E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models

Cheng Hsiao
University of Southern California

Qiankun Zhou
Louisiana State University

Working Paper 2017-11
http://faculty.bus.lsu.edu/papers/pap17_11.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models*

Cheng Hsiao† Qiankun Zhou‡

This version August 27, 2017

Abstract

We use a quasi-likelihood function approach to clarify the role of initial values and the relative size of the cross-section dimension N and the time series dimension T in the asymptotic distribution of dynamic panel data models with the presence of individual-specific effects. We show that the quasi-maximum likelihood estimator (QMLE) treating initial values as fixed constants is asymptotically biased of order $\sqrt{\frac{N}{T^3}}$ as T goes to infinity for a time series models and asymptotically biased of order $\sqrt{\frac{N}{T}}$ for a model that also contains other covariates that are correlated with the individual-specific effects. Using Mundlak-Chamberlain approach to condition the effects on the covariates can reduce the asymptotic bias to the order of $\sqrt{\frac{N}{T^3}}$, provided the data generating processes for the covariates are homogeneous across cross-sectional units. On the other hand, the QMLE combining the Mundlak-Chamberlain approach with the proper treatment of initial value distribution is asymptotically unbiased if N goes to infinity whether T is fixed or goes to infinity. Monte Carlo studies are conducted to demonstrate the importance of properly treating initial values in getting valid statistical inference. The results also suggest that when using the conditional approach to get around the issue of incidental parameters, in finite sample it is perhaps better to follow Mundlak's (1978) suggestion to simply condition the individual

*This paper was stimulated by the private communication with Jushan Bai. We would like to thank him for pointing out the terms that cancel the impact of individual-specific effects and the impact of correlation between the errors of the equation and the lagged dependent variables. We would also like to thank the editor Oliver Linton, an associate editor, two anonymous referees and Elie Tamer for helpful comments. Partial research support by China NSF #71103004 and #71631004 to the first author is also gratefully acknowledged.

†Department of Economics, University of Southern California, University Park, Los Angeles, California 90089, Email: chsiao@usc.edu, Department of Quantitative Finance, National Tsing Hua University and WISE, Xiamen University, China.

‡Department of Economics, Louisiana State University, Baton Rouge, LA 70803. Email: qzhou@lsu.edu.

effects or initial values on the time series average of individual's observed regressors under the assumption that our model is correctly specified.
Keywords: Dynamic panel models, Individual effects, Initial values, Projection method, Conditional or unconditional likelihood approach.
JEL classification: C01, C13, C23

1 Introduction

In the estimation of dynamic panel data models with the presence of time-invariant individual effects, three issues have arisen (e.g., Hsiao (2014)): (i) whether the unobserved individual-specific effects should be treated as fixed or random? (2) whether the initial values should be treated as fixed constants or random? (iii) does the relative size of cross-sectional dimension N and time series dimension T matter? We argue in this paper that all three issues matter in obtaining consistent estimation of unknown parameters and obtaining valid statistical inference. We illustrate our points using a quasi-likelihood function approach because it allows us to synthesize all these issues, also because many panel estimators such as the within estimator (e.g., Hsiao (2014)), the Bai (2013) factor estimator or the Phillips (2010, 2015) control function estimator can also be put in this framework.

Because the impact of the presence of time-invariant individual specific effects on the limiting distribution differ between a panel time series model and a model involving other explanatory variables, we consider these issues first in a panel time series setting, then for a general dynamic panel model containing exogenous explanatory variables in section 2 and 3, respectively. Section 4 discusses the implication of Chamberlain (1980)-Mundlak (1978) approach to get around the issue of incidental parameters. Section 5 considers the case of heteroscedatic errors. Section 6 provides a small scale Monte Carlo study to highlight the issues involved. Concluding remarks are in Section 7. All proofs are in the Appendix.

Throughout this paper, we use $(N,T) \to \infty$ to denote that both N and T jointly go to infinity, "$\to_p$" and "$\to_d$" to denote convergence in probability and in distribution, respectively.

2 A Panel Time Series Model

In this section, we discuss the asymptotic properties of the QMLE of a simple panel time series model. We distinguish two cases: inference based on fixed initial and random initial observations.

2.1 The model

There is no loss of generality to consider the following simple model,

$$y_{it} = \rho y_{it-1} + \eta_i + u_{it}, i = 1, \ldots, N; t = 1, \ldots, T, \tag{2.1}$$

where $|\rho| < 1$ and the initial value y_{i0} is also available for $i = 1, \ldots, N$. We assume

Assumption A1(a): The errors u_{it} are independent of η_i and are independently and identically distributed (i.i.d.) over i and t with mean zero and constant variance σ_u^2. For ease of notation, we let $\sigma_u^2 = 1$.

Assumption A2: The individual-specific effects η_i is i.i.d. over i with mean zero and variance σ_η^2.

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, $\mathbf{u}_i = (u_{i1}, \ldots, u_{iT})'$ and 1_T be a $T \times 1$ vector of ones, model (2.1) can be rewritten as a T-equation system of the form,

$$\mathbf{y}_i = \mathbf{y}_{i,-1}\rho + 1_T\eta_i + \mathbf{u}_i, \quad i = 1, \ldots, N. \tag{2.2}$$

2.2 Fixed Initial Observation

Under the assumption y_{i0} are fixed constants, the quasi-likelihood function takes the form

$$L = \prod_{i=1}^{N} (2\pi)^{-\frac{T}{2}} |\mathbf{V}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \rho\mathbf{y}_{i,-1})' \mathbf{V}^{-1} (\mathbf{y}_i - \rho\mathbf{y}_{i,-1})\right\}, \tag{2.3}$$

where

$$\mathbf{V} = \mathbf{I}_T + \sigma_\eta^2 1_T 1_T', \quad \mathbf{V}^{-1} = \mathbf{I}_T - \frac{\sigma_\eta^2}{1 + T\sigma_\eta^2} 1_T 1_T'. \tag{2.4}$$

The quasi-maximum likelihood estimator (QMLE) is obtained by maximizing the logarithm of (2.3). When σ_u^2 and σ_η^2 are known, the QMLE is the (naive) generalized least squares (GLS) estimator,

$$\hat{\rho}_{QMLE,f} = \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{V}^{-1} \mathbf{y}_{i,-1}\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{V}^{-1} \mathbf{y}_i\right). \tag{2.5}$$

where $_{QMLE,f}$ refers to QMLE treating y_{i0} as fixed constants.

Remark 2.1 *Bai (2013) derives (2.5) from the factor analytic framework by minimizing*[1,2]

$$\log |\Sigma_N(\boldsymbol{\theta})| + tr\left(\Sigma_N(\boldsymbol{\theta})^{-1} S_N\right), \tag{2.6}$$

[1] Bai (2013) derived (2.5) under the assumption that $y_{i0} = 0$. However, one may view $y_{i0} = 0$ as a special case of y_{i0} being a constant.

[2] Bai (2013) actually considers a model involving both the individual- and time- specific effects. However, taking the deviation of individual observation from the cross-section mean at time t, $(y_{it} - \bar{y}_t)$, removes the time-specific effects, where $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$. The transformed model no longer involves time-specific effects. The asymptotic distributions for Bai (2013) model or (2.1) are identical. So for ease of exposition, we just consider (2.1).

where $\boldsymbol{\theta} = \left(\rho, \sigma_\eta^2, \sigma_u^2\right)'$, $\Sigma_N(\boldsymbol{\theta}) = \Gamma\left(\sigma_u^2 \mathbf{I}_T + \left(\sigma_\eta^2 + \rho \frac{1}{N}\sum_{i=1}^N y_{i0}^2\right) 1_T 1_T'\right)\Gamma'$ *and* $S_N = \frac{1}{N}\sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$ *with* $\bar{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^N \mathbf{y}_i$,[3]

$$\Gamma_{T\times T} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \rho & 1 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \end{pmatrix}.$$

The difference between (2.6) and (2.3) is in the way of how the likelihood function $f(\mathbf{y}_i)$ *is written. There is not any fundamental difference between the QMLE and factor estimator. To see this, note that by continuous substitution,*

$$y_{it} = \rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}\eta_i + \sum_{j=0}^{t-1}\rho^j u_{i,t-j}. \tag{2.7}$$

Thus,

$$\mathbf{y}_i = \Gamma e_1 y_{i0}\rho + \Gamma 1_T \eta_i + \Gamma \mathbf{u}_i, \quad i = 1, 2, \ldots, N, \tag{2.8}$$

where $e_1 = (1, 0, \ldots, 0)'$ *is a* $T \times 1$ *vector. Under the assumption that* η_i *and* u_{it} *are i.i.d over* i*, and* $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N y_{i0}^2$ *converges to a constant, the logarithm of the quasi-likelihood function divided by* N *takes the form (2.6).*

We note that premultiplying (2.8) by the $T \times T$ *matrix* Λ

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}, \tag{2.9}$$

yields (2.2) and the quasi-likelihood function (2.3). In other words, the QMLE and factor estimator are different ways of obtaining the QMLE, not two different estimators based on different assumptions or inference procedures. So for ease of reference, we shall call the QMLE treating initial value fixed either the naive GLS or the Bai (2013) factor estimator.

Under the assumption that $y_{i0} = 0$ for all i, Bai (2013, Supplement) shows that the factor estimator (2.5) is fixed-T consistent. However, if $y_{i0} \neq 0$ and $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N y_{i0}\eta_i \neq 0$,

[3] For simplicity of exposition, we do not include an intercept term in (2.1). Thus, under our framework, S_N should be just $\frac{1}{N}\sum_{i=1}^N \mathbf{y}_i\mathbf{y}_i'$.

Lemma 2.1 *Under assumptions A1(a) and A2, the Bai (2013) factor estimator (or naive GLS) (2.5) for model (2.1) is inconsistent when T is fixed and $N \to \infty$. It is consistent when $T \to \infty$. Furthermore, if u_{it} and η_i are normally distributed or $E\left|u_{it}\right|^{4+\epsilon} < \infty$ and $E\left|\eta_i\right|^{4+\epsilon} < \infty$ for some $\epsilon > 0$, when $(N,T) \to \infty$,*

$$\sqrt{NT}\left(\hat{\rho}_{QMLE,f} - \rho - \frac{1}{T^2}d\right) \to_d N\left(0, 1-\rho^2\right), \tag{2.10}$$

where $d = \frac{1}{(1-\rho)\sigma_\eta^2} plim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N y_{i0}\eta_i$. If the process has been going on for a long time, then $d = \frac{1}{(1-\rho)^2}$.

Remark 2.2 *Lemma 2.1 says if $\frac{N}{T} \to a < \infty$ as $(N,T) \to \infty$, the QMLE or naive GLS treating initial observations as fixed constants is asymptotically unbiased. However, if $\frac{N}{T^3} \to c \neq 0$ as $T \to \infty$, the naive GLS is asymptotically biased and the bias is of order $\sqrt{\frac{N}{T^3}}$. Monte Carlo studies conducted by Hsiao and Zhang (2015) and Hsiao and Zhou (2015) show that valid statistical inference depends critically on the use of asymptotically unbiased estimators.*

Remark 2.3 *Treating η_i and y_{i0} as fixed constants, the QMLE is the within estimator. The within estimator is inconsistent if T is finite. It is consistent when $T \to \infty$. However, Hahn and Kuersteiner (2002) show that when $(N,T) \to \infty$ and $\frac{N}{T} \to a \neq 0 < \infty$, the within estimator is asymptotically biased and bias is of order $\sqrt{a}$.*

Remark 2.4 *The reason that the factor estimator (2.5) is asymptotically unbiased when $\frac{N}{T} \to a \neq 0 < \infty$ while the within estimator remains biased of order $\sqrt{a}$ is because the former treats η_i as random that allows the cancellation of correlations due to $\frac{1}{N}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{V}^{-1}\mathbf{u}_i$ with part of the correlations due to $\frac{1}{N}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{V}^{-1}1_T\eta_i$ (e.g., Appendix, equations (A.3) and (A.4)), while the within transformation removes η_i from the transformed equation and there is no term to cancel the bias due to $\frac{1}{N}\sum_{i=1}^N \mathbf{y}_{i,-1}'1_T1_T'\mathbf{u}_i$.*

2.3 Random Initial Observations

The starting date of collecting data is arbitrary. There is no reason to assume the data generating process of y_{i0} to be different from that of y_{it}. Under the assumption that the data generating process of y_{i0} is the same as that of y_{it},

$$y_{i0} = y_{i,-1}\rho + \eta_i + u_{i0} = \frac{1}{1-\rho}\eta_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}.$$

Then $E\left(y_{i0}v_{it}\right) \neq 0$ for all $v_{it} = \eta_i + u_{it}$, $t = 1, \ldots, T$. Rewrite y_{i0} in the form

$$y_{i0} = \mu + v_{i0}, \tag{2.11}$$

where $v_{i0} = \frac{1}{1-\rho}\eta_i + \sum_{j=0}^{\infty} \rho^j u_{i,-j}$. Under assumptions A1(a) and A2, we have $E(v_{i0}) = 0$ and $E(v_{i0}^2) = \sigma_0^2$ and $E(v_{i0}v_{it}) = \sigma_1^2 \left(= \frac{1}{1-\rho}\sigma_\eta^2\right)$.

Combining (2.11) and (2.2) yields a system of $(T+1)$ equations

$$\begin{aligned} y_{i0} &= \mu + v_{i0}, \\ \mathbf{y}_i &= \rho \mathbf{y}_{i,-1} + \eta_i 1_T + \mathbf{u}_i, \quad i = 1, \ldots, N, \end{aligned} \tag{2.12}$$

with variance-covariance matrix

$$\breve{\mathbf{V}} = \begin{pmatrix} \sigma_0^2 & \sigma_1^2 1_T' \\ \sigma_1^2 1_T & \mathbf{V} \end{pmatrix}, \tag{2.13}$$

where $\mathbf{V}$ is defined in (2.4).

The quasi-log-likelihood function of the system $(y_{i0}, \mathbf{y}_i)$ takes the form

$$\log L = -\frac{N}{2} \log \left| \breve{\mathbf{V}} \right| - \frac{1}{2} \sum_{i=1}^{N} \begin{pmatrix} y_{i0} - \mu \\ \mathbf{y}_i - \rho \mathbf{y}_{i,-1} \end{pmatrix}' \breve{\mathbf{V}}^{-1} \begin{pmatrix} y_{i0} - \mu \\ \mathbf{y}_i - \rho \mathbf{y}_{i,-1} \end{pmatrix}. \tag{2.14}$$

Conditional on $\breve{\mathbf{V}}$, the QMLE is the GLS of ρ

$$\hat{\rho}_{QMLE,r} = \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{C}^{-1} \mathbf{y}_{i,-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{C}^{-1} \mathbf{y}_i - \sum_{i=1}^{N} (y_{i0} - \mu) \sigma_0^{-2} \sigma_1^2 1_T' \mathbf{C}^{-1} \mathbf{y}_{i,-1} \right), \tag{2.15}$$

where ${}_{QMLE,r}$ refers to QMLE treating y_{i0} as a random variable, $\mathbf{C} = I_T + \tilde{\sigma}_\eta^2 1_T 1_T'$ with $\tilde{\sigma}_\eta^2 = \sigma_\eta^2 - \sigma_1^4 \sigma_0^{-2}$.

Lemma 2.2 *Under assumptions A1(a) and A2, and if y_{i0} is treated as a random variable, when $N \to \infty$, the QMLE estimator (2.15) for model (2.1) is consistent either T is fixed or $T \to \infty$ and*

$$\sqrt{NT} \left(\hat{\rho}_{QMLE,r} - \rho \right) \to_d N \left(0, 1 - \rho^2 \right). \tag{2.16}$$

Remark 2.5 *In the supplement material of Bai (2013), Bai assumes $y_{i0} = \delta_0 + \phi \eta_i + u_{i0}$, which is similar to (2.11). Rewrite the system (2.12) in the form,*

$$\begin{pmatrix} y_{i0} \\ \mathbf{y}_i \end{pmatrix} = \begin{pmatrix} 1 \\ \rho \Gamma e_1 \end{pmatrix} \delta_0 + \begin{pmatrix} \phi \\ \rho \Gamma e_1 \phi + \Gamma 1_T \end{pmatrix} \eta_i + \begin{pmatrix} 1 & 0 \\ \rho \Gamma e_1 & \Gamma \end{pmatrix} \begin{pmatrix} u_{i0} \\ \mathbf{u}_i \end{pmatrix}. \tag{2.17}$$

Premultiplying (2.17) by the $(T+1)\times(T+1)$ *matrix*

$$\tilde{\Lambda} = \begin{pmatrix} 1 & \mathbf{0}_{1\times T} \\ \mathbf{0}_{T\times 1} & \Lambda \end{pmatrix},$$

yields the system (2.12) and the quasi-likelihood function similar in the form to that of (2.14), where Λ *is given by (2.9). In other words, the GLS of (2.12) is identical to the Bai (2013) factor estimator when* y_{i0} *are treated as random variables.*

3 Panel Dynamic Models with Exogenous Explanatory Variables

We consider a dynamic model of the form

$$y_{it} = \rho y_{it-1} + x_{it}\beta + \eta_i + u_{it}, i = 1, \ldots, N; t = 1, \ldots, T, \tag{3.1}$$

where y_{i0} is observable, x_{it} is stationary and is strictly exogenous with respect to u_{it}, and $\text{plim}_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2$ is a nonzero constant, or nonsingular constant matrix if x_{it} is multidimension.

3.1 Fixed Initial Conditions

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})'$ and $\mathbf{Z}_i = (\mathbf{y}_{i,-1}, \mathbf{x}_i)$. Treating y_{i0} as fixed constants, the analogous estimator of (2.5) now becomes

$$\begin{pmatrix} \hat{\rho}_{nGLS} \\ \hat{\beta}_{nGLS} \end{pmatrix} = \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{Z}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{y}_i\right), \tag{3.2}$$

where $\mathbf{V}$ is defined in (2.4).

