written as[12]

$$\widehat{\alpha}_G = \frac{x^{*\prime} M y^*}{x^{*\prime} M x^*} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} M_t y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} M_t x_t^*} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} M_t x_t^* \cdot \widehat{\alpha}_{2SLS,t}}{\sum_{t=1}^{T-1} x_t^{*\prime} M_t x_t^*}, \tag{6}$$

where $x_t^* = (x_{1t}^*, ..., x_{Nt}^*)'$, $y_t^* = (y_{1t}^*, ..., y_{Nt}^*)'$, $M_t = Z_t(Z_t'Z_t)^{-1}Z_t'$, $Z_t = (z_{1t}, ..., z_{Nt})'$, and $\widehat{\alpha}_{2SLS,t} = (x_t^{*\prime} M_t y_t^*)/(x_t^{*\prime} M_t x_t^*)$.[13] Note that $\widehat{\alpha}_G$ can be written as a weighted sum of the

[11] Here, we limit ourselves to a simple stable AR(1) model. Possible extensions are discussed in the conclusion.

[12] We do not employ the level (Arellano and Bover, 1995) or the system (Arellano and Bover 1995; Blundell and Bond 1998) GMM estimators because these GMM estimators are known to be inconsistent under large N and fixed T asymptotics when initial conditions are nonstationary.

[13] When the invertibility condition of $Z'Z$, i.e., $N \geq T-1$, does not hold, we can use the Moore-Penrose inverse. For a detailed discussion of this problem, see Alvarez and Arellano (2003).

cross section 2SLS estimator at time t, $\widehat{\alpha}_{2SLS,t}$. This implies that the properties of $\widehat{\alpha}_G$ are closely related to those of $\widehat{\alpha}_{2SLS,t}$.

To derive the asymptotic properties of the GMM estimator, we impose the following assumptions that relax those of Alvarez and Arellano (2003):

Assumption 1. $\{v_{it}\}$ $(t = 1, ..., T; i = 1, ..., N)$ *are i.i.d across time and individuals and independent of* η_i *and* y_{i0} *with* $E(v_{it}) = 0$, $var(v_{it}) = \sigma_v^2$, *and finite moments up to the fourth order.*

Assumption 2. *The initial observations satisfy*

$$\begin{aligned} y_{i0} &= \delta \left(\frac{\eta_i}{1-\alpha} \right) + w_{i0} \qquad\qquad for \quad i = 1, ..., N & (7) \\ &= \delta \mu_i + w_{i0}, & (8) \end{aligned}$$

where w_{i0} *is* $w_{i0} = \sum_{j=0}^{\infty} \alpha^j v_{i,-j}$ *and is independent of* η_i, *and* $\mu_i = \eta_i/(1-\alpha)$.

Assumption 3. η_i *are i.i.d across individuals with* $E(\eta_i) = 0$, $var(\eta_i) = \sigma_\eta^2$, *and finite moments up to the fourth order.*

Assumptions 1 and 3 are identical to those of Alvarez and Arellano (2003). Assumption 2 allows y_{it} to be nonstationary in the sense that the conditional mean of y_{it} given η_i depends on t.[14] In fact, under Assumption 2, y_{it} can be expressed as

$$\begin{aligned} y_{it} &= \left[1 - (1-\delta)\alpha^t \right] \mu_i + w_{it} & (9) \\ &= \mu_{it}^* + w_{it}, & (10) \end{aligned}$$

where $w_{it} = \sum_{j=0}^{\infty} \alpha^j v_{i,t-j}$ and $\mu_{it}^* = \left[1 - (1-\delta)\alpha^t \right] \mu_i$. The conditional mean of y_{it} given η_i is

$$E(y_{it}|\eta_i) = \frac{1-(1-\delta)\alpha^t}{1-\alpha} \eta_i. \qquad (11)$$

Thus, we find that when $\delta \neq 1$, y_{it} is nonstationary due to the dependence on t, and when $\delta = 1$, y_{it} is stationary. Therefore, we extend Alvarez and Arellano (2003) to allow nonstationary initial conditions.

This extension to allow for nonstationary initial conditions has important implications for empirical analyses. For example, when we consider a cross-country panel data set that begins after a war or another large historical event, it is unlikely that initial conditions are distributed according to the steady state (Barro and Sala-i-Martin, 1995). Another example of nonstationary initial conditions is young workers or new firms, for whom initial conditions have little relation to steady state conditions (Hause, 1980).

[14] Note that this type of initial conditions is also used by Arellano (2003) and Kiviet (2007).

3 The asymptotic properties of GMM estimators with instruments in levels

3.1 The GMM estimator with all instruments in levels

Let us define $\widehat{\alpha}_{G,l1}$ as the GMM estimator with instruments $z_{it} = z_{it}^{l1} = (y_{i0}, ..., y_{i,t-1})'$ as follows:

$$\widehat{\alpha}_{G,l1} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l1} y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l1} x_t^*}, \tag{12}$$

where $M_t^{l1} = Z_t^{l1}(Z_t^{l1\prime} Z_t^{l1})^{-1} Z_t^{l1\prime}$ and $Z_t^{l1} = (z_{1t}^{l1}, ..., z_{Nt}^{l1})'$.

The following lemma is used to derive the asymptotic properties of $\widehat{\alpha}_{G,l1}$:

Lemma 1. *Let Assumptions 1–3 hold. Then, we have*

$$(a) \quad \frac{1}{\sqrt{NT_*}} \sum_{t=1}^{T-1} E(x_t^{*\prime} M_t^{l1} v_t^*) = -\frac{T}{\sqrt{NT_*}} \frac{\sigma_v^2}{1-\alpha}\left[1 - \frac{1}{T(1-\alpha)} \sum_{t=1}^{T} \frac{1-\alpha^t}{t}\right] = \mu_{G,l1},$$

$$(b) \quad \frac{1}{NT_*} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l1} x_t^* \xrightarrow[N\to\infty]{p} R_T^{G,l1}.$$

Moreover, as both N and T tend to infinity, provided $(\log T)^2/N \to 0$, we have

$$(c) \quad var\left(\frac{1}{\sqrt{NT_*}} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l1} v_t^*\right) \xrightarrow[N,T\to\infty]{} \frac{\sigma_v^4}{1-\alpha^2},$$

$$(d) \quad \frac{1}{NT_*} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l1} x_t^* \xrightarrow[N,T\to\infty]{p} \left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

where

$$\begin{aligned}
R_T^{G,l1} &= \frac{\sigma_v^2}{T_*} \sum_{t=1}^{T-1} \psi_t^2 \left[\frac{1}{1-\alpha^2} - 2\lambda(1-\delta)\alpha^{t-1}[1-(1-\delta)\alpha^{t-1}] + \lambda^2(1-\delta)^2\alpha^{2(t-1)} q_t \right. \\
&\qquad \left. - \frac{\lambda}{1+\lambda q_t}\left\{[1-(1-\delta)\alpha^{t-1}] - \lambda(1-\delta)\alpha^{t-1} q_t\right\}^2\right], \\
\sigma_\mu^2 &= var(\mu_i) = \frac{\sigma_\eta^2}{(1-\alpha)^2}, \\
\lambda &= \frac{\sigma_\mu^2}{\sigma_v^2} = \frac{1}{(1-\alpha)^2}\frac{\sigma_\eta^2}{\sigma_v^2}, \\
q_t &= 1-\alpha^2 + (t-1)(1-\alpha)^2 - (1-\delta^2)(1-\alpha^2), \\
\psi_t &= c_t\left(1 - \frac{\alpha\phi_{T-t}}{T-t}\right), \\
\phi_j &= \frac{1-\alpha^j}{1-\alpha}.
\end{aligned}$$

Theorem 1. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity, provided $(\log T)^2/N \to 0$, we have*

$$\widehat{\alpha}_{G,l1} \xrightarrow{p} \alpha. \tag{13}$$

Moreover, provided $T/N \to c$, $(0 \le c < \infty)$, we have

$$\sqrt{NT_*}(\widehat{\alpha}_{G,l1} - \alpha - B_{G,l1}) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \tag{14}$$

where

$$B_{G,l1} = \frac{1}{\sqrt{NT_*}} \frac{\mu_{G,l1}}{R_T^{G,l1}} = O\left(\frac{1}{N}\right). \tag{15}$$

In relation to the above, we provide the following remarks.

Remark 1. Although, as is well known, the initial conditions do not matter when T is large, this is not the case here, since $B_{G,l1}$ can be seen as a finite sample bias which is naturally affected by the initial conditions.

Remark 2. Hahn and Kuersteiner (2002) show that if we further assume normality on v_{it}, then $\mathcal{N}(0, 1-\alpha^2)$ is the minimal asymptotic distribution. Hence, $(1-\alpha^2)$ is the lower bound of the asymptotic variance when both N and T are large under the assumption of normality on v_{it}.

Remark 3. In the case of $\delta = 1$, we find that $R_T^{G,l1}$ is bounded when $\lambda \to \infty$.[15] This result is related to the evaluation of the asymptotic bias and variance. From (15), we find that the asymptotic bias is bounded, i.e., $0 < B_{G,l1} < \infty$ for any λ, including zero and infinity. Although the asymptotic variance under large N and large T asymptotics is not affected by the degree of heterogeneity, that under large N and fixed T asymptotics, given by $\sigma_v^2/R_T^{G,l1}$, depends on the degree of heterogeneity despite it being bounded even when λ is large.

Further, using Lemma 1(d), (14) can be alternatively expressed as follows:

$$\sqrt{NT_*}\left[\widehat{\alpha}_{G,l1} - \left(\alpha - \frac{1}{N}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}\left(0, 1-\alpha^2\right). \tag{16}$$

This is the result derived by Alvarez and Arellano (2003).

Remark 4. In the case of $\delta \neq 1$, we find that for given N and T, as $\lambda \to \infty$, $R_T^{G,l1} \to \infty$, $B_{l1} \to 0$. Therefore, large heterogeneity makes the GMM estimator, $\widehat{\alpha}_{G,l1}$, to have a small bias under nonstationary initial conditions. Note that this result is in conflict with the one obtained by Bun and Kiviet (2006), who imposed mean stationarity. Further, note that the asymptotic variance under large N and large T asymptotics is not affected by the degree of heterogeneity, unlike that under large N and fixed T asymptotics, which is given by $\sigma_v^2/R_T^{G,l1}$. In fact, we find that the asymptotic variance under large N and fixed T asymptotics tends to zero as $\lambda \to \infty$. This difference arises since the terms associated with λ vanish asymptotically as $T \to \infty$.

[15] Although λ is finite in practice, we use this notation to indicate that λ becomes large.

Remark 5. The intuition behind the result of Remark 3 is that when initial conditions are nonstationary, instruments become strong as λ gets larger. To see this, consider the following cross section 2SLS regression at time t:

$$
\begin{aligned}
y_{it}^* &= \alpha x_{it}^* + v_{it}^* \qquad\qquad i = 1, ..., N &(17)\\
x_{it}^* &= \pi_t' z_{it}^{l1} + \varepsilon_{it} &(18)
\end{aligned}
$$

As is well known, the correlation between x_{it}^* and z_{it}^{l1} plays a very important role for the performance of the 2SLS estimator $\widehat{\alpha}_{2SLS,t}^{l1} = x_t^{*\prime} M_t^{l1} y_t^* / x_t^{*\prime} M_t^{l1} x_t^*$. Since x_{it}^* can be written as

$$
x_{it}^* = \psi_t \left[w_{i,t-1} - (1-\delta)\alpha^{t-1}\mu_i \right] - c_t \tilde{v}_{itT} \tag{19}
$$

where

$$
\begin{aligned}
\psi_t &= c_t \left(1 - \frac{\alpha \phi_{T-t}}{T-t} \right), &(20)\\
\tilde{v}_{itT} &= \frac{\phi_{T-t} v_{it} + \cdots + \phi_1 v_{i,T-1}}{T-t}, &(21)\\
\phi_j &= \frac{1-\alpha^j}{1-\alpha} = 1 + \alpha + \cdots + \alpha^{j-1}, &(22)
\end{aligned}
$$

we find that the individual effect μ_i is removed from x_{it}^* when $\delta = 1$, while this is not the case when $\delta \neq 1$. Using (19), we have

$$
\begin{aligned}
E(x_{it}^* z_{it}^{l1}) &= \psi_t E[w_{i,t-1} z_{it}^{l1}] - (1-\delta)\psi_t \alpha^{t-1} E[\mu_i z_{it}^{l1}] &(23)\\
&= \text{“idiosyncratic part”} + \text{“individual effects part”} &(24)
\end{aligned}
$$

From (23), we find that the correlation between x_{it}^* and z_{it}^{l1} is composed of only the idiosyncratic term when $\delta = 1$, while it is composed of the “idiosyncratic part” *and* the “individual effects part” when $\delta \neq 1$. This implies that nonstationary initial conditions provide an additional correlation between x_{it}^* and z_{it}^{l1} through individual effects. However, we have to investigate the “individual effects part” carefully since it can be negative, while the “idiosyncratic part” is always positive. When $\delta > 1$, since the “individual effects part” is always positive, $E[x_{it}^* z_{it}^{l1}]$ gets larger as σ_μ^2 grows. However, when $\delta < 1$, $E(x_{it}^* z_{it}^{l1})$ might be close to zero since the “idiosyncratic part” is positive while the “individual effects part” is negative. In this case, the instruments may be weak. However, if λ is large enough, the “individual effects part” becomes much smaller than the “idiosyncratic part” and $E(x_{it}^* z_{it}^{l1})$ can be large in absolute value. Therefore, when λ is large, the instruments become strong regardless of $\delta > 1$ or $\delta < 1$.[16]

Finally, although these properties are obtained for the cross section 2SLS estimator at time t, $\widehat{\alpha}_{2SLS,t}^{l1}$, similar properties will carry over to $\widehat{\alpha}_{G,l1}$, since these properties do not depend on time t and $\widehat{\alpha}_{G,l1}$ is a weighted sum of cross section 2SLS estimators as in (6).

[16] Although, following Bun and Windmeijer (2007), this statement can be explained by deriving the concentration parameter that measures the strength of instruments, we do not report the results to save space.

Although $\widehat{\alpha}_{G,l1}$ has some desirable properties, that is, consistency and asymptotic normality, as shown in Okui (2005b) and Section 6 below, the size distortion of the test for the hypothesis $H_0 : \alpha = \alpha_o$, where α_o is the true value, is very large. We suspect that the source of the size distortion is the bias that results from using all the instruments. Therefore, we expect that reducing the number of instruments will mitigate this problem because as per the literature on cross-sectional data models, using fewer instruments reduces the bias of the estimator.

3.2 The GMM estimator with a reduced number of instruments in levels

Let us define $\widehat{\alpha}_{G,l2}$ as the GMM estimator with instruments $z_{it} = z_{it}^{l2} = y_{i,t-1}$ as follows:

$$\widehat{\alpha}_{G,l2} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l2} y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l2} x_t^*}, \tag{25}$$

where $M_t^{l2} = Z_t^{l2}(Z_t^{l2\prime} Z_t^{l2})^{-1} Z_t^{l2\prime}$ and $Z_t^{l2} = (z_{1t}^{l2}, ..., z_{Nt}^{l2})'$.

Lemma 2. *Let Assumptions 1–3 hold. Then, we have*

$$(a) \quad \frac{1}{\sqrt{NT_*}} \sum_{t=1}^{T-1} E(x_t^{*\prime} M_t^{l2} v_t^*) = -\frac{1}{\sqrt{NT_*}} \frac{\sigma_v^2}{1-\alpha} \left(1 - \frac{\phi_{T-1}}{T-1}\right)$$
$$= \mu_{G,l2},$$

$$(b) \quad \frac{1}{NT_*} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l2} x_t^* \xrightarrow[N \to \infty]{p} R_T^{G,l2}.$$

Moreover, as both N and T tend to infinity,

$$(c) \quad var\left(\frac{1}{\sqrt{NT_*}} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l2} v_t^*\right) \xrightarrow[N,T \to \infty]{} \rho_{l2}\left(\frac{\sigma_v^4}{1-\alpha^2}\right),$$

$$(d) \quad \frac{1}{NT_*} \sum_{t=1}^{T-1} x_t^{*\prime} M_t^{l2} x_t^* \xrightarrow[N,T \to \infty]{} \rho_{l2}\left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

where

$$R_T^{G,l2} = \frac{\sigma_v^2}{T_*} \sum_{t=1}^{T-1} \psi_t^2 \frac{\left[\frac{1}{1-\alpha^2} - \lambda(1-\delta)\alpha^{t-1}\left\{1-(1-\delta)\alpha^{t-1}\right\}\right]^2}{\frac{1}{1-\alpha^2} + \lambda\left[1-(1-\delta)\alpha^{t-1}\right]^2},$$
$$\rho_{l2} = \left[1 + \lambda(1-\alpha^2)\right]^{-1}.$$

Remark 6. In the case of $\delta \neq 1$, there is a substantial difference in the convergence of $x^{*\prime} M^{l2} x^* / NT$. Under large N and fixed T asymptotics, $R_T^{G,l2} \to \infty$ as $\lambda \to \infty$. However, under large N and large T asymptotics, $\rho_{l2}\sigma_v^2/(1-\alpha^2) \to 0$ as $\lambda \to \infty$. As shown below, this difference plays an important role in assessing the asymptotic variance.

Theorem 2. *Let Assumptions 1–3 hold. Then, we have*

$$\widehat{\alpha}_{G,l2} \xrightarrow{p} \alpha,$$

$$\sqrt{NT_*}(\widehat{\alpha}_{G,l2} - \alpha - B_{G,l2}) \xrightarrow{d} N\left(0, (1-\alpha^2)\rho_{l2}^{-1}\right), \tag{26}$$

where

$$B_{G,l2} = \frac{1}{\sqrt{NT_*}} \frac{\mu_{G,l2}}{R_T^{G,l2}} = O\left(\frac{1}{NT}\right).$$

Remark 7. Note that the conditions $(\log T)^2/N \to 0$ and $T/N \to c$, $(0 \leq c < \infty)$, which are imposed in Theorem 1, are unnecessary in Theorem 2. This is because the number of instruments z_{it}^{l2} grows at rate T, not T^2 like z_{it}^{l1}, and we do not have terms of order $(\log N)^2/T$ and T/N. This implies that we do not need to impose conditions on the relative speed of N and T. This is also true for the two cases that will be discussed in the next section.

Remark 8. In the case of $\delta = 1$, as $\lambda \to \infty$, $R_T^{G,l2} \to 0$ for given N and T; this indicates that as $\lambda \to \infty$, $B_{G,l2} \to \infty$. The asymptotic variance also becomes substantial when λ is large. This indicates that if we use a smaller number of instruments to reduce the bias arising from the use of many instruments, then a bias due to large heterogeneity may appear. To examine this more precisely, we derive the following alternative expression of (26), using Lemma 2(d):

$$\sqrt{NT_*}\left[\widehat{\alpha}_{G,l2} - \left(\alpha - \frac{1}{NT_*}(1+\alpha)\rho_{l2}^{-1}\right)\right] \to \mathcal{N}\left(0, (1-\alpha^2)\rho_{l2}^{-1}\right). \tag{27}$$

Comparing (16) and (27), we find that if $T_* < \rho_{l2}^{-1}$, the asymptotic bias of $\widehat{\alpha}_{G,l2}$ will be larger than that of $\widehat{\alpha}_{G,l1}$, although $\widehat{\alpha}_{G,l2}$ uses a smaller number of instruments than $\widehat{\alpha}_{G,l1}$. Hence, if a large degree of heterogeneity is present, reducing the number of instruments to decrease the bias may not work well. Furthermore, the asymptotic variance becomes quite large.

Remark 9. In the case of $\delta \neq 1$, as $\lambda \to \infty$, $R_T^{G,l2} \to \infty$; this indicates that, for given N and T, $B_{G,l2} \to 0$ as $\lambda \to \infty$. The intuitive reason for this is similar to the case of $\widehat{\alpha}_{G,l1}$.[17] Further, we find that the asymptotic variance under large N and large T asymptotics tends to infinity. However, the asymptotic variance under large N and fixed T asymptotics, given by $\sigma_v^2/R_T^{G,l2}$, tends to zero when $\lambda \to \infty$. This difference stems from the fact that the terms associated with λ vanish asymptotically when $T \to \infty$ as in the case of $\widehat{\alpha}_{G,l1}$. From the simulation studies in Okui (2005b), we expect that the asymptotic variance under large N and fixed T asymptotics captures the finite sample behavior more closely than the variance under large N and large T asymptotics.

4 Removing the individual effects from the instruments

Since the asymptotic distribution of $\widehat{\alpha}_{G,l2}$ under stationary initial conditions is heavily affected by the degree of heterogeneity arising from the instruments, we expect that if we

[17] See Remark 5.

use the instruments without the individual effects, the GMM estimator will not be affected by the degree of heterogeneity. Therefore, we consider the removal of the individual effects from the instruments. We employ two methods to do so. The first involves simply taking the first-difference. In the second method, we use a transformation known as the backward orthogonal deviation (BOD) transformation. The BOD transformation is a modification of the FOD transformation. While the FOD transformation induces a deviation from the mean of all future values, the BOD transformation induces a deviation from the mean of all past values. To rid the instruments of the individual effects, we only have to multiply the following matrix by x_i:

$$B = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & 1 & 0 \\ -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & 1 \end{bmatrix}. \tag{28}$$

By multiplying above matrix by x_i, we get the following expression:

$$y^b_{i,t-1} = y_{i,t-1} - \frac{y_{i,0} + \cdots + y_{i,t-2}}{t-1} \qquad t = 2, ..., T-1 \tag{29}$$

$$= \left(w_{i,t-1} - \frac{w_{i,t-2} + \cdots + w_{i,0}}{t-1} \right) - (1-\delta) \left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1} \right) \mu_i. \tag{30}$$

Thus, under the assumption of stationary initial conditions, i.e., $\delta = 1$, $y^b_{i,t-1}$ no longer has individual effects. This method is known as the recursive mean adjustment method in the context of a pure time series model (So and Shin, 1999), and is considered in Hayakawa (2007b) in the context of the instrumental variable (IV) estimation of panel AR(p) models.[18]

The GMM estimators with instruments $z_{it} = z^{d2}_{it} = \Delta y_{i,t-1}$ and $z_{it} = z^{b2}_{it} = y^b_{i,t-1}$ are defined as follows:[19]

$$\widehat{\alpha}_{G,d2} = \frac{\sum_{t=2}^{T-2} x^{*\prime}_t M^{d2}_t y^*_t}{\sum_{t=2}^{T-2} x^{*\prime}_t M^{d2}_t x^*_t}, \tag{31}$$

$$\widehat{\alpha}_{G,b2} = \frac{\sum_{t=2}^{T-2} x^{*\prime}_t M^{b2}_t y^*_t}{\sum_{t=2}^{T-2} x^{*\prime}_t M^{b2}_t x^*_t}, \tag{32}$$

where $M^{d2}_t = Z^{d2}_t (Z^{d2\prime}_t Z^{d2}_t)^{-1} Z^{d2\prime}_t$, $Z^{d2}_t = (z^{d2}_{1t}, ..., z^{d2}_{Nt})'$, $M^{b2}_t = Z^{b2}_t (Z^{b2\prime}_t Z^{b2}_t)^{-1} Z^{b2\prime}_t$, and $Z^{b2}_t = (z^{b2}_{1t}, ..., z^{b2}_{Nt})'$.

Asymptotic properties of these two estimators are given in the following lemmas and theorems.

[18] Hayakawa (2007b) shows that $y^b_{i,t-1}$ is asymptotically equivalent to the infeasible optimal instruments, and the IV estimator using $y^b_{i,t-1}$ as instruments has the same asymptotic distribution as the infeasible optimal IV estimator when both N and T are large.

[19] We do not consider the case where all instruments are used since it is suspected that inference would be unreliable as for the case of $\widehat{\alpha}_{G,l1}$.

