Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models with Cross-Sectional Heteroskedasticity*

Kazuhiko Hayakawa
Hiroshima University

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

January 27, 2014

Abstract

This paper extends the transformed maximum likelihood approach for estimation of dynamic panel data models by Hsiao, Pesaran, and Tahmiscioglu (2002) to the case where the errors are cross-sectionally heteroskedastic. This extension is not trivial due to the incidental parameters problem and its implications for estimation and inference. We approach the problem by working with a mis-specified homoskedastic model, and then show that the transformed maximum likelihood estimator continues to be consistent even in the presence of cross-sectional heteroskedasticity. We also obtain standard errors that are robust to cross-sectional heteroskedasticity of unknown form. By means of Monte Carlo simulations, we investigate the finite sample behavior of the transformed maximum likelihood estimator and compare it with various GMM estimators proposed in the literature. Simulation results reveal that, in terms of median absolute errors and accuracy of inference, the transformed likelihood estimator outperforms the GMM estimators in almost all cases.

Keywords: Dynamic Panels, Cross-sectional heteroskedasticity, Monte Carlo simulation, Transformed MLE, GMM estimation

JEL Codes: C12, C13, C23

*We are grateful to three referees and the participants at the 18th International Conference on Panel Data, and seminars at Osaka, Sogang and Nanyang Technological University for helpful comments. This paper was written whilst Hayakawa was visiting the University of Cambridge as a JSPS Postdoctoral Fellow for Research Abroad. He acknowledges the financial support from the JSPS Fellowship and the Grant-in-Aid for Scientific Research (KAKENHI 22730178) provided by the JSPS. Pesaran acknowledges financial support from the ESRC Grant No. ES/1031626/1. Elisa Tosetti contributed to a preliminary version of this paper. Her assistance in coding of the transformed ML estimator and some of the derivations is gratefully acknowledged.

1 Introduction

In dynamic panel data models where the time dimension (T) is short, the presence of lagged dependent variables among the regressors makes standard panel estimators inconsistent, and complicates statistical inference on the model parameters considerably. To deal with these difficulties a sizable literature has emerged, starting with the seminal papers of Anderson and Hsiao (1981, 1982) who proposed the application of the instrumental variable (IV) approach to the first-differenced form of the model. More recently, a large number of studies have been focusing on the generalized method of moments (GMM), see, among others, Holtz-Eakin, Newey, and Rosen (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995) and Blundell and Bond (1998). One important reason for the popularity of GMM in applied economic research is that it provides asymptotically valid inference under a minimal set of statistical assumptions. Arellano and Bond (1991) proposed GMM estimators based on moment conditions where lagged variables in levels are used as instruments. Blundell and Bond (1998) showed that the performance of this estimator deteriorates when the parameter associated with the lagged dependent variable is close to unity and/or the *variance ratio* of the individual effects to the idiosyncratic errors is large, since in such cases the instruments are only weakly related to the lagged dependent variables.[1] The poor finite sample properties of GMM estimators has been documented using Monte Carlo studies by Kiviet (2007), for example. To deal with the weak instrument problem, Arellano and Bover (1995) and Blundell and Bond (1998) proposed the use of extra moment conditions arising from the model in levels, which become available when the initial observations satisfy certain conditions. The resulting GMM estimator, known as system GMM, combines moment conditions for the model in first differences with moment conditions for the model in levels. We refer to Blundell, Bond, and Windmeijer (2000) for an extension to the multivariate case, and for a Monte Carlo study of the properties of GMM estimators using moment conditions from either the first differenced and/or levels models. Bun and Windmeijer (2010) proved that the equation in levels suffers from a weak instrument problem when the variance ratio is large. Hayakawa (2007) also shows that the finite sample bias of the system GMM estimator becomes large when the variance ratio is relatively large.