Lemma 3.1 *Under assumption A1(a), A2 and* $E(x_{it}\eta_i) = 0$*, the naive GLS (3.2) for* ρ *and* β *is inconsistent if* T *is fixed and* $N \to \infty$*. When* $(N,T) \to \infty$*, the naive GLS is consistent and is asymptotically unbiased if* $\frac{N}{T} \to a < \infty$*. However, it is asymptotically biased of order* $\sqrt{\frac{N}{T^3}}$ *if* $\frac{N}{T^3} \to c \neq 0 < \infty$*.*

It is often argued that the individual-specific effects η_i could be correlated with x_{it}, namely, $E(x_{it}\eta_i) \neq 0$. Then

Lemma 3.2 *Under assumption A1(a), A2, and the assumption that* $plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \bar{x}_i \eta_i = \varsigma \neq 0$ *where* $\bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it}$*, the naive GLS, (3.2), for* ρ *and* β *is inconsistent if* T *is fixed and* $N \to \infty$*. When* $T \to \infty$*, it is consistent. However, when* $(N,T) \to \infty$ *and* $\frac{N}{T} \to a \neq 0 < \infty$*, the naive GLS is asymptotically biased of order* $\sqrt{\frac{N}{T}}$*.*

3.2 Random Initial Observations

If the data generating process of y_{i0} is no different from that of y_{it} for $t \geq 1$. By continuous substitution of (3.1), it can be shown that y_{i0} is not only a function of η_i, but also past $x_{i,-j}$ and $u_{i,-j}$ $(j \geq 0)$,

$$\begin{aligned} y_{i0} &= \frac{1}{1-\rho}\eta_i + \beta\sum_{j=0}^{\infty}\rho^j x_{i,-j} + \sum_{j=0}^{\infty}\rho^j u_{i,-j} \\ &= \theta_{i0} + \frac{1}{1-\rho}\eta_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}, \end{aligned} \tag{3.3}$$

where $\theta_{i0} = \beta\sum_{j=0}^{\infty}\rho^j x_{i,-j}$ that varies with i. Bhargava and Sargan (1983) propose to eliminate the incidental parameters, θ_{i0}, through

$$\theta_{i0} = E\left(\theta_{i0}|\mathbf{x}_i\right) + w_i = \tilde{\mathbf{x}}_i'\mathbf{b} + w_i, \quad i = 1, \ldots, N, \tag{3.4}$$

where $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$.

Remark 3.1 *For* $\mathbf{b}$ *to be constant across* i*, the data generating process of* $\mathbf{x}_i$ *is stationary and homogeneous across* i *(Hsiao and Zhou (2015)), otherwise,* $E\left(\theta_{i0}|\mathbf{x}_i\right) = \tilde{\mathbf{x}}_i'\mathbf{b}_i$*. Issues of incidental parameters will arise.*

Substituting (3.4) into (3.3), we have

$$y_{i0} = \tilde{\mathbf{x}}_i'\mathbf{b} + v_{i0}, \tag{3.5}$$

where v_{i0} is now $w_i + \frac{1}{1-\rho}\eta_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}$. Combining (3.5) with the vector form of (3.1) yields a system of $(T+1)$ equations

$$\begin{pmatrix} y_{i0} \\ \mathbf{y}_i \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{x}}_i' & 0 \\ 0 & \mathbf{Z}_i \end{pmatrix}\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\delta} \end{pmatrix} + \begin{pmatrix} v_{i0} \\ \mathbf{v}_i \end{pmatrix}, \tag{3.6}$$

where $\mathbf{v}_i = 1_T\eta_i + \mathbf{u}_i$. The error term $\tilde{\mathbf{v}}_i = (v_{i0}, \tilde{\mathbf{v}}_i')'$ is i.i.d over i with variance-covariance matrix of the form

$$\tilde{\mathbf{V}} = \begin{pmatrix} \sigma_0^2 & \sigma_\tau^2 1_T' \\ \sigma_\tau^2 1_T & \mathbf{V} \end{pmatrix}, \tag{3.7}$$

where $\sigma_0^2 = Var\left(v_{i0}\right)$, $\sigma_\tau^2 = Cov\left(v_{i0}, v_{it}\right)$, and $\mathbf{V}$ is defined in (2.4).

Conditional on $\tilde{\mathbf{V}}$, the QMLE of the system (3.6) is the GLS,

$$\begin{pmatrix} \hat{\mathbf{b}}_{GLS} \\ \hat{\boldsymbol{\delta}}_{GLS} \end{pmatrix} = \left(\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{V}}^{-1}\tilde{\mathbf{Z}}_i\right)^{-1}\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\tilde{\mathbf{V}}^{-1}\tilde{\mathbf{y}}_i, \tag{3.8}$$

where $\tilde{\mathbf{y}}_i = (y_{i0}, \mathbf{y}_i')'$ and $\tilde{\mathbf{Z}}_i = \begin{pmatrix} \tilde{\mathbf{x}}_i' & 0 \\ 0 & \mathbf{Z}_i \end{pmatrix}$.

Lemma 3.3 *When $N \to \infty$, under the assumption A1(a) and A2, the GLS estimator (3.8):*

(i) is consistent and asymptotically unbiased whether T is fixed or goes to infinity if $E(x_{it}\eta_i) = 0$ (following the convention we shall call model (3.1) under $E(x_{it}\eta_i) = 0$ the random effects model);

(ii) is inconsistent if T is fixed when $E(x_{it}\eta_i) \neq 0$ (we shall call model (3.1) under $E(x_{it}\eta_i) \neq 0$ the fixed effects model). If $T \to \infty$ and $\frac{N}{T} \to a \neq 0 < \infty$, it is consistent, however, it is asymptotically biased of order $\sqrt{a}$.

4 Chamberlain-Mundlak Approach

When η_i are correlated with x_{it}, treating η_i as fixed constants introduces an incidental parameter issue. Mundlak (1978) has suggested to use the conditional mean of η_i conditional on the ith individual's time series average of observed explanatory variables, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, and Chamberlain (1980) has suggested to use the conditional mean of η_i conditional on the observed explanatory variables $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})'$ in place of η_i to get around the issue of incidental parameters. This approach has been very popular in both theoretical and empirical analysis (e.g., Abowd et al (1999), Bai (2013), and Islam (1995)). We consider the asymptotic properties of the QMLE under this formulation.

Let $\tilde{\mathbf{x}}_i = (1, \bar{x}_i)'$ if one follows the Mundlak (1978) formulation or $\tilde{\mathbf{x}}_i = (1, x_{i1}, \ldots, x_{iT})' = (1, \mathbf{x}_i')'$ if one follows the Chamberlain (1980) approach. Then[4]

$$\begin{aligned} \eta_i &= E(\eta_i | \tilde{\mathbf{x}}_i) + w_i^* \\ &= \tilde{\mathbf{x}}_i' \mathbf{b}^* + w_i^*, \end{aligned} \tag{4.1}$$

where $\mathbf{b}^* = (\mu, b_1^*)'$ if $\tilde{\mathbf{x}}_i = (1, \bar{x}_i)'$ or $\mathbf{b}^* = (\mu, b_1^*, \ldots, b_T^*)'$ if $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$.

Remark 4.1 *For (4.1) to hold, the data generating process for $\mathbf{x}_i$ is stationary and homogeneous across i (Hsiao et al (2002)). If $\mathbf{x}_i$ are generated from heterogenous process, then $E(\eta_i | \tilde{\mathbf{x}}_i) = \tilde{\mathbf{x}}_i' \mathbf{b}_i^*$, issues of incidental parameters will still arise even with the Chamberlain (1980) or Mundlak (1978) approach.*

[4] Note that if the data generating process is nonlinear, then (4.1) should be treated as a linear projection. The asymptotic properties of the estimator to be discussed remain holding as long as $\mathbf{b}^*$ is constant across i and w_i^* is uncorrelated with $\mathbf{x}_i$.

Substituting (4.1) back to (3.1) yields

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \tilde{\mathbf{x}}_i'\mathbf{b}^* + w_i^* + u_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{4.2}$$

which can be rewritten in vector form as

$$\begin{aligned} \mathbf{y}_i &= \mathbf{y}_{i,-1}\rho + \mathbf{x}_i\beta + 1_T\tilde{\mathbf{x}}_i'\mathbf{b}^* + w_i^*1_T + \mathbf{u}_i \\ &= \mathbf{Z}_i\boldsymbol{\delta} + 1_T\tilde{\mathbf{x}}_i'\mathbf{b}^* + w_i^*1_T + \mathbf{u}_i \\ &= \left[\mathbf{Z}_i, 1_T\tilde{\mathbf{x}}_i'\right]\boldsymbol{\theta} + w_i^*1_T + \mathbf{u}_i, \end{aligned} \tag{4.3}$$

where $\mathbf{Z}_i = (\mathbf{y}_{i,-1}, \mathbf{x}_i)$, $\boldsymbol{\delta} = (\rho, \beta)'$ and $\boldsymbol{\theta} = (\rho, \beta, \mathbf{b}^{*\prime})'$.

Under the assumption that w_i^* is independent of $\mathbf{x}_i$ and i.i.d over i with mean 0 and variance $\sigma_{w^*}^2$, the variance-covariance matrix of $(w_i^*1_T + \mathbf{u}_i)$ takes the form,

$$\begin{aligned} \bar{\mathbf{V}} &= E\left[(w_i^*1_T + \mathbf{u}_i)(w_i^*1_T + \mathbf{u}_i)'\right] \\ &= \sigma_u^2\mathbf{I}_T + \sigma_{w^*}^2 1_T 1_T'. \end{aligned} \tag{4.4}$$

then

$$\bar{\mathbf{V}}^{-1} = \sigma_u^{-2}\left(\mathbf{I}_T - \frac{\sigma_{w^*}^2\sigma_u^{-2}}{1 + T\sigma_{w^*}^2\sigma_u^{-2}}1_T 1_T'\right).$$

4.1 Fixed Initial Observations

Treating initial values y_{i0} as fixed constants, the naive generalized least squares estimator of (4.3) takes the form

$$\hat{\boldsymbol{\theta}}_{nGLS} = \left[\sum_{i=1}^{N}\begin{pmatrix} \mathbf{Z}_i' \\ \tilde{\mathbf{x}}_i 1_T' \end{pmatrix}\bar{\mathbf{V}}^{-1}\left(\mathbf{Z}_i, 1_T\tilde{\mathbf{x}}_i'\right)\right]^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix} \mathbf{Z}_i' \\ \tilde{\mathbf{x}}_i 1_T' \end{pmatrix}\bar{\mathbf{V}}^{-1}\mathbf{y}_i\right]. \tag{4.5}$$

For this naive GLS $\hat{\boldsymbol{\theta}}_{nGLS}$, including the $\hat{\boldsymbol{\delta}}_{nGLS} = \left(\hat{\rho}_{nGLS}, \hat{\beta}_{nGLS}\right)$, we have

Lemma 4.1 *Under assumption A1(a), A2 and (4.1), the naive GLS (4.5) is inconsistent if T is fixed and $N \to \infty$. When $(N,T) \to \infty$ and $\frac{N}{T} \to a \neq 0 < \infty$, it is consistent and asymptotically normally distributed with mean zero. However, if N tends to infinity faster than T so $\frac{N}{T^3} \to c \neq 0 < \infty$, (4.5) is asymptotically biased of order $\sqrt{c}$.*

4.2 Random Initial Observations

Combining (3.5) for the initial distribution y_{i0} and the vector form of (4.2), we have a system of $(T+1)$ equations. The $(T+1)\times(T+1)$ covariance matrix of this system takes the form

$$\Omega = \begin{pmatrix} \sigma_0^2 & \sigma_\tau^2 1_T' \\ \sigma_\tau^2 1_T & \sigma_u^2\mathbf{I}_T + \sigma_{w^*}^2 1_T 1_T' \end{pmatrix}. \tag{4.6}$$

Conditional on σ_0^2, σ_τ^2 and $\sigma_{w^*}^2$, the QMLE of the complete system $(y_{i0}, \mathbf{y}_i|\mathbf{x}_i)$ takes the form of

$$\hat{\tilde{\boldsymbol{\theta}}}_{GLS} = \left[\sum_{i=1}^{N} \mathbf{Z}_i^{*\prime}\boldsymbol{\Omega}^{-1}\mathbf{Z}_i^*\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{Z}_i^{*\prime}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{y}}_i\right]. \tag{4.7}$$

where $\tilde{\boldsymbol{\theta}} = \left(\mathbf{b}', \mathbf{b}^{*\prime}, \boldsymbol{\delta}'\right)'$,

$$\mathbf{Z}_i^* = \begin{pmatrix} \tilde{\mathbf{x}}_i' & 0 & 0 \\ 0 & 1_T\mathbf{x}_i & \mathbf{Z}_i \end{pmatrix}. \tag{4.8}$$

The QMLE (4.7) is asymptotically unbiased when $N \to \infty$ whether T is fixed or goes to infinity.

Conditional on y_{i0} and $\mathbf{x}_i$, the system of $(\mathbf{y}_i|y_{i0}, \mathbf{x}_i)$, $i = 1, \ldots, N$, is of the form

$$\mathbf{y}_i = \mathbf{y}_{i,-1}\rho + \mathbf{x}_i\beta + 1_T\tilde{\mathbf{x}}_i'\tilde{\mathbf{b}}^* + 1_T y_{i0}\gamma + \mathbf{v}_i^*, \quad i = 1, \ldots, N, \tag{4.9}$$

where $\tilde{\mathbf{b}}^* = \gamma\mathbf{b} + \mathbf{b}^*$ and $\gamma = -\frac{\sigma_\tau^2}{\sigma_0^2}$.

The covariance matrix of (4.9) is

$$E\left(\mathbf{v}_i^*\mathbf{v}_i^{*\prime}\right) = \sigma_u^2\mathbf{I}_T + \sigma_w^2 1_T 1_T' = \mathbf{V}^*, \tag{4.10}$$

where for notational ease, we now use σ_w^2 to indicate $\left(\sigma_{w*}^2 - \frac{\sigma_\tau^4}{\sigma_0^2}\right)$.

The GLS of (4.9) now takes the form

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}_C \\ \hat{\tilde{\mathbf{b}}}_C^* \\ \hat{\gamma} \end{pmatrix} = \left[\sum_{i=1}^{N} \begin{pmatrix} \mathbf{Z}_i' \\ \tilde{\mathbf{x}}_i 1_T' \\ y_{i0}1_T' \end{pmatrix} \mathbf{V}^{*-1}\left(\mathbf{Z}_i', 1_T\tilde{\mathbf{x}}_i', 1_T y_{i0}\right)\right]^{-1} \left[\sum_{i=1}^{N} \begin{pmatrix} \mathbf{Z}_i' \\ \tilde{\mathbf{x}}_i 1_T' \\ y_{i0}1_T' \end{pmatrix} \mathbf{V}^{*-1}\mathbf{y}_i\right]. \tag{4.11}$$

The GLS of $\boldsymbol{\delta}$, $\hat{\boldsymbol{\delta}}_C$, for (4.11) has essentially the same form as (4.5). Thus,

Lemma 4.2 *Under assumption A1(a), A2, (4.1) and (3.5), both the unconditional GLS (4.7) and the conditional GLS (4.11) are consistent and asymptotically unbiased when $N \to \infty$ whether T is fixed or goes to infinity.*

Remark 4.2 *The difference between $\hat{\boldsymbol{\delta}}_C$ and $\hat{\boldsymbol{\delta}}_{nGLS}$ is that $\hat{\boldsymbol{\delta}}_C$ is based on the conditional distribution of $(\mathbf{y}_i|y_{i0}, \mathbf{x}_i)$ while $\hat{\boldsymbol{\delta}}_{nGLS}$ is derived from the distribution of $(\mathbf{y}_i|\mathbf{x}_i)$ assuming $E(y_{i0}\eta_i) = 0$. If $E(y_{i0}\eta_i) \neq 0$, there is a bias term due to this. On the other hand, $\hat{\boldsymbol{\delta}}_C$ is also conditional on y_{i0}, so $plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i0}v_{it}^* = 0$, while for the system (4.3) $plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i0}\left(w_i^* + u_{it}\right) \neq 0$. In other words, the conditional GLS estimator is asymptotically unbiased when $N \to \infty$ independent of the size of T.*

Remark 4.3 *The system (4.9) is identical to the system of Phillip (2010, 2015) based on the control function approach. However, the derivation of the conditional system (4.9) shows that for the control function approach to avoid incidental parameters issue, the data generating process of* $\mathbf{x}_i$ *has to be homogeneous across* i.

Remark 4.4 *In some applications, one takes the approach of conditioning on* y_{i0} *to take count the endogeneity of* y_{i0}. *This is fine if the model is a time series model like (2.1). However, if the dynamic model also contains exogenous regressors like (3.1), conditioning on* y_{i0} *alone, but not also on* $\tilde{\mathbf{x}}_i$ *(eq (3.5)), cannot remove the asymptotic bias, even under the assumption that* $\mathbf{x}_i$ *are independent of* η_i. *As a matter of fact, if* T *is fixed, the resulting estimator is biased of order* $\frac{1}{T}$ *no matter how large* N *is. If* $\frac{N}{T} \to a \neq 0 < \infty$ *as* $T \to \infty$, *the estimator is consistent, but is asymptotically biased of order* $\sqrt{\frac{N}{T}}$.

For ease of notation, we assume x_{it} *and* η_i *are independent. We note that from (3.3),*

$$y_{i0} = \frac{1}{1-\rho}\eta_i + v_{i0}^*, \tag{4.12}$$

where $v_{i0}^* = \mu + \beta\sum_{j=0}^{\infty}\rho^j x_{i,-j} + \sum_{j=0}^{\infty}\rho^j u_{i,-j} = \theta_{i0} + v_{i0}$. *Combining (3.1) and (4.12) yields a system,*

$$\begin{aligned}\begin{pmatrix} y_{i0} \\ \mathbf{y}_i \end{pmatrix} &= \begin{pmatrix} 0 \\ \mathbf{y}_{i,-1} \end{pmatrix}\rho + \begin{pmatrix} 0 \\ \mathbf{x}_i \end{pmatrix}\beta + \begin{pmatrix} \frac{1}{1-\rho} \\ 1_T \end{pmatrix}\eta_i + \begin{pmatrix} v_{i0}^* \\ \mathbf{u}_i \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \mathbf{Z}_i \end{pmatrix}\boldsymbol{\delta} + \begin{pmatrix} \frac{1}{1-\rho}\eta_i + v_{i0}^* \\ 1_T\eta_i + \mathbf{u}_i \end{pmatrix}, \quad i = 1, \ldots, N. \end{aligned} \tag{4.13}$$

Thus, the conditional system of $\mathbf{y}_i$ *conditional on* y_{i0} *takes the form,*

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\delta} + 1_T y_{i0}\gamma + \tilde{\mathbf{v}}_i^*, \quad i = 1, \ldots, N, \tag{4.14}$$

where $\tilde{\mathbf{v}}_i^* = \mathbf{u}_i + \left\{\left(1 - \frac{\sigma_\eta^2}{(1-\rho)^2\sigma_{0*}^2}\right)\eta_i - \frac{\sigma_\eta^2}{(1-\rho)\sigma_{0*}^2}v_{i0}^*\right\}1_T$ *and* $\sigma_{0*}^2 = Var\left(\frac{1}{1-\rho}\eta_i + v_{i0}^*\right) = Var\left(\frac{1}{1-\rho}\eta_i + v_{i0}\right)$. *The GLS of (4.14) takes the form of*

$$\begin{pmatrix} \hat{\boldsymbol{\delta}}_C^* \\ \hat{\gamma}_C^* \end{pmatrix} = \left[\sum_{i=1}^N \begin{pmatrix} \mathbf{Z}_i' \\ y_{i0}1_T' \end{pmatrix}\tilde{\mathbf{V}}^{*-1}\left(\mathbf{Z}_i, 1_T y_{i0}\right)\right]^{-1}\left[\sum_{i=1}^N \begin{pmatrix} \mathbf{Z}_i' \\ y_{i0}1_T' \end{pmatrix}\tilde{\mathbf{V}}^{*-1}\mathbf{y}_i\right], \tag{4.15}$$

where

$$\tilde{\mathbf{V}}^* = Cov\left(\mathbf{v}_i^*\right) = \mathbf{I}_T + \sigma_{\tilde{w}}^2 1_T 1_T', \tag{4.16}$$

where for notational ease we let $\sigma_{\tilde{w}}^2 = \sigma_\eta^2\left(1 - \frac{\sigma_\eta^2}{(1-\rho)^2\sigma_{0*}^2}\right)$.