Lemma 3. *Let Assumptions 1–3 hold. Then, we have*

$$(a) \quad \frac{1}{\sqrt{NT_*}} \sum_{t=2}^{T-1} E(x_t^{*\prime} M_t^{d2} v_t^*) = -\frac{1}{\sqrt{NT_*}} \frac{\sigma_v^2}{1-\alpha} \left(1 - \frac{\phi_{T-2}}{T-2}\right) = \mu_{G,d2},$$

$$(b) \quad \frac{1}{NT_*} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{d2} x_t^* \xrightarrow[N \to \infty]{p} R_T^{G,l2}.$$

Moreover, as both N and T tend to infinity,

$$(c) \quad var\left(\frac{1}{\sqrt{NT_*}} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{d2} v_t^*\right) \xrightarrow[N,T \to \infty]{} \rho_{d2} \left(\frac{\sigma_v^4}{1-\alpha^2}\right),$$

$$(d) \quad \frac{1}{NT_*} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{d2} x_t^* \xrightarrow[N,T \to \infty]{p} \rho_{d2} \left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

where

$$R_T^{G,d2} = \frac{\sigma_v^2}{T_*} \sum_{t=2}^{T-1} \psi_t^2 \frac{\left[\frac{1}{1+\alpha} - \lambda(1-\delta)^2(1-\alpha)\alpha^{2t-3}\right]^2}{\frac{2}{1+\alpha} + \lambda(1-\delta)^2(1-\alpha^2)\alpha^{2(t-2)}}, \tag{33}$$

$$\rho_{d2} = \frac{1-\alpha}{2}. \tag{34}$$

Theorem 3. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity, we have*

$$\widehat{\alpha}_{G,d2} \xrightarrow{p} \alpha, \tag{35}$$

$$\sqrt{NT_*}(\widehat{\alpha}_{G,d2} - \alpha - B_{G,d2}) \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\rho_{d2}^{-1}\right), \tag{36}$$

where

$$B_{G,d2} = \frac{1}{\sqrt{NT_*}} \frac{\mu_{G,d2}}{R_T^{d2}} = O\left(\frac{1}{NT}\right). \tag{37}$$

Lemma 4. *Let Assumptions 1–3 hold. Then, we have*

$$(a) \quad \frac{1}{\sqrt{NT_*}} \sum_{t=2}^{T-1} E(x_t^{*\prime} M_t^{b2} v_t^*) = \mu_{G,d2} = \mu_{G,b2},$$

$$(b) \quad \frac{1}{NT_*} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{b2} x_t^* \xrightarrow[N \to \infty]{p} R_T^{G,b2}.$$

Moreover, as both N and T tend to infinity,

$$(c) \quad var\left(\frac{1}{\sqrt{NT_*}} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{b2} v_t^*\right) \xrightarrow[N,T \to \infty]{} \frac{\sigma_v^4}{1-\alpha^2},$$

$$(d) \quad \frac{1}{NT_*} \sum_{t=2}^{T-1} x_t^{*\prime} M_t^{b2} x_t^* \xrightarrow[N,T \to \infty]{p} \left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

where

$$R_T^{G,b2} = \frac{\sigma_v^2}{T_*} \sum_{t=2}^{T-1} \psi_t^2 \frac{\left[\frac{1}{1-\alpha^2}\left(1 - \frac{\alpha\phi_{t-1}}{t-1}\right) + \lambda(1-\delta)^2\alpha^{t-1}\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)\right]^2}{\frac{1}{1-\alpha^2}\left[1 - \frac{2\alpha\phi_{t-1}}{t-1} + \frac{1}{(t-1)^2}\left\{\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}\right] + \lambda(1-\delta)^2\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)^2}.$$

Theorem 4. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity, we have*

$$\widehat{\alpha}_{G,b2} \xrightarrow{p} \alpha, \tag{38}$$

$$\sqrt{NT_*}(\widehat{\alpha}_{G,b2} - \alpha - B_{G,b2}) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \tag{39}$$

where

$$B_{G,b2} = \frac{1}{\sqrt{NT_*}} \frac{\mu_{G,b2}}{R_T^{G,b2}} = O\left(\frac{1}{NT}\right). \tag{40}$$

Remark 10. When $\delta = 1$, we find that $R_T^{G,d2}$, $R_T^{G,b2}$, $B_{G,d2}$, and $B_{G,b2}$ are not affected by λ. Using Lemmas 3(d) and 4(d), we have

$$\sqrt{NT_*}\left[\widehat{\alpha}_{G,d2} - \left(\alpha - \frac{1}{NT_*}(1+\alpha)\rho_{d2}^{-1}\right)\right] \xrightarrow{d} \mathcal{N}\left(0, (1-\alpha^2)\rho_{d2}^{-1}\right), \tag{41}$$

$$\sqrt{NT_*}\left[\widehat{\alpha}_{G,b2} - \left(\alpha - \frac{1}{NT_*}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}\left(0, 1-\alpha^2\right). \tag{42}$$

We find that the asymptotic biases and variances of $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$ are not affected by the degree of heterogeneity. However, there is a notable difference both in the asymptotic biases and in the variances of $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$. Since ρ_{d2}^{-1} is strictly larger than one, both the asymptotic bias and variance of $\widehat{\alpha}_{G,d2}$ are strictly larger than those of $\widehat{\alpha}_{G,b2}$. Therefore, we can state that $\widehat{\alpha}_{G,b2}$ is superior to $\widehat{\alpha}_{G,d2}$. Furthermore, the asymptotic variance of $\widehat{\alpha}_{G,d2}$ is strictly larger than the lower bound and can never be efficient. However, the asymptotic variance of $\widehat{\alpha}_{G,b2}$ is equal to the lower bound, and hence $\widehat{\alpha}_{G,b2}$ is asymptotically efficient when v_{it} is normally distributed. However, it is noteworthy that although $\widehat{\alpha}_{G,l1}$ becomes asymptotically efficient by using all instruments, $\widehat{\alpha}_{G,b2}$ is asymptotically efficient by using a smaller number of instruments. This implies that instruments that are not used, i.e., $(y_{i,2}^b, ..., y_{i,t-2}^b)$, are asymptotically redundant.

Remark 11. When $\delta \neq 1$, as $\lambda \to \infty$, $R_T^{G,d2}, R_T^{G,b2} \to \infty$. This indicates that as $\lambda \to \infty$, $B_{G,d2}, B_{G,b2} \to 0$ for given N and T. The intuition behind this result is similar to the case of $\widehat{\alpha}_{G,l1}$. Further, note that the asymptotic variances of $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$ under large N and large T asymptotics are not affected by the degree of heterogeneity, unlike those under large N and fixed T asymptotics. In fact, the asymptotic variances under large N and fixed T asymptotics, given by $\sigma_v^2/R_T^{G,d2}$ and $\sigma_v^2/R_T^{G,b2}$, tend to zero as $\lambda \to \infty$.

5 A comparison of the GMM, WG and LIML estimators

In this section, we derive the asymptotic properties of the WG and LIML (analog) estimators with possibly nonstationary initial conditions and investigate whether the results are similar to those of the GMM estimators.

5.1 The (bias-corrected) within groups estimator

The WG estimator is the OLS estimator of (4) and can be written as follows:

$$\widehat{\alpha}_{wg} = \frac{x^{*\prime} y^*}{x^{*\prime} x^*} = \frac{\sum_{i=1}^N x_i' Q_T y_i}{\sum_{i=1}^N x_i' Q_T x_i}, \tag{43}$$

where $x^* = (x_1^{*\prime}, ..., x_N^{*\prime})'$ and $y^* = (y_1^{*\prime}, ..., y_N^{*\prime})'$.

Hahn and Kuersteiner (2002) proposed a bias-corrected WG estimator of the form

$$\widehat{\alpha}_{hk} = \frac{T+1}{T}\widehat{\alpha}_{wg} + \frac{1}{T}. \tag{44}$$

Note that $\widehat{\alpha}_{hk}$ corrects for the bias of order T^{-1}.

To derive the asymptotic properties of the WG estimator, the following lemma is useful.

Lemma 5. *Let Assumptions 1–3 hold. Then, we have*

$$(a) \quad E\left(\frac{x^{*\prime} x^*}{NT}\right) = \frac{\sigma_v^2}{1-\alpha^2} - \frac{1}{T}\frac{\sigma_v^2}{1-\alpha^2}\left[\frac{1+\alpha}{1-\alpha} - \frac{1}{T}\frac{2\alpha(1-\alpha^T)}{(1-\alpha)^2}\right] + \frac{\sigma_\mu^2(1-\delta)^2}{T}\left[\frac{1-\alpha^{2T}}{1-\alpha^2} - \frac{1}{T}\left(\frac{1-\alpha^T}{1-\alpha}\right)^2\right] \tag{45}$$

$$= R_T^{wg} \xrightarrow[T\to\infty]{} \frac{\sigma_v^2}{1-\alpha^2}, \tag{46}$$

$$(b) \quad E\left(\frac{x^{*\prime} v^*}{NT}\right) = -\frac{1}{T}\left(\frac{\sigma_v^2}{1-\alpha}\right)\left[1 - \frac{1}{T}\left(\frac{1-\alpha^T}{1-\alpha}\right)\right],$$

$$(c) \quad var\left(\frac{1}{\sqrt{NT}} x^{*\prime} v^*\right) \xrightarrow[T\to\infty]{} \frac{\sigma_v^4}{1-\alpha^2}.$$

The following theorem establishes the asymptotic properties of the WG estimator with possibly nonstationary initial conditions.

Theorem 5. *Let Assumptions 1–3 hold. Then, as $T \to \infty$, regardless of whether N is fixed or tends to infinity, we have*

$$\widehat{\alpha}_{wg} \xrightarrow{p} \alpha,$$

$$\sqrt{NT}(\widehat{\alpha}_{wg} - \alpha - B_{wg}) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2),$$

where

$$B_{wg} = -\frac{\frac{1+\alpha}{T} - \frac{1}{T^2}\frac{(1+\alpha)(1-\alpha^T)}{1-\alpha}}{1 - \frac{1}{T}\left\{\frac{1+\alpha}{1-\alpha} - \frac{1}{T}\frac{2\alpha(1-\alpha^T)}{(1-\alpha)^2}\right\} + \frac{1}{T}\lambda(1-\delta)^2\left\{1-\alpha^{2T} - \frac{1}{T}\frac{(1+\alpha)(1-\alpha^T)^2}{1-\alpha}\right\}}.$$

Theorem 5 extends the result of Arellano (2003a, p.86) who only considers the case of $T = 2$.

Remark 12. Under the assumption of stationary initial conditions, i.e., $\delta = 1$, we find that $\widehat{\alpha}_{wg}$ is not affected by the degree of heterogeneity. In this case, the bias-corrected estimator of Hahn and Kuersteiner (2002), $\widehat{\alpha}_{hk}$, works well since $\widehat{\alpha}_{wg}$ has a bias of order $O(T^{-1})$ for any value of λ.

Remark 13. Under the assumption of nonstationary initial conditions, i.e., $\delta \neq 1$, we find that as $\lambda \to \infty$, $B_{wg} \to 0$. This implies that large heterogeneity makes the WG estimator to have a smaller bias provided that the initial conditions are nonstationary. The intuition behind this result is that nonstationary initial conditions make $E(x^{*'}x^*/NT)$, which is the denominator of B_{wg}, larger as λ grows since the third term in (45) is positive. Therefore, when there is large heterogeneity, $\widehat{\alpha}_{hk}$ does not work well since it has the positive bias $1/T$ by construction. Further, note that this result conflicts with Bun and Kiviet's (2006) result, obtained under mean stationarity, that the bias of the WG estimator is not affected by the degree of heterogeneity. This difference comes from the fact that, when initial conditions are nonstationary, the within-group transformation does not completely remove the individual effects and the properties are affected by the distribution of the individual effects.[20]

5.2 The LIML estimator

Let z_{it} be a generic instruments vector that is orthogonal to v_{it}^*. Then, the non-robust LIML analog estimator considered by Alonso-Borrego and Arellano (1999) and Alvarez and Arellano (2003) takes the following form:

$$\widehat{\alpha}_L = \frac{x^{*'}My^* - \widehat{\ell}x^{*'}y^*}{x^{*'}Mx^* - \widehat{\ell}x^{*'}x^*} = \frac{\sum_{t=1}^{T-1} x_t^{*'}M_t y_t^* - \widehat{\ell}\sum_{t=1}^{T-1} x_t^{*'}y_t^*}{\sum_{t=1}^{T-1} x_t^{*'}M_t x_t^* - \widehat{\ell}\sum_{t=1}^{T-1} x_t^{*'}x_t^*}, \tag{47}$$

where $\widehat{\ell}$ is the minimum eigenvalue of $W^{*'}MW^*(W^{*'}W^*)^{-1}$ with $W^* = (y^*, x^*)$. If we set z_{it} to be z_{it}^{l1}, z_{it}^{l2}, z_{it}^{d2}, and z_{it}^{b2}, then the corresponding LIML estimators can be defined as follows:

$$\widehat{\alpha}_{L,l1} = \frac{\sum_{t=1}^{T-1} x_t^{*'}M_t^{l1}y_t^* - \widehat{\ell}_{l1}\sum_{t=1}^{T-1} y_t^{*'}x_t^*}{\sum_{t=1}^{T-1} x_t^{*'}M_t^{l1}x_t^* - \widehat{\ell}_{l1}\sum_{t=1}^{T-1} x_t^{*'}x_t^*}, \tag{48}$$

$$\widehat{\alpha}_{L,l2} = \frac{\sum_{t=1}^{T-1} x_t^{*'}M_t^{l2}y_t^* - \widehat{\ell}_{l2}\sum_{t=1}^{T-1} x_t^{*'}y_t^*}{\sum_{t=1}^{T-1} x_t^{*'}M_t^{l2}x_t^* - \widehat{\ell}_{l2}\sum_{t=1}^{T-1} x_t^{*'}x_t^*}, \tag{49}$$

$$\widehat{\alpha}_{L,d2} = \frac{\sum_{t=2}^{T-1} x_t^{*'}M_t^{d2}y_t^* - \widehat{\ell}_{d2}\sum_{t=2}^{T-1} x_t^{*'}y_t^*}{\sum_{t=2}^{T-2} x_t^{*'}M_t^{d2}x_t^* - \widehat{\ell}_{d2}\sum_{t=2}^{T-1} x_t^{*'}x_t^*}, \tag{50}$$

$$\widehat{\alpha}_{L,b2} = \frac{\sum_{t=2}^{T-1} x_t^{*'}M_t^{b2}y_t^* - \widehat{\ell}_{b2}\sum_{t=2}^{T-1} x_t^{*'}x_t^*}{\sum_{t=2}^{T-2} x_t^{*'}M_t^{b2}x_t^* - \widehat{\ell}_{b2}\sum_{t=2}^{T-1} x_t^{*'}x_t^*}. \tag{51}$$

First, we consider $\widehat{\alpha}_{L,l1}$. The probability limit of $\widehat{\ell}_{l1}$ is given in the following Lemma.

Lemma 6. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity with $T/N \to c$, $(0 \leq c \leq 2)$, we have*

$$\widehat{\ell}_{l1} \xrightarrow{p} \frac{c}{2}.$$

[20]This interpretation was suggested by a referee.

The following theorem establishes the consistency and asymptotic normality of $\widehat{\alpha}_{L,l1}$.

Theorem 6. *Let Assumptions 1–3 hold. Then as both N and T tend to infinity, provided $T/N \to c$ $(0 \le c \le 2)$, we have*

$$\widehat{\alpha}_{L,l1} \xrightarrow{p} \alpha, \tag{52}$$

$$\sqrt{NT_*}(\widehat{\alpha}_{L,l1} - \alpha - B_{L,l1}) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \tag{53}$$

where

$$B_{L,l1} = \frac{1}{\sqrt{NT_*}} \frac{\mu_{G,l1} - \frac{T_*}{2N}\mu_{wg}}{R_T^{G,l1} - \frac{T_*}{2N}R_T^{wg}}.$$

Remark 14. In the case of $\delta = 1$, since $R_T^{G,l1}$ is bounded even when $\lambda \to \infty$ and R_T^{wg} does not depend on λ, $B_T^{L,l1}$ is bounded as the GMM estimator $\widehat{\alpha}_{G,l1}$. Using Lemmas 1(d) and 5(a), we have

$$\sqrt{NT_*}\left[\widehat{\alpha}_{L,l1} - \left(\alpha - \frac{1}{2N - T_*}(1+\alpha)\right)\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2). \tag{54}$$

This is the result obtained by Alvarez and Arellano (2003).

Remark 15. In the case of $\delta \neq 1$, since both $R_T^{G,l1}$ and R_T^{wg} approach infinity as $\lambda \to \infty$, $B_{L,l1}$ tends to zero as $\lambda \to \infty$. This implies that, as in the case of the GMM and WG estimators, a large degree of heterogeneity makes the LIML estimators to have a small bias when initial conditions are nonstationary.

Next, we consider $\widehat{\alpha}_{L,l2}$, $\widehat{\alpha}_{L,d2}$, and $\widehat{\alpha}_{L,b2}$.

Lemma 7. *Let Assumptions 1–3 hold. Then, as both N and T tend to infinity,*

$$\widehat{\ell}_2 \xrightarrow{p} 0,$$

where $\widehat{\ell}_2$ denotes $\widehat{\ell}_{l2}$, $\widehat{\ell}_{d2}$, and $\widehat{\ell}_{b2}$.

Theorem 7. *Let Assumptions 1–3 hold. Then, when both N and T are large, $\widehat{\alpha}_{L,l2}$, $\widehat{\alpha}_{L,d2}$, and $\widehat{\alpha}_{L,b2}$ are asymptotically equivalent to $\widehat{\alpha}_{G,l2}$, $\widehat{\alpha}_{G,d2}$, and $\widehat{\alpha}_{G,b2}$, respectively.*

Remark 16. We find that the difference in the asymptotic distribution between the GMM and LIML estimators exists only when all the instruments are used. If we use a smaller number of instruments, for example, $\widehat{\alpha}_{G,l2}$ and $\widehat{\alpha}_{L,l2}$, the GMM and LIML estimators have the same asymptotic distribution as in the large N and fixed T case.

6 Numerical studies

6.1 Monte Carlo studies with stationary initial conditions

In this subsection, we conduct Monte Carlo experiments to examine the performance of the estimators with stationary initial conditions. We consider the following AR(1) model:

$$y_{i,t} = \alpha y_{i,t-1} + \eta_i + v_{it} \qquad (i = 1, ..., N; t = 2, ..., T+1), \tag{55}$$

where $\eta_i \sim iid\mathcal{N}(0, \sigma_\eta^2)$, $y_{i,0} \sim iid\mathcal{N}(\eta_i/(1-\alpha), \sigma_v^2/(1-\alpha^2))$, and $v_{it} \sim iid\mathcal{N}(0, \sigma_v^2)$. We consider $(T, N) = (10, 50)$, $(10, 100)$, $(10, 500)$, $(15, 50)$, $(15, 100)$, $(15, 300)$, $(25, 50)$, $(25, 100)$, $(50, 50)$, $\alpha = 0.5, 0.8, 0.9$, and $\sigma_\eta^2 = 0.2, 1, 10$. σ_v^2 is set to be 1. The number of replications is 5,000 for all cases.

For each estimator, we compute the median, the interquartile range (IQR), the median absolute error (MAE),[21] and the size of the Wald test for $H_0 : \alpha = \alpha_o$ with a 5% level of significance.[22]

Median. Table 1 reports the simulation results for the median of the estimators discussed in the previous sections. Among the GMM estimators, $\widehat{\alpha}_{G,b2}$ has the smallest bias in almost all cases except when $\sigma_\eta^2 = 0.2$ and T is less than 15. Although the biases of $\widehat{\alpha}_{G,b2}$ in the near unit root case, i.e., $\alpha = 0.9$, are somewhat large in small samples, for example, $(T, N) = (10, 100)$, $\widehat{\alpha}_{G,b2}$ performs quite well when N is as large as 500 or when T is larger than 25. Further, we find that $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$ are affected by the degree of heterogeneity, while $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$ are not. For example, when $\alpha = 0.9$, $T = 15$, and $N = 100$, the medians of $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$ are 0.790 and 0.867 in the case of $\sigma_\eta^2 = 0.2$ and 0.742 and 0.675 in the case of $\sigma_\eta^2 = 10$. This demonstrates that $\widehat{\alpha}_{G,l2}$ is more seriously affected than $\widehat{\alpha}_{G,l1}$ by large heterogeneity. It is also worth noting that when $\sigma_\eta^2 = 10$, $\widehat{\alpha}_{G,l2}$ has a larger bias than $\widehat{\alpha}_{G,l1}$ despite the fact that the former uses fewer instruments than the latter.

With regard to the bias-corrected WG estimator, we find that $\widehat{\alpha}_{hk}$ works well in the near unit root case when T is as large as 25. As for other features, we find that the bias of $\widehat{\alpha}_{hk}$ is not affected by the degree of heterogeneity and the cross-sectional sample size N. The latter feature is particularly important when we compare $\widehat{\alpha}_{hk}$ with the GMM estimators. For instance, when $\alpha = 0.9$, the medians of $\widehat{\alpha}_{G,b2}$ and $\widehat{\alpha}_{hk}$ are 0.734 and 0.810, respectively, in the case of $T = 10$ and $N = 50$, but 0.872 and 0.812 when $T = 10$ and $N = 500$. This implies that $\widehat{\alpha}_{G,b2}$ is preferable to $\widehat{\alpha}_{hk}$ especially when N is large. Further, note that this result is consistent with the theoretical result of $\widehat{\alpha}_{G,l1}$ being consistent when N is large regardless of T, whereas $\widehat{\alpha}_{wg}$ can never be consistent unless T is large.

As for the LIML estimators, similar comments as those regarding the GMM estimators apply. $\widehat{\alpha}_{L,b2}$ performs best in many cases, particularly when σ_η^2 is larger than one. Further, we find that $\widehat{\alpha}_{L,l1}$ and $\widehat{\alpha}_{L,l2}$ are negatively affected by the degree of heterogeneity, and the magnitude of the effects is more serious than in the case of the GMM estimators. Even with a large sample, say, $T = 50$ and $N = 50$, when $\alpha = 0.9$ and $\sigma_\eta^2 = 10$, $\widehat{\alpha}_{L,l1}$ and $\widehat{\alpha}_{L,l2}$ have quite large biases.

Among all the estimators, $\widehat{\alpha}_{L,b2}$ has the smallest bias for a wide range of sample sizes, although the difference between $\widehat{\alpha}_{G,b2}$ and $\widehat{\alpha}_{L,b2}$ is quite small when N is as large as 300

[21] We use these robust statistics since the LIML estimators we compute are suspected to have no moments.

[22] The standard errors are calculated under large N and fixed T asymptotics, i.e., $se(\widehat{\alpha}) = \sqrt{\widehat{\sigma}_v^2 (x^{*\prime} M x^*)^{-1}}$ for the GMM and LIML estimators and $se(\widehat{\alpha}) = \sqrt{\widehat{\sigma}_v^2 (x^{*\prime} x^*)^{-1}}$ for the (bias-corrected) WG estimators. Although the standard error of the bias-corrected WG estimators cannot be estimated consistently under large N and fixed T asymptotics, it is expected that it will work well if the bias-correction works well.

or T is as large as 25. This result is quite different from the cross-sectional case where the GMM estimator has a much larger bias than the LIML estimators (see, e.g. Anderson, Kunitomo, and Matsushita, 2005).

Interquartile range. Table 2 shows the simulation results for the interquartile range of the estimators. Among the GMM estimators, the variability of $\widehat{\alpha}_{G,l1}$ is the smallest in all cases. Although both $\widehat{\alpha}_{G.l1}$ and $\widehat{\alpha}_{G,b2}$ are asymptotically efficient, they differ in a finite sample, and this difference becomes smaller as N or T becomes large. We also find that the effect of a large degree of heterogeneity is serious for $\widehat{\alpha}_{G,l2}$. As σ_η^2 gets larger, $\widehat{\alpha}_{G,l2}$ becomes more dispersed. If we compare $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$, we observe that the former is more dispersed than the latter; this result is consistent with the theoretical results. For the bias-corrected WG estimator, the IQR is quite small in all cases, and as N or T becomes large, the IQR becomes small. With regard to the LIML estimators, we observe that they have quite large dispersion, particularly when α is large, for example, $\alpha = 0.8, 0.9$. Moreover, we find that the IQRs of the LIML estimators are larger than those of the GMM estimators in all cases. In particular, $\widehat{\alpha}_{L,l1}$ and $\widehat{\alpha}_{L,l2}$ are greatly affected by a large degree of heterogeneity. In addition, we find that the LIML estimators have quite large dispersion when N is as small as 50.

Median absolute error. The results for the median absolute error of the estimators are summarized in Table 3. Among the GMM estimators, we observe that the MAE of $\widehat{\alpha}_{G,b2}$ is the smallest, except for some cases of $\sigma_\eta^2 = 0.2, 1$. Particularly, in the range of $T \geq 15$, improvements of $\widehat{\alpha}_{G,b2}$ compared to $\widehat{\alpha}_{G,l1}$ are significant, and the performance of $\widehat{\alpha}_{G,b2}$ is the best in almost all the cases when $\sigma_\eta^2 = 1, 10$. As for the bias-corrected WG estimator, its MAE is not affected by the degree of heterogeneity and is smaller than that of $\widehat{\alpha}_{G,b2}$, except for the case of large N. With regard to the LIML estimators, $\widehat{\alpha}_{L,b2}$ performs well when $\sigma_\eta^2 = 10$, although $\widehat{\alpha}_{L,l1}$ and $\widehat{\alpha}_{L,l2}$ may be preferable in the case of $\sigma_\eta^2 = 0.2$. We also find that in all cases, the MAEs of the LIML estimators are larger than those of the GMM estimators. Hence, in terms of the MAE, $\widehat{\alpha}_{G,b2}$ is preferable to $\widehat{\alpha}_{L,b2}$. Although $\widehat{\alpha}_{hk}$ performs quite well, this estimator is not recommended since it is not robust to nonstationary initial conditions, as will be discussed in the next subsection.