The GMM estimators have been used in a large number of empirical studies to investigate problems in areas such as labour, development, health, macroeconomics and finance. Theoretical and applied research on dynamic panels has mostly focused on the GMM, and has by and large neglected the maximum likelihood (ML) approach though there are several theoretical advances such as Hsiao, Pesaran, and Tahmiscioglu (2002), Binder, Hsiao, and Pesaran (2005), Alvarez and Arellano (2004) and Kruiniger (2008). Hsiao, Pesaran, and Tahmiscioglu (2002) propose the transformed likelihood approach while Binder, Hsiao, and Pesaran (2005) have extended the approach to estimating panel VAR (PVAR) models. Alvarez and Arellano (2004) have studied ML estimation of autoregressive panels in the presence of time-specific heteroskedasticity (see also Bhargava and Sargan (1983)). Kruiniger

[1] See also the discussion in Binder, Hsiao, and Pesaran (2005), who proved that the asymptotic variance of the Arellano and Bond (1991) GMM estimator depends on the variance of the individual effects.

(2008) considers ML estimation of a stationary/unit root AR(1) panel data models.

There are several reasons why the GMM approach is preferred compared to the ML approach. First, the regularity conditions required to prove consistency and asymptotic normality of the GMM type estimators are relatively mild and allow for the presence of cross-sectional heteroskedasiticity of the errors. In particular, see Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998). Second, for the ML approach, the incidental parameters problem and the initial values problem lead to a violation of the standard regularity conditions, which causes inconsistency. Although Hsiao, Pesaran, and Tahmiscioglu (2002) developed a transformed likelihood approach to overcome some of the weaknesses of the GMM approach (particularly the weak IV problem), their analysis still requires the idiosyncratic errors to be homoskedastic, which is likely to be restrictive in many empirical applications.[2]

It is therefore highly desirable to extend the transformed ML approach of Hsiao, Pesaran and Tahmiscioglu (HPT) so that heteroskedastic errors can be permitted.[3]. This is accomplished in this paper. The extension is not trivial due to the incidental parameters problem that arises, in particular its implications for inference. We follow the time series literature, and initially ignore the error variance heterogeneity and work with a misspecified homoskedastic model, but show that the transformed maximum likelihood estimator by Hsiao, Pesaran, and Tahmiscioglu (2002) continues to be consistent. We then derive, under fairly general conditions, a covariance matrix estimator for the quasi-ML (QML) estimator which is robust to cross-sectional heteroskedasticity. Using Monte Carlo simulations, we investigate the finite sample performance of the transformed QML estimator and compare it with a range of GMM estimators. Simulation results reveal that, in terms of median absolute errors and accuracy of inference, the transformed likelihood estimator outperforms the GMM estimators in *almost all* cases when the model contains an exogenous regressor, and in many cases if we consider pure autoregressive panels.

The rest of the paper is organized as follows. Section 2 describes the model and its underlying assumptions. Section 3 proposes the transformed QML estimator for cross-sectionally heteroskedastic errors. Section 4 provides an overview of the GMM estimators used in the simulation exercise. Section 5 describes the Monte Carlo design and comments on the small sample properties of the transformed likelihood and GMM estimators. Finally, Section 6 ends with some concluding remarks.

2 The dynamic panel data model

Consider the following dynamic panel data model

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + u_{it}, \qquad i = 1, 2, ..., N, \tag{1}$$

[2] In the application of the GMM approach to dynamic panels, it is generally difficult to avoid the so-called many/weak instruments problem, which is shown to result in biased estimates and substantially distorted test outcomes. See Section 5 for further evidence.