Lemma 4.3 *The system (4.14) is the conditional system of (4.13), conditional on y_{i0}. The resulting estimator (4.15) is inconsistent if T is fixed because $E\left(v_{i0}^*\right) \neq 0$ under (3.4)*

$$\hat{\boldsymbol{\delta}}_C^* = \boldsymbol{\delta} + O_p\left(\frac{1}{T}\right). \tag{4.17}$$

In other words, if T is fixed, conditional on y_{i0} alone cannot yield a consistent estimator no matter how large N is. The bias is of order $\frac{1}{T}$. If $T \rightarrow \infty$, $\hat{\boldsymbol{\delta}}_C^*$ is consistent. However, if $\frac{N}{T} \rightarrow a \neq 0 < \infty$ as $T \rightarrow \infty$, $\hat{\boldsymbol{\delta}}_C^*$ is asymptotically biased of order $\sqrt{\frac{N}{T}}$.

Remark 4.5 *The difference between (4.9) and (4.14) is that the former is a legitimate conditional system whether $E\left(x_{it}\eta_i\right) = 0$ or not, while the latter is not a legitimate system even under $E\left(x_{it}\eta_i\right) = 0$.*

Remark 4.6 *When $E\left(x_{it}\eta_i\right) = 0$, treating y_{i0} as fixed constants, the naive GLS is asymptotically biased of order $\sqrt{\frac{N}{T^3}}$. On the other hand, conditional on y_{i0} ignoring the fact that y_{i0} is also a function of past $x_{i,-s}$ $(s \geq 0)$ yields an estimator that is asymptotically biased of order $\sqrt{\frac{N}{T}}$, worse than treating initial value y_{i0} as fixed constants (order of $\sqrt{\frac{N}{T^3}}$). The reason is that treating y_{i0} as a fixed constant, the system (3.1) of T equations has the initial value, y_{i0}, appearing only in the y_{i1} equation. The other $(T-1)$ y_{it} equation does not contain y_{i0} as a regressor. As T goes to infinity, the error of ignoring the correlation between y_{i0} and η_i becomes increasingly negligible. On the other hand, the conditional system (4.14) has each y_{it} equation containing y_{i0} as a regressor, yet $E\left(y_{i0}v_{it}^*\right) \neq 0$ for $t = 1, \ldots, T$.*

Remark 4.7 *Arellano and Bond (1990) suggest to take the generalized method of moments (GMM) approach to estimate the unknown parameters. The advantages of GMM are (i) there is no need to consider if x_{it} are correlated with η_i, and (ii) there is no need to consider the initial value distribution. When $N \rightarrow \infty$, the GMM is asymptotically unbiased when T is fixed. However, if $(N, T) \rightarrow \infty$, Alvarez and Arellano (2003) show that the GMM is asymptotically biased of order $\sqrt{\frac{T}{N}}$. If an estimator is asymptotically biased, there could be significant size distortion (see, for example, Hsiao and Zhang (2015) or Hsiao and Zhou (2015, 2017)).*

5 Model with Heteroscedastic Errors

Although the above results are based on homoscedastic errors (Assumption A1(a)), we show in this section that the statistical properties with regard to the consistency and order of asymptotic bias for model (2.1) with fixed or random initial conditions remain valid with the heteroscedastic errors, u_{it}, Assumption A1(b): The errors u_{it} are independent of η_i and are independently distributed over i and t with mean zero and constant variance σ_{ui}^2, where $0 < \sigma_{ui}^2 < \infty$ for all $i = 1, \ldots, N$, and $\frac{1}{N}\sum_{i=1}^{N}\sigma_{ui}^2 = \bar{\sigma}_u^2 < \infty$.

For model (2.1), the analogues estimator of (2.5) now takes the form

$$\hat{\rho}_{heter,f} = \left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}_i^{-1}\mathbf{y}_{i,-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}_i^{-1}\mathbf{y}_i\right), \tag{5.1}$$

where $_{heter,f}$ refers to heteroscedastic errors and y_{i0} is treated as fixed constant, $\mathbf{y}_{i,-1}$ and $\mathbf{y}_i$ are defined before and

$$\mathbf{V}_i = \sigma_{ui}^2\mathbf{I}_T + \sigma_\eta^2 1_T 1_T', \quad \mathbf{V}_i^{-1} = \sigma_{ui}^{-2}\left(\mathbf{I}_T - \frac{\varkappa_i}{1 + T\varkappa_i}1_T 1_T'\right). \tag{5.2}$$

where $\varkappa_i = \frac{\sigma_\eta^2}{\sigma_{ui}^2}$.

Similarly, the unconditional system (2.12) now has the variance-covariance matrix $\check{\mathbf{V}}_i$, where

$$\check{\mathbf{V}}_i = \begin{pmatrix} \sigma_{0i}^2 & \sigma_1^2 1_T' \\ \sigma_1^2 1_T & \mathbf{V}_i \end{pmatrix}, \tag{5.3}$$

where σ_1^2 is the same as before and $\sigma_{0i}^2 = Var(y_{i0})$, $\mathbf{V}_i = \sigma_{ui}^2 I_T + \sigma_\eta^2 1_T 1_T'$.

Thus, the analogues estimator of (2.15) is

$$\hat{\rho}_{heter,r} = \left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{y}_i - \sum_{i=1}^{N}(y_{i0} - \mu)\,\sigma_{0i}^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right), \tag{5.4}$$

where $_{heter,r}$ refers to heteroscedastic errors and y_{i0} is treated as a random variable, and $\mathbf{C}_i = \sigma_{ui}^2 I_T + \tilde{\sigma}_{\eta i}^2 1_T 1_T'$ with $\tilde{\sigma}_{\eta i}^2 = \sigma_\eta^2 - \sigma_1^4\sigma_{0i}^{-2}$.

The asymptotics of the GLS estimators (5.1) and (5.4) are summarized in the following lemma.

Lemma 5.1 *Under assumption A1(b), A2, when $N \to \infty$, the naive GLS estimator (5.1) is asymptotically biased of order $\sqrt{\frac{N}{T^3}}$ as $(N, T) \to \infty$, but the GLS estimator (5.4) is asymptotically unbiased whether T is fixed for goes to infinity.*

Remark 5.1 *The above lemma states that whether the idiosyncratic errors u_{it} is heteroscedastic doesn't affect the asymptotic properties of the QMLE when treating y_{i0} as fixed constant or random variable. This result can also be generalized to model with exogenous variables.*

Remark 5.2 *Similarly, the consistency and the order of asymptotic bias for the system (3.1) with fixed or random initial conditions remain valid with time heteroscedasticity. Suppose* u_{it} *is independently distributed over* t *with variance* σ_t^2 *and* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 = \bar{\sigma}^2 < \infty$. *The* $NT \times NT$ *covariance matrix of* $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2', \ldots .\mathbf{v}_N')'$ *has the form of*

$$E\left(\mathbf{v}\mathbf{v}'\right) = I_N \otimes \Psi + I_N \otimes \sigma_\eta^2 1_T 1_T' = I_N \otimes \left(\Psi + \sigma_\eta^2 1_T 1_T'\right),$$

where $\Psi = diag\left(\sigma_1^2, \ldots, \sigma_T^2\right)$ *and* $\otimes$ *denotes the Kronecker product. Similar, but more laborious manipulations, show that the order of asymptotic bias is the same as the homoscedastic case*[5].

6 Monte Carlo Simulation

In this section, we investigate the finite sample properties for the estimator conditioning on the initial values being fixed (Bai (2013) factor estimator) or random (conditional GLS and the unconditional GLS) for dynamic panel model. We consider the following data generating processes.

DGP1: Panel time series model

$$y_{it} = \eta_i + \rho y_{it-1} + u_{it}. \tag{6.1}$$

DGP2: Dynamic panel with exogenous variables

$$y_{it} = \eta_i + \rho y_{it-1} + x_{it}\beta + u_{it}, \tag{6.2}$$

where the exogenous variables x_{it} are generated as

$$x_{it} = 0.5x_{i,t-1} + 0.4\eta_i + v_{it},$$

where $v_{it} \sim IID\chi^2(1)$ for all i and t.

DGP3: Random effects dynamic model with the same DGP of (6.2), but x_{it} are generated as

$$x_{it} = 0.5x_{i,t-1} + v_{it}, \tag{6.3}$$

where $v_{it} \sim IID\chi^2(1)$ for all i and t.

For these three DGPs, we assume that $\eta_i \sim IIDN(0,1)$ for all i. For the values of ρ and β, we let $\rho_0 = 0.5$ and $\beta = 1$.[6] We also let $T = 10, 100, 200$ and $N = 100, 200, 500$. We generate $T + 100$ observations, and the first 100 observations are discarded.

[5] The derivation for the asymptotic properties of the naive GLS (2.5) when u_{it} is time series heteroscedastic is avaiable upon request.

[6] To save the space, here we only report the simulation results for $\rho_0 = 0.5$. Additional simulation results for $\rho_0 = 0.2$ and 0.8 are presented in the Appendix.

In order to examine the finite sample performance of the QMLE or GLS, we generate the idiosyncratic errors as homoscedastic or heteroscedastic as follows:

Case 1: Homoscedasitic errors

We generate $u_{it} \sim IIDN(0,1)$ for all i,t.

Case 2: Heteroscedasitic errors

We generate $u_{it} \sim IIDN\left(0,\sigma_i^2\right)$ for all i,t, where σ_i^2 is iid draw from $0.5\left(1+0.5\chi^2(2)\right)$ for all i, with $\chi^2(2)$ being the Chi-squared distribution variable with degree of freedom of 2.

For DGP (6.1), we consider the factor estimator or naive GLS (2.5), and the conditional GLS conditional on $y_{i0}-\mu$ only (formula (4.11) with $\mathbf{x}_i = 0$).

The DGP (6.2), $E(x_{it}\eta_i) \neq 0$, and DGP (6.3), $E(x_{it}\eta_i) = 0$, we consider estimator (3.2), (4.5), (4.7) (4.11) and (4.15).[7] For comparison, we also consider Arellano-Bond type GMM estimation (e.g., Arellano and Bond (1990)) for DGP (6.1)-(6.3). The number of replication is set at 1000 times. Simulation results are summarized in Table 1-3 for homoscedastic errors and Table 4-6 for heteroscedastic errors.

The GLS or conditional GLS requires the knowledge of σ_u^2 and σ_η^2 or σ_w^2. We obtain initial estimators for σ_u^2 and σ_w^2 (or σ_η^2) first using Anderson and Hsiao (1981, 1982) simple instrumental variable estimator of ρ and β ($\beta = 0$ for DGP1) to obtain

$$\hat{e}_{it} = \Delta y_{it} - \hat{\rho}\Delta y_{it-1} - \hat{\beta}\Delta x_{it}, \quad i = 1,2,\ldots,N; t = 1,2,\ldots,T, \tag{6.4}$$

then estimate σ_u^2 by

$$\hat{\sigma}_u^2 = \frac{1}{2N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2. \tag{6.5}$$

To obtain initial estimator for σ_w^2, we also need initial estimator of γ and $\mathbf{a}$, which can be estimated by

$$\begin{pmatrix} \hat{\gamma} \\ \widehat{\tilde{\mathbf{a}}}^* \end{pmatrix} = \left[\sum_{i=1}^{N}\begin{pmatrix} y_{i0}^2 & y_{i0}\tilde{\mathbf{x}}_i' \\ y_{i0}\tilde{\mathbf{x}}_i & \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i' \end{pmatrix}\right]^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix} y_{i0} \\ \tilde{\mathbf{x}}_i \end{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}\right], \tag{6.6}$$

where $\hat{e}_{it}$ is defined in (6.4). Then calculate the residuals

$$\bar{e}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \hat{\rho}y_{it-1} - \hat{\beta}x_{it} - \hat{\gamma}y_{i0} - \tilde{\mathbf{x}}_i'\widehat{\tilde{\mathbf{a}}}^*\right), \tag{6.7}$$

and

$$\hat{\sigma}_w^2 = \frac{1}{N}\sum_{i=1}^{N}\bar{e}_i^2 - \frac{\hat{\sigma}_u^2}{T}. \tag{6.8}$$

[7] GLS estimators (4.7) and (4.11) performs quite similar to each other. So we only display simulation results for GLS estimator (4.7). Estimation reults of GLS estimator (4.11) are avaiable upon request.

However, the two step procedure of estimating $\hat{\sigma}_w^2$ could yield negative value. In the event that an estimator $\hat{\sigma}_w^2 < 0$, we let $\hat{\sigma}_w^2 = 0$ in estimating $\tilde{\mathbf{V}}$.

Tables 1 and 2 summarize the bias, root mean squared error (RMSE) and actual size based on the critical value of 1.96 for $\rho = 0.5$ for different combination of N (100, 200, 500) and T (10, 100, 500) when the idiosyncratic errors are homoscedastic and heteroscedastic for DGP1, respectively. As expected, when N and T are of similar magnitude, the Bai (2013) factor estimator or naive GLS has negligible bias and the actual size is close to the nominal value of 5%. However, if N is much larger than T, then there is significant size distortion and the distortion increases with the value of ρ. For instance, for $N = 100$, $T = 10$, the actual size is 38.5% for $\rho = 0.5$ (9.1% for $\rho = 0.2$ and 100% when $\rho = 0.8$ in the Table A1 and A2). Moreover, the size distortion increases with the ratio N/T. When $N = 500$ and $T = 10$, the actual size for nominal value 5% significance level is 36.2% when $\rho = 0.5$ (24.7% for $\rho = 0.2$ and 100% when $\rho = 0.8$ in the Table A1 and A2). On the other hand, the conditional GLS estimator conditional $y_{i0} - \mu$, is indeed asymptotically unbiased, and the actual size is close to the nominal size for whatever combination of N and T and for whatever value of ρ. Furthermore, the bias and root mean square error (RMSE) of the conditional GLS are substantially smaller than the estimator treating initial values as fixed constants. Also, similar findings can be observed in Table 4 when the idiosyncratic errors are heteroscedastic. For the GMM estimation, it is obvious that GMM is asymptotically biased of order $\frac{T}{N}$, the bias is quite significant when N is relatively small, and the size distortion is quite significant if the ratio of $\frac{T}{N}$ is large.

Tables 3-4 summarize the results for the different estimators for DGP2 (fixed effects model) when $\tilde{\mathbf{x}}_i = (1, \bar{x}_i)'$ for homoscedastic and heteroscedastic errors, respectively. As expected, ignoring the correlation between the individual effects and regressors lead to significant bias and size distortion for the factor estimator (3.2). However, even with the Mundlak-Chamberlain approach of correction of the correlations between the individual effects and regressors, it could still lead to significant size distortion when N is much larger than T if initial values are treated as fixed constants (GLS (4.5)). The distortion increases with the ratio $\frac{N}{T}$ and the absolute value of ρ. For instance, for a nominal significant level 5% test, when $T = 10$, and $N = 200$, the actual size is 22.7% when $\rho = 0.5$. When $T = 10$, and $N = 500$, the size distortion increases to 40% when $\rho = 0.5$. The size distortion will disappear only if N and T are of similar magnitude. On the other hand, the feasible unconditional GLS or feasible conditional GLS (GLS (4.7) and (4.11)) have the empirical size close to the nominal size for whatever combination of N and T. Moreover, they have smaller bias and RMSE than the estimator treating initial values as fixed constants. Tables 3-4 also show that although y_{i0} are treated as random but conditional on y_{i0} alone to take account the randomness of y_{i0} can also yield significant size distortion if N

is much larger than T. While for the GMM estimator, it is still asymptotically biased, but the size distortion is quite small.

Tables 5-6 summarize the results for the different estimators for DGP3 (random effects model) or homoscedastic and heteroscedastic errors, respectively. When the individual effects and regressors are uncorrelated, $\tilde{\mathbf{b}} = 0$. The size distortion of the factor estimator are of similar magnitude to the pure time series model (DGP1) when N is much larger than T. There will be no size distortion only when N and T are of similar magnitude. However, implementing the unneeded Munlak-Chamberlain adjustment (GLS (4.5)) further increases the size distortion. On the other hand, the performance of unconditional and conditional feasible GLS is not affected by implementing the unneeded Mundlak-Chamberlain adjustment (GLS (4.7) and (4.11)). The actual size is close to nominal size for whatever combination of N and T. The GMM estimation performs similarly to DGP 2 and still is still asymptotically biased.