Size. The results of the empirical size of the Wald test are given in Table 4. Among the GMM estimators, we find that the size distortion of $\widehat{\alpha}_{G,l1}$ is substantial, particularly when $\alpha = 0.9$. However, the empirical sizes of the GMM estimators with a smaller number of instruments, $\widehat{\alpha}_{G,l2}$, $\widehat{\alpha}_{G,d2}$, and $\widehat{\alpha}_{G,b2}$ are close to the nominal size, with a few exceptions in the case of small samples, for instance, when $T = 10$ and $N = 50$. With regard to the bias-corrected WG estimators, the size distortion is quite large in many cases. For the LIML estimators, we observe that the size distortion of $\widehat{\alpha}_{L,l1}$ is substantial when $\alpha = 0.9$; $\widehat{\alpha}_{L,l2}$ is also severely oversized, particularly when σ_η^2/σ_v^2 is large. For $\widehat{\alpha}_{L,b2}$, although its empirical size is closer to the nominal size than in the case of other LIML estimators, it is not very close to the size of the GMM estimator $\widehat{\alpha}_{G,b2}$.

6.2 Numerical studies with nonstationary initial conditions

The theoretical analysis showed that large heterogeneity makes the GMM, WG, and LIML estimators to have small bias if initial conditions are nonstationary. In this subsection, we confirm this theoretical prediction numerically for some value of δ and λ.

We computed the median and the interquartile range of each estimator with $(\bar{T}, N) = (10, 200), (15, 100)$ where $\bar{T} = T + 1$ for the case of $\alpha = 0.9$. As for δ, we use $\delta = (1-\alpha)/(1-\bar{\alpha})$ with $\bar{\alpha}$ from 0.800 to 0.995 in steps of 0.0025.[23] Note that $\bar{\alpha} = 0.9$ corresponds to the stationary case. $(1-\alpha)^2\lambda = \sigma_\eta^2/\sigma_v^2$ is set to be $0.2, 0.5, 1, 3, 10$. The number of replications is $2,500$ for each $\bar{\alpha}$.

Median. Figures 1 to 20 depict the simulation values of the medians of the GMM, (bias-corrected) WG, and LIML estimators. Comparing the cases of $\bar{T} = 10, 15$, we find that the shapes of the graphs for each of the estimators are very similar, but their magnitudes (as indicated by the different scales of the vertical axes) differ. Thus, in what follows we consider each estimator in turn. We begin with $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$. What is common to both estimators is that the bias becomes very small as $\bar{\alpha}$ approaches one. In particular, we find that for the range of $\bar{\alpha} > 0.9$, i.e., $\delta > 1$, the biases are smaller than the case where $\bar{\alpha} = 0.9$, i.e., $\delta = 1$. This is consistent with the theoretical result that the instruments become strong when $\delta > 1$ as discussed in Remark 5. We also find that $\widehat{\alpha}_{G,l2}$ is more sensitive to the degree of heterogeneity than $\widehat{\alpha}_{l1}$. For example, in the case of $\bar{T} = 15$ and $\sigma_\eta^2 = 10$, the simulation value of $\widehat{\alpha}_{G,l1}$ changes from 0.742 to 0.870 when $\bar{\alpha}$ moves from 0.9 to 0.91, while that of $\widehat{\alpha}_{G,l2}$ changes from 0.675 to 0.895. This illustrates the sensitivity of $\widehat{\alpha}_{G,l2}$ to the degree of heterogeneity. With regard to $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$, we find that the shape of the graphs is different from that of $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$. While the graphs for $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$ are "U" or "V" shaped, those for $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$ are "W" shaped. In both cases, although the magnitude of the local maximum bias is almost the same for any value of σ_η^2/σ_v^2, $\bar{\alpha}$ that takes the local maximum bias approaches $\bar{\alpha} = 0.9$ as σ_η^2/σ_v^2 becomes larger. We also find that when $\bar{T} = 15$ and $\sigma_\eta^2/\sigma_v^2 = 10$, $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,l2}$ are almost unbiased when $\bar{\alpha} \le 0.87$ and $\bar{\alpha} \ge 0.92$.

We proceed to consider $\widehat{\alpha}_{wg}$. We find that when the degree of heterogeneity is small, i.e., for instance, $\sigma_\eta^2 = 0.2$, the estimate of $\widehat{\alpha}_{wg}$ is nearly flat around $\bar{\alpha} = 0.9$. However, as the degree of heterogeneity gets larger, the bias of $\widehat{\alpha}_{wg}$ becomes small. Therefore, it follows that large heterogeneity makes the WG estimator to have small bias. However, this feature exacerbates the result of $\widehat{\alpha}_{hk}$. Since $\widehat{\alpha}_{hk}$ corrects for the negative bias by adding $1/T$ as shown in (44), it becomes upwardly biased by construction when the initial conditions are nonstationary and heterogeneity is large. This can be seen by observing that the scale of the vertical axis in Figures 6 and 16 is different from the figures for other

[23] Note that y_{i0} can be written as

$$y_{i0} = \delta\left(\frac{1}{1-\alpha}\right)\eta_i + w_{i0} = \frac{1}{1-\bar{\alpha}}\eta_i + w_{i0},$$

where $\delta = (1-\alpha)/(1-\bar{\alpha})$.

estimators.

With regard to the LIML estimators, the effect of the degree of heterogeneity is quite large. In the case of $\widehat{\alpha}_{L,l1}$, as σ_η^2/σ_v^2 becomes small, the magnitude of the local maximum bias becomes large, although $\widehat{\alpha}_{L,l1}$ is not very much affected at $\bar{\alpha} = 0.9$. On the other hand, $\widehat{\alpha}_{L,l2}$, as mentioned above, has substantial bias when $\bar{\alpha} = 0.9$ and σ_η^2/σ_v^2 is large. We also find that $\widehat{\alpha}_{L,l2}$ has quite a large local maximum bias. Although the movement of $\widehat{\alpha}_{G,l2}$ and $\widehat{\alpha}_{L,l2}$ is similar, the magnitude of the bias is quite different. The shape of the graphs for $\widehat{\alpha}_{L,d2}$ and $\widehat{\alpha}_{L,b2}$ is quite similar to that of the graphs for $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$, except that the former has a smaller bias at $\bar{\alpha} = 0.9$ than the latter.

Interquartile range. Figures 21 to 40 depict the simulation values for the IQR of the GMM, (bias-corrected) WG, and LIML estimators. As in the case of the median, the IQR of $\widehat{\alpha}_{G,l1}$ and $\widehat{\alpha}_{G,l2}$ becomes small as the degree of heterogeneity becomes large under nonstationary initial conditions. We also find that although $\widehat{\alpha}_{G,l2}$ is quite dispersed when $\sigma_\eta^2/\sigma_v^2 = 10$ and $\bar{\alpha} = 0.9$, its variation becomes quite small as $\bar{\alpha}$ moves away from 0.9, particularly as $\bar{\alpha}$ approaches one. For example, in the case of $\bar{T} = 15$ and $\sigma_\eta^2/\sigma_v^2 = 10$, the IQR of $\widehat{\alpha}_{G,l1}$ decreases from 0.103 to 0.005 when $\bar{\alpha}$ moves from 0.9 to 0.95 and that of $\widehat{\alpha}_{G,l2}$ falls from 0.305 to 0.005. The shape of of the graphs for $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$ is quite similar although $\widehat{\alpha}_{G,b2}$ has a smaller variation than $\widehat{\alpha}_{G,d2}$ at $\bar{\alpha} = 0.9$.

For the (bias-corrected) WG estimators, we find that they have the largest IQR around $\bar{\alpha} = 0.9$, and as $\bar{\alpha}$ moves away from 0.9, the IQR becomes quite small. Moreover, the IQRs of $\widehat{\alpha}_{wg}$ and $\widehat{\alpha}_{hk}$ are much smaller than those of the GMM estimators (note the different scale of the vertical axis in the graphs for these estimators).

With regard to the LIML estimators, note that the scale of the vertical axis in the graph is much larger than those of the GMM and (bias-corrected) WG estimators. We find that the IQR of $\widehat{\alpha}_{L,l1}$ becomes substantially large when $\bar{\alpha} = 0.8$ and $\sigma_\eta^2/\sigma_v^2 = 0.2$. In the case of $\widehat{\alpha}_{L,l2}$, the sensitivity of the IQR to large heterogeneity is noteworthy. If $\bar{T} = 15$ and $\sigma_\eta^2/\sigma_v^2 = 10$, the IQR of $\widehat{\alpha}_{L,l2}$ with $\bar{\alpha} = 0.9$ is 1.907, while that with $\bar{\alpha} = 0.91$ is 0.043. The shapes of $\widehat{\alpha}_{L,d2}$ and $\widehat{\alpha}_{L,b2}$ are quite similar to those of $\widehat{\alpha}_{G,d2}$ and $\widehat{\alpha}_{G,b2}$, although the scale is quite different.

7 Conclusion

In this paper, we considered the asymptotic properties of GMM estimators with various kinds of instruments in a dynamic panel data model with possibly nonstationary initial conditions. We showed that the GMM estimators with instruments in levels perform poorly under stationary initial conditions and that they perform well under nonstationary initial conditions if the degree of heterogeneity is large. We demonstrated that this result comes from the fact that nonstationary initial conditions provide an additional correlation between the lagged dependent variable and the instruments, and found that, as the degree of heterogeneity gets larger, the instruments become strong.

For the purpose of comparison, we also derived the asymptotic properties of WG and LIML estimators. We showed that under stationary initial conditions, the performance of the WG estimator is not affected by the degree of heterogeneity, while, under nonstationary initial conditions, the bias of the WG estimator becomes small if the degree of heterogeneity is large. For the LIML estimators, we found that the results are similar to those of the GMM estimators. We conducted Monte Carlo simulations to assess the estimators. The simulation results indicate that the LIML estimator with the BOD-transformed instruments has smaller bias than the GMM estimator, although the difference becomes small as the sample size becomes larger. However, in terms of the median absolute error, the GMM estimator outperforms the LIML estimator in almost all cases.

Finally, we note some possible extensions. First, although the model considered in this paper is limited to a stable AR(1) panel model, for practical application, it is important to extend the analysis to models with additional regressors and/or unobserved heterogeneous time trends, or unit root models. In the case of the former, however, it is likely that the points made in Remark 5 will apply. In the case of the latter, it is well known from the literature on pure time series models that initial conditions affect the performance of unit root tests (see, e.g. Elliott and Müller, 2003; Müller and Elliott, 2006). However, to the best of our knowledge, there are no studies that investigate the effect of initial conditions on the performance of panel unit root tests. These extensions are particularly important in practice. Second, it would be interesting to investigate the properties of inconsistent estimators. Although Arellano (2003) discusses an inconsistent IV estimator of models in levels, it would be interesting to extend it to the level and system GMM estimators which are known to be inconsistent when initial conditions are nonstationary. In terms of practical application, it would be particularly important to investigate the power of the over-identification test. These topics are left for future research.

References

[1] Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business and Economic Statistics*, 17, 36-49.

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

[3] Anderson, T.W., and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

[4] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2005): "A New Light from Old Wisdoms . Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models," CIRJE Discussion Paper, F-321, University of Tokyo.

[5] Andrews, D. W. and J. H. Stock (2006): "Inference with Weak Instruments, " in *Advances in Economics and Econometrics, Theory and Applications*, Ninth World

Congress of the Econometric Society, Vol. III, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press, Cambridge, UK.

[6] Angrist, J.D., and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106, 979-1014.

[7] Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press, New York.

[8] Arellano, M. and S.R. 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.

[9] Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Component Models," *Journal of Econometrics*, 68, 29-45.

[10] Baltagi, B.H (2005): *Econometric Analysis of Panel Data*, 3rd ed., John Wiley and Sons, Chichester.

[11] Barro, R. J. and X. Sala-i-Martin (1995): *Economic Growth*, McGraw-Hill, New York.

[12] Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-43.

[13] Bun, M. and J. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132, 409-444.

[14] Elliott, G. and U. K. Müller (2006): "Minimizing the Impact of the Initial Condition on Testing for Unit Roots," *Journal of Econometrics*, 135, 285-310.

[15] 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.

[16] Hause, J.C. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, 48, 1013-1029.

[17] Hayakawa, K. (2007a): "Small Sample Bias Properties of the System GMM estimators in Dynamic Panel Data Models," *Economics Letters*, 95, 32-38.

[18] Hayakawa, K. (2007b): "A Simple Efficient Instrumental Variable Estimator in Panel AR(p) Models," mimeo.

[19] Hsiao, C.(2003): *Analysis of Panel Data*, Cambridge University Press, Cambridge.

[20] Judson, R.A. and A. L. Owen (1999): "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists," *Economics Letters*, 65, 9-15.

[21] Kiviet J.F. (2007): "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models," in *The Refinement of Econometric Estimation and Test Procedures*, ed. by G.D.A. Phillips and E. Tzavalis, Cambridge University Press, Cambridge, UK.

[22] Müller, U. K., and G. Elliott (2003): "Tests for Unit Roots and the Initial Condition," *Econometrica*, 71, 1269-1286.

[23] Okui, R. (2005a): "Instrumental Variable Estimation in the Presence of Many Moment Conditions," unpublished manuscript.

[24] Okui, R. (2005b): "The Optimal Choice of Moments in Dynamic Panel Data Models," unpublished manuscript.

[25] Ziliak, J.P. (1997): "Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics*, 15, 419-431.

A Mathematical Proofs

A.1 The GMM estimator

Throughout the appendix, let us denote $t_0 = 1$ for the case of $M_t = M_t^{l1}, M_t^{l2}$, and $t_0 = 2$ for the case of $M_t = M_t^{d2}, M_t^{b2}$.

We collect some results which are useful to prove the main results.

Lemma A1. *Let a, b, c, d be constants satisfying $a > 0$, and $a + b\rho_1^t > 0$ for all t. We also assume $|\rho_1|, |\rho_2| < 1$. Then as $T \to \infty$,*

$$S = \frac{1}{T_*} \sum_{t=t_0}^{T_*} \frac{c + d\rho_2^t}{a + b\rho_1^t} \to \frac{c}{a}. \tag{56}$$

Proof: Since $0 < 1 + (b/a)\rho_1^t$ for $t = t_0, ..., T_*$, it follows that

$$\begin{aligned} S &= \frac{c}{a}\frac{1}{T_*}\sum_{t=t_0}^{T_*} \frac{1 + \frac{d}{c}\rho_2^t}{1 + \frac{b}{a}\rho_1^t} = \frac{c}{a}\frac{1}{T_*}\sum_{t=t_0}^{T_*}\left[1 - \frac{\frac{b}{a}\rho_1^t - \frac{d}{c}\rho_2^t}{1 + \frac{b}{a}\rho_1^t}\right] \\ &= \frac{c}{a}\left[1 - \frac{1}{T_*}\sum_{t=t_0}^{T_*}\frac{\frac{b}{a}\rho_1^t - \frac{d}{c}\rho_2^t}{1 + \frac{b}{a}\rho_1^t}\right] = \frac{c}{a}\left[1 - O\left(\frac{1}{T}\right)\right] \to \frac{c}{a}. \end{aligned}$$

Lemma A2. *Let Assumption 1-3 hold. Then, we have*

$$(a) \quad E(y_{i,t-1}^2) = \sigma_v^2\left[\frac{1}{1-\alpha^2} + \lambda\left[1 - (1-\delta)\alpha^{t-1}\right]^2\right],$$

$$(b) \quad E(\mu_i y_{i,t-1}) = \sigma_\mu^2\left[1 - (1-\delta)\alpha^{t-1}\right],$$

$$(c) \quad E(w_{i,t-1}y_{i,t-1}) = \frac{\sigma_v^2}{1-\alpha^2},$$

$$(d) \quad E(\Delta y_{i,t-1}^2) = \sigma_v^2\left[\frac{2}{1+\alpha} + \lambda(1-\delta)^2(1-\alpha)^2\alpha^{2(t-2)}\right],$$

$$(e) \quad E(y_{i,t-1}\Delta y_{i,t-1}) = \sigma_v^2\left[\frac{1}{1+\alpha} + \lambda(1-\delta)(1-\alpha)\left[1-(1-\delta)\alpha^{t-1}\right]\alpha^{t-2}\right],$$

$$(f) \quad E(w_{i,t-1}\Delta y_{i,t-1}) = \frac{\sigma_v^2}{1+\alpha},$$

$$(g) \quad E(\mu_i \Delta y_{i,t-1}) = \sigma_\mu^2(1-\delta)(1-\alpha)\alpha^{t-2},$$

$$\begin{aligned} (h) \quad E[(y_{i,t-1}^b)^2] = \sigma_v^2\Bigg[&\frac{1}{1-\alpha^2}\left(1 - \frac{2\alpha\phi_{t-1}}{t-1} + \frac{1}{(t-1)^2}\left\{\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}\right) \\ &+\lambda(1-\delta)^2\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)^2\Bigg], \end{aligned}$$

$$(i) \; E(y_{i,t-1}y_{i,t-1}^b) = \sigma_v^2\left[\frac{1}{1-\alpha^2}\left(1 - \frac{\alpha\phi_{t-1}}{t-1}\right) - \lambda(1-\delta)\left[1-(1-\delta)\alpha^{t-1}\right]\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)\right],$$

$$(j) \; E(w_{i,t-1}y_{i,t-1}^b) = \frac{\sigma_v^2}{1-\alpha^2}\left(1 - \frac{\alpha\phi_{t-1}}{t-1}\right),$$

(k) $E(\mu_i y^b_{i,t-1}) = -\sigma^2_\mu(1-\delta)\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right).$

Proof: Using

$$\begin{aligned} y_{i,t-1} &= \left[1-(1-\delta)\alpha^{t-1}\right]\mu_i + w_{i,t-1}, \\ \Delta y_{i,t-1} &= (1-\delta)(1-\alpha)\alpha^{t-2}\mu_i + w_{i,t-1} - w_{i,t-2}, \\ y^b_{i,t-1} &= \left[-(1-\delta)\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)\mu_i + \left(w_{i,t-1} - \frac{w_{i,t-2}+\cdots+w_{i,0}}{t-1}\right)\right], \end{aligned}$$

and

$$\begin{aligned} & E\left(w_{i,t-1} - \frac{w_{i,t-2}+\cdots+w_{i,0}}{t-1}\right)^2 \\ = & \frac{\sigma^2_v}{1-\alpha^2}\left[1 - \frac{2\alpha\phi_{t-1}}{t-1} + \frac{1}{(t-1)^2}\left\{\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}\right], \end{aligned}$$

it is straightforward to prove.

Lemma A3. *Let κ_3 and κ_4 denote the third and fourth-order cumulants of v_{it}. Also let d_t and d_s be $(N\times 1)$ vectors containing the diagonal elements of M_t and M_s, respectively, so that $tr(M_t) = d_t'\iota_t = 1$, $tr(M_s) = d_s'\iota_t = 1$, and $d_t'd_s \leq 1$. Then under Assumption 1, for $l \geq r \geq t$, $p \geq q \geq s$, and $t \geq s$, we have*

$$\begin{aligned} cov(v_l' M_t v_r, v_p' M_s v_q) &= \begin{cases} 2\sigma_v^4 tr(M_t M_s) + \kappa_4 E(d_t' d_s) \leq (2\sigma_v^4 + \kappa_4) & l=r=p=q \\ \kappa_3 E(d_t' M_s v_q) & l=r=p\neq q<t \\ \sigma_v^4 tr(M_t M_s) \leq \sigma_v^4 & l=p\neq r=q \\ 0 & otherwise, \end{cases} \\ |E(d_t' M_s v_q)| &\leq \sigma_v \end{aligned}$$

Proof Using the similar arguments to Alvarez and Arellano (2003), it is straightforward to show.

Lemma A4. *Let Assumptions 1, 2, 3 hold. Then as both N and T tend to infinity,*

$$(a) \quad \frac{1}{NT_*}\sum_{t=1}^{T-1} w_{t-1}' M_t^{l2} w_{t-1} \xrightarrow[N,T\to\infty]{p} \rho_{l2}\left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

$$(b) \quad \frac{1}{NT_*}\sum_{t=2}^{T-1} w_{t-1}' M_t^{d2} w_{t-1} \xrightarrow[N,T\to\infty]{p} \rho_{d2}\left(\frac{\sigma_v^2}{1-\alpha^2}\right).$$

As T tends to infinity, regardless of whether N is fixed or tends to infinity,

$$(c) \quad \frac{1}{NT_*}\sum_{t=1}^{T-1} w_{t-1}' M_t^{l1} w_{t-1} \xrightarrow[T\to\infty]{p} \left(\frac{\sigma_v^2}{1-\alpha^2}\right),$$

$$(d) \quad \frac{1}{NT_*}\sum_{t=2}^{T-1} w_{t-1}' M_t^{b2} w_{t-1} \xrightarrow[T\to\infty]{p} \left(\frac{\sigma_v^2}{1-\alpha^2}\right)$$

where

$$\rho_{l2} = \left[1+\lambda(1-\alpha^2)\right]^{-1},$$
$$\rho_{d2} = \left(\frac{1-\alpha}{2}\right).$$

Proof: (a), (b): They are obtained using $w'_{t-1}Z_t/N = E(w_{i,t-1}z_{it}) + O_p(1/\sqrt{N})$, $Z'_tZ_t/N = E(z_{it}^2) + O_p(1/\sqrt{N})$, and Lemma A1, and A2 as follows:

$$\begin{aligned}\frac{1}{NT_*}\sum_{t=t_0}^{T-1} w'_{t-1}M_t w_{t-1} &= \frac{1}{T_*}\sum_{t=t_0}^{T-1} E(w_{i,t-1}z_{it})[E(z_{it}^2)]^{-1}E(z_{it}w_{i,t-1}) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &\to \rho_2\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\end{aligned}$$

where z_{it} denotes z_{it}^{l2} or z_{it}^{d2}, and ρ_2 denotes ρ_{l2} or ρ_{d2}.