[3] Note, however, that since the transformed ML approach does not impose any restrictions on the individual effects, the errors of the original panel (before differencing) can have any arbitrary degree of cross-sectional heteroskedasticity.

where α_i, $i = 1, 2, ..., N$ are the unobserved individual effects, u_{it} is an idiosyncratic error term, x_{it} is observed regressor assumed to vary over time (t) and across the individuals (i). It is further assumed that x_{it} is a scaler variable to simplify the notations.[4] The coefficients of interest are γ and β, which are assumed to be fixed finite constants. No restrictions are placed on the individual effects, α_i. They can be heteroskedastic, correlated with x_{jt} and u_{jt}, for all i and j, and can be cross-sectionally dependent. In contrast, the idiosyncratic errors, u_{it}, are assumed to be uncorrelated with $x_{it'}$ for all i, t and t'. However, we allow the variance of u_{it} to vary across i, and let the variance ratio, $\tau^2 = \left[N^{-1}\Sigma_{i=1}^{N} Var(\alpha_i)\right] / \left[N^{-1}\Sigma_{i=1}^{N} Var(u_{it})\right]$ to take any positive value. We shall investigate the robustness of the QML and GMM estimators to choices of τ and γ.

Follwoing the literature we take first differences of (1) to eliminate the individual effects[5]

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta u_{it}, \tag{2}$$

and make the following assumptions:

Assumption 1 *(Starting period) The dynamic processes (1) have started at time $t = -m$, (m being a positive constant) but only the time series data, $y_{it}, x_{it}; i = 1, 2, ..., N; t = 0, 1, ..., T$, are observed.*

Assumption 2 *(Exogenous variable) It is assumed that x_{it} is generated either by*

$$x_{it} = \mu_i + \phi t + \sum_{j=0}^{\infty} a_j \varepsilon_{i,t-j}, \qquad \sum_{j=0}^{\infty} |a_j| < \infty \tag{3}$$

or

$$\Delta x_{it} = \phi + \sum_{j=0}^{\infty} d_j \varepsilon_{i,t-j}, \qquad \sum_{j=0}^{\infty} |d_j| < \infty \tag{4}$$

where μ_i can either be fixed or random. ε_{it} are independently distributed over i and t with $E(\varepsilon_{it}) = 0$, and $var(\varepsilon_{it}) = \sigma_{\varepsilon i}^2$ with $0 < \sigma_{\varepsilon i}^2 < K < \infty$, and independent of u_{is} for all s and t.

Assumption 3 *(Initialization) Depending on whether the y_{it} process has reached stationarity, one of the following two assumptions holds:*
(i) $|\gamma| < 1$, and the process has been going on for a long time, namely $m \to \infty$;
(ii) The process has started from a finite period in the past not too far back from the 0th period, namely for given values of $y_{i,-m}$ with m finite, such that

$$E(\Delta y_{i,-m+1} | \Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT}) = b_m + \boldsymbol{\pi}_m' \Delta \mathbf{x}_i, \text{ for all } i,$$

where b_m is a finite constant, $\boldsymbol{\pi}_m$ is a T-dimensional vector of constants, and $\Delta \mathbf{x}_i = (\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT})'$.

[4] Extension to the case of multiple regressors is straightforward at the expense of notational complexity.

[5] As shown in Appendix A of Hsiao, Pesaran, and Tahmiscioglu (2002), other transformations can be used to eliminate the individual effects and the QML estimator proposed in this paper is invariant to the choice of such transformations.

Assumption 4 *(idiosyncratic shocks) Disturbances u_{it} are serially and cross-sectionally independently distributed, with $E(u_{it}) = 0$, and $E(u_{it}^2) = \sigma_i^2$ such that $0 < \sigma_i^2 < K < \infty$, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$.*

Remark 1 *Assumption 3.(ii) constrains the expected changes in the initial values to be the same across all individuals, but does not necessarily require that $\mid \gamma \mid< 1$. Note that this does not require the initial values, $y_{i,-m}$, to have the same mean across i, and allows $y_{i,-m}$ to vary both with α_i and μ_i. It is only required that $y_{i,-m+1} - y_{i,-m}$ is free of the incidental paramter problem.*

Remark 2 *Assumptions 2, and 4 allow for heteroskedastic disturbances in the equations for y_{it} and x_{it}. Also Assumption 2 requires x_{it} to be strictly exogenous. These restrictions can be relaxed by considering a panel vector autoregressive specification of the type considered in Binder, Hsiao, and Pesaran (2005). However, these further developments are beyond the scope of the present paper. See also the remarks in Section 6 .*