Table 7 summarizes the results of naive GLS and conditional or unconditional GLS for DGP2 following the Chamberlain (1980) approach of considering the individual effects or initial observations on all observed regressors $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$ when N and T are of similar magnitude. As one can see if $N = T$, then using the Chamberlain approach can still lead to significant size distortion due to bad approximation of the variance of $(\eta_i - E(\eta_i \mathbf{x}_i)$ using the Chamberlain approach when T greater than or equal to N. On the other hand, if $N > T$, using either the Chamberlain approach or Mundlak approach can lead to asymptotically unbiased inference. However, the Mundlak (1978) approach yields smaller bias and RMSE. This suggests that perhaps empirically one should just use the time series average $\bar{x}_i$ instead of $\mathbf{x}_i$ to condition η_i or y_{i0}.[8]

7 Concluding Remarks

Whether an estimator is asymptotically biased or not plays a crucial role in obtaining valid statistical inference as shown in our simulations as well as those of Hsiao and Zhang (2015), Hsiao and Zhou (2015). For a dynamic panel data model, in addition to the issue of fixed vs random effects specification, there is also an issue of whether the initial values should be treated as fixed constants or random variables. Because many dynamic panel data estimators can be viewed as the quasi-maximum likelihood estimator (QMLE) under different initial value assumptions, we take a quasi-likelihood approach to illustrate the sensitivity of valid statistical inference to the initial value distribution assumptions and cross-sectional dimension N and the

[8] This result is based on the model is correctly specified. If a model is misspecified, the Chamberlain approach may perform better.

time series dimension T. When N is fixed and T is large, it does not matter whether the initial values are treated as fixed constants or random variables. When N is large, how the initial values are treated becomes very important. Treating initial values as fixed constants, the QMLE is inconsistent if T is fixed and $N \to \infty$. When both N and T are large, it is consistent but asymptotically biased of order $\sqrt{\frac{N}{T^3}}$ for a panel time series model whether the individual-specific effects are fixed or random or a general dynamic panel model containing strictly exogenous explanatory variables when the effects are random and independent of the explanatory variables. However, if the effects are correlated with the included exogenous variables, the estimator (3.2) is asymptotically biased of order $\sqrt{\frac{N}{T}}$ as $(N, T) \to \infty$. Using the the Chamberlain (1980)-Mundlak (1978) approach to condition the effects on observed explanatory variables can reduce the order of asymptotic bias to $\sqrt{\frac{N}{T^3}}$. On the other hand, the QMLE with a properly modeled initial value distribution combined with Chamberlain-Mundlak approach is consistent and asymptotically unbiased whether T is fixed or goes to infinity as long as $N \to \infty$. We summarize the main conclusions in Table 8.

There is also an important difference in formulating the initial value distribution for a pure time series model and the model that also contains other covariates. For a pure time series model, modeling $y_{i0} = \mu + v_{i0}$ is sufficient. For a dynamic model containing other covariates, then y_{i0} is also a function of past covariates. Conditioning on observed covariates can get around the incidental parameters issue only if the data generating process for the covariates is homogeneous across i. Finally, although asymptotically there is no difference between the Chamberlain (1980) approach of conditioning the individual effects or initial values on individual's observed regressors and Mundlak's (1978) approach of simply conditioning on their time series means, our simulation results appear to favor the Mundlak's (1978) approach, in particular if T is not small based on the assumption that the panel data model is correctly specified.

References

[1] Abowd, J.M., F. Kramarz, and D. Margolis, 1999, High Wage Workers and High Wage Firms, *Econometrica* 67, 251–333

[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, 1981 Estimation of dynamic models with error components, *Journal of the American Statistical Association*, 589-606.

[4] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82.

[5] Arellano, M., and S. Bond, 1991. Some tests of specification for panel data: monte carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277-297.

[6] Bai, J., 2013, Fixed-Effects Dynamic Panel Models, a Factor Analytical Method, *Econometrica* 81, 285-314.

[7] Bai, J., 2013, Fixed-Effects Dynamic Panel Models, a Factor Analytical Method: Online supplement, *Econometrica*.

[8] Bhargava, A., and J.D., Sargan, 1983, Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, *Econometrica* 51, 1635-1659.

[9] Chamberlain, G. 1980, Analysis of covariance with qualitative data, *Review of Economic Studies* 47, 225-238;

[10] Hahn, J., and G. K, Kuersteiner, 2002, Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large, *Econometrica* 70, 1639–1657.

[11] Hsiao, C., 2014, *Analysis of Panel Data*, 3rd Edition, Cambridge University Press.

[12] Hsiao, C., M.H., Pesaran, and A.K Tahmiscioglu, 2002, Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods, *Journal of Econometrics* 109, 107-150.

[13] Hsiao, C., and A.K Tahmiscioglu, 2008, Estimation of Dynamic Panel Data Models with Both Individual and Time Specific Effects. *Journal of Statistical Planning and Inference* 138, 2698-2721.

[14] Hsiao, C., and J. Zhang, 2015, IV, GMM or likelihood approach to estimate dynamic panel models when either N or T or both are large, *Journal of Econometrics* 187, 312–322.

[15] Hsiao, C., and Q. Zhou, 2015, Statistical inference for panel dynamic simultaneous equations models, *Journal of Econometrics* 189, 383–396.

[16] Hsiao, C., and Q. Zhou, 2016, Asymptotic distribution of quasi-maximum likelihood estimation of dynamic panels using long difference transformation when both N and T are large, *Statistical Methods & Applications* 25, 675-683.

[17] Hsiao, C., and Q. Zhou, 2017, First difference or forward demeaning: Implications for the method of moments estimators, *Econometric Reviews* 36, 883-897.

[18] Islam, N., 1995, Growth Empirics: A Panel Data Approach, *The Quarterly Journal of Economics* 110, 1127-1170.

[19] Mundlak, Y. 1978, On the pooling of time series and cross section data, *Econometrica* 46, 69-85;

[20] Phillips, R.F., 2010, Iterated Feasible Generalized Least-Squares Estimation of Augmented Dynamic Panel Data Models, *Journal of Business & Economic Statistics* 28, 410-422.

[21] Phillips, R.F., 2015, On quasi-maximum likelihood estimation of dynamic panel models, *Economic Letters* 137, 91-94.

Table 1: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP1 (6.1) and Case 1

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
10	estim	0.7043	0.5014	0.4653	0.6976	0.5005	0.4810	0.6828	0.4990	0.4914
	bias	0.2043	0.0014	-0.0347	0.1976	0.0005	-0.0190	0.1828	-0.0010	-0.0086
	rmse	0.2501	0.0399	0.0698	0.2333	0.0274	0.0487	0.2071	0.0174	0.0296
	size	38.5%	5.2%	9.1%	34.4%	5.1%	7.8%	36.2%	4.4%	7%
100	estim	0.5002	0.4998	0.4843	0.5001	0.4997	0.4916	0.5003	0.4999	0.4922
	bias	0.0002	-0.0002	-0.0157	0.0001	-0.0003	-0.0084	0.0003	-0.0001	-0.0078
	rmse	0.0088	0.0088	0.0182	0.0063	0.0063	0.0108	0.0042	0.0042	0.0091
	size	5%	5%	41.7%	5.6%	5.5%	22.2%	6.1%	6%	40%
200	estim	0.4998	0.4997	0.4918	0.4999	0.4998	0.4964	0.5000	0.4999	0.4967
	bias	-0.0002	-0.0003	-0.0082	-0.0001	-0.0002	-0.0036	0.0000	-0.00001	-0.0033
	rmse	0.0061	0.0062	0.0103	0.0044	0.0045	0.0057	0.0028	0.0028	0.0044
	size	4.4%	4.7%	26.1%	5.2%	5.2%	12.3%	5.1%	5%	20.2%

Note: 1. "nGLS" refers to the naive GLS (2.5) treating y_{i0} as fixed, "GLS" refers to the naive GLS (2.15) treating y_{i0} as random. "GMM" refers to the Arellano and Bond (1990) type GMM.

2. size is calculated for $H_0 : \rho = 0.5$.

Table 2: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP1 (6.1) and Case 2

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
10	estim	0.6991	0.4999	0.4585	0.6988	0.4988	0.4744	0.6893	0.4990	0.4887
	bias	0.1991	-0.0001	-0.0415	0.1988	-0.0012	-0.0256	0.1893	-0.0010	-0.0113
	rmse	0.2494	0.0435	0.0776	0.2379	0.0294	0.0548	0.2168	0.0191	0.0332
	size	37.5%	5.3%	9.8%	34.2%	5.3%	8.4%	35.2%	4.7%	6.6%
100	estim	0.5005	0.5001	0.4843	0.5002	0.4999	0.4907	0.5004	0.5000	0.4960
	bias	0.0005	0.0001	-0.0157	0.0002	-0.0001	-0.0093	0.0004	0.0000	-0.0040
	rmse	0.0096	0.0096	0.0187	0.0071	0.0071	0.0120	0.0044	0.0045	0.0062
	size	5.2%	5.1%	34%	5.1%	5.4%	24.2%	5.2%	4.9%	13.2%
200	estim	0.4998	0.4997	0.4918	0.5000	0.4999	0.4922	0.4998	0.4997	0.4961
	bias	-0.0002	-0.0003	-0.0082	0.0000	-0.0001	-0.0078	-0.0002	-0.0003	-0.0039
	rmse	0.0067	0.0067	0.0106	0.0050	0.0050	0.0093	0.0030	0.0030	0.0050
	size	5.3%	5.4%	22.8%	5.3%	5.4%	32.2%	5.2%	5.4%	22.2%

Note: size is calculated for $H_0 : \rho = 0.5$. See note 1 of Table 1.

Table 3: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP2 (6.2) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
10	estim	0.5324	0.5147	0.4988	0.4884	0.4921	0.5318	0.5142	0.4983	0.4879	0.4957	0.5314	0.5137	0.4981	0.4880	0.4980
	bias	0.0324	0.01457	-0.0012	-0.0116	-0.0079	0.0318	0.0142	-0.0017	-0.0121	-0.0043	0.0314	0.0137	-0.0019	-0.0120	-0.0020
	rmse	0.0360	0.0228	0.0162	0.0191	0.0287	0.0335	0.0183	0.0109	0.0159	0.0207	0.0322	0.0159	0.0077	0.0139	0.0133
	size	53.1%	13%	5.5%	12.4%	6.5%	85.9%	22.7%	5.4%	21.9%	6%	99.6%	40.1%	6.5%	38.4%	6%
100	estim	0.5020	0.5001	0.5000	0.4990	0.4964	0.5020	0.5002	0.5001	0.4991	0.4983	0.5020	0.5002	0.5001	0.4991	0.4993
	bias	0.0020	0.0001	0.0000	-0.0010	-0.0036	0.0020	0.0002	0.0001	-0.0009	-0.0017	0.0020	0.0002	0.0001	-0.0009	-0.0007
	rmse	0.0045	0.0040	0.0040	0.0041	0.0056	0.0035	0.0029	0.0029	0.0030	0.0035	0.0027	0.0018	0.0018	0.0020	0.0021
	size	7.7%	5.2%	5.3%	6%	13.5%	10.4%	5.1%	4.9%	6.1%	8.9%	20.5%	4.8%	4.6%	7.8%	6.8%
200	estim	0.5009	0.5001	0.5000	0.4995	0.4983	0.5010	0.5000	0.5000	0.4996	0.4983	0.5008	0.5000	0.5000	0.4995	0.4993
	bias	0.0009	0.0001	0.0000	-0.0005	-0.0017	0.0010	0.0000	0.0000	-0.0004	-0.0017	0.0008	0.0000	0.0000	-0.0005	-0.0007
	rmse	0.0030	0.0029	0.0029	0.0029	0.0034	0.0024	0.0022	0.0022	0.0022	0.0028	0.0016	0.0013	0.0013	0.0014	0.0016
	size	6.3%	4.6%	4.7%	4.6%	9.1%	7.1%	4.8%	4.6%	5.1%	10.6%	11.1%	5.1%	4.8%	6.5%	7.1%

Note: 1. "nGLS" refers to the naive GLS (3.2), "GLS1" referes to the naive GLS (4.5), "GLS2" refers to the GLS estimator (4.7), "GLS3" refers to the conditional GLS conditional on y_{i0} alone ((4.11)). "GMM" refers to Arellano-Bond type GMM estimator.

2. size is calculated for $H_0 : \rho = 0.5$.

Table 4: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP2 (6.2) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
10	estim	0.5316	0.5144	0.4987	0.4880	0.4927	0.5306	0.5129	0.4974	0.4874	0.4943	0.5314	0.5140	0.4985	0.4885	0.4974
	bias	0.0316	0.0144	-0.0013	-0.0120	-0.0073	0.0306	0.0129	-0.0026	-0.0126	-0.0057	0.0314	0.0140	-0.0015	-0.0115	-0.0026
	rmse	0.0352	0.0226	0.0166	0.0197	0.0316	0.0325	0.0175	0.0114	0.0165	0.0213	0.0322	0.0159	0.0073	0.0134	0.0138
	size	54.2%	13%	4.3%	11.2%	5.5%	79.6%	18.7%	6.2%	22.2%	5.7%	99.3%	45.7%	6.4%	38%	5.5%
100	estim	0.5019	0.5001	0.5000	0.4990	0.4965	0.5018	0.5000	0.4999	0.4989	0.4980	0.5019	0.5001	0.5000	0.4991	0.4993
	bias	0.0019	0.0001	0.0000	-0.0010	-0.0035	0.0018	0.0000	-0.0001	-0.0011	-0.0020	0.0019	0.0001	0.0000	-0.0009	-0.0007
	rmse	0.0048	0.0044	0.0044	0.0045	0.0059	0.0034	0.0029	0.0029	0.0031	0.0039	0.0027	0.0019	0.0018	0.0021	0.0022
	size	6.3%	4.9%	5%	6.1%	12%	9.1%	5.8%	6%	6.8%	9%	18.2%	4.9%	5.2%	7.9%	5.9%
200	estim	0.5008	0.4999	0.4999	0.4994	0.4982	0.5008	0.5000	0.4999	0.4994	0.4982	0.5009	0.5000	0.5000	0.4995	0.4993
	bias	0.0008	-0.0001	-0.0001	-0.0006	-0.0018	0.0008	0.0000	-0.0001	-0.0006	-0.0018	0.0009	0.0000	0.0000	-0.0005	-0.0007
	rmse	0.0030	0.0029	0.0029	0.0030	0.0035	0.0022	0.0020	0.0020	0.0021	0.0028	0.0016	0.0013	0.0000	0.0014	0.0015
	size	6%	5.6%	5.6%	5.6%	11%	7.1%	5.4%	5.5%	6.1%	12%	10.8%	4.9%	4.9%	6.2%	8.3%

Note: size is calculated for $H_0 : \rho = 0.5$. See note 1 of Table 2.

Table 5: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP3 (6.3) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.5086	0.5106	0.5008	0.4826	0.4926	0.5079	0.5101	0.4998	0.4821	0.4953	0.5076	0.5097	0.4994	0.4821	0.4981
10	bias	0.0086	0.0106	0.0008	-0.0174	-0.0074	0.0079	0.0101	-0.0002	-0.0179	-0.0047	0.0076	0.0097	-0.0006	-0.0179	-0.0019
	rmse	0.0173	0.0201	0.0147	0.0232	0.0279	0.0128	0.0152	0.0098	0.0207	0.0199	0.0102	0.0125	0.0066	0.0193	0.0128
	size	8.8%	9.3%	5%	19.9%	5.8%	11.7%	14.4%	5.2%	42.2%	5.6%	19.3%	22.9%	5.7%	71%	5.9%
	estim	0.5001	0.5001	0.5001	0.4985	0.4964	0.5002	0.5002	0.5001	0.4986	0.4983	0.5002	0.5001	0.5001	0.4985	0.4993
100	bias	0.0001	0.0001	0.0001	-0.0015	-0.0036	0.0002	0.0002	0.0001	-0.0014	-0.0017	0.0002	0.0001	0.0001	-0.0015	-0.0007
	rmse	0.0040	0.0040	0.0040	0.0043	0.0056	0.0029	0.0029	0.0029	0.0032	0.0035	0.0018	0.0018	0.0018	0.0023	0.0022
	size	5.4%	5.2%	5.1%	6.7%	13.5%	4.5%	5%	4.7%	7.5%	9%	4.8%	4.8%	4.9%	13.3%	6.7%
	estim	0.5000	0.5001	0.5001	0.4993	0.4983	0.5001	0.5001	0.5001	0.4993	0.4983	0.5000	0.5000	0.5000	0.4992	0.4993
200	bias	0.0000	0.0001	0.0001	-0.0007	-0.0017	0.0001	0.0001	0.0001	-0.0007	-0.0017	0.0000	0.0000	0.0000	-0.0002	-0.0007
	rmse	0.0029	0.0029	0.0029	0.0030	0.0034	0.0022	0.0022	0.0022	0.0022	0.0028	0.0013	0.0013	0.0013	0.0015	0.0022
	size	4.9%	4.6%	5.1%	4.6%	8.7%	4.7%	4.7%	4.7%	6.2%	10.3%	4.8%	5.1%	5.1%	8.5%	6.7%

Note: 1. "nGLS" refers to the naive GLS (3.2), "GLS1" refers to the naive generalized least squares (GLS) estimator with η_i conditional on x_i (4.5), "GLS2" refers to the GLS estimator with η_i conditional on x_i (4.7), "GLS3" refers to the conditional GLS estimator (4.15) and "GMM" refers to Arellano-Bond type GMM estimator.

2. size is calculated for $H_0 : \rho = 0.5$.

Table 6: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.5$ for DGP3 (6.3) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.5079	0.5103	0.5003	0.4822	0.4932	0.5070	0.5089	0.4991	0.4816	0.4947	0.5078	0.5100	0.4997	0.4827	0.497
10	bias	0.0079	0.0103	0.0003	-0.0178	-0.0068	0.0070	0.0089	-0.0009	-0.0184	-0.0053	0.0078	0.0100	-0.0003	-0.0173	-0.003
	rmse	0.0168	0.0200	0.0146	0.0237	0.0303	0.0126	0.0147	0.0102	0.0212	0.0205	0.0103	0.0125	0.0065	0.0186	0.013
	size	7.8%	9.7%	5.1%	20.5%	5.4%	8.5%	11.3%	5.1%	40.2%	5.5%	21.8%	27.7%	4.5%	73%	5.9%
	estim	0.5001	0.5001	0.5001	0.4985	0.4965	0.5000	0.5000	0.4999	0.4984	0.4980	0.5001	0.5001	0.5001	0.4985	0.499
100	bias	0.0001	0.0001	0.0001	-0.0015	-0.0035	0.0000	0.0000	-0.0001	-0.0016	-0.0020	0.0001	0.0001	0.0001	-0.0015	-0.000
	rmse	0.0044	0.0044	0.0001	0.0047	0.0059	0.0029	0.0029	0.0029	0.0033	0.0039	0.0018	0.0018	0.0018	0.0024	0.002
	size	5.1%	5.1%	5.2%	7.8%	12.1%	5.9%	5.9%	5.8%	8.2%	8.8%	5.5%	5.1%	5.2%	12.2%	5.8%
	estim	0.4999	0.4999	0.4999	0.4991	0.4982	0.4999	0.4999	0.4999	0.4992	0.4982	0.5000	0.5000	0.5000	0.4993	0.499
200	bias	-0.0001	-0.0001	-0.0001	-0.0009	-0.0018	-0.0001	-0.0001	-0.0001	-0.0008	-0.0018	0.0000	0.0000	0.0000	-0.0007	-0.000
	rmse	0.0029	0.0029	0.0029	0.0031	0.0035	0.0020	0.0020	0.0020	0.0022	0.0028	0.0013	0.0013	0.0013	0.0015	0.001
	size	5%	5.6%	5%	6.4%	10.7%	5.7%	5.4%	5.7%	7.6%	11.8%	5.4%	4.9%	5.2%	8.8%	8.1%

Note: size is calculated for $H_0 : \rho = 0.5$. See note 1 of Table 3.