(c) As in Alvarez and Arellano (2003), let e_t be the $(N\times 1)$ vector of errors of the population linear projection of μ_t^* on Z_t^{l1} as follows:

$$e_t = \mu_t^* - Z_t^{l1}\gamma_t^{l1} \tag{57}$$

where $\mu_t^* = (\mu_{1t}^*, ..., \mu_{Nt}^*)'$ and $\gamma_t^{l1} = [E(z_{it}^{l1}z_{it}^{l1\prime})]^{-1}E(z_{it}^{l1}\mu_{it}^*)$. We derive the explicit expression of γ_t. First, note that z_{it}^{l1} can be expressed as

$$z_{it}^{l1} = \left[\iota_t - (1-\delta)\psi_\alpha\right]\mu_i + \psi_{w,i} = b_t\mu_i + \psi_{w,i} \tag{58}$$

where ι_t is a $(t\times 1)$ vector of ones, $\psi_\alpha = (1, \alpha, ..., \alpha^{t-1})'$, and $\psi_{w,i} = (w_{i,0}, ..., w_{i,t-1})'$. Let V_t denote a $(t\times t)$ matrix whose (j,k) element is $\alpha^{|j-k|}/(1-\alpha^2)$. Then, we have

$$E(z_{it}^{l1}z_{it}^{l1\prime}) = \sigma_\mu^2 b_t b'_t + \sigma_v^2 V_t. \tag{59}$$

Using the formula $(A+bb')^{-1} = A^{-1} - A^{-1}bb'A^{-1}/(1+b'A^{-1}b)$, it follows that

$$\begin{aligned}\left[E(z_{it}^{l1}z_{it}^{l1\prime})\right]^{-1} &= \left[\sigma_v^2V_t + \sigma_\mu^2 b_t b'_t\right]^{-1} = \sigma_v^{-2}\left[V_t + (\sqrt{\lambda}b_t)(\sqrt{\lambda}b_t)'\right]^{-1} \\ &= \sigma_v^{-2}\left[V_t^{-1} - \frac{\lambda}{1+\lambda b'_tV_t^{-1}b_t}V_t^{-1}b_tb'_tV_t^{-1}\right].\end{aligned}$$

Next, we have

$$E(z_{it}^{l1}\mu_{it}^*) = \sigma_\mu^2\left[\iota_t - (1-\delta)\psi_\alpha\right]\left[1-(1-\delta)\alpha^t\right] = \sigma_\mu^2 b_t\left[1-(1-\delta)\alpha^t\right].$$

Then, we have

$$\gamma_t = \frac{\lambda}{1+\lambda b'_tV_t^{-1}b_t}V_t^{-1}b_t[1-(1-\delta)\alpha^t]. \tag{60}$$

Using the expression of γ_t, the i-th component of e_t is given by

$$e_{it} = \mu_{it}^* - z_{it}^{l1\prime}\gamma_t = \frac{\left[\mu_i - \lambda\delta(1-\alpha^2)w_{i0} - \lambda(1-\alpha)(v_{i1}+\cdots+v_{i,t-1})\right]\left[1-(1-\delta)\alpha^t\right]}{1+\lambda\{(1-\alpha^2)+(t-1)(1-\alpha)^2-(1-\delta^2)(1-\alpha^2)\}}$$

where we use the results that $b_t' V_t^{-1} b_t = (1-\alpha^2) + (t-1)(1-\alpha)^2 - (1-\delta^2)(1-\alpha^2)$, and $\psi_{wi}' V_t^{-1} b_t = \delta(1-\alpha^2) w_{i0} + (1-\alpha)(v_{i1} + \cdots + v_{i,t-1})$. Since e_{it} is composed of $(t+1)$ independent variables, its variance is given by

$$
\begin{aligned}
E(e_{it}^2) &= \frac{\left[\sigma_\mu^2 - \lambda^2\delta^2\sigma_v^2(1-\alpha^2) + (1-\alpha)^2(t-1)\sigma_v^2\right]\left[1-(1-\delta)\alpha^t\right]^2}{\left[1+\lambda\{(1-\alpha^2)+(t-1)(1-\alpha)^2-(1-\delta^2)(1-\alpha^2)\}\right]^2} \\
&= O\left(\frac{1}{t}\right).
\end{aligned}
$$

Now we consider the decomposition:

$$
\begin{aligned}
w_{t-1}' M_t^{l1} w_{t-1} &= w_{t-1}' w_{t-1} - w_{t-1}'(I_N - M_t^{l1}) w_{t-1} \\
&= w_{t-1}' w_{t-1} - e_t'(I_N - M_t^{l1}) e_t
\end{aligned}
$$

where the second equality comes from the fact that $w_{t-1} = y_{t-1} - Z_t^{l1}\gamma_t^{l1} - e_t$ and $(I_N - M_t^{l1})(y_{t-1} - Z_t^{l1}) = 0$. Hence we have

$$
\frac{1}{NT_*}\sum_{t=1}^{T-1} E(w_{t-1}' M_t^{l1} w_{t-1}) = E(w_{i,t-1}^2) - \frac{1}{NT_*}\sum_{t=1}^{T-1} E(e_t'(I_N - M_t^{l1}) e_t).
$$

Since the maximum eigenvalue of $(I_N - M_t^{l1})$ is equal to 1,

$$
\frac{1}{NT_*}\sum_{t=1}^{T-1} E(e_t'(I_N - M_t^{l1}) e_t) \le \frac{1}{NT_*}\sum_{t=1}^{T-1} E(e_t' e_t) = \frac{1}{T_*}\sum_{t=1}^{T-1} E(e_{i,t}^2) = \frac{1}{T_*} O(\log T) \to 0.
$$

Hence, as $T \to \infty$,

$$
\frac{1}{NT_*}\sum_{t=1}^{T} E(w_{t-1}' M_t^{l1} w_{t-1}) \to E(w_{i,t-1}^2) = \frac{\sigma_v^2}{1-\alpha^2}.
$$

With regards to the proofs that the variance of $(NT_*)^{-1}\sum_{t=1}^{T-1} w_{t-1}' w_{t-1}$ and $(NT_*)^{-1}\sum_{t=1}^{T-1} e_t' e_t$ tend to zero, see Alvarez and Arellano (2003).

(d) The flow of the proof is the same as that of the proof of (c). Let ε_t denote the $(N \times 1)$ vector of errors of the population linear projection of w_{t-1} on Z_t^{b2}:

$$
w_{t-1} = Z_t^{b2}\gamma_t^{b2} + \varepsilon_t
$$

where $\gamma_t^{b2} = E(z_{it}^{b2} w_{i,t-1})/E(z_{it}^{b2})^2 = E(y_{i,t-1}^b w_{i,t-1})/[E(y_{i,t-1}^b)^2]$. Using Lemma A2, the i-th component of ε_t can be expressed as

$$
\varepsilon_{it} = w_{i,t-1} - \delta y_{i,t-1}^b = \frac{C}{D}
$$

where

$$
\begin{aligned}
C &= \frac{[\lambda_{12} w_{i,t-1} + \lambda_3(w_{i,t-2} + \cdots + w_{i0})]}{t-1} + \lambda_4\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)^2 w_{i,t-1} + \lambda_5\left(\alpha^{t-1} - \frac{\phi_{t-1}}{t-1}\right)\mu_i \\
&= C_1 + C_2 + C_3, \\
D &= E[(y_{i,t-1}^b)^2] = O(1),
\end{aligned}
$$

$$
\lambda_1 = -\frac{\sigma_v^2 \alpha \phi_{t-1}}{1-\alpha^2},
$$

$$
\begin{aligned}
\lambda_2 &= \frac{1}{t-1}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left\{\frac{(t-1)(1+\alpha)}{1-\alpha} - \frac{2\alpha(1-\alpha^{t-1})}{(1-\alpha)^2}\right\}, \\
\lambda_3 &= \left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left(1 - \frac{\alpha\phi_{t-1}}{t-1}\right), \\
\lambda_4 &= (1-\delta)^2\sigma_\mu^2, \\
\lambda_5 &= (1-\delta)\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left(1 - \frac{\alpha\phi_{t-1}}{t-1}\right), \\
\lambda_{12} &= \lambda_1 + \lambda_2.
\end{aligned}
$$

We derive the order of $E(\varepsilon_{it}^2)$. Since $D = O(1)$, we consider only C. Since μ_i and $w_{i,t-1}$ are independent, we have

$$
\begin{aligned}
var(C) &= var(C_1) + var(C_2) + var(C_3) + 2cov(C_1, C_2) + 2cov(C_2, C_3) + 2cov(C_1, C_3) \\
&= var(C_1) + var(C_2) + var(C_3) + 2cov(C_1, C_2) \\
&= O\left(\frac{1}{t}\right) + \lambda_4^2\left[\alpha^{2(t-1)} + O\left(\frac{1}{t}\right)\right]\left(\frac{\sigma_v^2}{1-\alpha^2}\right) + \lambda_5^2\left[\alpha^{2(t-1)} + O\left(\frac{1}{t}\right)\right]\sigma_\mu^2 + O\left(\frac{1}{t}\right) \\
&= O\left(\frac{1}{t}\right) + \left(\frac{\lambda_4^2\sigma_v^2}{1-\alpha^2} + \lambda_5^2\sigma_\mu^2\right)\alpha^{2(t-1)} \\
&= O\left(\frac{1}{t}\right) + \lambda_6\alpha^{2(t-1)}.
\end{aligned}
$$

Hence,

$$
E(\varepsilon_{it}^2) = \frac{var(C)}{D^2} = O\left(\frac{1}{t}\right) + \frac{\lambda_6}{D^2}\alpha^{2(t-1)}.
$$

Given the existence of the fourth order moments of ε_{it}, we also have

$$
E(\varepsilon_{it}^4) = O\left(\frac{1}{t^2}\right) + O\left(\frac{1}{t}\right)\alpha^{2(t-1)} + \frac{\lambda_6^2}{D^4}\alpha^{4(t-1)}.
$$

As in the proof of (c), we consider the decomposition:

$$
\begin{aligned}
w_{t-1}' M_t^{b2} w_{t-1} &= w_{t-1}'w_{t-1} - w_{t-1}'(I_N - M_t^{b2})w_{t-1} \\
&= w_{t-1}'w_{t-1} - \varepsilon_t'(I_N - M_t^{b2})\varepsilon_t.
\end{aligned}
$$

Since the maximum eigenvalue of $(I_N - M_t^{b2})$ is equal to 1,

$$
\begin{aligned}
\frac{1}{NT_*}\sum_{t=2}^{T-1} E(\varepsilon_t'(I_N - M_t^{b2})\varepsilon_t) &\le \frac{1}{NT_*}\sum_{t=2}^{T-1} E(\varepsilon_t'\varepsilon_t) = \frac{1}{T_*}\sum_{t=2}^{T-1} E(\varepsilon_{it}^2) \\
&= \frac{1}{T_*}O(\log T) + \frac{\lambda_6}{T_* D^2}\frac{\alpha^2(1-\alpha^{2(T-2)})}{1-\alpha^2} \to 0.
\end{aligned}
$$

Hence, as $T \to \infty$,

$$
\frac{1}{NT_*}\sum_{t=2}^{T} E(w_{t-1}'M_t^{b2}w_{t-1}) \to E(w_{i,t-1}^2) = \frac{\sigma_v^2}{1-\alpha^2}.
$$

Finally, the variance of $(NT_*)^{-1}\sum_{t=2}^{T-1}\varepsilon_t'\varepsilon_t$ is shown to tend to zero as follows:

$$
var\left(\frac{1}{NT_*}\sum_{t=2}^{T-1}\varepsilon_t'\varepsilon_t\right) = \frac{1}{N}var\left(\frac{1}{T_*}\sum_{t=2}^{T-1}\varepsilon_{it}^2\right)
$$

$$
\begin{aligned}
&= \frac{1}{N}\left[\frac{1}{T_*^2}\sum_t var(\varepsilon_{it}^2) + \frac{2}{T_*^2}\sum_s\sum_{t>s} cov(\varepsilon_{it}^2, \varepsilon_{is}^2)\right] \\
&\leq \frac{1}{N}\left[\frac{1}{T_*^2}\sum_t \left\{O\left(\frac{1}{t^2}\right) + O\left(\frac{1}{t}\right)\alpha^{2(t-1)} + \frac{\lambda_6^2}{D^4}\alpha^{4(t-1)}\right\}\right. \\
&\quad + \frac{2}{T_*^2}\sum_t \left\{O\left(\frac{1}{t}\right) + O\left(\frac{1}{\sqrt{t}}\right)\alpha^{(t-1)} + \frac{\lambda_6}{D^2}\alpha^{2(t-1)}\right\} \\
&\quad \times \sum_s \left\{O\left(\frac{1}{s}\right) + O\left(\frac{1}{\sqrt{s}}\right)\alpha^{(s-1)} + \frac{\lambda_6}{D^2}\alpha^{2(s-1)}\right\} \\
&\to 0.
\end{aligned}
$$

Proof of Lemma 1(a), 2(a), 3(a), 4(a)

We shall use the decomposition as follows:

$$
\begin{aligned}
x_{it}^* &= \psi_t\left[w_{i,t-1} - (1-\delta)\alpha^{t-1}\mu_i\right] - c_t\tilde{v}_{itT} && (61) \\
&= \psi_t(y_{i,t-1} - \mu_i) - c_t\tilde{v}_{itT} && (62)
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_t &= c_t\left(1 - \frac{\alpha\phi_{T-t}}{T-t}\right), && (63) \\
\tilde{v}_{itT} &= \frac{\phi_{T-t}v_{it} + \cdots + \phi_1 v_{i,T-1}}{T-t}, && (64) \\
\phi_j &= \frac{1-\alpha^j}{1-\alpha} = 1 + \alpha + \cdots + \alpha^{j-1}. && (65)
\end{aligned}
$$

Then, following Alvarez and Arellano (2003), we have

$$
\begin{aligned}
E(x^{*'}Mv^*) &= -\sum_{t=t_0}^{T-1} E(c_t\tilde{v}_{tT}'M_t v_t^*) && (66) \\
&= -\sum_{t=t_0}^{T-1} \frac{\sigma_v^2 tr(M_t)}{1-\alpha}\left(\frac{\phi_{T-t}}{T-t} - \frac{\phi_{T-t+1}}{T-t+1}\right). && (67)
\end{aligned}
$$

Proof of Lemma 1(a) See Alvarez and Arellano (2003).
Proof of Lemma 2(a), 3(a), 4(a) Since $tr(M_t) = 1$, the results follow from a simple calculation.

Proof of Lemma 1(b), 2(b), 3(b), 4(b),

It is straightforward to show from Lemma A2 and a simple manipulation.

Proof of Lemma 1(c), 2(c), 3(c), 4(c),

Using $v_t^* = (v_t - \bar{v}_{tT})/c_t$, we have the following decomposition:

$$
\frac{x^{*'}Mv^*}{\sqrt{NT_*}} = \frac{1}{\sqrt{NT_*}}\sum_{t=t_0}^{T-1} w_{t-1}'M_t v_t - \Upsilon_{11NT} - \Upsilon_{12NT} - \Upsilon_{13NT} + \Upsilon_{14NT} + \Upsilon_{15NT}
$$

$$
\begin{aligned}
& -(\Upsilon_{21NT} - \Upsilon_{22NT}) \\
= \; & \Upsilon_{11NT}^{*} - \Upsilon_{12NT} - \Upsilon_{13NT} + \Upsilon_{14NT} + \Upsilon_{15NT} - (\Upsilon_{21NT} - \Upsilon_{22NT})
\end{aligned}
$$

where

$$\Upsilon_{11NT} = \frac{1}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} w_{t-1}' M_t \bar{v}_{tT}, \tag{68}$$

$$\Upsilon_{11NT}^{*} = \frac{1}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} c_t w_{t-1}' M_t v_t^{*}, \tag{69}$$

$$\Upsilon_{12NT} = \frac{1}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \frac{c_t \alpha \phi_{T-t}}{T-t} w_{t-1}' M_t v_t^{*}, \tag{70}$$

$$\Upsilon_{13NT} = \frac{(1-\delta)}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \alpha^{t-1} \mu' M_t v_t, \tag{71}$$

$$\Upsilon_{14NT} = \frac{(1-\delta)}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \alpha^{t-1} \mu' M_t \bar{v}_{tT}, \tag{72}$$

$$\Upsilon_{15NT} = \frac{(1-\delta)}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \frac{c_t \alpha \phi_{T-t}}{T-t} \alpha^{t-1} \mu' M_t \bar{v}_t^{*}, \tag{73}$$

$$\Upsilon_{21NT} = \frac{1}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \tilde{v}_{tT}' M_t v_t, \tag{74}$$

$$\Upsilon_{22NT} = \frac{1}{\sqrt{NT_*}} \sum_{t=t_0}^{T-1} \tilde{v}_{tT}' M_t \bar{v}_{tT}, \tag{75}$$

$$\bar{v}_{tT} = \frac{v_t + \cdots + v_T}{T-t+1}. \tag{76}$$

Proof of Lemma 1(b) For the case of $M_t = M_t^{l1}$, Alvarez and Arellano (2003) showed that the variance of the leading term converges to $\sigma_v^4/(1-\alpha^2)$, and those of Υ_{11NT}, Υ_{12NT}, Υ_{21NT}, and Υ_{22NT} tend to zero if $(\log T)^2/N \to 0$. Hence, to complete the proof, we show that the variances of Υ_{13NT}, Υ_{14NT}, and Υ_{15NT} tend to zero. First, we consider Υ_{13NT}:

$$
\begin{aligned}
var(\Upsilon_{13NT}) &= \frac{(1-\delta)^2}{NT_*} var\left(\sum_{t=1}^{T-1} \alpha^{t-1} \mu' M_t^{l1} v_t\right) = \frac{(1-\delta)^2}{NT_*} \sum_{t=1}^{T-1} \alpha^{2(t-1)} var\left(\mu' M_t^{l1} v_t\right) \\
&= \frac{\sigma_\mu^2 (1-\delta)^2}{NT_*} \sum_{t=1}^{T-1} \alpha^{2(t-1)} E(v_t' M_t^{l1} v_t) = \frac{\sigma_v^2 \sigma_\mu^2 (1-\delta)^2}{NT_*} \sum_{t=1}^{T-1} \alpha^{2(t-1)} tr(M_t^{l1}) \\
&\to 0.
\end{aligned}
$$

Next, the variance of Υ_{14NT} can be decomposed into two parts as follows:

$$
\begin{aligned}
var(\Upsilon_{14NT}) &= \frac{(1-\delta)^2}{NT_*} var\left(\sum_{t=1}^{T-1} \alpha^{t-1} \mu' M_t^{l1} \bar{v}_{tT}\right) \\
&= \frac{(1-\delta)^2}{NT_*} \left[\sum_{t=1}^{T-1} var\left(\alpha^{t-1} \mu' M_t^{l1} \bar{v}_{tT}\right)\right.
\end{aligned}
$$

$$+2\sum_{s}\sum_{t>s}cov(\alpha^{t-1}\mu' M_t^{l1}\bar{v}_{tT},\alpha^{s-1}\mu' M_s^{l1}\bar{v}_{sT})\Bigg].$$

For the first term, since $t/(T-t+1)<T$ for $t=1,...,T-1$, we have

$$var(\alpha^{t-1}\mu' M_t^{l1}\bar{v}_{tT})=\alpha^{2(t-1)}\frac{\sigma_v^2\sigma_\mu^2 tr(M_t^{l1})}{T-t+1}<T\alpha^{2(t-1)}\sigma_v^2\sigma_\mu^2.$$

Then, it follows that

$$\frac{\sigma_v^2\sigma_\mu^2(1-\delta)^2}{NT_*}\sum_{t=1}^{T-1}\alpha^{2(t-1)}\frac{t}{T-t+1} \quad < \quad \frac{\sigma_v^2\sigma_\mu^2(1-\delta)^2}{N}\sum_{t=1}^{T-1}\alpha^{2(t-1)}=O\left(\frac{1}{N}\right)\to 0.$$

For the second term, since

$$\begin{aligned}|cov(\alpha^{t-1}\mu' M_t^{l1}\bar{v}_{tT},\alpha^{s-1}\mu' M_s^{l1}\bar{v}_{sT})| &\le \sqrt{var(\alpha^{t-1}\mu' M_t^{l1}\bar{v}_{tT})}\sqrt{var(\alpha^{s-1}\mu' M_s^{l1}\bar{v}_{sT})}\\ &\le T\sigma_v^2\sigma_\mu^2\alpha^{t-1}\alpha^{s-1},\end{aligned}$$

it follows that

$$\begin{aligned}&\frac{(1-\delta)^2}{NT_*}\sum_{s}\sum_{t>s}|cov(\alpha^{t-1}\mu' M_t^{l1}\bar{v}_{tT},\alpha^{s-1}\mu' M_s^{l1}\bar{v}_{sT})|\\ \le\ &\frac{\sigma_v^2\sigma_\mu^2(1-\delta)^2}{N}\sum_{s}\sum_{t>s}\alpha^{t-1}\alpha^{s-1}\\ \le\ &\frac{\sigma_v^2\sigma_\mu^2(1-\delta)^2}{N}\sum_{t}\alpha^{t-1}\sum_{s}\alpha^{s-1}=O\left(\frac{1}{N}\right)\to 0.\end{aligned}$$

Thus, the variance of Υ_{14NT} is shown to tend to zero. Finally, we consider Υ_{15NT}. The variance of Υ_{15NT} is shown to tend to zero as follows:

$$\begin{aligned}var(\Upsilon_{15NT}) &= \frac{(1-\delta)^2}{NT_*}var\left(\sum_{t=1}^{T-1}\frac{c_t\alpha\phi_{T-t}}{T-t}\alpha^{t-1}\mu' M_t^{l1}v_t^*\right)\\ &= \frac{(1-\delta)^2}{NT_*}\sum_{t=1}^{T-1}\frac{c_t^2\alpha^2\phi_{T-t}^2}{(T-t)^2}\alpha^{2(t-1)}var\left(\mu' M_t^{l1}v_t^*\right)\\ &< \frac{(1-\delta)^2}{NT_*}\frac{\alpha^2}{(1-\alpha)^2}\left(\sum_{t=1}^{T-1}\frac{\alpha^{2(t-1)}}{(T-t)^2}var\left(\mu' M_t^{l1}v_t^*\right)\right)\\ &= \frac{(1-\delta)^2}{NT_*}\frac{\alpha^2\sigma_v^2\sigma_\mu^2}{(1-\alpha)^2}\left(\sum_{t=1}^{T-1}\frac{t\alpha^{2(t-1)}}{(T-t)^2}\right)\to 0.\end{aligned}$$

To prove the result, we used the fact that $E_t(v_t^*v_t^{*'})=\sigma_v^2 I_N$, $E_t(v_t^*v_s^{*'})=0$ for $t>s$, $c_t^2<1$, and $\phi_{T-t}^2<1/(1-\alpha)^2$, where $E_t(\cdot)$ denotes an expectation conditional on η_i and $\{v_{i,t-j}\}_{j=1}^{\infty}$.

Proof of Lemma 2(b), 3(b), 4(b) First, we consider Υ_{11NT}^*. Its variance is given by

$$\begin{aligned}var(\Upsilon_{11NT}^*) &= \frac{1}{NT_*}var\left(\sum_{t=t_0}^{T-1}c_t w_{t-1}' M_t v_t^*\right)=\frac{1}{NT_*}\sum_{t=t_0}^{T-1}\sum_{s=t_0}^{T-1}c_t c_s E(w_{t-1}' M_t v_t^* v_s^{*'} M_s w_{s-1})\\ &= \frac{\sigma_v^2}{NT_*}\sum_{t=t_0}^{T-1}c_t^2 E(w_{t-1}' M_t w_{t-1})+\frac{2}{NT}\sum_{s}\sum_{t>s}c_t c_s E(w_{t-1}' M_t v_t^* v_s^{*'} M_s w_{s-1})\end{aligned}$$

$$\rightarrow \quad \rho_2 \left(\frac{\sigma_v^4}{1-\alpha^2} \right)$$

where ρ_2 denotes ρ_{l2}, ρ_{d2}, 1 for the case of $M_t = M_t^{l2}, M_t^{d2}$, and M_t^{b2}, respectively. The last convergence comes from Lemma A4, $c_t^2 = 1 - O(1/(T-t))$ and $E_t(v_t^* v_s^*) = 0$ for $t > s$.

Next, the variance of Υ_{12NT} is shown to tend to zero as follows:

$$\begin{aligned}
var(\Upsilon_{12NT}) &= \frac{1}{NT_*} var\left(\sum_{t=t_0}^{T-1} \frac{c_t \alpha \phi_{T-t}}{T-t} w_{t-1}' M_t v_t^* \right) \\
&= \frac{\sigma_v^2}{NT_*} \sum_{t=t_0}^{T-1} \frac{\alpha^2 \phi_{T-t}^2}{(T-t)(T-t+1)} E(w_{t-1}' M_t w_{t-1}) \\
&\leq \frac{\sigma_v^2}{T_*} \sum_{t=t_0}^{T-1} \frac{\alpha^2 \phi_{T-t}^2}{(T-t)(T-t+1)} E(w_{i,t-1}^2) \\
&\leq \frac{constant}{T_*} \sum_{t=t_0}^{T-1} \frac{1}{(T-t)(T-t+1)} \rightarrow 0.
\end{aligned}$$

The variances of Υ_{13NT} and Υ_{14NT} are shown to tend to zero in similar way to the case of M_t^{l1}.