3 Transformed likelihood estimation

The first-differenced model (2) is well defined for $t = 2, 3, ..., T$, and can be used to derive the joint distribution of $(\Delta y_{i2}, \Delta y_{i3}, ..., \Delta y_{iT})$ conditional on Δy_{i1}. To obtain the (unconditional) distribution of Δy_{i1}, starting from $\Delta y_{i,-m+1}$, and by continuous substitution, we note that

$$\Delta y_{i1} = \gamma^m \Delta y_{i,-m+1} + \beta \sum_{j=0}^{m-1} \gamma^j \Delta x_{i,1-j} + \sum_{j=0}^{m-1} \gamma^j \Delta u_{i,1-j}. \tag{5}$$

Note that the mean of Δy_{i1} conditional on $\Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...$, is given by

$$\eta_{i1} = E(\Delta y_{i1} | \Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...) = \gamma^m \Delta y_{i,-m+1} + \beta \sum_{j=0}^{m-1} \gamma^j \Delta x_{i,1-j}, \tag{6}$$

which depends on the unknown values $\Delta y_{i,-m+1}$, and $\Delta x_{i,1-j}$, for $j = 1, 2, ..., m-1$, $i = 1, 2, ..., N$. To solve this problem, we need to express the expected value of η_{i1}, conditional on the observables, in a way that it only depends on a finite number of parameters. The following theorem provides the conditions under which it is possible to derive a marginal model for Δy_{i1}, which is a function of a finite number of unknown parameters.

Theorem 1 *Consider model (2), where x_{it} follows either (3) or (4). Suppose that Assumptions 1, 2, 3, and 4 hold. Then Δy_{i1} can be expressed as:*

$$\Delta y_{i1} = b + \boldsymbol{\pi}' \Delta \mathbf{x}_i + v_{i1}, \tag{7}$$

where b is a constant, $\boldsymbol{\pi}$ is a T-dimensional vector of constants, $\Delta \mathbf{x}_i = (\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT})'$, and v_{i1}

is independently distributed across i*, such that* $E(v_{i1}) = 0$*, and* $E(v_{i1}^2) = \omega_i \sigma_i^2$*, with* $0 < \omega_i < K < \infty$*, for all* i*.*

It is now possible to derive the likelihood function of the *transformed model* given by equations (2) for $t = 2, 3, ..., T$ and (7). Let $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$,

$$\underset{T\times(T+3)}{\Delta \mathbf{W}_i} = \begin{pmatrix} 1 & \Delta \mathbf{x}_i' & 0 & 0 \\ 0 & \mathbf{0} & \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{0} & \Delta y_{i,T-1} & \Delta x_{iT} \end{pmatrix}, \tag{8}$$

and note that the transformed model can be rewritten as

$$\Delta \mathbf{y}_i = \Delta \mathbf{W}_i \boldsymbol{\varphi} + \mathbf{r}_i, \tag{9}$$

with $\boldsymbol{\varphi} = (b, \boldsymbol{\pi}', \gamma, \beta)'$. The covariance matrix of $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$ has the form:

$$E(\mathbf{r}_i \mathbf{r}_i') = \sigma_i^2 \begin{pmatrix} \omega_i & -1 & & & 0 \\ -1 & 2 & \ddots & & \\ & & \ddots & & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} = \sigma_i^2 \boldsymbol{\Omega}(\omega_i), \tag{10}$$

where $\omega_i > 0$ is a free parameter defined in Theorem 1.