Table 7: Sample mean, bias, RMSE and size of $\hat{\rho}$ using Chamberlain approach when $N \geq T$ for DGP2 (6.2) and Case 1

		$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.8$		
		GLS1*	GLS2*	GLS3*	GLS1*	GLS2*	GLS3*	GLS1*	GLS2*	GLS3*
$N = 100$	estim	0.1966	0.1966	0.1966	0.4967	0.4967	0.4967	0.7972	0.7972	0.7972
$T = 100$	bias	-0.0034	-0.0034	-0.0034	-0.0033	-0.0033	-0.0033	-0.0028	-0.0028	-0.0028
	rmse	0.0062	0.0062	0.0062	0.0052	0.0052	0.0052	0.0037	0.0037	0.0037
	size	10.6%	10.6%	10.6%	12.6%	12.6%	12.6%	21.1%	21.1%	21.1%
$N = 200$	estim	0.1983	0.1983	0.1983	0.4984	0.4984	0.4984	0.7986	0.7987	0.7986
$T = 200$	bias	-0.0017	-0.0017	-0.0017	-0.0016	-0.0016	-0.0016	-0.0014	-0.0013	-0.0014
	rmse	0.0032	0.0032	0.0032	0.0027	0.0027	0.0027	0.0019	0.0018	0.0019
	size	9%	9%	9%	10.6%	10.6%	10.6%	20.7%	20.7%	20.7%
$N = 200$	estim	0.2002	0.2002	0.2002	0.5002	0.5001	0.5001	0.8002	0.8000	0.8000
$T = 100$	bias	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0002	0.0000	0.0000
	rmse	0.0037	0.0037	0.0037	0.0029	0.0029	0.0029	0.0018	0.0017	0.0017
	size	4.8%	5%	5%	5.2%	4.9%	4.9%	4.3%	4.5%	4.5%
$N = 500$	estim	0.2000	0.2000	0.2000	0.5000	0.5000	0.5000	0.8001	0.8000	0.8000
$T = 200$	bias	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
	rmse	0.0017	0.0017	0.0017	0.0013	0.0013	0.0013	0.0008	0.0008	0.0008
	size	5.8%	5.8%	5.8%	5.1%	4.8%	4.8%	4.8%	4.8%	4.8%

Note: 1. GLS1*, GLS2* and GLS3* use (4.5), (4.7) and (4.11), respectively, with $\tilde{\mathbf{x}}_i = (1, \mathbf{x}_i')'$.

2. size calculated for $H_0 : \rho = \rho_0$ where $\rho_0 = 0.2, 0.5$ and 0.8.

Table 8: Properties of QMLE for dynamic panel model when $N \to \infty$

	T fixed		$(N,T) \to \infty, \frac{N}{T} \to a \neq 0 < \infty$		$(N,T) \to \infty, \frac{N}{T^3} \to c \neq 0 < \infty$	
	Consistency	Asymptotic Bias	Consistency	Asymptotic Bias	Consistency	Asymptotic Bias
RE model						
y_{i0} fixed	No	∞	Yes	No	Yes	$O(\sqrt{c})$
y_{i0} random	Yes	No	Yes	No	Yes	No
FE model						
y_{i0} fixed	No	∞	Yes	$O(\sqrt{a})$	Yes	∞
y_{i0} random	No	∞	Yes	$O(\sqrt{a})$	Yes	∞
FE model with Mundlak-Chammberlain Adjustment						
y_{i0} fixed	No	∞	Yes	No	Yes	$O(\sqrt{c})$
y_{i0} random	Yes	No	Yes	No	Yes	No

Note: When $E(x_{it}\eta_i) = 0$, we call the model random effects model (RE), when $E(x_{it}\eta_i) \neq 0$, we call the model fixed effects model (FE). For the FE model, we consider the GLS ((3.8)) and naive GLS ((3.2)) of model (3.1). For the FE model with Mundlak-Chamberlin approach, we consider the GLS ((4.7))and naive GLS ((4.5)) of model (4.2).

Appendix

A.1 Derivation of Lemma 2.1

We first note that by substituting (2.2) into (2.5) yields

$$\hat{\rho}_{QMLE,f}-\rho=\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}^{-1}\mathbf{y}_{i,-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\right). \tag{A.1}$$

Making use of (2.7), we obtain

$$\begin{aligned}\sum_{t=1}^{T}y_{i,t-1} &= \frac{1-\rho^T}{1-\rho}y_{i0}+\frac{1}{1-\rho}\left[(T-1)-\rho\frac{1-\rho^{T-1}}{1-\rho}\right]\eta_i \\ &+\left\{\frac{1-\rho^{T-1}}{1-\rho}u_{i1}+\frac{1-\rho^{T-2}}{1-\rho}u_{i2}+\cdots+\frac{1-\rho^2}{1-\rho}u_{i,T-2}+u_{i,T-1}\right\}.\end{aligned} \tag{A.2}$$

Following Alvarez and Arellano (2003), it can be shown that the denominator of (A.1) divided by NT converges to $\frac{1}{1-\rho^2}$ as $(N,T)\to\infty$. When $N\to\infty$ or $(N,T)\to\infty$,

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\left(\mathbf{I}_T-\frac{\sigma_\eta^2}{1+T\sigma_\eta^2}1_T1_T'\right)1_T\eta_i \\ =&\sqrt{\frac{N}{T}}\frac{T\sigma_\eta^2}{\left(1+T\sigma_\eta^2\right)(1-\rho)^2}\frac{(T-1)-T\rho+\rho^T}{T} \\ &+\sqrt{\frac{N}{T^3}}\frac{T}{\left(1+T\sigma_\eta^2\right)(1-\rho)}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}y_{i0}\eta_i+o(1),\end{aligned} \tag{A.3}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\left(\mathbf{I}_T-\frac{\sigma_\eta^2}{1+T\sigma_\eta^2}1_T1_T'\right)\mathbf{u}_i \\ =&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}u_{it}-\sqrt{\frac{N}{T}}\frac{T\sigma_\eta^2}{\left(1+T\sigma_\eta^2\right)(1-\rho)^2}\frac{(T-1)-T\rho+\rho^T}{T}.\end{aligned} \tag{A.4}$$

The second term of (A.4) cancels out with the first term of (A.3)[9]. The first term of the numerator of (A.4) converges to a normal distribution under fairly general conditions (e.g., u_{it} and η_i are normally distributed or $E\left|u_{it}\right|^{4+\epsilon}<\infty$ and $E\left|\eta_i\right|^{4+\epsilon}<\infty$ for some $\epsilon>0$, see, for

[9] We owe this point to the private communication with Jushan Bai.

example, Hsiao and Zhang (2015) or Hsiao and Zhou (2016)). The first term of the numerator of (A.3) is $O_p\left(\sqrt{\frac{N}{T^3}}\right)$ if $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} y_{i0}\eta_i$ is a nonzero constant. Thus, if N and T are of similar order, $\frac{N}{T} \to a < \infty$ as Bai (2013) has assumed, the estimator (2.5) is asymptotically unbiased. On the other hand, if N is much larger than T such that $\frac{N}{T^3} \to c \neq 0$, the second term of the numerator multiplied by $\sqrt{NT}$ converges to a constant that is proportional to $\sqrt{c}$.

A.2 Derivation of Lemma 2.2

The inverse of variance-covariance matrix $\check{\mathbf{V}}$ is given by

$$\check{\mathbf{V}}^{-1} = \begin{pmatrix} \omega^{-1} & -\sigma_0^{-2}\sigma_1^2 1_T' \mathbf{C}^{-1} \\ -\sigma_0^{-2}\mathbf{C}^{-1}\sigma_1^2 1_T & \mathbf{C}^{-1} \end{pmatrix}, \tag{A.5}$$

where

$$\begin{aligned} \omega &= \sigma_0^2 - \sigma_1^2 1_T' \left(I_T + \sigma_\eta^2 1_T 1_T'\right)^{-1} \sigma_1^2 1_T \\ &= \sigma_0^2 - \frac{\sigma_1^4 T}{1 + T\sigma_\eta^2}, \end{aligned}$$

and

$$\mathbf{C} = I_T + \tilde{\sigma}_\eta^2 1_T 1_T', \text{ with } \mathbf{C}^{-1} = \mathbf{I}_T - \frac{\tilde{\sigma}_\eta^2}{1 + T\tilde{\sigma}_\eta^2} 1_T 1_T'. \tag{A.6}$$

where $\tilde{\sigma}_\eta^2 = \sigma_\eta^2 - \sigma_1^4\sigma_0^{-2}$.

Consequently, we have

$$\begin{aligned} \sqrt{NT}\left(\hat{\rho}_{QMLE,r} - \rho\right) &= \left(\frac{1}{NT}\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{C}^{-1} \mathbf{y}_{i,-1}\right)^{-1} \\ &\quad \times \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{y}_{i,-1}'\mathbf{C}^{-1}\left(\eta_i 1_T + \mathbf{u}_i\right) - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(y_{i0} - \mu\right)\sigma_0^{-2}\sigma_1^2 1_T' \mathbf{C}^{-1}\mathbf{y}_{i,-1}\right) \end{aligned} \tag{A.7}$$

We note that $\mathbf{C}$ and $\mathbf{C}^{-1}$ are of the same form as $\mathbf{V}$ and $\mathbf{V}^{-1}$, then

$$\begin{aligned} E\left(\mathbf{y}_{i,-1}'\mathbf{C}^{-1}\eta_i 1_T\right) &= \frac{1}{1 + T\tilde{\sigma}_\eta^2} E\left(\mathbf{y}_{i,-1}' 1_T \eta_i\right) = \frac{1}{1 + T\tilde{\sigma}_\eta^2} E\left(\sum_{t=1}^{T} y_{i,t-1}\eta_i\right) \\ &= \frac{1}{1 + T\tilde{\sigma}_\eta^2} E\left(y_{i0}\eta_i + \sum_{t=1}^{T-1} y_{it}\eta_i\right) \\ &= \frac{1}{1 + T\tilde{\sigma}_\eta^2}\left(\sigma_1^2 + \sum_{t=1}^{T-1}\left(\gamma^t E\left(y_{i0}\eta_i\right) + E\left(\eta_i^2\right)\frac{1-\rho^t}{1-\rho}\right)\right) \\ &= \frac{1}{1 + T\tilde{\sigma}_\eta^2}\left(\frac{1-\rho^T}{1-\rho}\sigma_1^2 + \frac{\sigma_\eta^2}{1-\rho}\left(T - 1 - \frac{\rho - \rho^T}{1-\rho}\right)\right), \end{aligned} \tag{A.8}$$

and

$$
\begin{aligned}
E\left(\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{u}_i\right) &= E\left(\mathbf{y}_{i,-1}'\mathbf{u}_i\right) - \frac{\tilde{\sigma}_\eta^2}{1+T\tilde{\sigma}_\eta^2}E\left(\mathbf{y}_{i,-1}'1_T1_T'\mathbf{u}_i\right) = -\frac{\tilde{\sigma}_\eta^2}{1+T\tilde{\sigma}_\eta^2}E\left(\sum_{t=1}^{T} y_{i,t-1}\sum_{t=1}^{T}u_{it}\right) \\
&= -\frac{\tilde{\sigma}_\eta^2}{1+T\tilde{\sigma}_\eta^2}\left(\sum_{t=1}^{T-1}E\left(y_{it}u_{it}\right) + \sum_{s>t}^{T-1}E\left(y_{is}u_{it}\right)\right) \\
&= -\frac{\tilde{\sigma}_\eta^2}{1+T\tilde{\sigma}_\eta^2}\left((T-1) + \sum_{t=1}^{T-2}\sum_{s=t+1}^{T-1}\rho^{s-t}\right) \\
&= -\frac{\tilde{\sigma}_\eta^2}{1+T\tilde{\sigma}_\eta^2}\frac{1}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right). \qquad \text{(A.9)}
\end{aligned}
$$

Also,

$$
\begin{aligned}
& E\left[\left(y_{i0}-\mu\right)\sigma_0^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right] \\
= \;& \sigma_0^{-2}\sigma_1^2\frac{1}{1+T\tilde{\sigma}_\eta^2}E\left(v_{i0}1_T'\mathbf{y}_{i,-1}\right) \\
= \;& \sigma_0^{-2}\sigma_1^2\frac{1}{1+T\tilde{\sigma}_\eta^2}\left[\frac{\sigma_1^2}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) + \frac{1-\rho^T}{1-\rho}\sigma_0^2\right] \qquad \text{(A.10)}
\end{aligned}
$$

since

$$
\begin{aligned}
E\left(v_{i0}1_T'\mathbf{y}_{i,-1}\right) &= E\left(\sum_{t=1}^{T} y_{i,t-1}v_{i0}\right) = E\left(y_{i0}v_{i0}\right) + \sum_{t=1}^{T-1}E\left(y_{it}v_{i0}\right) \\
&= \sigma_0^2 + \sum_{t=1}^{T-1}E\left(\alpha_i\frac{1-\rho^t}{1-\rho} + \rho^t y_{i0} + \sum_{s=0}^{t-1}\rho^s u_{t-s}\right)v_{i0} \\
&= \sigma_0^2 + \sum_{t=1}^{T-1}\frac{1-\rho^t}{1-\rho}\sigma_1^2 + \sum_{t=1}^{T-1}\rho^t\sigma_0^2, \\
&= \frac{\sigma_1^2}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) + \frac{1-\rho^T}{1-\rho}\sigma_0^2.
\end{aligned}
$$

Combining (A.8)-(A.10), it follows that the numerator of (A.7) has expected value 0, thus the QMLE of γ treating y_{i0} as random variable is asymptotically unbiased as long as $N \to \infty$.

A.3 Derivation of Lemma 3.1 and Lemma 3.2

Rewrite (3.2) as

$$\begin{aligned}\begin{pmatrix}\hat{\rho}_{nGLS}\\ \hat{\beta}_{nGLS}\end{pmatrix} &= \left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{V}^{-1}\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{V}^{-1}\mathbf{y}_i\right)\\ &= \begin{pmatrix}\rho\\ \beta\end{pmatrix}+\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{V}^{-1}\mathbf{Z}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\right)\\ &= \begin{pmatrix}\rho\\ \beta\end{pmatrix}+\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{V}^{-1}\mathbf{Z}_{i,-1}\right)^{-1}\begin{pmatrix}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\\ \sum_{i=1}^{N}\mathbf{x}_i'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\end{pmatrix},\end{aligned} \tag{A.11}$$

where $\mathbf{V}$ is defined in (2.4).

For the numerator of the second term of (A.11), we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{V}^{-1}\eta_i 1_T=\sqrt{\frac{N}{T}}\frac{T}{1+T\sigma_\eta^2}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\eta_i. \tag{A.12}$$

Under the assumption that $E\left(x_{it}\eta_i\right)=0$, $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\eta_i=0$. The limiting distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)$ is normally distributed with mean zero. Lemma 3.1 follows from the proof of Lemma 2.1.

When $E\left(x_{it}\eta_i\right)\neq 0$, $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\eta_i\neq 0$, (A.12) converges to

$$\frac{T\varsigma}{1+T\sigma_\eta^2}=O\left(1\right), \tag{A.13}$$

where $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\bar{x}_i\eta_i=\varsigma$. Thus, when $T\to\infty$, (A.11) converges to $\left(\rho,\beta\right)'$, However, the numerator of the last term of (A.11) divided by $\sqrt{NT}$ reduces to

$$\begin{aligned}&\begin{pmatrix}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{V}^{-1}\left(\eta_i 1_T+\mathbf{u}_i\right)\end{pmatrix}\\ =&\begin{pmatrix}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}u_{it}+\frac{\sqrt{NT}}{T\left(1+T\sigma_\eta^2\right)}\frac{1-\rho^T}{1-\rho}\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}y_{i0}\eta_i\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}u_{it}+\sqrt{\frac{N}{T}}\frac{T\varsigma}{1+T\sigma_\eta^2}\end{pmatrix}+o_p\left(1\right)\end{aligned} \tag{A.14}$$

Under assumptions A1(a) and A2, it can be easily verified that the terms $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}u_{it}$ and $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}u_{it}$ converge to normal distributions with zero means if u_{it} and η_i are normally distributed or $E\left|u_{it}\right|^{4+\epsilon}<\infty$ and $E\left|\eta_i\right|^{4+\epsilon}<\infty$ for some $\epsilon>0$ (see, for example, Hsiao and Zhang (2015) or Hsiao and Zhou (2016)). However, the second term in the first element of the last vector on the right hand side of (A.14) is of order $\sqrt{\frac{N}{T^3}}$ and the second term of the second element is of order $\sqrt{\frac{N}{T}}$. Therefore, the naive GLS for the general fixed effects estimator is asymptotically biased of order $\sqrt{\frac{N}{T}}$ as $(N,T)\to\infty$.

A.4 Derivation of Lemma 3.3

The proof follows that of Lemma 3.2, here we only provide the sketch of the proof. We note that

$$\begin{pmatrix} \hat{\mathbf{b}}_{GLS} \\ \hat{\boldsymbol{\delta}}_{GLS} \end{pmatrix} - \begin{pmatrix} \mathbf{b} \\ \boldsymbol{\delta} \end{pmatrix} = \left(\sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{Z}}_i \right)^{-1} \sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \tilde{\mathbf{V}}^{-1} \begin{pmatrix} v_{i0} \\ 1_T \eta_i + \mathbf{u}_i \end{pmatrix}, \tag{A.15}$$

where $\tilde{\mathbf{V}}^{-1}$ essentially takes the similar form as (A.5) except for few notation changes, i.e.,

$$\tilde{\mathbf{V}}^{-1} = \begin{pmatrix} \tilde{\omega}^{-1} & -\frac{\sigma_\tau^2}{\sigma_0^2} 1_T' \tilde{\mathbf{C}}^{-1} \\ -\frac{\sigma_\tau^2}{\sigma_0^2} \tilde{\mathbf{C}}^{-1} 1_T & \tilde{\mathbf{C}}^{-1} \end{pmatrix}, \tag{A.16}$$

where $\tilde{\omega} = \sigma_0^2 - \frac{\sigma_\tau^4}{1+T\sigma_\eta^2}$ and

$$\tilde{\mathbf{C}}^{-1} = \left(I_T + \check{\sigma}_\eta^2 1_T 1_T' \right)^{-1} = \mathbf{I}_T - \frac{\check{\sigma}_\eta^2}{1 + T\check{\sigma}_\eta^2} 1_T 1_T'. \tag{A.17}$$

with $\check{\sigma}_\eta^2 = \sigma_\eta^2 - \sigma_\tau^4 \sigma_0^{-2}$.