We then turn to consider $var(\Upsilon_{21NT})$:

$$\begin{aligned}
var(\Upsilon_{21NT}) &= \frac{1}{NT_*} var\left[\sum_{t=t_0}^{T-1} \frac{1}{T-t} v_t' M_t (\phi_{T-t} v_t + \cdots + \phi_1 v_{T-1}) \right] \\
&= a_{0NT} + a_{1NT}
\end{aligned}$$

where

$$\begin{aligned}
a_{0NT} &= \frac{1}{NT_*} \sum_{t=t_0}^{T-1} \frac{\phi_{T-t}^2 var(v_t' M_t v_t) + \cdots + \phi_1^2 var(v_t' M_t v_{T-1})}{(T-t)^2} \\
&= \frac{1}{NT_*} \sum_{t=t_0}^{T-1} \frac{\phi_{T-t}^2 [2\sigma_v^4 tr(M_t M_s) + \kappa_4 E(d_t' d_s)] + (\phi_{T-t-1}^2 + \cdots + \phi_1^2) tr(M_t M_s) \sigma_v^4}{(T-t)^2}
\end{aligned}$$

and

$$\begin{aligned}
a_{1NT} &= \frac{2}{NT_*} \sum_{t=t_0}^{T-2} \left[\frac{\phi_{T-t-1}^2 cov(v_t' M_t v_{t+1}, v_{t+1}' M_{t+1} v_{t+1})}{(T-t)(T-t-1)} \right. \\
&\qquad \left. + \cdots + \frac{\phi_1^2 cov(v_t' M_t v_{T-1}, v_{T-1}' M_{T-1} v_{T-1})}{(T-t)} \right] \\
&= \frac{2}{NT_*} \sum_{t=t_0}^{T-2} \left[\frac{\phi_{T-t-1}^2 \kappa_3 E(d_{t+1}' M_t v_t)}{(T-t)(T-t-1)} + \cdots + \frac{\phi_1^2 \kappa_3 E(d_{T-1}' M_t v_t)}{(T-t)} \right].
\end{aligned}$$

Using Lemma A3 and the fact that $\phi_j^2 < 1/(1-\alpha)^2$ for all j,

$$\begin{aligned}
a_{0NT} &\leq \frac{1}{NT_*} \sum_{t=t_0}^{T-1} \frac{\phi_{T-t}^2 [2\sigma_v^4 + \kappa_4] + (\phi_{T-t-1}^2 + \cdots + \phi_1^2) \sigma_v^4}{(T-t)^2} \\
&\leq \frac{1}{(1-\alpha)^2} \frac{1}{NT_*} \sum_{t=t_0}^{T-1} \frac{[2\sigma_v^4 + \kappa_4] + (T-t-1) \sigma_v^4}{(T-t)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2\sigma_v^4+\kappa_4}{(1-\alpha)^2}\frac{1}{NT_*}\sum_{t=t_0}^{T-1}\frac{1}{(T-t)^2}+\frac{\sigma_v^4}{(1-\alpha)^2}\frac{1}{NT_*}\sum_{t=t_0}^{T-1}\frac{T-t-1}{(T-t)^2} \\
&\to 0.
\end{aligned}
$$

From the triangle inequality, Lemma A3, and the fact that $|E(d'_{t+j}M_tv_t)| \le \sigma_v$,

$$
\begin{aligned}
|a_{1NT}| &\le \frac{2|\kappa_3|\sigma_v}{(1-\alpha)^2}\frac{1}{NT_*}\sum_{t=t_0}^{T-2}\frac{1}{T-t}\left[\frac{1}{(T-t-1)}+\cdots+\frac{1}{1}\right] \\
&< \frac{2|\kappa_3|\sigma_v}{(1-\alpha)^2}\frac{1}{NT_*}\sum_{t=t_0}^{T-2}\frac{1}{T-t}\left[\sum_{s=1}^{T-1}\frac{1}{s}\right]=O\left(\frac{(\log T)^2}{NT}\right)\to 0.
\end{aligned}
$$

Lastly, we consider the term Υ_{22NT}. We decompose the variance of Υ_{22NT} as follows:

$$
var(\Upsilon_{22NT})=\frac{1}{NT_*}var\left(\sum_{t=t_0}^{T-1}\bar{v}'_{tT}M_t\tilde{v}_{tT}\right)=b_{0NT}+b_{1NT}
$$

where

$$
b_{0NT}=\frac{1}{NT_*}\sum_{t=t_0}^{T-1}var(\bar{v}'_{tT}M_t\tilde{v}_{tT})
$$

and

$$
b_{1NT}=\frac{2}{NT_*}\sum_s\sum_{s>t}cov(\bar{v}'_{tT}M_t\tilde{v}_{tT},\bar{v}'_{sT}M_s\tilde{v}_{sT}).
$$

From (A73) in Alvarez and Arellano (2003), we have

$$
var(\bar{v}'_{t,T}M_t\tilde{v}_{t,T})=O\left(\frac{1}{(T-t)^2}\right).
$$

Hence, $b_{0NT}\to 0$. Next, with regard to the term b_{1NT}, we have

$$
\begin{aligned}
|b_{1NT}| &\le \frac{2}{NT_*}\sum_s\sum_{s>t}|cov(\bar{v}'_{t,T}M_t\tilde{v}_{t,T},\bar{v}'_{s,T}M_s\tilde{v}_{s,T})| \\
&\le \frac{2}{NT_*}\sum_s\sum_{s>t}\sqrt{var(\bar{v}'_{t,T}M_t\tilde{v}_{t,T})}\sqrt{var(\bar{v}'_{s,T}M_s\tilde{v}_{s,T})} \\
&\le \frac{2}{NT_*}\sum_s O\left(\frac{1}{T-t}\right)\sum_t O\left(\frac{1}{T-s}\right)=O\left(\frac{(\log T)^2}{NT}\right)\to 0.
\end{aligned}
$$

Proof of Lemma 1(d), 2(d), 3(d), 4(d),

We use the decomposition as follows:

$$
\begin{aligned}
\frac{x^{*'}Mx^*}{NT_*} &= \frac{1}{NT_*}\sum_{t=1}^{T-1}\psi_t^2w'_{t-1}M_tw_{t-1}+\frac{(1-\delta)^2}{NT_*}\sum_{t=1}^{T-1}\psi_t^2\alpha^{2(t-1)}\mu'M_t\mu \\
&\quad -\frac{2(1-\delta)}{NT_*}\sum_{t=1}^{T-1}\psi_t^2\alpha^{t-1}\mu'M_tw_{t-1} \\
&\quad +\frac{1}{NT_*}\sum_{t=1}^{T-1}c_t^2\tilde{v}'_{tT}M_t\tilde{v}_{tT}-\frac{2}{NT_*}\sum_{t=1}^{T-1}c_t\psi_t\tilde{v}'_{tT}M_t(y_{t-1}-\mu).
\end{aligned}
$$

Since the last four terms are easily shown to tend to zero, we consider only the first term. With regards to the first term, using Lemma A4 and noting $\psi_t^2=1-O[1/(T-t)]$, the results directly follow.

Proof of Theorem 1

It is straightforward to show consistency from Lemma 1. Next, we show the asymptotic normality of $\widehat{\alpha}_{G,l1}$. Alvarez and Arellano (2003) showed that

$$\frac{1}{\sqrt{NT_*}}\sum_{t=1}^{T-1} x_t^{*'} M_t^{l1} v_t^* - \mu_{G,l1} = \frac{1}{\sqrt{NT_*}}\sum_{t=1}^{T-1} w_{t-1}' v_t + o_p(1) \xrightarrow{d} N\left(0, \frac{\sigma_v^4}{1-\alpha^2}\right).$$

Hence, by Cramer's theorem, we have

$$\begin{aligned}
\left(\frac{x^{*'}M^{l1}x^*}{NT_*}\right)^{-1}\left(\frac{x^{*'}M^{l1}v^*}{\sqrt{NT_*}} - \mu_{G,l1}\right) &= \sqrt{NT_*}\left[\widehat{\alpha}_{G,l1} - \alpha - \frac{\mu_{G,l1}}{\sqrt{NT_*}}\left(\frac{x^{*'}M^{l1}x^*}{NT_*}\right)^{-1}\right] \\
&= \sqrt{NT_*}\left[\widehat{\alpha}_{G,l1} - \alpha - B_{G,l1}\right] + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, 1-\alpha^2)
\end{aligned}$$

where

$$B_{G,l1} = \frac{1}{\sqrt{NT_*}}\frac{\mu_{G,l1}}{R_T^{G,l1}}.$$

Proof of Theorem 2, 3

Consistency directly follows from Lemma 2 and 3. Next, we show the asymptotic normality. Using $w_{t-1}'Z_t/N = E(w_{i,t-1}z_{it}) + O_p(1/\sqrt{N})$, and $Z_t'Z_t/N = E(z_{it}^2) + O_p(1/\sqrt{N})$,

$$\begin{aligned}
\frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} x_t^{*'} M_t^{l2} v_t^* - \mu_{G,l2} &= \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} w_{t-1}' M_t^{l2} v_t + o_p(1) \\
&= \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1}\sum_{i=1}^{N} \frac{E(w_{i,t-1}z_{it})}{E(z_{it}^2)} z_{it} v_{it} + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&\xrightarrow{d} N\left(0, \rho_{l2}\left(\frac{\sigma_v^4}{1-\alpha^2}\right)\right).
\end{aligned}$$

Hence, by Cramer's theorem, we have

$$\begin{aligned}
\left(\frac{x^{*'}M^{l2}x^*}{NT_*}\right)^{-1}\left(\frac{x^{*'}M^{l2}v^*}{\sqrt{NT_*}} - \mu_{G,l2}\right) &= \sqrt{NT_*}\left[\widehat{\alpha}_{l2} - \alpha - \frac{\mu_{G,l2}}{\sqrt{NT_*}}\left(\frac{x^{*'}M^{l2}x^*}{NT_*}\right)^{-1}\right] \\
&= \sqrt{NT_*}\left[\widehat{\alpha}_{l2} - \alpha - B_{G,l2}\right] + o_p(1) \\
&\xrightarrow{d} N\left(0, (1-\alpha^2)\rho_{l2}^{-1}\right)
\end{aligned}$$

where

$$B_{G,l2} = \frac{1}{\sqrt{NT_*}}\frac{\mu_{G,l2}}{R_T^{G,l2}}.$$

The proof of Theorem 3 can be done exactly in the same way as Theorem 2.

Proof of Theorem 4

It is straightforward to show consistency from Lemma 4. Next, we show the asymptotic normality of $\widehat{\alpha}_{G,b2}$. To begin with, since the variances of Υ_{11NT}, Υ_{12NT}, Υ_{13NT}, Υ_{14NT}, Υ_{21NT}, and Υ_{22NT} are shown to tend to zero, note that

$$\begin{aligned} \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} x_t^{*'} M_t^{b2} v_t^* - \mu_{G,b2} &= \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} w_{t-1}' M_t^{b2} v_t + o_p(1) \\ &= \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} w_{t-1}' v_t - \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} w_{t-1}'(I_N - M_t^{b2}) v_t + o_p(1). \end{aligned} \tag{77}$$

The second term in (77) is shown to be $o_p(1)$ using the same argument as Alvarez and Arellano (2003). Therefore,

$$\frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} x_t^{*'} M_t^{b2} v_t^* - \mu_{G,b2} = \frac{1}{\sqrt{NT_*}}\sum_{t=2}^{T-1} w_{t-1}' v_t + o_p(1) \xrightarrow{d} N\left(0, \frac{\sigma_v^4}{(1-\alpha^2)}\right).$$

Using Cramer's theorem, we get the following result:

$$\begin{aligned} \left(\frac{x^{*'} M^{b2} x^*}{NT_*}\right)^{-1}\left(\frac{x^{*'} M^{b2} v^*}{\sqrt{NT_*}} - \mu_{G,b2}\right) &= \sqrt{NT_*}\left[\widehat{\alpha}_{G,b2} - \alpha - \frac{\mu_{G,b2}}{\sqrt{NT_*}}\left(\frac{x^{*'} M^{b2} x^*}{NT_*}\right)^{-1}\right] \\ &= \sqrt{NT_*}\left[\widehat{\alpha}_{G,b2} - \alpha - B_{G,b2}\right] + o_p(1) \\ &\to^d \mathcal{N}(0, 1-\alpha^2) \end{aligned}$$

where

$$B_{G,b2} = \frac{1}{\sqrt{NT_*}}\frac{\mu_{G,b2}}{R_T^{G,b2}}.$$

A.2 The WG estimator

Proof of Lemma 6 (a) ,

$$\begin{aligned} E\left(\frac{x^{*'} x^*}{NT}\right) &= E\left(\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} x_{it}^2 - \left(\frac{1}{T}\sum_{t=1}^{T} x_{it}\right)^2\right]\right) \\ &= \frac{1}{T}\sum_{t=1}^{T} E(x_{it}^2) - E\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}\right)^2. \end{aligned}$$

With regards to the first term, we have

$$\frac{1}{T}\sum_{t=1}^{T} E(x_{it}^2) = \frac{\sigma_v^2}{1-\alpha^2} + \sigma_\mu^2\left[1 - \frac{2(1-\delta)}{T}\frac{1-\alpha^T}{1-\alpha} + \frac{(1-\delta)^2}{T}\frac{1-\alpha^{2T}}{1-\alpha^2}\right].$$

For the second term, we have

$$E\left(\frac{1}{T}\sum_{t=1}^{T} x_{it}\right)^2 = E\left(\frac{1}{T}\sum_{t=1}^{T} \mu_{i,t-1}^*\right)^2 + E\left(\frac{1}{T}\sum_{t=1}^{T} w_{i,t-1}\right)^2$$

$$= \left[1 - \frac{(1-\delta)}{T}\frac{1-\alpha^T}{1-\alpha}\right]^2 \sigma_\mu^2 + \frac{1}{T}\left(\frac{\sigma_v^2}{1-\alpha^2}\right)\left[\frac{1+\alpha}{1-\alpha} - \frac{1}{T}\frac{2\alpha(1-\alpha^T)}{(1-\alpha)^2}\right].$$

Then the result follows. (b) See Alvarez and Arellano (2003). (c) We shall decompose as follows:

$$\begin{aligned} \frac{x^{*\prime} v^*}{\sqrt{NT}} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{i,t-1}v_{it} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_{i,t-1}^{*}v_{it} \\ &\quad -\sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{\mu}_{i(-1)}^{*}\bar{v}_i - \sqrt{\frac{T}{N}}\sum_{i=1}^{N}\bar{w}_{i(-1)}\bar{v}_i \\ &= \Psi_1 + \Psi_2 - \Psi_3 - \Psi_4. \end{aligned}$$

where $\bar{\mu}_{i(-1)}^{*} = \left(\sum_{t=1}^{T}\mu_{i,t-1}^{*}\right)/T$, $\bar{v}_i = \left(\sum_{t=1}^{T} v_{it}\right)/T$, and $\bar{w}_{i(-1)} = \left(\sum_{t=1}^{T} w_{i,t-1}\right)/T$. Since η_i and v_{it} are independent, we have

$$var\left(\frac{x^{*\prime} v^*}{\sqrt{NT}}\right) = var(\Psi_1) + var(\Psi_2) + var(\Psi_3) + var(\Psi_4) - 2cov(\Psi_2,\Psi_3) - 2cov(\Psi_1,\Psi_4)$$

where

$$var(\Psi_1) = var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_{i,t-1}v_{it}\right) = \frac{\sigma_v^4}{1-\alpha^2},$$

$$var(\Psi_2) = \frac{1}{T}\sum_{t=1}^{T} E(\mu_{i,t-1}^{*2}v_{it}^2) = \left[1 - \frac{2(1-\delta)}{T}\frac{1-\alpha^T}{1-\alpha} + \frac{(1-\delta)^2}{T}\frac{1-\alpha^{2T}}{1-\alpha^2}\right]\sigma_\mu^2\sigma_v^2,$$

$$var(\Psi_3) = Tvar(\bar{\mu}_{i(-1)}^{*}\bar{v}_i) = TE(\bar{\mu}_{i(-1)}^{*2})E(\bar{v}_i^2) = \sigma_v^2\sigma_\mu^2\left(1 - \frac{(1-\delta)}{T}\frac{1-\alpha^T}{1-\alpha}\right)^2,$$

$$var(\Psi_4) = Tvar(\bar{v}_i\bar{w}_{i(-1)}) = O(T^{-1}),$$

$$|cov(\Psi_1,\Psi_4)| \le \sqrt{var(\Psi_1)}\sqrt{var(\Psi_4)} = O\left(\frac{1}{\sqrt{T}}\right) \to 0,$$

$$\begin{aligned} cov(\Psi_2,\Psi_3) &= E\left[\left(\sum_{t=1}^{T}\mu_{i,t-1}^{*}v_{it}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mu_{i,t-1}^{*}\right)\left(\frac{1}{T}\sum_{t=1}^{T}v_{it}\right)\right] \\ &= \sigma_\mu^2\sigma_v^2\left(1 - \frac{(1-\delta)}{T}\frac{1-\alpha^T}{1-\alpha}\right)^2. \end{aligned}$$

Collecting these terms, the result follows.

Proof of Theorem 5 Let us define μ_{wg} as follows:

$$\mu_{wg} = \frac{1}{\sqrt{NT}}E(x^{*\prime}v^*) = -\sqrt{\frac{N}{T}}\frac{\sigma_v^2}{1-\alpha} + \sqrt{\frac{N}{T^3}}\frac{\sigma_v^2(1-\alpha^T)}{(1-\alpha)^2}. \tag{78}$$

Then using the similar arguments to Alvarez and Arellano (2003), it follows that

$$\frac{1}{\sqrt{NT}}x^{*\prime}v^* - \mu_{wg} \xrightarrow{d} N\left(0, \frac{\sigma_v^4}{1-\alpha^2}\right). \tag{79}$$

Hence from Cramer's theorem,

$$\left(\frac{x^{*\prime}x^*}{NT}\right)^{-1}\left[\frac{1}{\sqrt{NT}}x^{*\prime}v^* - \mu_{wg}\right] \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2), \tag{80}$$

or,

$$\sqrt{NT}\left(\widehat{\alpha}_{wg} - \alpha + \widehat{B}_{wg}\right) \xrightarrow{d} \mathcal{N}(0, 1-\alpha^2) \tag{81}$$

where $\widehat{B}_{wg} = (\mu_{wg}/\sqrt{NT})/(x^{*\prime}x^*/NT)$. Since $\widehat{B}_{wg} = plim_{N\to\infty}\widehat{B}_{wg} + o_p(1) = E(x^{*\prime}v^*)/E(x^{*\prime}x^*) + o_p(1)$, the result follows.

A.3 The LIML estimator

Proof of Lemma 6

From Lemma 1 and Alvarez and Arellano (2003, p.1149-1150), it is straightforward to show.

Proof of Theorem 6

It is straightforward to show consistency from Lemma 6. Next, we show the asymptotic normality of $\widehat{\alpha}_{L,l1}$. Alvarez and Arellano (2003) showed that

$$\begin{aligned}
&\frac{1}{\sqrt{NT_*}}\left(x^{*\prime}M^{l1}v^* - \widehat{\ell}_{l1}x^{*\prime}v^*\right) - (\mu_{G,l1} - \widehat{\ell}_{l1}\mu_{wg}) \\
&= \left(1-\frac{c}{2}\right)\frac{1}{\sqrt{NT_*}}\sum_{t=1}^{T-1} w_{t-1}'v_t + o_p(1) \xrightarrow{d} N\left(0, \left(1-\frac{c}{2}\right)^2 \frac{\sigma_v^4}{1-\alpha^2}\right).
\end{aligned}$$

Hence, using the similar arguments to Alvarez and Arellano (2003), we have

$$\begin{aligned}
&\left(\frac{x^{*\prime}M^{l1}x^* - \widehat{\ell}_{l1}x^{*\prime}x^*}{NT_*}\right)^{-1}\left(\frac{x^{*\prime}M^{l1}v^* - \widehat{\ell}_{l1}x^{*\prime}x^*}{\sqrt{NT_*}} - (\mu_{G,l1} - \widehat{\ell}_{l1}\mu_{wg})\right) \\
= &\quad \sqrt{NT_*}\left[\widehat{\alpha}_{L,l1} - \alpha - \frac{\mu_{G,l1} - \widehat{\ell}_{l1}\mu_{wg}}{\sqrt{NT_*}}\left(\frac{x^{*\prime}M^{l1}x^* - \widehat{\ell}_{l1}x^{*\prime}x^*}{NT_*}\right)^{-1}\right] \\
= &\quad \sqrt{NT_*}\left[\widehat{\alpha}_{L,l1} - \alpha - B_{L,l1}\right] + o_p(1) \\
\xrightarrow{d} &\quad \mathcal{N}(0, 1-\alpha^2)
\end{aligned}$$

where

$$B_{L,l1} = \frac{1}{\sqrt{NT_*}}\frac{\mu_{G,l1} - \frac{T_*}{2N}\mu_{wg}}{R_T^{G,l1} - \frac{T_*}{2N}R_T^{wg}}.$$

Proof of Lemma 7

Using $y_t^* = \alpha x_t^* + v_t^*$ and Lemma 2,3, and 4, we have

$$\frac{1}{NT_*}\sum_{t=t_0}^{T-1} y_t^{*\prime}M_t y_t^* \quad = \quad \frac{\alpha^2}{NT_*}\sum_{t=t_0}^{T-1} x_t^{*\prime}M_t x_t^* + \frac{2\alpha}{NT_*}\sum_{t=t_0}^{T-1} x_t^{*\prime}M_t v_t^* + \frac{1}{NT_*}\sum_{t=t_0}^{T-1} v_t^{*\prime}M_t v_t^*$$

$$= \frac{\alpha^2}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t x_t^* + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right).$$

$$\begin{aligned} \frac{1}{NT_*} \sum_{t=t_0}^{T-1} y_t^{*'} M_t x_t^* &= \frac{\alpha}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t x_t^* + \frac{1}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t v_t^* \\ &= \frac{\alpha}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t x_t^* + O_p\left(\frac{1}{\sqrt{NT}}\right). \end{aligned}$$

Then, as both N and T go to infinity, it follows that

$$\frac{W^{*'} M W}{NT_*} = \begin{pmatrix} \frac{1}{NT_*} \sum_{t=t_0}^{T-1} y_t^{*'} M_t y_t^* & \frac{1}{NT_*} \sum_{t=t_0}^{T-1} y_t^{*'} M_t x_t^* \\ \frac{1}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t y_t^* & \frac{1}{NT_*} \sum_{t=t_0}^{T-1} x_t^{*'} M_t x_t^* \end{pmatrix} \xrightarrow{\text{p}} \rho_2 \left(\frac{\sigma_v^2}{1-\alpha^2}\right) \begin{pmatrix} \alpha^2 & \alpha \\ \alpha & 1 \end{pmatrix},$$

$$\frac{W^{*'} W^*}{NT} \xrightarrow{\text{p}} \frac{\sigma_v^2}{1-\alpha^2} \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix} \tag{82}$$

where ρ_2 denotes ρ_{l2}, ρ_{d2}, and 1 for the case of $M_t = M_t^{l2}, M_t^{d2}$, and M_t^{b2}, respectively.

After some manipulation, it follows that the smallest eigenvalue of the probability limit of $W^{*'} M W^* (W^{*'} W^*)^{-1}$ is 0.

Proof of Theorem 7

From Lemma 7, it is straightforward to show.