The log-likelihood function of the transformed model (9) is given by

$$\begin{aligned} \ell(\boldsymbol{\psi}_N) &= -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^{N} \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{N} \ln[1 + T(\omega_i - 1)] \\ & \quad - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (\Delta \mathbf{y}_i - \Delta \mathbf{W}_i \boldsymbol{\varphi})' \boldsymbol{\Omega}(\omega_i)^{-1} (\Delta \mathbf{y}_i - \Delta \mathbf{W}_i \boldsymbol{\varphi}), \end{aligned} \tag{11}$$

where $\boldsymbol{\psi}_N = (\boldsymbol{\varphi}', \omega_1, \omega_2 ..., \omega_N, \sigma_1^2, \sigma_2^2, ... \sigma_N^2)'$. Unfortunately, the maximum likelihood estimation based on $\ell(\boldsymbol{\psi}_N)$ encounters the incidental parameters problem of Neyman and Scott (1948) since the number of parameters grows with the sample size, N. Following the mis-specification literature in econometrics, (White, 1982; Kent, 1982), we examine the asymptotic properties of the ML estimators of the parameters of interest, $\boldsymbol{\varphi}$, using a mis-specified model where the heteroskedastic nature of the errors is ignored.

Accordingly, suppose that it is incorrectly assumed that the regression errors u_{it} are homoskedastic, i.e., $\sigma_i^2 = \sigma^2$ and $\omega_i = \omega$, for $i = 1, 2, ..., N$. Then under this mis-specification, the pseudo log-likelihood function of the transformed model (9), is given by

$$\ell_p(\boldsymbol{\theta}) = -\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln\left(\sigma^2\right) - \frac{N}{2}\ln\left[1+T(\omega-1)\right] \\ -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\varphi}\right)'\boldsymbol{\Omega}(\omega)^{-1}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\varphi}\right), \quad (12)$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\varphi}', \omega, \sigma^2\right)'$ is the vector of unknown parameters. Let $\widehat{\boldsymbol{\theta}}$ be the estimator obtained by maximizing the pseudo log-likelihood function in (12).

To characterize the relationship between the true parameter values $\boldsymbol{\psi}_{N0} = \left(\boldsymbol{\varphi}_0', \omega_{10}, ..., \omega_{N0}, \sigma_{10}^2, ...\sigma_{N0}^2\right)'$ and the pseudo true values $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_*', \omega_*, \sigma_*^2\right)'$ that maximizes the pseudo log-likelihood function (12), we introduce the following average parameter measures.

Assumption 5 *The average true parameter values*

$$\bar{\sigma}_{N,0}^2 = N^{-1}\sum_{i=1}^{N}\sigma_{i0}^2, \quad \textit{and } \bar{\omega}_{N,0} = \frac{N^{-1}\sum_{i=1}^{N}\omega_{i0}\sigma_{i0}^2}{N^{-1}\sum_{i=1}^{N}\sigma_{i0}^2}, \quad (13)$$

have finite limits (as $N \to \infty$) given by

$$\bar{\sigma}_0^2 = \lim_{N\to\infty}\left(\bar{\sigma}_{N,0}^2\right), \quad \textit{and } \ \bar{\omega}_0 = \frac{\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\omega_{i0}\sigma_{i0}^2}{\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_{i0}^2}. \quad (14)$$

The following theorem establishes the relationship between the true value and the pseudo true value.

Theorem 2 *Suppose that Assumptions 1, 2, 3, 4, and 5 hold, and let $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_*', \omega_*, \sigma_*^2\right)'$ be the solution of the first-order condition $E\left[\partial\ell_p(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}\right] = \mathbf{0}$ of the pseudo log-likelihood function, (12), where expectations are taken with respect to the true probability measures. Then,*

$$\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_0', \bar{\omega}_0, \bar{\sigma}_0^2\right)'.$$

This theorem summarizes one of the key results of the paper, and holds under fairly general conditions. Assumptions 1, 2, 3 are identical to those used in Hsiao et. al. (2002). Assumption 4 allows the error variances of the error terms to be heteroskedastic in an arbitrary manner. Assumption 5 only requires the indiviudal error variances and their ratios to be finite. The possible non-uniqeness of the pseudo true values in the case of heterogenous ω_i, is analogous to the non-uniqueness of the ML estimators encountered in the case of the random effect models as demonstarted initially by Maddala (1971) and further discussed by Breusch (1987) who proposes a practical approach to detecting the presence of local maxima.[6] Theoretically, it is quite complicated to demonsrate which solution leads