Now the numerator of (A.15) divided by $\sqrt{NT}$ takes the form,

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \begin{pmatrix} \tilde{\mathbf{x}}_i & 0 \\ 0 & \mathbf{Z}_i' \end{pmatrix} \begin{pmatrix} \tilde{\omega}^{-1} & -\sigma_0^{-2} \sigma_\tau^2 1_T' \tilde{\mathbf{C}}^{-1} \\ -\sigma_0^{-2} \tilde{\mathbf{C}}^{-1} \sigma_\tau^2 1_T & \tilde{\mathbf{C}}^{-1} \end{pmatrix} \begin{pmatrix} v_{i0} \\ 1_T \eta_i + \mathbf{u}_i \end{pmatrix} \\ = & \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \begin{pmatrix} \tilde{\mathbf{x}}_i & 0 \\ 0 & \mathbf{Z}_i' \end{pmatrix} \begin{pmatrix} \tilde{\omega}^{-1} v_{i0} - \sigma_0^{-2} \sigma_\tau^2 1_T' \tilde{\mathbf{C}}^{-1} (1_T \eta_i + \mathbf{u}_i) \\ -\sigma_0^{-2} \tilde{\mathbf{C}}^{-1} \sigma_\tau^2 1_T v_{i0} + \tilde{\mathbf{C}}^{-1} (1_T \eta_i + \mathbf{u}_i) \end{pmatrix} \\ = & \sqrt{\frac{N}{T}} \frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} \tilde{\mathbf{x}}_i \left[\tilde{\omega}^{-1} v_{i0} - \sigma_0^{-2} \sigma_\tau^2 1_T' \tilde{\mathbf{C}}^{-1} (1_T \eta_i + \mathbf{u}_i) \right] \\ \mathbf{y}_{i,-1}' \left[-\sigma_0^{-2} \tilde{\mathbf{C}}^{-1} \sigma_\tau^2 1_T v_{i0} + \tilde{\mathbf{C}}^{-1} (1_T \eta_i + \mathbf{u}_i) \right] \\ \mathbf{x}_i' \left[-\sigma_0^{-2} \tilde{\mathbf{C}}^{-1} \sigma_\tau^2 1_T v_{i0} + \tilde{\mathbf{C}}^{-1} (1_T \eta_i + \mathbf{u}_i) \right] \end{pmatrix}. \end{aligned} \tag{A.18}$$

For the first case when $E(x_{it}\eta_i) = 0$, we observe the expectation of the first and third component of (A.18) is zero, and the second component also has zero mean by following the derivation in (A.8)-(A.10). As a result, if $E(x_{it}\eta_i) = 0$, (A.18) will have zero expectation either T is fixed or goes to infinity, i.e., the GLS estimator (3.8) is consistent and asymptotically unbiased as long as $N \to \infty$.

For the second case when $E(x_{it}\eta_i) \neq 0$, then we can observe that the expectation of the first and third component of (A.18) is no longer zero, and the expectation will be a finite constant depending on $E(x_{it}\eta_i)$. If T is fixed, in view of (A.14), $\frac{1}{N}\sum_{i=1}^{N} \tilde{\mathbf{Z}}_i' \tilde{\mathbf{V}}^{-1} \begin{pmatrix} v_{i0} \\ 1_T \eta_i + \mathbf{u}_i \end{pmatrix}$ will not converge to zero as $N \to \infty$, and will be $O_p(1)$, i.e., the GLS estimator (3.8) is inconsistent. If $T \to \infty$ and $\frac{N}{T} \to a \neq 0 < \infty$, it is asymptotically biased of order $\sqrt{a}$.

A.5 Derivation of Lemma 4.1

It follows from (4.5) that

$$\hat{\boldsymbol{\theta}}_{nGLS}-\boldsymbol{\theta}=\left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ \tilde{\mathbf{x}}_i 1_T'\end{pmatrix}\bar{\mathbf{V}}^{-1}\left(\mathbf{Z}_i,1_T\tilde{\mathbf{x}}_i'\right)\right]^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ \tilde{\mathbf{x}}_i 1_T'\end{pmatrix}\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)\right], \quad (A.19)$$

where the numerator on the right hand side (RHS) divided by $\sqrt{NT}$ is

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ \tilde{\mathbf{x}}_i 1_T'\end{pmatrix}\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)\\ \tilde{\mathbf{x}}_i 1_T'\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)\end{pmatrix}, \quad (A.20)$$

using the relation

$$\bar{\mathbf{V}}^{-1}1_T=\frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}1_T,$$

we obtain

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)\\ \tilde{\mathbf{x}}_i 1_T'\bar{\mathbf{V}}^{-1}\left(w_i^*1_T+\mathbf{u}_i\right)\end{pmatrix}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\begin{pmatrix}\frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\mathbf{y}_{i,-1}'1_Tw_i^*+\mathbf{y}_{i,-1}'\bar{\mathbf{V}}^{-1}\mathbf{u}_i\\ \frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\mathbf{x}_i'1_Tw_i^*+\mathbf{x}_i'\bar{\mathbf{V}}^{-1}\mathbf{u}_i\\ \frac{T}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\tilde{\mathbf{x}}_iw_i^*+\tilde{\mathbf{x}}_i1_T'\mathbf{u}_i\end{pmatrix}. \quad (A.21)$$

For model (4.2), under the projection (4.1), we have

$$E\left(\tilde{\mathbf{x}}_iw_i^*\right)=0,$$

also, under the assumption of strict exogeneity of $\mathbf{x}_i$, we have

$$E\left(\mathbf{x}_i'\mathbf{u}_i\right)=0.$$

Thus, the second and third elements of the RHS of (A.21) have zero expectation.

For the first element of the RHS of (A.21), we first note that by continuous substitution, model (4.2) can also be rewritten as

$$y_{it}=\rho^ty_{i0}+\frac{1-\rho^t}{1-\rho}w_i^*+\frac{1-\rho^t}{1-\rho}\tilde{\mathbf{x}}_i'\mathbf{b}^*+\sum_{j=1}^{t-1}\rho^jx_{i,t-j}\beta+\sum_{j=0}^{t-1}\rho^ju_{i,t-j}. \quad (A.22)$$

Then

$$\begin{aligned}&\frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\mathbf{y}_{i,-1}'1_Tw_i^*+\mathbf{y}_{i,-1}'\bar{\mathbf{V}}^{-1}\mathbf{u}_i\\ =\;&\frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\sum_{t=1}^{T}y_{i,t-1}w_i^*+\mathbf{y}_{i,-1}'\mathbf{u}_i-\frac{\sigma_{w^*}^2\sigma_u^{-2}}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\mathbf{y}_{i,-1}'1_T1_T'\mathbf{u}_i.\end{aligned} \quad (A.23)$$

The expected value of (A.23) in view of the derivation in (A.9) is equal to

$$\begin{aligned}
& E\left(\frac{1}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\mathbf{y}_{i,-1}'1_T w_i^* + \mathbf{y}_{i,-1}'\bar{\mathbf{V}}^{-1}\mathbf{u}_i\right) \\
= & \frac{\sigma_{w^*}^2}{1+T\sigma_{w^*}^2\sigma_u^{-2}}\sum_{t=1}^{T}\frac{1-\rho^{t-1}}{1-\rho} - \frac{\sigma_{w^*}^2\sigma_u^{-2}}{1+T\sigma_{w^*}^2\sigma_u^{-2}}E\left(\mathbf{y}_{i,-1}'1_T 1_T'\mathbf{u}_i\right) \\
= & \frac{\sigma_{w^*}^2}{(1-\rho)\left(1+T\sigma_{w^*}^2\sigma_u^{-2}\right)}\left[(T-1) - \frac{1-\rho^T}{1-\rho} - \left(T-1-\frac{\rho-\rho^T}{1-\rho}\right)\right] \\
= & -\frac{\sigma_{w^*}^2}{(1-\rho)\left(1+T\sigma_{w^*}^2\sigma_u^{-2}\right)} \\
= & O\left(\frac{1}{T}\right).
\end{aligned}$$

As a result, the first term (A.19) of A.21 is of order $\sqrt{\frac{N}{T^3}}$. Consequently, if T is fixed, then $\hat{\boldsymbol{\theta}}_{nGLS}$ is inconsistent. If $\frac{N}{T} \rightarrow a < \infty$ as T increases, $\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{nGLS} - \boldsymbol{\delta}\right)$ converges to a normally distributed variable centered at zero. However, if N tends to infinity faster than T so $\frac{N}{T^3} \rightarrow c \neq 0$, $\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{nGLS} - \boldsymbol{\delta}\right)$ is not centered at zero, and using the Chamberlain (1980)-Mundlak (1978) approach to get around the correlation between the effects and exogenous variable without proper consideration of initial value distribution will still yield estimators that are asymptotically biased of order $\sqrt{c}$.

A.6 Derivation of Lemma 4.2

We note (4.7) implies

$$\hat{\tilde{\boldsymbol{\theta}}}_{GLS} - \tilde{\boldsymbol{\theta}} = \left[\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\boldsymbol{\Omega}^{-1}\tilde{\mathbf{Z}}_i\right]^{-1}\left[\sum_{i=1}^{N}\tilde{\mathbf{Z}}_i'\boldsymbol{\Omega}^{-1}\begin{pmatrix} v_{i0} \\ w_i^* 1_T + \mathbf{u}_i \end{pmatrix}\right], \tag{A.24}$$

and $\boldsymbol{\Omega}^{-1}$ is similar in form to (A.16).

When $E\left(x_{it}\eta_i\right) = 0$, if the numerator of (A.24) have zero expectation, $\hat{\tilde{\boldsymbol{\theta}}}_{GLS}$ is asymptotically unbiased when $N \to \infty$. We note that the numerator of (A.24) divided by $\sqrt{NT}$ becomes

$$\frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\tilde{\mathbf{x}}_i' & 0\\ 0 & \tilde{\mathbf{x}}_i 1_T\\ 0 & \mathbf{x}_i'\\ 0 & \mathbf{y}_{i,-1}'\end{pmatrix}\begin{pmatrix}\tilde{\omega}^{-1} & -\sigma_0^{-2}\sigma_\tau^2 1_T'\tilde{\mathbf{C}}^{-1}\\ -\sigma_0^{-2}\tilde{\mathbf{C}}^{-1}\sigma_\tau^2 1_T & \tilde{\mathbf{C}}^{-1}\end{pmatrix}\begin{pmatrix}w_i + \frac{1}{1-\rho}\eta_i + \sum_{j=0}^{\infty}\rho^j u_{i,-j}\\ w_i^* 1_T + \mathbf{u}_i\end{pmatrix}$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\begin{pmatrix}\tilde{\mathbf{x}}_i' & 0\\ 0 & \tilde{\mathbf{x}}_i 1_T\\ 0 & \mathbf{x}_i'\\ 0 & \mathbf{y}_{i,-1}'\end{pmatrix}\begin{pmatrix}\tilde{\omega}^{-1} & -\sigma_0^{-2}\sigma_\tau^2 1_T'\tilde{\mathbf{C}}^{-1}\\ -\sigma_0^{-2}\tilde{\mathbf{C}}^{-1}\sigma_\tau^2 1_T & \tilde{\mathbf{C}}^{-1}\end{pmatrix}\begin{pmatrix}v_{i0}\\ \mathbf{v}_i\end{pmatrix}. \tag{A.25}$$

Since $E(x_{it}u_{it}) = 0$ and $E(x_{it}w_i^*) = 0$, whether $E(x_{it}\eta_i) = 0$ or not, then the expectation of (A.25) is zero as long as N is large, i.e., (A.24) is asymptotically unbiased.

For the conditional QMLE (4.11), we have

$$\begin{pmatrix}\hat{\boldsymbol{\delta}}_C\\ \widehat{\tilde{\mathbf{b}}}_C^*\\ \hat{\gamma}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\delta}_C\\ \tilde{\mathbf{b}}_C^*\\ \gamma\end{pmatrix} = \left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ \tilde{\mathbf{x}}_i 1_T'\\ y_{i0}1_T'\end{pmatrix}\mathbf{V}^{*-1}\left(\mathbf{Z}_i', 1_T\tilde{\mathbf{x}}_i', 1_T y_{i0}\right)\right]^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ \tilde{\mathbf{x}}_i 1_T'\\ y_{i0}1_T'\end{pmatrix}\mathbf{V}^{*-1}\mathbf{v}_i^*\right]. \tag{A.26}$$

By construction of $\mathbf{v}_i^*$ in (4.9), we have that

$$\begin{aligned} E\left(\mathbf{Z}_i'\mathbf{v}_i^*\right) &= 0,\\ E\left(\tilde{\mathbf{x}}_i'\mathbf{v}_i^*\right) &= 0,\\ E\left(y_{i0}\mathbf{v}_i^*\right) &= 0,\end{aligned}$$

as a result, the numerator of (4.11) divided by $\sqrt{NT}$ will have expectation zero as long as $N \to \infty$, i.e.,, the conditional QMLE (4.11) asymptotically unbiased as long as $N \to \infty$.

A.7 Derivation of Lemma 4.3

The derivation is the same as before, here we only sketch of the derivation. From (4.15), we have

$$\begin{pmatrix}\hat{\boldsymbol{\delta}}_C^*\\ \hat{\gamma}_C^*\end{pmatrix} - \begin{pmatrix}\boldsymbol{\delta}_C^*\\ \gamma_C^*\end{pmatrix} = \left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ y_{i0}1_T'\end{pmatrix}\tilde{\mathbf{V}}^{*-1}\left(\mathbf{Z}_i, 1_T y_{i0}\right)\right]^{-1}\left[\sum_{i=1}^{N}\begin{pmatrix}\mathbf{Z}_i'\\ y_{i0}1_T'\end{pmatrix}\tilde{\mathbf{V}}^{*-1}\tilde{\mathbf{v}}_i^*\right]. \tag{A.27}$$

We note that

$$E\left(y_{i0}v_{i0}^*\right) = E\left(y_{i0}\left(\mu + \beta\sum_{j=0}^{\infty}\rho^j x_{i,-j} + \sum_{j=0}^{\infty}\rho^j u_{i,-j}\right)\right) \neq 0,$$

so is $E\left(x_{it}\tilde{\mathbf{v}}_i^*\right) \neq 0$ if x_{it} is not independently distributed over t. Therefore, when T is fixed and $N \rightarrow \infty$,

$$\begin{aligned} \text{plim}_{N\rightarrow\infty} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \tilde{\mathbf{V}}^{*-1} \tilde{\mathbf{v}}_i^* &= \frac{1}{1+T\sigma_{\tilde{w}}^2} \text{plim}_{N\rightarrow\infty} \frac{1}{N} \sum_{i=1}^{N} v_{i0}^* \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_{it} \\ &= O\left(\frac{1}{T}\right), \end{aligned} \tag{A.28}$$

and

$$\begin{aligned} \text{plim}_{N\rightarrow\infty} \frac{1}{NT} \sum_{i=1}^{N} y_{i0} 1_T' \tilde{\mathbf{V}}^{*-1} \tilde{\mathbf{v}}_i^* &= \frac{1}{1+T\sigma_{\tilde{w}}^2} \text{plim}_{N\rightarrow\infty} \frac{1}{N} \sum_{i=1}^{N} v_{i0}^* y_{i0} \\ &= O\left(\frac{1}{T}\right). \end{aligned} \tag{A.29}$$

In other words, when T is fixed and $N \rightarrow \infty$, $\hat{\boldsymbol{\delta}}_C^*$ is inconsistent.

When $(N, T) \rightarrow \infty$,

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{Z}_i' \tilde{\mathbf{V}}^{*-1} \tilde{\mathbf{v}}_i^* = \sqrt{\frac{N}{T}} \frac{T}{1+T\sigma_{\tilde{w}}^2} \frac{1}{N} \sum_{i=1}^{N} v_{i0}^* \frac{1}{T} \sum_{t=1}^{T} \mathbf{z}_{it} = O_p\left(\sqrt{\frac{N}{T}}\right), \tag{A.30}$$

and

$$\frac{1}{\sqrt{NT}} \frac{1}{NT} \sum_{i=1}^{N} y_{i0} 1_T' \tilde{\mathbf{V}}^{*-1} \tilde{\mathbf{v}}_i^* = \sqrt{\frac{N}{T}} \frac{T}{1+T\sigma_{\tilde{w}}^2} \frac{1}{N} \sum_{i=1}^{N} v_{i0}^* y_{i0} = O_p\left(\sqrt{\frac{N}{T}}\right). \tag{A.31}$$

If $T \rightarrow \infty$, $\hat{\boldsymbol{\delta}}_C^*$ is consistent. However, if $\frac{N}{T} \rightarrow a \neq 0 < \infty$ as $T \rightarrow \infty$, $\hat{\boldsymbol{\delta}}_C^*$ is asymptotically biased of order $\sqrt{\frac{N}{T}}$.