Table 1: Median of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$						$\sigma_\eta^2 = 1$						$\sigma_\eta^2 = 10$					
α	T	N	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$
0.5	10	50	0.451	0.489	0.478	0.483	0.320	0.466	0.434	0.471	0.478	0.483	0.320	0.466	0.421	0.355	0.478	0.483	0.320	0.466
0.5	10	100	0.474	0.495	0.487	0.492	0.318	0.464	0.464	0.484	0.487	0.492	0.318	0.464	0.458	0.402	0.487	0.492	0.318	0.464
0.5	10	500	0.495	0.499	0.497	0.498	0.319	0.465	0.492	0.497	0.497	0.498	0.319	0.465	0.490	0.473	0.497	0.498	0.319	0.465
0.5	15	50	0.458	0.494	0.487	0.491	0.385	0.484	0.449	0.483	0.487	0.491	0.385	0.484	0.443	0.420	0.487	0.491	0.385	0.484
0.5	15	100	0.477	0.497	0.494	0.496	0.385	0.484	0.472	0.492	0.494	0.496	0.385	0.484	0.469	0.442	0.494	0.496	0.385	0.484
0.5	15	300	0.491	0.498	0.498	0.498	0.386	0.485	0.489	0.496	0.498	0.498	0.386	0.485	0.488	0.473	0.498	0.498	0.386	0.485
0.5	25	50	0.462	0.496	0.492	0.496	0.434	0.494	0.459	0.488	0.492	0.496	0.434	0.494	0.457	0.451	0.492	0.496	0.434	0.494
0.5	25	100	0.480	0.498	0.497	0.498	0.435	0.494	0.478	0.495	0.497	0.498	0.435	0.494	0.477	0.471	0.497	0.498	0.435	0.494
0.5	50	50	0.468	0.498	0.498	0.499	0.468	0.498	0.467	0.496	0.498	0.499	0.468	0.498	0.466	0.477	0.498	0.499	0.468	0.498
0.8	10	50	0.675	0.760	0.694	0.732	0.557	0.729	0.623	0.666	0.694	0.732	0.557	0.729	0.595	0.508	0.694	0.732	0.557	0.729
0.8	10	100	0.727	0.778	0.736	0.760	0.557	0.730	0.685	0.725	0.736	0.760	0.557	0.730	0.657	0.537	0.736	0.760	0.557	0.730
0.8	10	500	0.783	0.796	0.785	0.792	0.557	0.730	0.770	0.783	0.785	0.792	0.557	0.730	0.761	0.673	0.785	0.792	0.557	0.730
0.8	15	50	0.709	0.778	0.747	0.772	0.646	0.764	0.682	0.730	0.747	0.772	0.646	0.764	0.672	0.621	0.747	0.772	0.646	0.764
0.8	15	100	0.745	0.788	0.769	0.784	0.646	0.763	0.725	0.759	0.769	0.784	0.646	0.763	0.715	0.642	0.769	0.784	0.646	0.763
0.8	15	300	0.779	0.795	0.788	0.794	0.646	0.764	0.769	0.783	0.788	0.794	0.646	0.764	0.764	0.696	0.788	0.794	0.646	0.764
0.8	25	50	0.734	0.786	0.775	0.790	0.713	0.784	0.725	0.761	0.775	0.790	0.713	0.784	0.721	0.694	0.775	0.790	0.713	0.784
0.8	25	100	0.761	0.794	0.786	0.795	0.714	0.785	0.754	0.781	0.786	0.795	0.714	0.785	0.751	0.718	0.786	0.795	0.714	0.785
0.8	50	50	0.756	0.795	0.791	0.797	0.759	0.795	0.753	0.785	0.791	0.797	0.759	0.795	0.752	0.756	0.791	0.797	0.759	0.795
0.9	10	50	0.670	0.779	0.686	0.734	0.629	0.810	0.610	0.631	0.686	0.734	0.629	0.810	0.589	0.541	0.686	0.734	0.629	0.810
0.9	10	100	0.738	0.835	0.734	0.782	0.630	0.811	0.668	0.699	0.734	0.782	0.630	0.811	0.633	0.561	0.734	0.782	0.630	0.811
0.9	10	500	0.853	0.887	0.846	0.872	0.630	0.812	0.815	0.842	0.846	0.872	0.630	0.812	0.790	0.628	0.846	0.872	0.630	0.812
0.9	15	50	0.744	0.839	0.788	0.828	0.726	0.849	0.711	0.747	0.788	0.828	0.726	0.849	0.701	0.672	0.788	0.828	0.726	0.849
0.9	15	100	0.790	0.867	0.820	0.855	0.725	0.849	0.755	0.790	0.820	0.855	0.725	0.849	0.742	0.675	0.820	0.855	0.725	0.849
0.9	15	300	0.851	0.887	0.862	0.883	0.726	0.849	0.826	0.852	0.862	0.883	0.726	0.849	0.815	0.709	0.862	0.883	0.726	0.849
0.9	25	50	0.800	0.869	0.844	0.876	0.800	0.875	0.788	0.817	0.844	0.876	0.800	0.875	0.785	0.766	0.844	0.876	0.800	0.875
0.9	25	100	0.833	0.886	0.863	0.887	0.801	0.876	0.820	0.849	0.863	0.887	0.801	0.876	0.815	0.776	0.863	0.887	0.801	0.876
0.9	50	50	0.844	0.889	0.880	0.894	0.854	0.892	0.841	0.869	0.880	0.894	0.854	0.892	0.841	0.841	0.880	0.894	0.854	0.892

Table 1(cont.): Median of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$				$\sigma_\eta^2 = 1$				$\sigma_\eta^2 = 10$			
α	T	N	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$
0.5	10	50	0.482	0.498	0.499	0.495	0.474	0.491	0.499	0.495	0.467	0.393	0.499	0.495
0.5	10	100	0.492	0.499	0.498	0.497	0.487	0.495	0.498	0.497	0.485	0.443	0.498	0.497
0.5	10	500	0.498	0.500	0.499	0.499	0.498	0.499	0.499	0.499	0.497	0.488	0.499	0.499
0.5	15	50	0.481	0.499	0.498	0.497	0.476	0.495	0.498	0.497	0.471	0.452	0.498	0.497
0.5	15	100	0.492	0.500	0.500	0.499	0.489	0.498	0.500	0.499	0.486	0.467	0.500	0.499
0.5	15	300	0.497	0.499	0.500	0.499	0.496	0.498	0.500	0.499	0.495	0.486	0.500	0.499
0.5	25	50	0.477	0.498	0.498	0.498	0.474	0.494	0.498	0.498	0.472	0.465	0.498	0.498
0.5	25	100	0.490	0.500	0.500	0.499	0.489	0.498	0.500	0.499	0.488	0.486	0.500	0.499
0.5	50	50	0.469	0.499	0.500	0.500	0.466	0.498	0.500	0.500	0.464	0.485	0.500	0.500
0.8	10	50	0.764	0.792	0.780	0.786	0.724	0.737	0.780	0.786	0.679	0.323	0.780	0.786
0.8	10	100	0.784	0.796	0.791	0.790	0.769	0.777	0.791	0.790	0.753	0.492	0.791	0.790
0.8	10	500	0.797	0.799	0.799	0.799	0.794	0.795	0.799	0.799	0.793	0.738	0.799	0.799
0.8	15	50	0.765	0.795	0.795	0.793	0.738	0.771	0.795	0.793	0.720	0.522	0.795	0.793
0.8	15	100	0.784	0.797	0.796	0.796	0.773	0.786	0.796	0.796	0.766	0.634	0.796	0.796
0.8	15	300	0.794	0.798	0.799	0.798	0.791	0.794	0.799	0.798	0.789	0.738	0.799	0.798
0.8	25	50	0.758	0.794	0.798	0.797	0.743	0.780	0.798	0.797	0.737	0.612	0.798	0.797
0.8	25	100	0.783	0.799	0.799	0.799	0.778	0.793	0.799	0.799	0.775	0.727	0.799	0.799
0.8	50	50	0.748	0.798	0.799	0.799	0.737	0.793	0.799	0.799	0.733	0.743	0.799	0.799
0.9	10	50	0.777	0.859	0.780	0.826	0.501	0.646	0.780	0.826	0.285	0.028	0.780	0.826
0.9	10	100	0.866	0.889	0.836	0.868	0.773	0.775	0.836	0.868	0.650	0.231	0.836	0.868
0.9	10	500	0.895	0.898	0.895	0.896	0.889	0.888	0.895	0.896	0.883	0.622	0.895	0.896
0.9	15	50	0.800	0.884	0.868	0.884	0.626	0.789	0.868	0.884	0.497	0.156	0.868	0.884
0.9	15	100	0.866	0.893	0.887	0.891	0.825	0.842	0.887	0.891	0.793	0.325	0.887	0.891
0.9	15	300	0.891	0.897	0.895	0.896	0.883	0.883	0.895	0.896	0.878	0.662	0.895	0.896
0.9	25	50	0.802	0.889	0.894	0.895	0.705	0.845	0.894	0.895	0.643	0.262	0.894	0.895
0.9	25	100	0.871	0.897	0.896	0.897	0.855	0.879	0.896	0.897	0.846	0.604	0.896	0.897
0.9	50	50	0.786	0.896	0.897	0.898	0.728	0.883	0.897	0.898	0.698	0.704	0.897	0.898

Table 2: Interquartile range of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$						$\sigma_\eta^2 = 1$						$\sigma_\eta^2 = 10$					
α	T	N	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$
0.5	10	50	0.097	0.108	0.180	0.139	0.066	0.073	0.114	0.169	0.180	0.139	0.066	0.073	0.124	0.368	0.180	0.139	0.066	0.073
0.5	10	100	0.071	0.076	0.127	0.095	0.046	0.051	0.083	0.120	0.127	0.095	0.046	0.051	0.090	0.307	0.127	0.095	0.046	0.051
0.5	10	500	0.032	0.035	0.060	0.043	0.020	0.022	0.037	0.055	0.060	0.043	0.020	0.022	0.042	0.152	0.060	0.043	0.020	0.022
0.5	15	50	0.066	0.074	0.120	0.083	0.049	0.053	0.072	0.119	0.120	0.083	0.049	0.053	0.075	0.265	0.120	0.083	0.049	0.053
0.5	15	100	0.048	0.054	0.089	0.060	0.035	0.038	0.053	0.083	0.089	0.060	0.035	0.038	0.056	0.210	0.089	0.060	0.035	0.038
0.5	15	300	0.028	0.031	0.052	0.033	0.020	0.021	0.030	0.051	0.052	0.033	0.020	0.021	0.033	0.135	0.052	0.033	0.020	0.021
0.5	25	50	0.044	0.053	0.080	0.050	0.036	0.037	0.046	0.085	0.080	0.050	0.036	0.037	0.047	0.188	0.080	0.050	0.036	0.037
0.5	25	100	0.031	0.037	0.058	0.037	0.026	0.027	0.034	0.058	0.058	0.037	0.026	0.027	0.036	0.147	0.058	0.037	0.026	0.027
0.5	50	50	0.027	0.034	0.051	0.030	0.025	0.026	0.027	0.053	0.051	0.030	0.025	0.026	0.028	0.120	0.051	0.030	0.025	0.026
0.8	10	50	0.134	0.170	0.292	0.233	0.062	0.069	0.161	0.312	0.292	0.233	0.062	0.069	0.178	0.426	0.292	0.233	0.062	0.069
0.8	10	100	0.102	0.116	0.223	0.177	0.043	0.048	0.128	0.224	0.223	0.177	0.043	0.048	0.140	0.424	0.223	0.177	0.043	0.048
0.8	10	500	0.047	0.055	0.113	0.080	0.020	0.022	0.062	0.104	0.113	0.080	0.020	0.022	0.072	0.313	0.113	0.080	0.020	0.022
0.8	15	50	0.082	0.103	0.176	0.121	0.045	0.048	0.093	0.198	0.176	0.121	0.045	0.048	0.097	0.312	0.176	0.121	0.045	0.048
0.8	15	100	0.060	0.076	0.139	0.089	0.031	0.033	0.070	0.142	0.139	0.089	0.031	0.033	0.077	0.292	0.139	0.089	0.031	0.033
0.8	15	300	0.035	0.044	0.083	0.049	0.018	0.019	0.042	0.082	0.083	0.049	0.018	0.019	0.046	0.239	0.083	0.049	0.018	0.019
0.8	25	50	0.046	0.065	0.106	0.059	0.029	0.030	0.049	0.121	0.106	0.059	0.029	0.030	0.050	0.208	0.106	0.059	0.029	0.030
0.8	25	100	0.034	0.045	0.082	0.044	0.021	0.022	0.038	0.086	0.082	0.044	0.021	0.022	0.039	0.194	0.082	0.044	0.021	0.022
0.8	50	50	0.023	0.036	0.062	0.028	0.020	0.020	0.024	0.065	0.062	0.028	0.020	0.020	0.024	0.120	0.062	0.028	0.020	0.020
0.9	10	50	0.186	0.278	0.366	0.339	0.061	0.067	0.207	0.405	0.366	0.339	0.061	0.067	0.208	0.416	0.366	0.339	0.061	0.067
0.9	10	100	0.152	0.192	0.323	0.274	0.042	0.047	0.181	0.372	0.323	0.274	0.042	0.047	0.189	0.426	0.323	0.274	0.042	0.047
0.9	10	500	0.073	0.091	0.187	0.138	0.019	0.021	0.099	0.183	0.187	0.138	0.019	0.021	0.117	0.414	0.187	0.138	0.019	0.021
0.9	15	50	0.107	0.161	0.219	0.174	0.043	0.046	0.119	0.281	0.219	0.174	0.043	0.046	0.122	0.310	0.219	0.174	0.043	0.046
0.9	15	100	0.084	0.115	0.190	0.134	0.030	0.032	0.098	0.234	0.190	0.134	0.030	0.032	0.103	0.305	0.190	0.134	0.030	0.032
0.9	15	300	0.051	0.065	0.126	0.078	0.017	0.018	0.063	0.139	0.126	0.078	0.017	0.018	0.069	0.296	0.126	0.078	0.017	0.018
0.9	25	50	0.053	0.090	0.124	0.078	0.027	0.028	0.057	0.168	0.124	0.078	0.027	0.028	0.058	0.205	0.124	0.078	0.027	0.028
0.9	25	100	0.041	0.062	0.107	0.058	0.019	0.020	0.046	0.128	0.107	0.058	0.019	0.020	0.048	0.193	0.107	0.058	0.019	0.020
0.9	50	50	0.023	0.043	0.068	0.032	0.016	0.016	0.023	0.083	0.068	0.032	0.016	0.016	0.023	0.114	0.068	0.032	0.016	0.016

Table 2(cont.): Interquartile range of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$				$\sigma_\eta^2 = 1$				$\sigma_\eta^2 = 10$			
α	T	N	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$
0.5	10	50	0.108	0.111	0.203	0.146	0.137	0.183	0.203	0.146	0.161	0.780	0.203	0.146
0.5	10	100	0.075	0.078	0.136	0.099	0.091	0.125	0.136	0.099	0.102	0.451	0.136	0.099
0.5	10	500	0.032	0.035	0.061	0.044	0.038	0.055	0.061	0.044	0.043	0.165	0.061	0.044
0.5	15	50	0.074	0.076	0.134	0.085	0.088	0.128	0.134	0.085	0.096	0.520	0.134	0.085
0.5	15	100	0.050	0.055	0.094	0.061	0.058	0.087	0.094	0.061	0.064	0.306	0.094	0.061
0.5	15	300	0.028	0.031	0.053	0.034	0.032	0.051	0.053	0.034	0.033	0.151	0.053	0.034
0.5	25	50	0.051	0.055	0.088	0.051	0.057	0.091	0.088	0.051	0.060	0.363	0.088	0.051
0.5	25	100	0.033	0.037	0.060	0.038	0.038	0.060	0.060	0.038	0.040	0.202	0.060	0.038
0.5	50	50	0.033	0.034	0.055	0.031	0.037	0.057	0.055	0.031	0.039	0.211	0.055	0.031
0.8	10	50	0.207	0.187	0.488	0.304	0.382	0.499	0.488	0.304	0.544	1.789	0.488	0.304
0.8	10	100	0.128	0.122	0.293	0.202	0.199	0.278	0.293	0.202	0.258	1.656	0.293	0.202
0.8	10	500	0.050	0.056	0.121	0.081	0.069	0.109	0.121	0.081	0.082	0.488	0.121	0.081
0.8	15	50	0.128	0.111	0.265	0.137	0.199	0.282	0.265	0.137	0.253	1.659	0.265	0.137
0.8	15	100	0.075	0.079	0.171	0.096	0.102	0.165	0.171	0.096	0.117	1.322	0.171	0.096
0.8	15	300	0.038	0.044	0.091	0.050	0.047	0.085	0.091	0.050	0.052	0.467	0.091	0.050
0.8	25	50	0.077	0.069	0.146	0.063	0.105	0.166	0.146	0.063	0.119	1.426	0.146	0.063
0.8	25	100	0.042	0.046	0.097	0.045	0.053	0.097	0.097	0.045	0.056	0.822	0.097	0.045
0.8	50	50	0.048	0.037	0.078	0.029	0.061	0.083	0.078	0.029	0.066	0.851	0.078	0.029
0.9	10	50	0.737	0.402	0.956	0.695	1.576	1.594	0.956	0.695	1.795	1.986	0.956	0.695
0.9	10	100	0.296	0.231	0.649	0.428	0.771	1.036	0.649	0.428	1.204	1.985	0.649	0.428
0.9	10	500	0.084	0.094	0.238	0.152	0.134	0.215	0.238	0.152	0.173	1.641	0.238	0.152
0.9	15	50	0.480	0.208	0.571	0.278	1.247	1.221	0.571	0.278	1.463	1.964	0.571	0.278
0.9	15	100	0.156	0.128	0.353	0.171	0.328	0.520	0.353	0.171	0.482	1.908	0.353	0.171
0.9	15	300	0.062	0.068	0.162	0.083	0.089	0.167	0.162	0.083	0.106	1.616	0.162	0.083
0.9	25	50	0.274	0.109	0.265	0.095	0.712	0.649	0.265	0.095	0.894	1.932	0.265	0.095
0.9	25	100	0.076	0.067	0.165	0.064	0.118	0.201	0.165	0.064	0.138	1.765	0.165	0.064
0.9	50	50	0.172	0.048	0.119	0.034	0.325	0.173	0.119	0.034	0.399	1.731	0.119	0.034

Table 3: Median absolute error of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$						$\sigma_\eta^2 = 1$						$\sigma_\eta^2 = 10$					
α	T	N	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$
0.5	10	50	0.060	0.055	0.095	0.070	0.180	0.043	0.075	0.085	0.095	0.070	0.180	0.043	0.086	0.196	0.095	0.070	0.180	0.043
0.5	10	100	0.040	0.038	0.064	0.048	0.182	0.038	0.048	0.060	0.064	0.048	0.182	0.038	0.055	0.158	0.064	0.048	0.182	0.038
0.5	10	500	0.017	0.017	0.030	0.022	0.181	0.035	0.019	0.027	0.030	0.022	0.181	0.035	0.022	0.077	0.030	0.022	0.181	0.035
0.5	15	50	0.046	0.037	0.061	0.042	0.115	0.028	0.053	0.058	0.061	0.042	0.115	0.028	0.059	0.130	0.061	0.042	0.115	0.028
0.5	15	100	0.028	0.027	0.045	0.030	0.115	0.021	0.033	0.042	0.045	0.030	0.115	0.021	0.036	0.104	0.045	0.030	0.115	0.021
0.5	15	300	0.015	0.016	0.026	0.017	0.114	0.016	0.017	0.025	0.026	0.017	0.114	0.016	0.018	0.068	0.026	0.017	0.114	0.016
0.5	25	50	0.039	0.027	0.041	0.025	0.066	0.019	0.042	0.041	0.041	0.025	0.066	0.019	0.044	0.091	0.041	0.025	0.066	0.019
0.5	25	100	0.022	0.018	0.029	0.019	0.065	0.014	0.024	0.029	0.029	0.019	0.065	0.014	0.025	0.072	0.029	0.019	0.065	0.014
0.5	50	50	0.032	0.017	0.026	0.015	0.032	0.013	0.033	0.026	0.026	0.015	0.032	0.013	0.034	0.058	0.026	0.015	0.032	0.013
0.8	10	50	0.126	0.086	0.169	0.127	0.243	0.071	0.177	0.166	0.169	0.127	0.243	0.071	0.205	0.302	0.169	0.127	0.243	0.071
0.8	10	100	0.077	0.060	0.123	0.093	0.243	0.070	0.116	0.117	0.123	0.093	0.243	0.070	0.144	0.277	0.123	0.093	0.243	0.070
0.8	10	500	0.026	0.027	0.057	0.040	0.243	0.070	0.037	0.053	0.057	0.040	0.243	0.070	0.046	0.164	0.057	0.040	0.243	0.070
0.8	15	50	0.091	0.052	0.096	0.062	0.154	0.037	0.118	0.098	0.096	0.062	0.154	0.037	0.128	0.188	0.096	0.062	0.154	0.037
0.8	15	100	0.055	0.037	0.073	0.045	0.154	0.037	0.075	0.070	0.073	0.045	0.154	0.037	0.085	0.169	0.073	0.045	0.154	0.037
0.8	15	300	0.024	0.022	0.043	0.025	0.154	0.036	0.033	0.042	0.043	0.025	0.154	0.036	0.037	0.125	0.043	0.025	0.154	0.036
0.8	25	50	0.066	0.032	0.056	0.030	0.087	0.019	0.075	0.060	0.056	0.030	0.087	0.019	0.079	0.116	0.056	0.030	0.087	0.019
0.8	25	100	0.039	0.023	0.042	0.022	0.086	0.016	0.046	0.042	0.042	0.022	0.086	0.016	0.049	0.096	0.042	0.022	0.086	0.016
0.8	50	50	0.044	0.018	0.032	0.014	0.041	0.010	0.047	0.032	0.032	0.014	0.041	0.010	0.048	0.061	0.032	0.014	0.041	0.010
0.9	10	50	0.230	0.153	0.250	0.208	0.271	0.090	0.290	0.280	0.250	0.208	0.271	0.090	0.311	0.361	0.250	0.208	0.271	0.090
0.9	10	100	0.162	0.104	0.209	0.162	0.270	0.089	0.232	0.216	0.209	0.162	0.270	0.089	0.267	0.344	0.209	0.162	0.270	0.089
0.9	10	500	0.051	0.046	0.103	0.073	0.270	0.088	0.087	0.097	0.103	0.073	0.270	0.088	0.111	0.283	0.103	0.073	0.270	0.088
0.9	15	50	0.156	0.085	0.141	0.100	0.174	0.051	0.189	0.161	0.141	0.100	0.174	0.051	0.199	0.231	0.141	0.100	0.174	0.051
0.9	15	100	0.110	0.059	0.112	0.073	0.175	0.051	0.145	0.123	0.112	0.073	0.175	0.051	0.158	0.227	0.112	0.073	0.175	0.051
0.9	15	300	0.049	0.034	0.070	0.041	0.174	0.051	0.074	0.073	0.070	0.041	0.174	0.051	0.085	0.195	0.070	0.041	0.174	0.051
0.9	25	50	0.100	0.046	0.075	0.042	0.100	0.025	0.112	0.089	0.075	0.042	0.100	0.025	0.115	0.137	0.075	0.042	0.100	0.025
0.9	25	100	0.067	0.031	0.058	0.030	0.099	0.024	0.080	0.063	0.058	0.030	0.099	0.024	0.085	0.126	0.058	0.030	0.099	0.024
0.9	50	50	0.056	0.021	0.037	0.016	0.046	0.010	0.059	0.040	0.037	0.016	0.046	0.010	0.059	0.062	0.037	0.016	0.046	0.010

Table 3(Cont.): Median absolute error of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$				$\sigma_\eta^2 = 1$				$\sigma_\eta^2 = 10$			
α	T	N	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$
0.5	10	50	0.056	0.056	0.101	0.073	0.070	0.091	0.101	0.073	0.082	0.323	0.101	0.073
0.5	10	100	0.037	0.039	0.068	0.049	0.046	0.062	0.068	0.049	0.053	0.209	0.068	0.049
0.5	10	500	0.016	0.017	0.030	0.022	0.019	0.028	0.030	0.022	0.021	0.082	0.030	0.022
0.5	15	50	0.039	0.038	0.067	0.043	0.047	0.063	0.067	0.043	0.052	0.219	0.067	0.043
0.5	15	100	0.026	0.027	0.047	0.031	0.030	0.043	0.047	0.031	0.033	0.138	0.047	0.031
0.5	15	300	0.014	0.016	0.026	0.017	0.016	0.025	0.026	0.017	0.017	0.075	0.026	0.017
0.5	25	50	0.030	0.027	0.044	0.025	0.034	0.045	0.044	0.025	0.036	0.153	0.044	0.025
0.5	25	100	0.019	0.019	0.030	0.019	0.020	0.030	0.030	0.019	0.021	0.095	0.030	0.019
0.5	50	50	0.032	0.017	0.027	0.015	0.035	0.028	0.027	0.015	0.036	0.097	0.027	0.015
0.8	10	50	0.106	0.093	0.244	0.151	0.189	0.224	0.244	0.151	0.264	0.939	0.244	0.151
0.8	10	100	0.064	0.061	0.148	0.101	0.099	0.138	0.148	0.101	0.130	0.717	0.148	0.101
0.8	10	500	0.025	0.028	0.061	0.041	0.034	0.054	0.061	0.041	0.041	0.222	0.061	0.041
0.8	15	50	0.065	0.055	0.132	0.068	0.100	0.125	0.132	0.068	0.123	0.713	0.132	0.068
0.8	15	100	0.038	0.039	0.085	0.048	0.053	0.079	0.085	0.048	0.063	0.437	0.085	0.048
0.8	15	300	0.019	0.022	0.045	0.025	0.025	0.043	0.045	0.025	0.028	0.181	0.045	0.025
0.8	25	50	0.047	0.034	0.073	0.031	0.062	0.073	0.073	0.031	0.069	0.438	0.073	0.031
0.8	25	100	0.024	0.023	0.048	0.023	0.029	0.047	0.048	0.023	0.031	0.214	0.048	0.023
0.8	50	50	0.052	0.019	0.039	0.014	0.063	0.039	0.039	0.014	0.067	0.184	0.039	0.014
0.9	10	50	0.323	0.191	0.495	0.348	0.801	0.603	0.495	0.348	1.034	1.266	0.495	0.348
0.9	10	100	0.146	0.117	0.334	0.219	0.326	0.363	0.334	0.219	0.512	1.176	0.334	0.219
0.9	10	500	0.042	0.047	0.119	0.076	0.066	0.108	0.119	0.076	0.088	0.651	0.119	0.076
0.9	15	50	0.191	0.098	0.286	0.137	0.465	0.350	0.286	0.137	0.672	1.147	0.286	0.137
0.9	15	100	0.079	0.063	0.176	0.086	0.150	0.181	0.176	0.086	0.203	1.002	0.176	0.086
0.9	15	300	0.032	0.034	0.081	0.042	0.047	0.081	0.081	0.042	0.055	0.563	0.081	0.042
0.9	25	50	0.116	0.051	0.132	0.047	0.241	0.163	0.132	0.047	0.321	1.038	0.132	0.047
0.9	25	100	0.040	0.033	0.082	0.032	0.059	0.085	0.082	0.032	0.070	0.656	0.082	0.032
0.9	50	50	0.114	0.023	0.059	0.017	0.172	0.060	0.059	0.017	0.202	0.571	0.059	0.017