[6] This result follows since, as established by Grassetti (2011), the transformed likelihood function can be written equivalently in the form of a random effects model with endogenous regressors. For further details see Section A.3.

to the global maximum. However, the Monte Carlo simulation results in Section 5 suggest that the solution $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_0', \bar{\omega}_0, \bar{\sigma}_0^2\right)'$ is assocaited with the global maximum. In what follows we assume that the global maximum of the probability limit of the pseudo log-likelihood function is atatined at $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_0', \bar{\omega}_0, \bar{\sigma}_0^2\right)' = \bar{\boldsymbol{\theta}}_0$.

The following theorem establishes the asymptotic distribution of the ML estimator of the transformed model.

Theorem 3 *Suppose that Assumptions 1, 2, 3, 4, and 5 hold and let* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\varphi}}', \widehat{\omega}, \widehat{\sigma}^2\right)'$ *be the QML estimator obtained by maximizing the pseudo log-likelihood function in (12). Then as N tends to infinity,* $\widehat{\boldsymbol{\theta}}$ *is asymptotically normal with*

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}^{*-1}\mathbf{B}^*\mathbf{A}^{*-1}\right) \tag{15}$$

where $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_0', \bar{\omega}_0, \bar{\sigma}_0^2\right)'$,

$$\mathbf{A}^* = \lim_{N\to\infty} E\left[-\frac{1}{N}\frac{\partial^2 \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right], \quad \textit{and} \quad \mathbf{B}^* = \lim_{N\to\infty} E\left[\frac{1}{N}\frac{\partial \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}}\frac{\partial \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}'}\right]. \tag{16}$$

A consistent estimator of $\mathbf{A}^*$, denoted by $\widehat{\mathbf{A}}^*$ which is robust to unknown heteroskedasticity, is given by

$$\widehat{\mathbf{A}}^* = \begin{pmatrix} \frac{1}{N\widehat{\sigma}^2}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\Delta\mathbf{W}_i & \frac{1}{Ng(\widehat{\omega})^2\widehat{\sigma}^2}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Phi}\widehat{\mathbf{r}}_i & \mathbf{0} \\ \frac{1}{Ng(\widehat{\omega})^2\widehat{\sigma}^2}\sum_{i=1}^N \widehat{\mathbf{r}}_i'\boldsymbol{\Phi}\Delta\mathbf{W}_i & \frac{T^2}{2g(\widehat{\omega})^2} & \frac{T}{2g(\widehat{\omega})\widehat{\sigma}^2} \\ \mathbf{0} & \frac{T}{2g(\widehat{\omega})\widehat{\sigma}^2} & \frac{T}{2\widehat{\sigma}^4} \end{pmatrix}, \tag{17}$$

where $\widehat{\mathbf{r}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\widehat{\boldsymbol{\varphi}}$, $g(\hat{\omega}) = 1 + T(\hat{\omega} - 1)$, $\boldsymbol{\Phi}$ is defined by (47) in the Appendix, and $\widehat{\sigma}^2 = N^{-1}\sum_{i=1}^N \widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\widehat{\mathbf{r}}_i$. Partitioning $\mathbf{B}^*$, and its estimator $\widehat{\mathbf{B}}^*$, accordingly to the above partioned form of $\widehat{\mathbf{A}}^*$, we have

$$\widehat{\mathbf{B}}_{11}^* = \frac{1}{N\widehat{\sigma}^4}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\widehat{\mathbf{r}}_i\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\Delta\mathbf{W}_i,\ \widehat{\mathbf{B}}_{22}^* = \frac{T^2}{4g(\widehat{\omega})^4\widehat{\sigma}^4}\left\{\frac{1}{N}\sum_{i=1}^N\left(\frac{\widehat{\mathbf{r}}_i'\boldsymbol{\Phi}\widehat{\mathbf{r}}_i}{T}\right)^2 - g(\widehat{\omega})^2\widehat{\sigma}^4\right\}, \tag{18}$$