A.8 Derivation of Lemma 5.1

First, let's consider the GLS estimator (5.1). For this result, following the previous derivation for the homoscedastic errors, we have

$$\sqrt{NT}\left(\hat{\rho}_{heter,f} - \rho\right) = \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{V}_i^{-1} \mathbf{y}_{i,-1}\right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{V}_i^{-1} \left(\eta_i 1_T + \mathbf{u}_i\right)\right). \tag{A.32}$$

It follows from the previous derivation of Lemma 2.1 that

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sigma_{ui}^{-2}\mathbf{y}_{i,-1}'\mathbf{V}_i^{-1}1_T\eta_i &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sigma_{ui}^{-2}}{1+T\varkappa_i}\mathbf{y}_{i,-1}'1_T\eta_i \\
&= \frac{1-\rho^T}{1-\rho}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sigma_{ui}^{-2}}{1+T\varkappa_i}y_{i0}\eta_i \\
&\quad +\frac{(T-1)-\rho\frac{1-\rho^{T-1}}{1-\rho}}{1-\rho}\sigma_\eta^2\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sigma_{ui}^{-2}}{1+T\varkappa_i} \\
&= \frac{1-\rho^T}{1-\rho}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sigma_{ui}^{-2}}{1+T\varkappa_i}y_{i0}\eta_i \\
&\quad +\frac{(T-1)-\rho\frac{1-\rho^{T-1}}{1-\rho}}{1-\rho}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\varkappa_i}{1+T\varkappa_i}, \qquad \text{(A.33)}
\end{aligned}$$

where the last equation holds by using the definition of $\varkappa_i = \frac{\sigma_\eta^2}{\sigma_{ui}^2}$, and

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}_i^{-1}\mathbf{u}_i \\
=& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_{ui}^{-2}y_{i,t-1}u_{it}-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\sigma_{ui}^{-2}\varkappa_i}{1+T\varkappa_i}\mathbf{y}_{i,-1}'1_T1_T'\mathbf{u}_i \\
=& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_{ui}^{-2}y_{i,t-1}u_{it}-\frac{(T-1)-\rho\frac{1-\rho^{T-1}}{1-\rho}}{1-\rho}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sigma_{ui}^2\frac{\sigma_{ui}^{-2}\varkappa_i}{1+T\varkappa_i} \\
=& \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_{ui}^{-2}y_{i,t-1}u_{it}-\frac{(T-1)-\rho\frac{1-\rho^{T-1}}{1-\rho}}{1-\rho}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\frac{\varkappa_i}{1+T\varkappa_i}. \qquad \text{(A.34)}
\end{aligned}$$

It is obvious that the second term of (A.33) cancels out with the second term of (A.34), thus

$$\begin{aligned}
\sqrt{NT}\left(\hat{\rho}_{heter,f}-\rho\right) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{V}_i^{-1}\mathbf{y}_{i,-1}\right)^{-1} \\
&\quad\times\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_{ui}^{-2}y_{i,t-1}u_{it}+\frac{1-\rho^T}{1-\rho}\sqrt{\frac{N}{T^3}}\frac{1}{N}\sum_{i=1}^{N}\frac{T\sigma_{ui}^{-2}}{1+T\varkappa_i}y_{i0}\eta_i\right). \qquad \text{(A.35)}
\end{aligned}$$

Consequently, the first term of the numerator of (A.35) converges to a normal distribution under fairly general conditions as stated in derivation of Lemma 2.1. The second term of the numerator of (A.35) is $O_p\left(\sqrt{\frac{N}{T^3}}\right)$ if $\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{T\sigma_{ui}^{-2}}{1+T\varkappa_i}y_{i0}\eta_i$ converges to a fixed constant.

Thus, if N and T are of similar order, $\frac{N}{T} \to a < \infty$ as Bai (2013) has assumed, the GLS estimator (5.1) is asymptotically unbiased. On the other hand, if N is much larger than T such that $\frac{N}{T^3} \to c \neq 0$, treating initial value y_{i0} as fixed constants yields an estimator that is asymptotically biased of order $\sqrt{c}$.

Let's now consider the asymptotic properties of the QMLE estimator (5.4). To begin with, for the inverse of the variance-covariance matrix (5.3), by using the formula of the block matrix, we have

$$\breve{\mathbf{V}}_i^{-1} = \begin{pmatrix} \omega_i^{-1} & -\sigma_{0i}^{-2}\sigma_1^2 1_T' \mathbf{C}_i^{-1} \\ -\sigma_{0i}^{-2}\mathbf{C}_i^{-1}\sigma_1^2 1_T & \mathbf{C}_i^{-1} \end{pmatrix}, \tag{A.36}$$

where

$$\omega_i = \sigma_{0i}^2 - \frac{\sigma_1^4\sigma_{ui}^{-2}T}{1+T\varkappa_i}, \text{ and } \mathbf{C}_i = \sigma_{ui}^2 I_T + \tilde{\sigma}_{\eta i}^2 1_T 1_T', \tag{A.37}$$

with $\tilde{\sigma}_{\eta i}^2 = \sigma_\eta^2 - \sigma_1^4\sigma_{0i}^{-2}$, and

$$\mathbf{C}_i^{-1} = \left(\sigma_{ui}^2 I_T + \tilde{\sigma}_{\eta i}^2 1_T 1_T'\right)^{-1} = \sigma_{ui}^{-2}\left(\mathbf{I}_T - \frac{\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i} 1_T 1_T'\right), \tag{A.38}$$

with $\tilde{\varkappa}_i = \frac{\tilde{\sigma}_{\eta i}^2}{\sigma_{ui}^2}$.

Following the previous derivation, it is obvious that

$$\begin{aligned} \sqrt{NT}\left(\hat{\rho}_{heter,r} - \rho\right) &= \left(\frac{1}{NT}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right)^{-1} \\ &\quad \times \left(\frac{1}{\sqrt{NT}}\sum_{i=1}^N \left[\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\left(\eta_i 1_T + \mathbf{u}_i\right) - \left(y_{i0} - \mu\right)\sigma_{0i}^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right]\right) \end{aligned} \tag{A.39}$$

The first term of the numerator of (A.39) has

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\left(\eta_i 1_T + \mathbf{u}_i\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\eta_i 1_T + \frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{u}_i, \tag{A.40}$$

with

$$\begin{aligned} \mathbf{C}_i^{-1} &= \left(\sigma_{ui}^2 I_T + \tilde{\sigma}_{\eta i}^2 1_T 1_T'\right)^{-1} = \sigma_{ui}^{-2}\left(\mathbf{I}_T - \frac{\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i} 1_T 1_T'\right), \\ \mathbf{C}_i^{-1} 1_T &= \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i} 1_T, \end{aligned}$$

where $\tilde{\sigma}_{\eta i}^2 = \sigma_\eta^2 - \sigma_1^4\sigma_{0i}^{-2}$ and $\tilde{\varkappa}_i = \frac{\tilde{\sigma}_{\eta i}^2}{\sigma_{ui}^2}$.

The expectation of the first term of (A.40) is

$$\begin{aligned}
E\left(\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\eta_i 1_T\right) &= \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}E\left(\mathbf{y}_{i,-1}'1_T\eta_i\right) = \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}E\left(\sum_{t=1}^{T}y_{i,t-1}\eta_i\right) \\
&= \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}E\left(y_{i0}\eta_i+\sum_{t=1}^{T-1}y_{it}\eta_i\right) \\
&= \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}\left(\sigma_1^2+\sum_{t=1}^{T-1}\left(\rho^t E\left(y_{i0}\eta_i\right)+E\left(\eta_i^2\right)\frac{1-\rho^t}{1-\rho}\right)\right) \\
&= \frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}\left(\frac{1-\rho^T}{1-\rho}\sigma_1^2+\frac{\sigma_\eta^2}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right)\right), \qquad \text{(A.41)}
\end{aligned}$$

by using the iteration y_{it} in (A.2) and

$$\begin{aligned}
E\left(\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\mathbf{u}_i\right) &= \sigma_{ui}^{-2}\left[E\left(\mathbf{y}_{i,-1}'\mathbf{u}_i\right)-\frac{\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i}E\left(\mathbf{y}_{i,-1}'1_T 1_T'\mathbf{u}_i\right)\right] \\
&= -\frac{\sigma_{ui}^{-2}\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i}E\left(\sum_{t=1}^{T}y_{i,t-1}\sum_{t=1}^{T}u_{it}\right) \\
&= -\frac{\sigma_{ui}^{-2}\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i}\left(\sum_{t=1}^{T-1}E\left(y_{it}u_{it}\right)+\sum_{s>t}^{T-1}E\left(y_{is}u_{it}\right)\right) \\
&= -\frac{\tilde{\varkappa}_i}{1+T\tilde{\varkappa}_i}\left((T-1)+\frac{1}{1-\rho}\rho\left(T-2\right)-\frac{\rho^2-\rho^T}{\left(1-\rho\right)^2}\right). \qquad \text{(A.42)}
\end{aligned}$$

The expectation of the second term of (A.40), we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}E\left[\left(y_{i0}-\mu\right)\sigma_{0i}^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right] \\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sigma_{0i}^{-2}\sigma_1^2\frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}E\left(v_{i0}1_T'\mathbf{y}_{i,-1}\right) \\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sigma_{0i}^{-2}\sigma_1^2\frac{\sigma_{ui}^{-2}}{1+T\tilde{\varkappa}_i}\left[\frac{\sigma_1^2}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right)+\frac{1-\rho^T}{1-\rho}\sigma_{0i}^2\right] \qquad \text{(A.43)}
\end{aligned}$$

since

$$\begin{aligned}
E\left(v_{i0}1_T'\mathbf{y}_{i,-1}\right) &= E\left(\sum_{t=1}^{T}y_{i,t-1}v_{i0}\right) = E\left(y_{i0}v_{i0}\right)+\sum_{t=1}^{T-1}E\left(y_{it}v_{i0}\right) \\
&= \frac{\sigma_1^2}{1-\rho}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right)+\frac{1-\rho^T}{1-\rho}\sigma_{0i}^2.
\end{aligned}$$

Substituting the above results (A.41)-(A.43) into (A.40) yields

$$E\left[\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\left(\eta_i 1_T + \mathbf{u}_i\right)\right] - E\left[\left(y_{i0}-\mu\right)\sigma_{0i}^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right] = 0,$$

since

$$\begin{aligned}
& \sigma_{ui}^2\left(1-\rho\right)\left(1+T\tilde{\varkappa}_i\right)\left\{E\left[\mathbf{y}_{i,-1}'\mathbf{C}_i^{-1}\left(\eta_i 1_T + \mathbf{u}_i\right)\right] - E\left[\left(y_{i0}-\mu\right)\sigma_{0i}^{-2}\sigma_1^2 1_T'\mathbf{C}_i^{-1}\mathbf{y}_{i,-1}\right]\right\} \\
= \;& \left(1-\rho^T\right)\sigma_1^2 + \sigma_\eta^2\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) - \sigma_\eta^2\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) \\
& +\sigma_1^4\sigma_{0i}^{-2}\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) - \sigma_{0i}^{-2}\sigma_1^4\left(T-1-\frac{\rho-\rho^T}{1-\rho}\right) - \left(1-\rho^T\right)\sigma_1^2 \\
= \;& 0.
\end{aligned}$$

Thus, the QMLE of γ (5.4) treating initial value as a random variable is asymptotically unbiased as long as $N \to \infty$.

A.9 Additional simulation results

This section contains the additional simulation results for DGP1-3 when $\rho = 0.2$ and 0.8.

Table A1: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP1 (6.1) and Case 1

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
	estim	0.2260	0.2019	0.1832	0.2241	0.2005	0.1904	0.2228	0.1994	0.1958
10	bias	0.0260	0.0019	-0.0168	0.0241	0.0005	-0.0096	0.0228	-0.0006	-0.0042
	rmse	0.0482	0.0371	0.0534	0.0370	0.0260	0.0377	0.0291	0.0168	0.0235
	size	9.1%	4.6%	6.6%	14.1%	4.9%	5.5%	24.7%	4.5%	5%
	estim	0.2001	0.1999	0.1879	0.2000	0.1998	0.1937	0.2000	0.1999	0.1973
100	bias	0.0001	-0.0001	-0.0121	0.0000	-0.0002	-0.0063	0.0000	-0.0001	-0.0027
	rmse	0.0099	0.0099	0.0158	0.0070	0.0070	0.0096	0.0046	0.0046	0.0055
	size	4.1%	4.1%	23.8%	5.6%	5.8%	13.8%	5%	5.2%	8.6%
	estim	0.1999	0.1999	0.1937	0.1998	0.1997	0.1937	0.1999	0.1998	0.1974
200	bias	-0.0001	-0.0001	-0.0063	-0.0002	-0.0003	-0.0063	-0.0001	-0.0002	-0.0026
	rmse	0.0070	0.0070	0.0094	0.0049	0.0049	0.0080	0.0032	0.0032	0.0042
	size	4.5%	4.5%	14.4%	5.4%	5.4%	24.1%	5%	5.1%	12.7%

Note: size is calculated for $H_0 : \rho = 0.2$. See note 1 of Table 1.

Table A2: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP1 (6.1) and Case 1

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
	estim	0.9799	0.7970	0.6819	0.9801	0.7977	0.7313	0.9799	0.7958	0.7680
10	bias	0.1799	-0.0030	-0.1181	0.1801	-0.0023	-0.0687	0.1799	-0.0042	-0.0320
	rmse	0.1799	0.0475	0.1529	0.1801	0.0334	0.1012	0.1799	0.0213	0.0579
	size	100%	4%	21.1%	100%	5.5%	14.6%	100%	5%	10%
	estim	0.8009	0.7997	0.7790	0.8008	0.7996	0.7880	0.8011	0.7999	0.7948
100	bias	0.0009	-0.0003	-0.0210	0.0008	-0.0004	-0.0120	0.0011	-0.0001	-0.0052
	rmse	0.0067	0.0067	0.0223	0.0048	0.0049	0.0132	0.0033	0.0031	0.0064
	size	5.3%	5.5%	80%	5.7%	4.9%	55.6%	6.6%	4.8%	30.4%
	estim	0.7999	0.7996	0.7896	0.8002	0.7999	0.7902	0.8002	0.7999	0.7957
200	bias	-0.0001	-0.0004	-0.0104	0.0002	-0.0001	-0.0098	0.0002	-0.0001	-0.0043
	rmse	0.0044	0.0045	0.0114	0.0032	0.0033	0.0104	0.0020	0.0021	0.0049
	size	4.8%	4.8%	62.1%	4.8%	4.3%	80.2%	4.5%	4.5%	49.1%

Note: size is calculated for $H_0 : \rho = 0.8$. See note 1 of Table 1.

Table A3: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP1 (6.1) and Case 2

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
10	estim	0.2240	0.2003	0.1788	0.2220	0.1985	0.1856	0.2222	0.1987	0.1936
	bias	0.0240	0.0003	-0.0112	0.0220	-0.0015	-0.0144	0.0222	-0.0013	-0.0064
	rmse	0.0505	0.0408	0.0580	0.0304	0.0282	0.0418	0.0297	0.0183	0.0267
	size	8.3%	5.3%	6.1%	10.4%	5.4%	7.1%	20.5%	5.3%	5.5%
100	estim	0.2001	0.1999	0.1877	0.1999	0.1999	0.1928	0.2002	0.2001	0.1970
	bias	0.0001	-0.0001	-0.0123	-0.0001	-0.0001	-0.0072	0.0002	0.0001	-0.0030
	rmse	0.0109	0.0109	0.0168	0.0079	0.0079	0.0109	0.0049	0.0049	0.0059
	size	5.1%	5.1%	18.6%	5.1%	5.1%	13.6%	5.2%	5%	8.7%
200	estim	0.1998	0.1997	0.1936	0.1999	0.1999	0.1939	0.1997	0.1997	0.1969
	bias	-0.0002	-0.0003	-0.0064	-0.0001	-0.0001	-0.0061	-0.0003	-0.00033	-0.0031
	rmse	0.0076	0.0076	0.0100	0.0055	0.0055	0.0083	0.0033	0.0033	0.0046
	size	5.3%	5.3%	13%	4.9%	4.9%	18.6%	4.6%	4.7%	14.7%

Note: size is calculated for $H_0 : \rho = 0.2$. See note 1 of Table 1.

Table A4: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP1 (6.1) and Case 2

	N	100			200			500		
T		nGLS	GLS	GMM	nGLS	GLS	GMM	nGLS	GLS	GMM
10	estim	0.9802	0.7962	0.6662	0.9803	0.7970	0.7149	0.9803	0.7971	0.7618
	bias	0.1802	-0.0038	-0.1338	0.1803	-0.0030	-0.0851	0.1803	-0.0029	-0.0382
	rmse	0.1802	0.0512	0.1698	0.1803	0.0361	0.1157	0.1804	0.0239	0.0637
	size	100%	4.4%	22.2%	100%	5.7%	19.1%	100%	4.7%	11.3%
100	estim	0.8013	0.8002	0.7792	0.8011	0.7999	0.7871	0.8012	0.8001	0.7941
	bias	0.0013	0.0002	-0.0208	0.0011	-0.0001	-0.0129	0.0012	0.0001	-0.0059
	rmse	0.0072	0.0070	0.0223	0.0054	0.0052	0.0143	0.0035	0.0034	0.0071
	size	4.9%	4.2%	71.7%	4.3%	4.7%	55.1%	6.9%	3.8%	31.8%
200	estim	0.8000	0.7997	0.7897	0.8002	0.7999	0.7902	0.8001	0.7998	0.7951
	bias	0.0000	-0.0003	-0.0103	0.0002	-0.0001	-0.0098	0.0001	-0.0002	-0.0049
	rmse	0.0048	0.0048	0.0114	0.0036	0.0036	0.0105	0.0022	0.0022	0.0055
	size	4.8%	5.2%	53.5%	5.2%	5.5%	73.3%	5.2%	5.2%	55.1%

Note: size is calculated for $H_0 : \rho = 0.8$. See note 1 of Table 1.

Table A5: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP2 (6.2) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.2260	0.2076	0.1998	0.1943	0.1953	0.2249	0.2067	0.1989	0.1934	0.1965	0.2246	0.2065	0.1988	0.1935	0.1988
10	bias	0.0260	0.0076	-0.0002	-0.0057	-0.0047	-0.0249	0.0067	-0.0011	-0.0066	-0.0035	0.0246	0.0065	-0.0012	-0.0065	-0.0012
	rmse	0.0324	0.0207	0.0188	0.0191	0.0268	0.0279	0.0144	0.0125	0.0139	0.0191	0.0261	0.0109	0.0087	0.0106	0.0122
	size	27%	7.4%	5.3%	6.5%	5.2%	49%	8.4%	5.5%	7.8%	5.2%	82.5%	11%	5.4%	11.7%	5.4%
	estim	0.2017	0.2001	0.2001	0.1995	0.1964	0.2018	0.2002	0.2002	0.1997	0.1984	0.2017	0.2002	0.2001	0.1996	0.1993
100	bias	0.0017	0.0001	0.0001	-0.0005	-0.0036	0.0018	0.0002	0.0002	-0.0003	-0.0016	0.0017	0.0002	0.0001	-0.0004	-0.0007
	rmse	0.0054	0.0052	0.0052	0.0052	0.0065	0.0041	0.0037	0.0037	0.0037	0.0042	0.0029	0.0023	0.0023	0.0023	0.0026
	size	6.3%	5.4%	5.3%	4.9%	10.5%	7.5%	4.9%	4.9%	5%	6.7%	12.2%	4.8%	4.8%	4.6%	5.8%
	estim	0.2008	0.2001	0.2000	0.1998	0.1982	0.2009	0.2001	0.2001	0.1998	0.1983	0.2008	0.2000	0.2000	0.1998	0.1993
200	bias	0.0008	0.0001	0.0000	-0.0002	-0.0018	0.0009	0.0001	0.0001	-0.0002	-0.0017	0.0008	0.0000	0.0000	-0.0002	-0.0007
	rmse	0.0039	0.0038	0.0038	0.0038	0.0043	0.0029	0.0027	0.0027	0.0028	0.0033	0.0019	0.0017	0.0017	0.0017	0.0019
	size	5.6%	5.1%	5%	4.6%	7.4%	6%	4.4%	4.5%	4.7%	8.7%	7.8%	5.8%	5.8%	6%	7.5%

Note: size calculated for $H_0 : \rho = 0.2$. See note 1 of Table 2.