Table 4: Empirical size of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$						$\sigma_\eta^2 = 1$						$\sigma_\eta^2 = 10$					
α	T	N	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$	$\widehat{\alpha}_{G,l1}$	$\widehat{\alpha}_{G,l2}$	$\widehat{\alpha}_{G,d2}$	$\widehat{\alpha}_{G,b2}$	$\widehat{\alpha}_{wg}$	$\widehat{\alpha}_{hk}$
0.5	10	50	0.108	0.054	0.054	0.056	0.982	0.165	0.123	0.056	0.054	0.056	0.982	0.165	0.138	0.033	0.054	0.056	0.982	0.165
0.5	10	100	0.086	0.056	0.061	0.052	1.000	0.236	0.095	0.059	0.061	0.052	1.000	0.236	0.096	0.049	0.061	0.052	1.000	0.236
0.5	10	500	0.058	0.052	0.057	0.056	1.000	0.665	0.057	0.054	0.057	0.056	1.000	0.665	0.059	0.062	0.057	0.056	1.000	0.665
0.5	15	50	0.139	0.051	0.055	0.051	0.906	0.095	0.157	0.054	0.055	0.051	0.906	0.095	0.166	0.045	0.055	0.051	0.906	0.095
0.5	15	100	0.104	0.052	0.050	0.049	0.996	0.125	0.115	0.053	0.050	0.049	0.996	0.125	0.119	0.052	0.050	0.049	0.996	0.125
0.5	15	300	0.072	0.054	0.049	0.054	1.000	0.211	0.072	0.060	0.049	0.054	1.000	0.211	0.075	0.067	0.049	0.054	1.000	0.211
0.5	25	50	0.220	0.055	0.047	0.051	0.723	0.078	0.239	0.060	0.047	0.051	0.723	0.078	0.242	0.057	0.047	0.051	0.723	0.078
0.5	25	100	0.128	0.052	0.048	0.053	0.944	0.071	0.145	0.052	0.048	0.053	0.944	0.071	0.154	0.053	0.048	0.053	0.944	0.071
0.5	50	50	0.376	0.058	0.049	0.050	0.438	0.060	0.387	0.058	0.049	0.050	0.438	0.060	0.397	0.056	0.049	0.050	0.438	0.060
0.8	10	50	0.280	0.074	0.076	0.070	1.000	0.440	0.373	0.072	0.076	0.070	1.000	0.440	0.417	0.051	0.076	0.070	1.000	0.440
0.8	10	100	0.194	0.065	0.068	0.064	1.000	0.676	0.267	0.080	0.068	0.064	1.000	0.676	0.315	0.049	0.068	0.064	1.000	0.676
0.8	10	500	0.086	0.054	0.063	0.059	1.000	0.998	0.105	0.061	0.063	0.059	1.000	0.998	0.117	0.070	0.063	0.059	1.000	0.998
0.8	15	50	0.383	0.065	0.071	0.063	0.999	0.283	0.467	0.076	0.071	0.063	0.999	0.283	0.502	0.052	0.071	0.063	0.999	0.283
0.8	15	100	0.256	0.057	0.065	0.056	1.000	0.438	0.329	0.075	0.065	0.056	1.000	0.438	0.367	0.054	0.065	0.056	1.000	0.438
0.8	15	300	0.130	0.062	0.053	0.055	1.000	0.828	0.158	0.071	0.053	0.055	1.000	0.828	0.173	0.078	0.053	0.055	1.000	0.828
0.8	25	50	0.570	0.069	0.054	0.055	0.991	0.143	0.623	0.083	0.054	0.055	0.991	0.143	0.640	0.064	0.054	0.055	0.991	0.143
0.8	25	100	0.374	0.058	0.050	0.058	1.000	0.199	0.434	0.067	0.050	0.058	1.000	0.199	0.455	0.063	0.050	0.058	1.000	0.199
0.8	50	50	0.797	0.061	0.056	0.059	0.865	0.078	0.819	0.062	0.056	0.059	0.865	0.078	0.829	0.065	0.056	0.059	0.865	0.078
0.9	10	50	0.522	0.090	0.111	0.095	1.000	0.637	0.616	0.084	0.111	0.095	1.000	0.637	0.649	0.076	0.111	0.095	1.000	0.637
0.9	10	100	0.393	0.081	0.102	0.094	1.000	0.869	0.524	0.084	0.102	0.094	1.000	0.869	0.585	0.067	0.102	0.094	1.000	0.869
0.9	10	500	0.158	0.058	0.077	0.069	1.000	1.000	0.232	0.082	0.077	0.069	1.000	1.000	0.288	0.080	0.077	0.069	1.000	1.000
0.9	15	50	0.688	0.089	0.107	0.086	1.000	0.490	0.762	0.085	0.107	0.086	1.000	0.490	0.787	0.075	0.107	0.086	1.000	0.490
0.9	15	100	0.541	0.077	0.097	0.074	1.000	0.736	0.658	0.090	0.097	0.074	1.000	0.736	0.696	0.073	0.097	0.074	1.000	0.736
0.9	15	300	0.273	0.068	0.065	0.059	1.000	0.990	0.390	0.091	0.065	0.059	1.000	0.990	0.443	0.083	0.065	0.059	1.000	0.990
0.9	25	50	0.864	0.092	0.079	0.070	1.000	0.312	0.901	0.109	0.079	0.070	1.000	0.312	0.910	0.085	0.079	0.070	1.000	0.312
0.9	25	100	0.708	0.072	0.070	0.065	1.000	0.488	0.785	0.094	0.070	0.065	1.000	0.488	0.809	0.078	0.070	0.065	1.000	0.488
0.9	50	50	0.977	0.073	0.069	0.062	0.994	0.142	0.981	0.088	0.069	0.062	0.994	0.142	0.983	0.083	0.069	0.062	0.994	0.142

Table 4(Cont.): Empirical size of the GMM, (bias corrected) WG, and LIML estimators

			$\sigma_\eta^2 = 0.2$				$\sigma_\eta^2 = 1$				$\sigma_\eta^2 = 10$			
α	T	N	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$	$\widehat{\alpha}_{L,l1}$	$\widehat{\alpha}_{L,l2}$	$\widehat{\alpha}_{L,d2}$	$\widehat{\alpha}_{L,b2}$
0.5	10	50	0.089	0.058	0.075	0.066	0.119	0.077	0.075	0.066	0.151	0.245	0.075	0.066
0.5	10	100	0.071	0.058	0.070	0.057	0.083	0.066	0.070	0.057	0.087	0.180	0.070	0.057
0.5	10	500	0.055	0.051	0.059	0.058	0.055	0.053	0.059	0.058	0.053	0.077	0.059	0.058
0.5	15	50	0.098	0.057	0.074	0.056	0.125	0.078	0.074	0.056	0.149	0.288	0.074	0.056
0.5	15	100	0.072	0.055	0.061	0.050	0.085	0.057	0.061	0.050	0.098	0.185	0.061	0.050
0.5	15	300	0.062	0.054	0.054	0.054	0.063	0.061	0.054	0.054	0.063	0.097	0.054	0.054
0.5	25	50	0.164	0.060	0.066	0.056	0.197	0.079	0.066	0.056	0.210	0.300	0.066	0.056
0.5	25	100	0.086	0.053	0.056	0.054	0.102	0.063	0.056	0.054	0.111	0.168	0.056	0.054
0.5	50	50	0.389	0.062	0.065	0.052	0.442	0.077	0.065	0.052	0.459	0.265	0.065	0.052
0.8	10	50	0.246	0.092	0.193	0.140	0.440	0.226	0.193	0.140	0.540	0.380	0.193	0.140
0.8	10	100	0.137	0.069	0.122	0.086	0.239	0.148	0.122	0.086	0.315	0.344	0.122	0.086
0.8	10	500	0.065	0.053	0.068	0.063	0.080	0.064	0.068	0.063	0.090	0.227	0.068	0.063
0.8	15	50	0.276	0.084	0.173	0.087	0.423	0.234	0.173	0.087	0.495	0.547	0.173	0.087
0.8	15	100	0.147	0.062	0.104	0.068	0.219	0.126	0.104	0.068	0.266	0.485	0.104	0.068
0.8	15	300	0.084	0.062	0.062	0.059	0.096	0.079	0.062	0.059	0.100	0.320	0.062	0.059
0.8	25	50	0.396	0.082	0.136	0.069	0.488	0.211	0.136	0.069	0.523	0.597	0.136	0.069
0.8	25	100	0.185	0.063	0.086	0.062	0.250	0.106	0.086	0.062	0.269	0.477	0.086	0.062
0.8	50	50	0.755	0.072	0.120	0.064	0.792	0.150	0.120	0.064	0.802	0.562	0.120	0.064
0.9	10	50	0.595	0.207	0.290	0.244	0.752	0.347	0.290	0.244	0.801	0.414	0.290	0.244
0.9	10	100	0.382	0.129	0.254	0.185	0.601	0.292	0.254	0.185	0.693	0.384	0.254	0.185
0.9	10	500	0.097	0.061	0.102	0.078	0.159	0.121	0.102	0.078	0.215	0.346	0.102	0.078
0.9	15	50	0.625	0.206	0.399	0.243	0.766	0.466	0.399	0.243	0.809	0.591	0.399	0.243
0.9	15	100	0.371	0.109	0.275	0.134	0.576	0.351	0.275	0.134	0.651	0.576	0.275	0.134
0.9	15	300	0.139	0.073	0.105	0.068	0.213	0.159	0.105	0.068	0.256	0.516	0.105	0.068
0.9	25	50	0.686	0.162	0.345	0.113	0.785	0.463	0.345	0.113	0.816	0.682	0.345	0.113
0.9	25	100	0.394	0.092	0.202	0.086	0.517	0.296	0.202	0.086	0.567	0.647	0.202	0.086
0.9	50	50	0.902	0.112	0.255	0.073	0.925	0.349	0.255	0.073	0.933	0.694	0.255	0.073

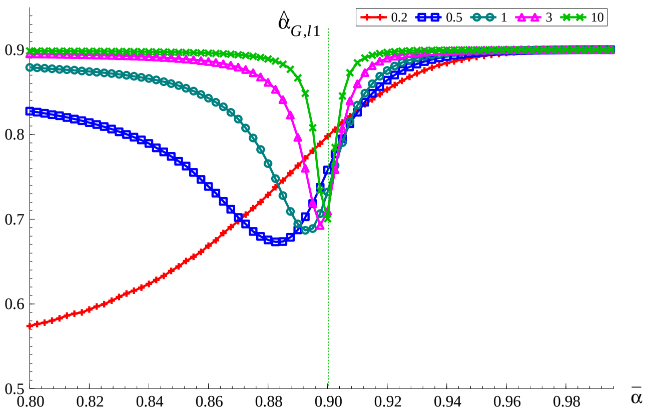

Figure 1: Median of $\widehat{\alpha}_{G,l1}$ $(T = 10, N = 200)$

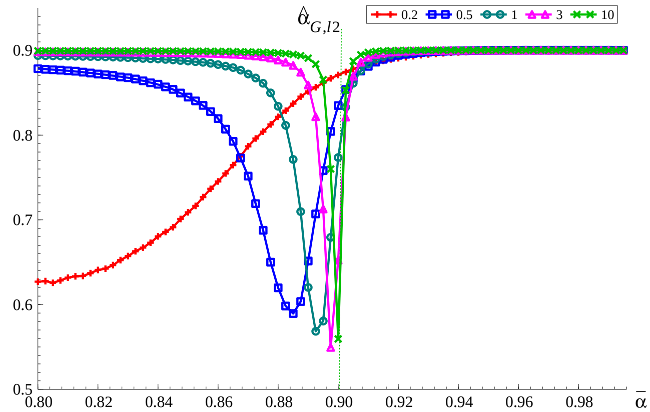

Figure 2: Median of $\widehat{\alpha}_{G,l2}$ $(T = 10, N = 200)$

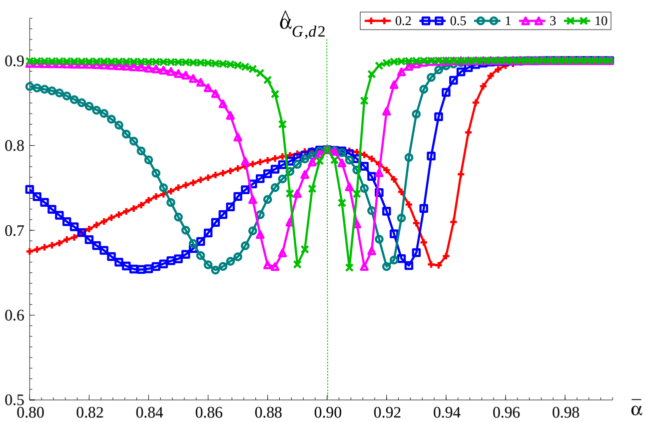

Figure 3: Median of $\widehat{\alpha}_{G,d2}$ $(T = 10, N = 200)$

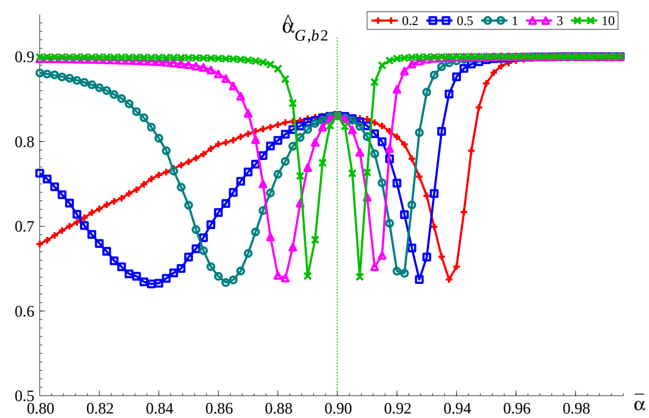

Figure 4: Median of $\widehat{\alpha}_{G,b2}$ $(T = 10, N = 200)$

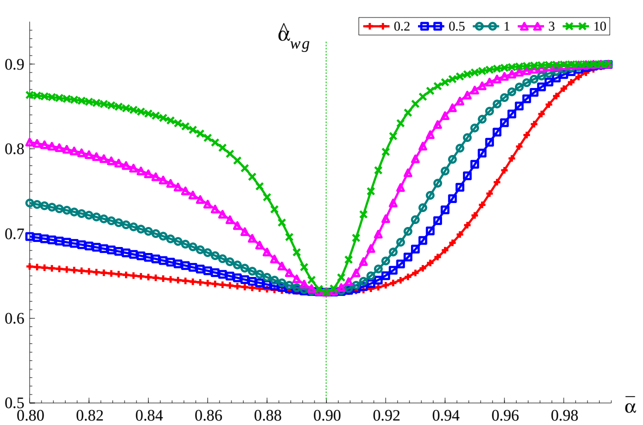

Figure 5: Median of $\widehat{\alpha}_{wg}$ $(T = 10, N = 200)$

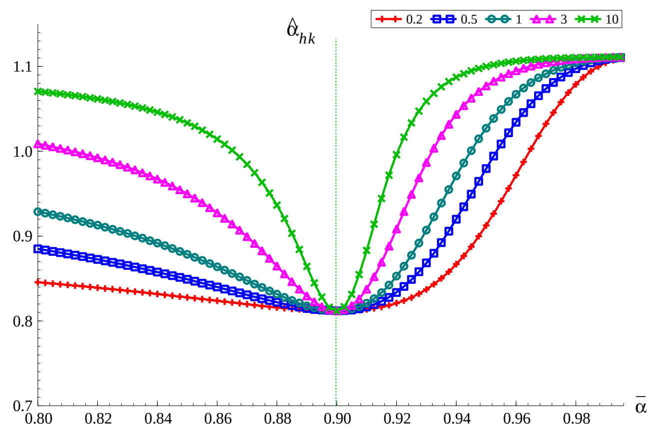

Figure 6: Median of $\widehat{\alpha}_{hk}$ $(T = 10, N = 200)$

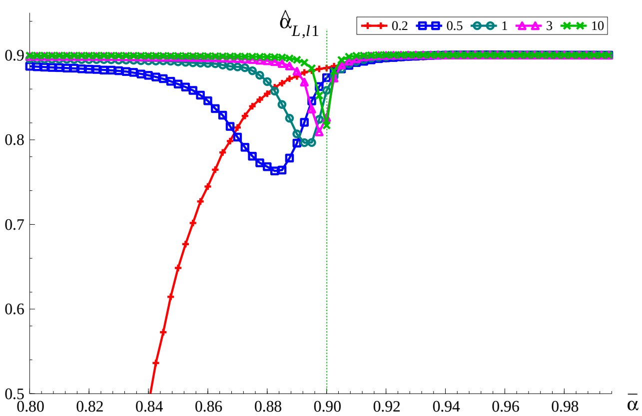

Figure 7: Median of $\widehat{\alpha}_{L,l1}$ ($T = 10, N = 200$)

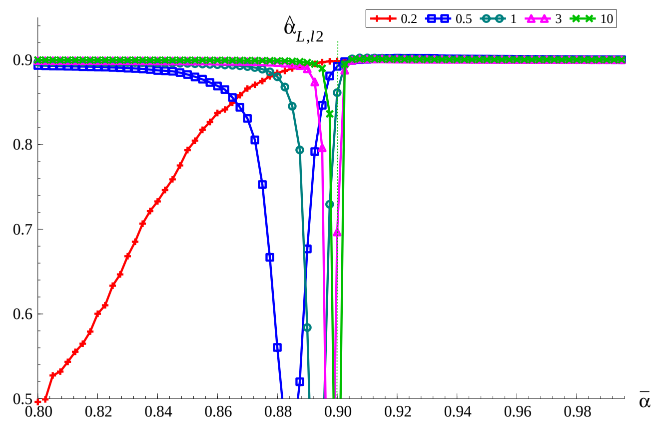

Figure 8: Median of $\widehat{\alpha}_{L,l2}$ ($T = 10, N = 200$)

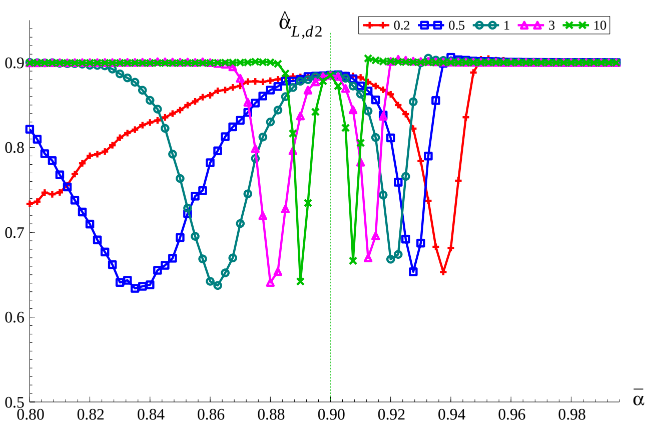

Figure 9: Median of $\widehat{\alpha}_{L,d2}$ ($T = 10, N = 200$)

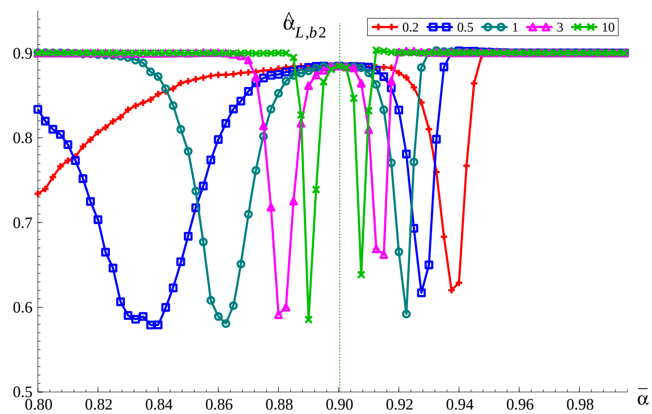

Figure 10: Median of $\widehat{\alpha}_{L,b2}$ ($T = 10, N = 200$)

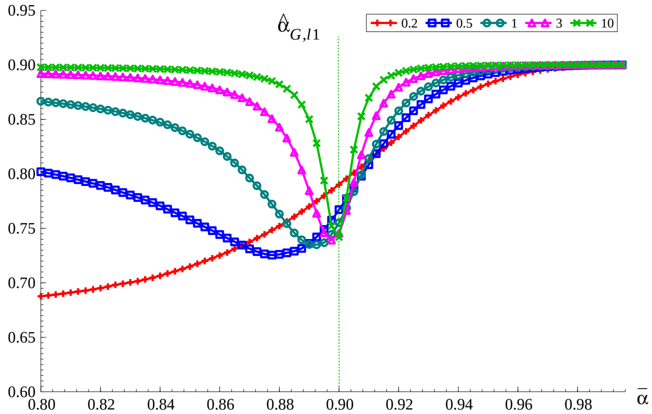

Figure 11: Median of $\widehat{\alpha}_{G,l1}$ $(T = 15, N = 100)$

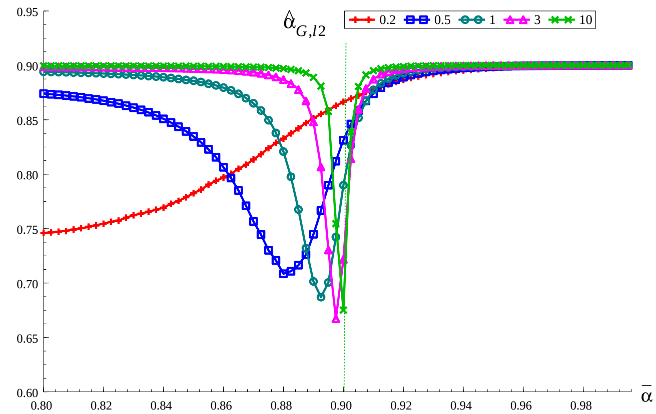

Figure 12: Median of $\widehat{\alpha}_{G,l2}$ $(T = 15, N = 100)$

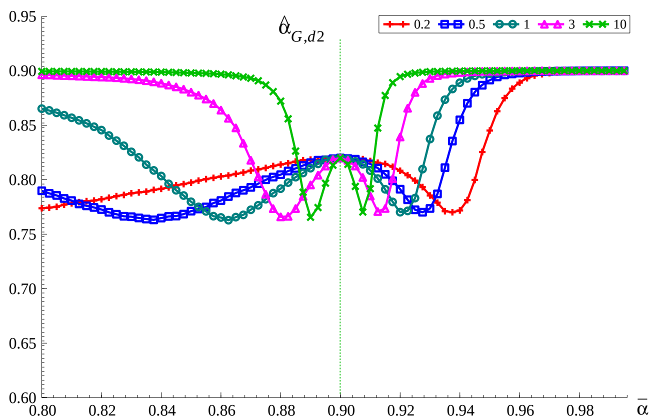

Figure 13: Median of $\widehat{\alpha}_{G,d2}$ $(T = 15, N = 100)$

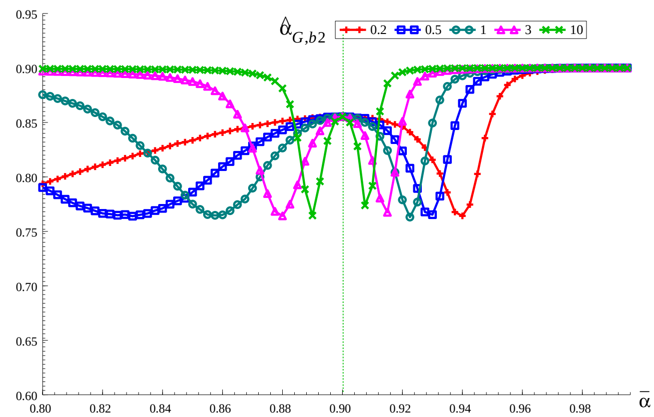

Figure 14: Median of $\widehat{\alpha}_{G,b2}$ $(T = 15, N = 100)$

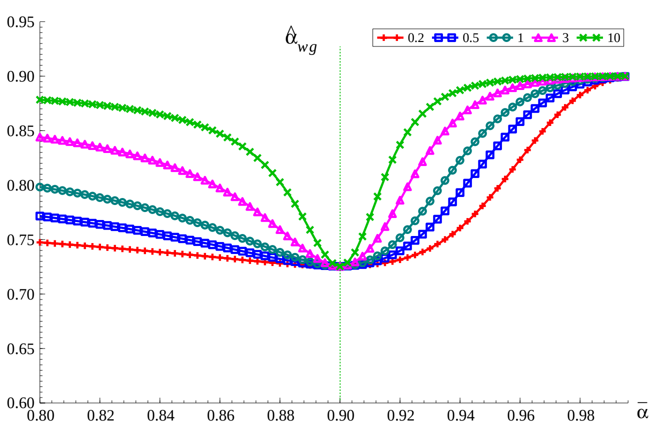

Figure 15: Median of $\widehat{\alpha}_{wg}$ $(T = 15, N = 100)$

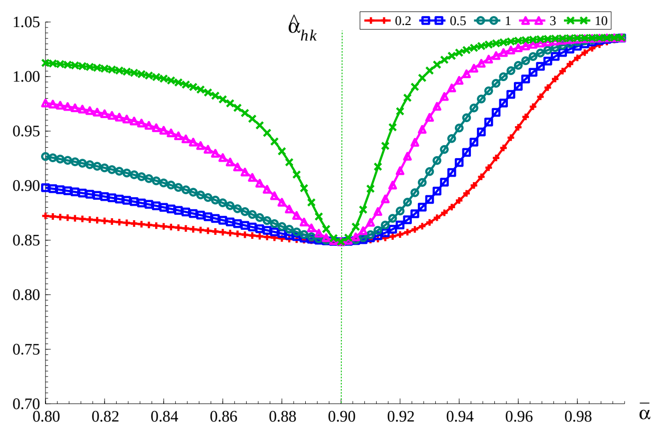

Figure 16: Median of $\widehat{\alpha}_{hk}$ $(T = 15, N = 100)$

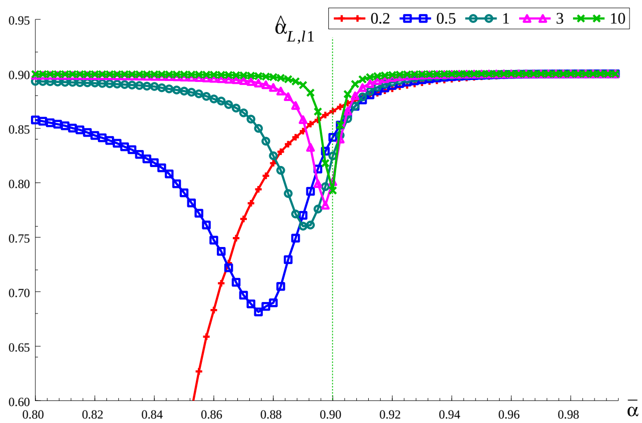

Figure 17: Median of $\widehat{\alpha}_{L,l1}$ ($T = 15, N = 100$)