Table A6: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP2 (6.2) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.8313	0.8636	0.7972	0.7800	0.7814	0.8313	0.8656	0.7972	0.7800	0.7880	0.8313	0.8651	0.7970	0.7800	0.7945
10	bias	0.0313	0.0646	-0.0028	-0.0200	-0.0186	0.0313	0.0656	-0.028	-0.0200	-0.0120	0.0313	0.0651	-0.0030	-0.0200	-0.0055
	rmse	0.0321	0.0667	0.0130	0.0231	0.0385	0.0316	0.0666	0.0089	0.0215	0.0293	0.0314	0.0655	0.0065	0.0206	0.0198
	size	98.5%	96.4%	5.7%	39.5%	8.8%	100%	100%	6.2%	71.4%	6.5%	100%	100%	8.2%	96.8%	5.5%
	estim	0.8022	0.8003	0.8000	0.7984	0.7967	0.8022	0.8002	0.8000	0.7983	0.7982	0.8022	0.8003	0.8000	0.7984	0.7992
100	bias	0.0022	0.0003	0.0000	-0.0016	-0.0033	0.0022	0.0002	0.0000	-0.0017	-0.0018	0.0022	0.0003	0.0000	-0.0016	-0.0008
	rmse	0.0033	0.0024	0.0024	0.0029	0.0043	0.0028	0.0018	0.0017	0.0024	0.0027	0.0025	0.0011	0.0011	0.0020	0.0016
	size	15.8%	4.4%	4.8%	10.8%	22.7%	22.8%	4.4%	4.5%	16.4%	14%	51.4%	5.8%	4.8%	31.9%	8%
	estim	0.8010	0.8001	0.8000	0.7992	0.7985	0.8010	0.8001	0.8000	0.7992	0.7985	0.8010	0.8001	0.8000	0.7992	0.7994
200	bias	0.0010	0.0001	0.0000	-0.0008	-0.0015	0.0010	0.0001	0.0000	-0.0008	-0.0015	0.0010	0.0001	0.0000	-0.0002	-0.0006
	rmse	0.0019	0.0016	0.0016	0.0018	0.0023	0.0016	0.0013	0.0013	0.0015	0.0021	0.0013	0.0008	0.0008	0.0008	0.0011
	size	8.9%	5.3%	5.5%	6.9%	13.8%	13.8%	3.6%	3.6%	10.4%	20.6%	26.5%	4.8%	4.8%	17.7%	11.9%

Note: size calculated for $H_0 : \rho = 0.8$. See note 1 of Table 2.

Table A7: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP2 (6.2) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.2253	0.2074	0.1997	0.1940	0.1958	0.2236	0.2055	0.1979	0.1927	0.1959	0.2246	0.2067	0.1991	0.1939	0.1981
10	bias	0.0253	0.0074	-0.0003	-0.0060	-0.0042	0.0236	0.0055	-0.0021	-0.0073	-0.0041	0.0246	0.0067	-0.0009	-0.0061	-0.0019
	rmse	0.0322	0.0216	0.0199	0.0202	0.0297	0.0270	0.0142	0.0130	0.0146	0.0201	0.0261	0.0108	0.0084	0.0102	0.0132
	size	24.2%	6.4%	4.7%	5.4%	5%	43.8%	6.7%	5.4%	8.3%	5.7%	82.6%	11.6%	5.2%	11.6%	4.9%
	estim	0.2016	0.2001	0.2000	0.1995	0.1964	0.2015	0.1999	0.1999	0.1994	0.1979	0.2017	0.2001	0.2001	0.1996	0.1993
100	bias	0.0016	0.0001	0.0000	-0.0005	-0.0036	0.0015	-0.0001	-0.0001	-0.0006	-0.0021	0.0017	0.0001	0.0001	-0.0004	-0.0007
	rmse	0.0059	0.0057	0.0057	0.0057	0.0070	0.0041	0.0039	0.0039	0.0039	0.0047	0.0029	0.0024	0.0001	0.0024	0.0027
	size	5.5%	4.1%	4.2%	4.8%	10.1%	6.8%	5%	4.9%	5%	7.5%	11.2%	4.6%	4.7%	5.3%	5.7%
	estim	0.2006	0.1998	0.1998	0.1996	0.1980	0.2007	0.1999	0.1999	0.1996	0.1981	0.2008	0.2000	0.2000	0.1998	0.1992
200	bias	0.0006	-0.0002	-0.0002	-0.0004	-0.0020	0.0007	-0.0001	-0.0001	-0.0004	-0.0019	0.0008	0.0000	0.0000	-0.0002	-0.0008
	rmse	0.0038	0.0038	0.0038	0.0038	0.0043	0.0028	0.0027	0.0027	0.0027	0.0033	0.0018	0.0016	0.0016	0.0017	0.0019
	size	5.3%	4.8%	4.8%	4.9%	8.9%	5.7%	4.9%	4.9%	5.1%	11.1%	7.2%	5.1%	5%	5.4%	7.1%

Note: size calculated for $H_0 : \rho = 0.2$. See note 1 of Table 2.

Table A8: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP2 (6.2) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.8311	0.8648	0.7973	0.7796	0.7818	0.8309	0.8639	0.7966	0.7797	0.7874	0.8312	0.8646	0.7973	0.7805	0.7943
10	bias	0.0311	0.0648	-0.0027	-0.0204	-0.0182	0.0309	0.0639	-0.0034	-0.0203	-0.0126	0.0312	0.0646	-0.0027	-0.0195	-0.0057
	rmse	0.0318	0.0668	0.0129	0.0235	0.0416	0.0313	0.0650	0.0094	0.0218	0.0296	0.0313	0.0651	0.0063	0.0202	0.0186
	size	99.6%	96%	6.4%	41.1%	8.2%	100%	99.8%	7.3%	71.9%	8.3%	100%	100%	7.6%	96.2%	6.8%
	estim	0.8021	0.8002	0.8000	0.7983	0.7968	0.8021	0.8002	0.7999	0.7983	0.7980	0.8021	0.8002	0.8000	0.7984	0.7992
100	bias	0.0021	0.0002	0.0000	-0.0017	-0.0032	0.0021	0.0002	-0.0001	-0.0017	-0.0020	0.0021	0.0002	0.0000	-0.0016	-0.0008
	rmse	0.0034	0.0026	0.0026	0.0031	0.0044	0.0027	0.0017	0.0017	0.0024	0.0029	0.0024	0.0011	0.0011	0.0020	0.0016
	size	11%	4.5%	4.8%	10.5%	20.8%	24.7%	4.9%	5.2%	17.8%	16.4%	50.9%	5.3%	5.1%	32.3%	9.8%
	estim	0.8009	0.8000	0.7999	0.7991	0.7984	0.8010	0.8000	0.8000	0.7992	0.7985	0.8010	0.8001	0.8000	0.7992	0.7993
200	bias	0.0009	0.0000	-0.0001	-0.0009	-0.0016	0.0010	0.0000	0.0000	-0.0008	-0.0015	0.0010	0.0001	0.0000	-0.0008	-0.0007
	rmse	0.0020	0.0017	0.0017	0.0019	0.0024	0.0015	0.0012	0.0012	0.0014	0.0020	0.0013	0.0008	0.0008	0.0011	0.0011
	size	8.4%	5%	5%	7.7%	14.2%	13.5%	5.3%	5.2%	10.6%	21%	26.8%	4.3%	4.6%	18.3%	12.3%

Note: size calculated for $H_0 : \rho = 0.8$. See note 1 of Table 2.

Table A9: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP3 (6.3) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.2059	0.2062	0.2016	0.1882	0.1955	0.2048	0.2052	0.2001	0.1873	0.1966	0.2046	0.2051	0.1998	0.1873	0.1989
10	bias	0.0059	0.0062	0.0016	-0.0118	-0.0045	0.0048	0.0052	0.0001	-0.0127	-0.0034	0.0046	0.0051	-0.0002	-0.0127	-0.011
	rmse	0.0194	0.0202	0.0183	0.0216	0.0263	0.0130	0.0137	0.0120	0.0176	0.0186	0.0095	0.0101	0.0082	0.0151	0.0119
	size	6.7%	6.9%	5.6%	20.2%	5.1%	7.6%	7.2%	5.6%	12.8%	5.3%	7.5%	7.9%	5.8%	31.6%	5.5%
	estim	0.2001	0.2001	0.2001	0.1989	0.1964	0.2002	0.2002	0.2002	0.1991	0.1984	0.2002	0.2001	0.2001	0.1990	0.1993
100	bias	0.0001	0.0001	0.0001	-0.0011	-0.0036	0.0002	0.0002	0.0002	-0.0009	-0.0016	0.0002	0.0001	0.0001	-0.0010	-0.0007
	rmse	0.0051	0.0052	0.0051	0.0053	0.0065	0.0037	0.0037	0.0037	0.0038	0.0042	0.0023	0.0023	0.0023	0.0025	0.0026
	size	4.9%	5.4%	4.7%	5.7%	10.3%	5%	4.8%	4.8%	5.3%	7.2%	4.9%	4.8%	4.9%	7.3%	5.8%
	estim	0.2001	0.2001	0.2001	0.1995	0.1982	0.2001	0.2001	0.2001	0.1995	0.1982	0.2000	0.2000	0.2000	0.1994	0.1993
200	bias	0.0001	0.0001	0.0001	-0.0005	-0.0018	0.0001	0.0001	0.0001	-0.0005	-0.0018	0.0000	0.0000	0.0000	-0.0006	-0.0007
	rmse	0.0038	0.0038	0.0038	0.0038	0.0043	0.0028	0.0028	0.0028	0.0028	0.0033	0.0017	0.0017	0.0017	0.0018	0.0019
	size	5.3%	5.1%	5.2%	4.7%	7.7%	4.5%	4.4%	4.5%	5.6%	8.9%	5.8%	5.8%	5.9%	7.1%	7.7%

Note:. size is calculated for $H_0 : \rho = 0.2$. See note 1 of Table 3.

Table A10: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP3 (6.3) and Case 1

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.8146	0.8458	0.7998	0.7752	0.7826	0.8144	0.8452	0.7994	0.7752	0.7889	0.8143	0.8437	0.7992	0.7752	0.7951
10	bias	0.0146	0.0458	-0.0002	-0.0248	-0.0174	0.0144	0.0452	-0.0006	-0.0248	-0.0111	0.0143	0.0437	-0.0008	-0.0248	-0.0049
	rmse	0.0169	0.0507	0.0086	0.0275	0.0370	0.0156	0.0477	0.0060	0.0261	0.0277	0.0148	0.0447	0.0040	0.0254	0.0184
	size	40.6%	52.8%	6.1%	53.5%	8.5%	68%	84.7%	5.9%	89%	6.5%	96.8%	99.8%	5%	99.6%	5.9%
	estim	0.8002	0.8002	0.8001	0.7980	0.7967	0.8002	0.8002	0.8000	0.7980	0.7982	0.8002	0.8000	0.8000	0.7980	0.7992
100	bias	0.0002	0.0002	0.0001	-0.0020	-0.0033	0.0002	0.0002	0.0000	-0.0020	-0.0018	0.0002	0.0000	0.0000	-0.0020	-0.0008
	rmse	0.0023	0.0024	.00023	0.0031	0.0043	0.0017	0.0018	0.0017	0.0026	0.0027	0.0011	0.0011	0.0011	0.0022	0.0016
	size	4.5%	5%	4.5%	13.7%	22.3%	4.1%	4.3%	4.3%	21.1%	14.3%	5.3%	5.6%	4.9%	42.9%	8%
	estim	0.8000	0.8001	0.8000	0.7990	0.7985	0.8000	0.8000	0.8000	0.7990	0.7985	0.8000	0.8000	0.8000	0.7990	0.7994
200	bias	0.0000	0.0001	0.0000	-0.0010	-0.0015	0.0000	0.0000	0.0000	-0.0010	-0.0015	0.0000	0.0000	0.0000	-0.0010	-0.0006
	rmse	0.0016	0.0016	0.0016	0.0016	0.0023	0.0013	0.0013	0.0013	0.0016	0.0021	0.0008	0.0008	0.0008	0.0012	0.0011
	size	5.6%	5.3%	5.6%	13.5%	13.9%	4.4%	3.5%	4.4%	13.4%	20.5%	5.1%	4.8%	5.2%	24.1%	11.3%

Note:. size is calculated for $H_0 : \rho = 0.8$. See note 1 of Table 3.

Table A11: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.2$ for DGP3 (6.3) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
	estim	0.2054	0.2059	0.2011	0.1879	0.1960	0.2038	0.2041	0.1992	0.1866	0.1960	0.2049	0.2053	0.2000	0.1878	0.1982
10	bias	0.0054	0.00059	0.0011	-0.0121	-0.0040	0.0038	0.0041	-0.0008	-0.0134	-0.0040	0.0049	0.0053	0.0000	-0.0122	-00018
	rmse	0.0198	0.0211	0.0190	0.0227	0.0290	0.0132	0.0137	0.0125	0.0184	0.0196	0.0095	0.0100	0.0081	0.0147	0.0127
	size	5.5%	5.7%	5.2%	9.4%	5%	5.1%	6.2%	4.4%	18.8%	5.9%	7.8%	9.2%	4.8%	31%	4.5%
	estim	0.2000	0.2000	0.2001	0.1989	0.1964	0.1999	0.1999	0.1999	0.1987	0.1979	0.2001	0.2001	0.2001	0.1989	0.1993
100	bias	0.0000	0.0000	0.0001	-0.0011	-0.0036	-0.0001	-0.0001	-0.0001	-0.0013	-0.0021	0.0001	0.0001	0.0001	-0.0011	-0.0007
	rmse	0.0057	0.0057	0.0057	0.0058	0.0070	0.0038	0.0039	0.0038	0.0041	0.0047	0.0024	0.0024	0.0024	0.0026	0.0027
	size	4.5%	4.1%	4.4%	5.6%	10.1%	5.2%	5%	5.2%	6.1%	7.5%	5.2%	4.5%	5.2%	7.1%	5.8%
	estim	0.1998	0.1998	0.1999	0.1993	0.1980	0.1999	0.1999	0.1999	0.1993	0.1981	0.2000	0.2000	0.2000	0.1995	0.1992
200	bias	-0.0002	-0.0002	-0.0001	-0.0007	-0.0020	-0.0001	-0.0001	-0.0001	-0.0007	-0.0019	0.0000	0.0000	0.0000	-0.0005	-0.0008
	rmse	0.0038	0.0038	0.0038	0.0039	0.0043	0.0027	0.0027	-0.0001	0.0028	0.0033	0.0016	0.0016	0.0016	0.0017	0.0019
	size	4.8%	4.8%	4.8%	5.5%	8.7%	4.7%	4.9%	4.6%	5.5%	11.1%	4.7%	5%	4.8%	6.5%	6.9%

Note: size calculated for $H_0 : \rho = 0.2$. See note 1 of Table 3.

Table A12: Sample mean, bias, RMSE and size of $\hat{\rho}$ when $\rho = 0.8$ for DGP3 (6.3) and Case 2

	N	100					200					500				
T		nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM	nGLS	GLS1	GLS2	GLS3	GMM
10	estim	0.8142	0.8459	0.7994	0.7748	0.7830	0.8139	0.8428	0.7991	0.7750	0.7884	0.8143	0.8434	0.7993	0.7757	0.7949
	bias	0.0142	0.0459	-0.0006	-0.0252	-0.0170	0.0139	0.0428	-0.0009	-0.0250	-0.0116	0.0143	0.0434	-0.0007	-0.0243	-0.0051
	rmse	0.0164	0.0509	0.0084	0.0279	0.0396	0.0151	0.0454	0.0061	0.0263	0.0281	0.0148	0.0445	0.0040	0.0248	0.0174
	size	40.8%	51.2%	5.6%	55.2%	7.7%	63.4%	79.8%	5.7%	87.2%	7.4%	95.9%	99.8%	5.4%	99.6%	6.2%
100	estim	0.8001	0.8001	0.8000	0.7980	0.7968	0.8001	0.8001	0.8000	0.7980	0.7980	0.8000	0.8002	0.8000	0.7981	0.7992
	bias	0.0001	0.0001	0.0000	-0.0020	-0.0032	0.0001	0.0001	0.0000	-0.0020	-0.0020	0.0000	0.0002	0.0000	-0.0019	-0.0008
	rmse	0.0026	0.0026	0.0026	0.0033	0.0044	0.0017	0.0017	0.0016	0.0026	0.0029	0.0012	0.0011	0.0011	0.0022	0.0016
	size	5.3%	4.7%	5.4%	12.8%	20.5%	4.8%	4.7%	4.9%	23%	15.9%	5.3%	4.7%	5%	44.1%	9.7%
200	estim	0.8000	0.8000	0.8000	0.7990	0.7984	0.8001	0.8000	0.8000	0.7990	0.7985	0.8001	0.8001	0.8000	0.7991	0.7993
	bias	0.0000	0.0000	0.0000	-0.0010	-0.0016	0.0001	0.0000	0.0000	-0.0010	-0.0015	0.0001	0.0001	0.0000	-0.0009	-0.0007
	rmse	0.0017	0.0017	0.0017	0.0020	0.0024	0.0017	0.0012	0.0012	0.0015	0.0020	0.0007	0.0008	0.0007	0.0012	0.0011
	size	4.7%	5%	4.6%	9.1%	13.9%	5.3%	5.5%	5.2%	5.3%	20.8%	4.5%	4.2%	4.4%	24.4%	12.3%

Note: size calculated for $H_0 : \rho = 0.8$. See note 1 of Table 3.