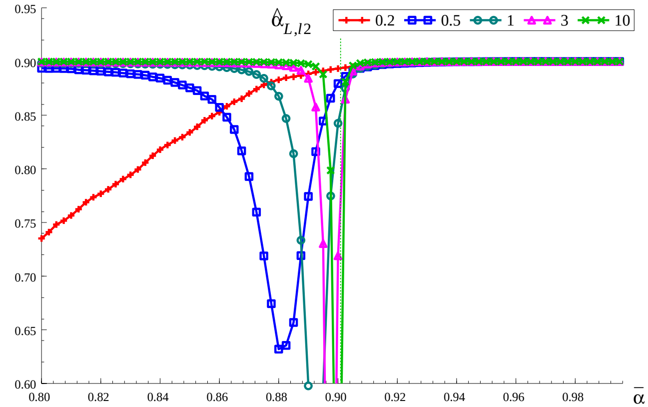

Figure 18: Median of $\widehat{\alpha}_{L,l2}$ ($T = 15, N = 100$)

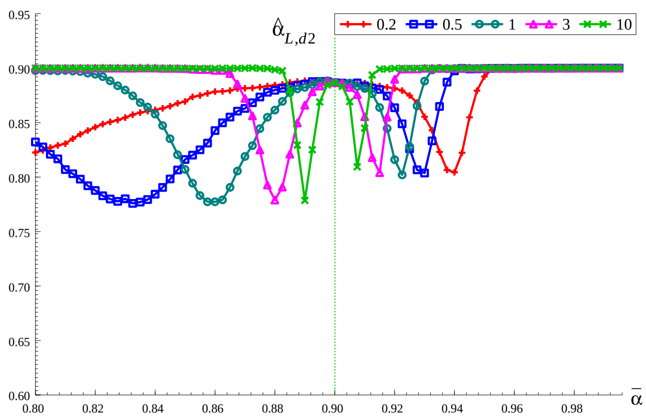

Figure 19: Median of $\widehat{\alpha}_{L,d2}$ ($T = 15, N = 100$)

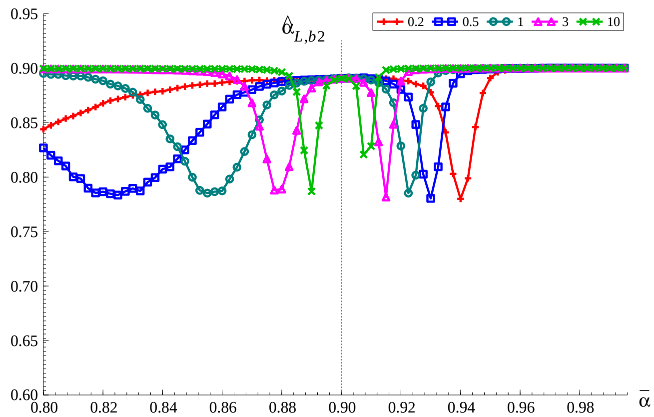

Figure 20: Median of $\widehat{\alpha}_{L,b2}$ ($T = 15, N = 100$)

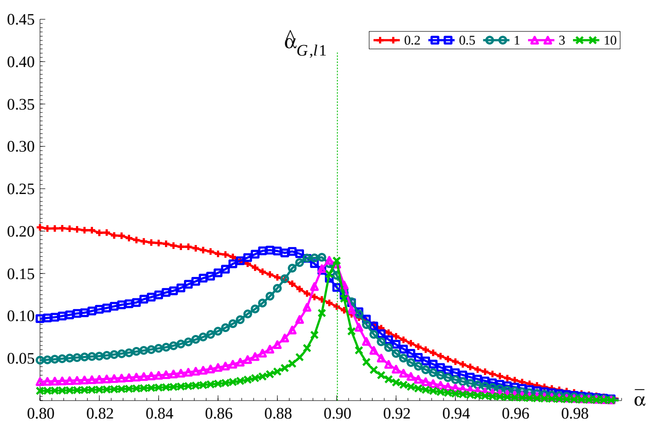

Figure 21: IQR of $\widehat{\alpha}_{G,l1}$ $(T = 10, N = 200)$

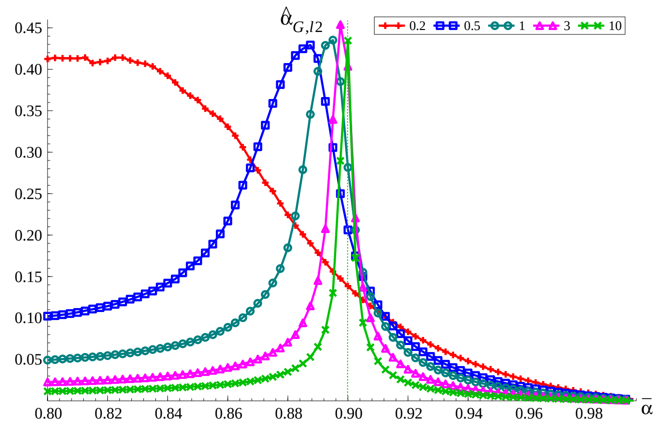

Figure 22: IQR of $\widehat{\alpha}_{G,l2}$ $(T = 10, N = 200)$

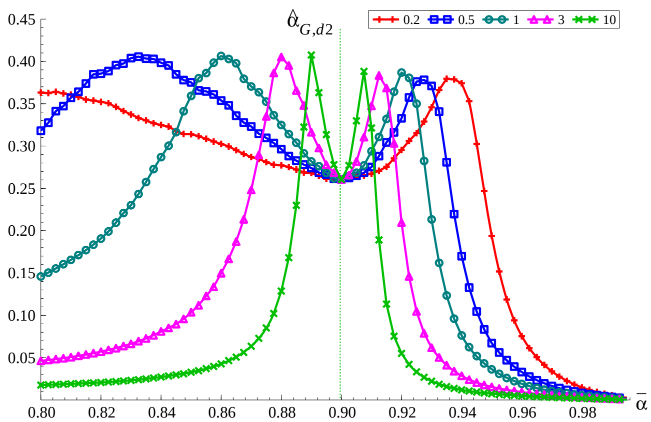

Figure 23: IQR of $\widehat{\alpha}_{G,d2}$ $(T = 10, N = 200)$

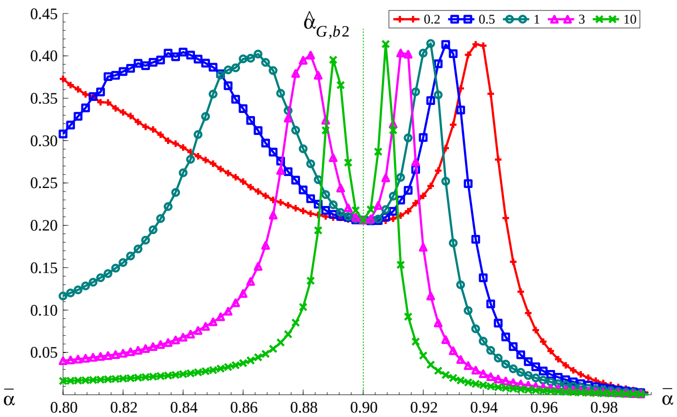

Figure 24: IQR of $\widehat{\alpha}_{G,b2}$ $(T = 10, N = 200)$

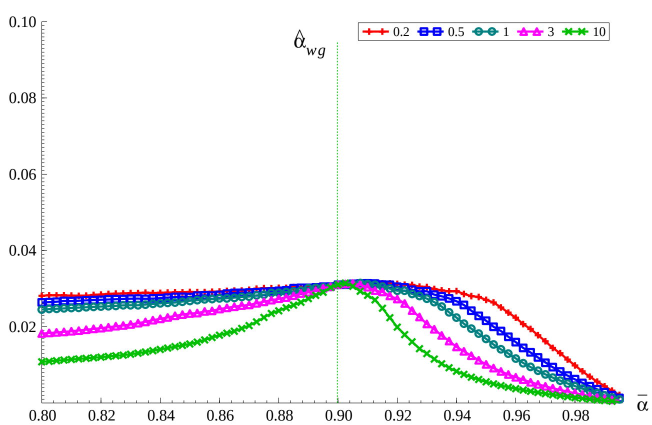

Figure 25: IQR of $\widehat{\alpha}_{wg}$ $(T = 10, N = 200)$

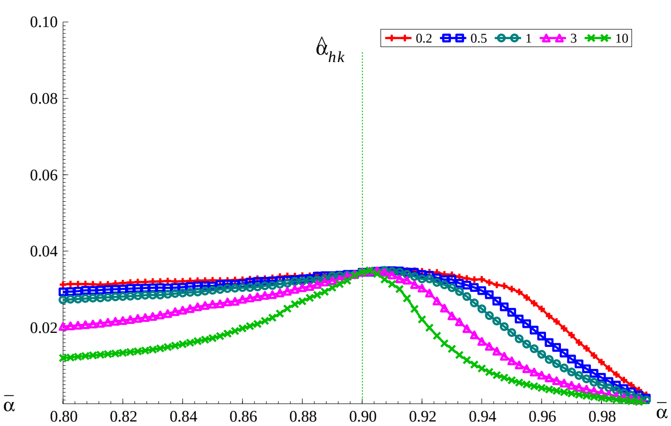

Figure 26: IQR of $\widehat{\alpha}_{hk}$ $(T = 10, N = 200)$

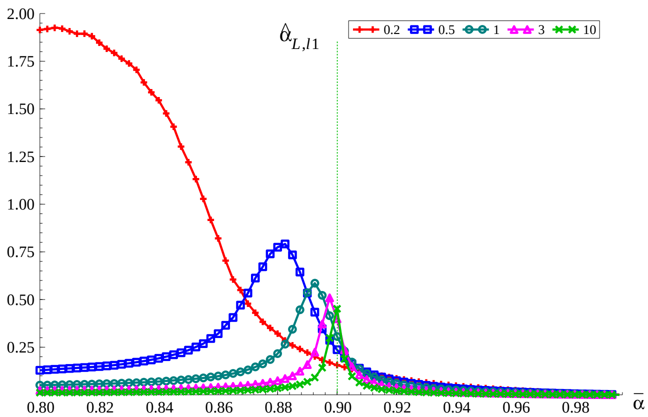

Figure 27: IQR of $\widehat{\alpha}_{L,l1}$ ($T = 10, N = 200$)

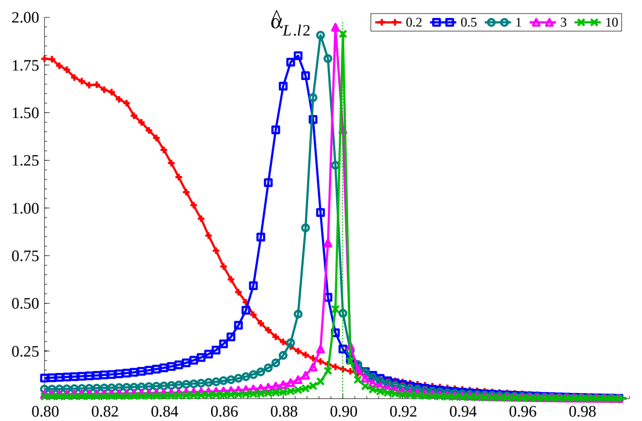

Figure 28: IQR of $\widehat{\alpha}_{L,l2}$ ($T = 10, N = 200$)

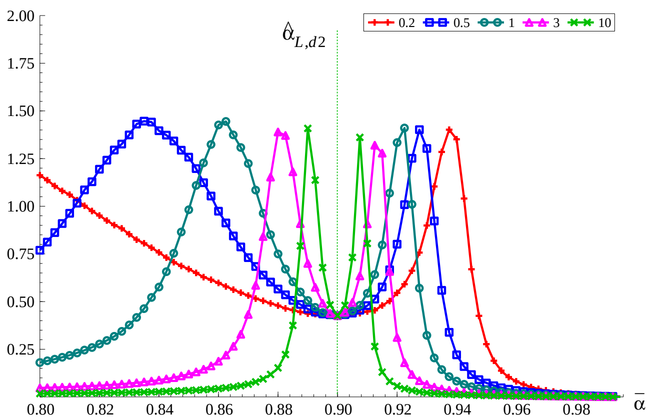

Figure 29: IQR of $\widehat{\alpha}_{L,d2}$ ($T = 10, N = 200$)

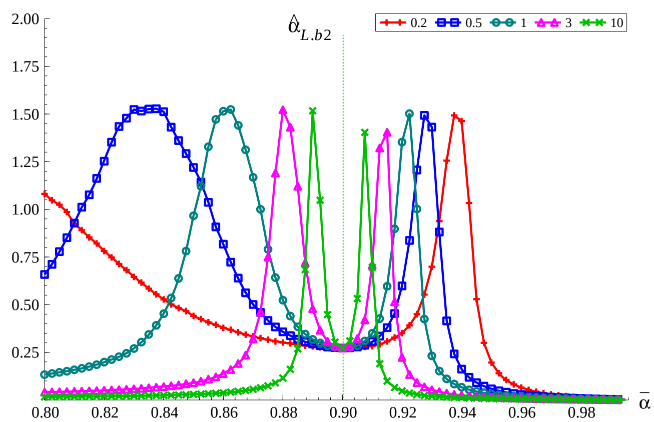

Figure 30: IQR of $\widehat{\alpha}_{L,b2}$ ($T = 10, N = 200$)

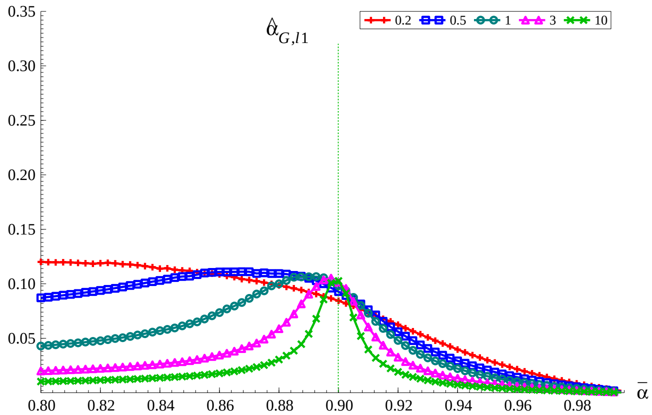

Figure 31: IQR of $\widehat{\alpha}_{G,l1}$ ($T = 15, N = 100$)

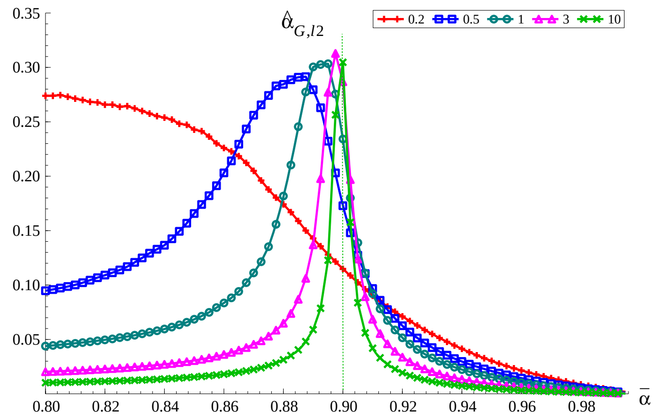

Figure 32: IQR of $\widehat{\alpha}_{G,l2}$ ($T = 15, N = 100$)

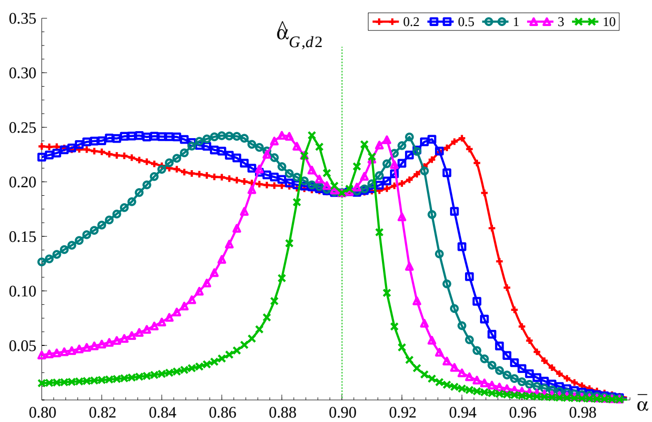

Figure 33: IQR of $\widehat{\alpha}_{G,d2}$ ($T = 15, N = 100$)

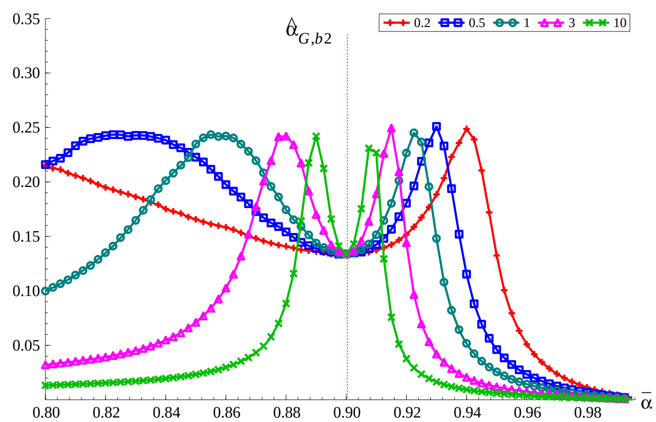

Figure 34: IQR of $\widehat{\alpha}_{G,b2}$ ($T = 15, N = 100$)

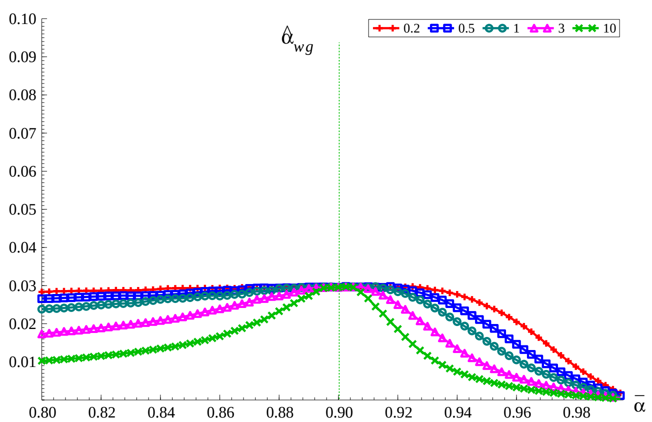

Figure 35: IQR of $\widehat{\alpha}_{wg}$ ($T = 15, N = 100$)

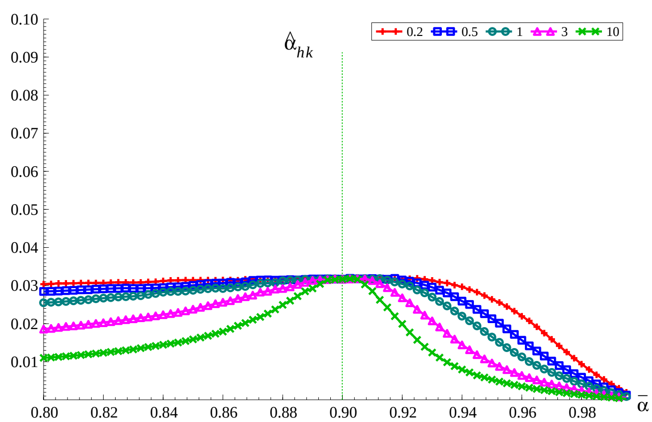

Figure 36: IQR of $\widehat{\alpha}_{hk}$ ($T = 15, N = 100$)

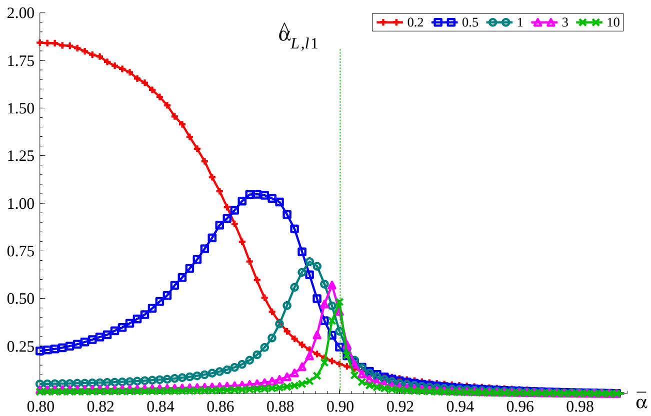

Figure 37: IQR of $\widehat{\alpha}_{L,l1}$ ($T = 15, N = 100$)

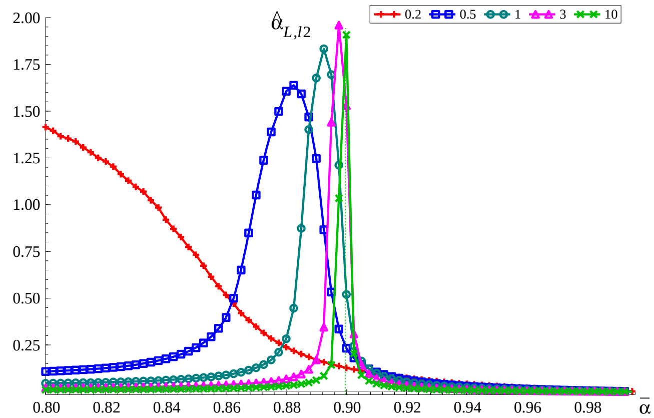

Figure 38: IQR of $\widehat{\alpha}_{L,l2}$ ($T = 15, N = 100$)

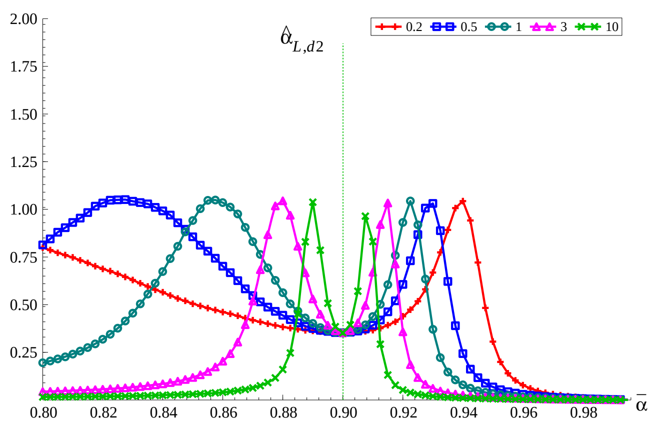

Figure 39: IQR of $\widehat{\alpha}_{L,d2}$ ($T = 15, N = 100$)

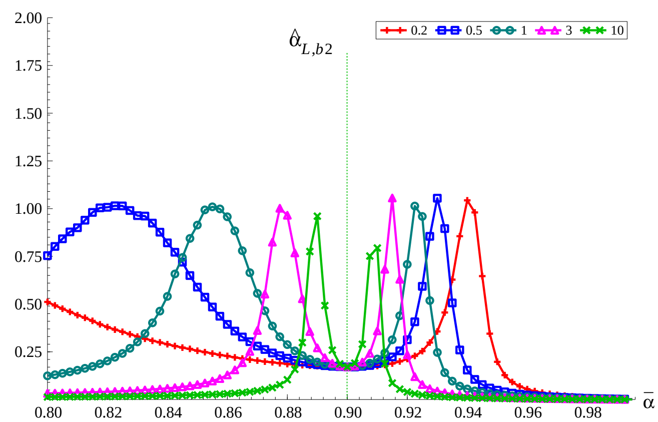

Figure 40: IQR of $\widehat{\alpha}_{L,b2}$ ($T = 15, N = 100$)