$$\widehat{\mathbf{B}}_{33}^* = \frac{T^2}{4\widehat{\sigma}^8}\left\{\frac{1}{N}\sum_{i=1}^N\left(\frac{\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\widehat{\mathbf{r}}_i}{T}\right)^2 - \widehat{\sigma}^4\right\},\ \widehat{\mathbf{B}}_{21}^* = \frac{1}{2g(\widehat{\omega})^2\widehat{\sigma}^4}\frac{1}{N}\sum_{i=1}^N\left(\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\Delta\mathbf{W}_i\right)\left(\widehat{\mathbf{r}}_i'\boldsymbol{\Phi}\widehat{\mathbf{r}}_i\right), \tag{19}$$

$$\widehat{\mathbf{B}}_{31}^* = \frac{1}{2\widehat{\sigma}^6}\frac{1}{N}\sum_{i=1}^N\left(\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\Delta\mathbf{W}_i\right)\left(\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\widehat{\mathbf{r}}_i\right), \tag{20}$$

$$\widehat{\mathbf{B}}_{32}^* = \frac{T^2}{4g(\widehat{\omega})^2\widehat{\sigma}^6}\left[\frac{1}{N}\sum_{i=1}^N\left(\frac{\widehat{\mathbf{r}}_i'\boldsymbol{\Phi}\widehat{\mathbf{r}}_i}{T}\right)\left(\frac{\widehat{\mathbf{r}}_i'\boldsymbol{\Omega}(\widehat{\omega})^{-1}\widehat{\mathbf{r}}_i}{T}\right) - g(\widehat{\omega})\widehat{\sigma}^4\right]. \tag{21}$$

See also lemmas A5 and A6 and section A.4.

4 GMM Estimators: an overview

In this section, we review, and for completeness, define the GMM type estimators which are included in our simulation exercise.

The GMM approach assumes that α_i and u_{it} have an error components structure,

$$E(\alpha_i) = 0, \quad E(u_{it}) = 0, \quad E(\alpha_i u_{it}) = 0, \qquad (i = 1, .., N;\ t = 1, 2, ..., T), \tag{22}$$

and the errors are uncorrelated with the initial values

$$E(y_{i0} u_{it}) = 0, \qquad (i = 1, .., N;\ t = 1, 2, ..., T). \tag{23}$$

As with the transformed likelihood approach, it is also assumed that the errors, u_{it}, are serially and cross-sectionally independent:

$$E(u_{it} u_{is}) = 0, \qquad (i = 1, .., N;\ t = 1, 2, ..., T). \tag{24}$$

However, note that under the transformed QML no restrictions are placed on $E(\alpha_i u_{it})$, and $E(\alpha_i u_{it})$ are allowed to be non-zero and heterogenous across i.

4.1 Estimation

4.1.1 The first-difference GMM estimator

Under (22)-(24), and focusing on the equation in first differences, (2), Arellano and Bond (1991) suggest the following $T(T-1)/2$ moment conditions:

$$E(y_{is} \Delta u_{it}) = 0, \qquad (s = 0, 1, ..., t-2, t = 2, 3, ..., T). \tag{25}$$

If regressors, x_{it}, are strictly exogenous, i.e., if $E(x_{is} u_{it}) = 0$, for all t and s, then the following additional moments can also be used

$$E(x_{is} \Delta u_{it}) = 0, \qquad (s, t = 2, ..., T). \tag{26}$$

The moment conditions (25) and (26) can be written compactly as:

$$E\left[\dot{\mathbf{Z}}_i' \dot{\mathbf{u}}_i\right] = \mathbf{0},$$

where $\dot{\mathbf{u}}_i = \dot{\mathbf{q}}_i - \dot{\mathbf{W}}_i\boldsymbol{\delta}$, $\boldsymbol{\delta} = (\gamma, \beta)' = (\delta_1, \delta_2)'$ and

$$\dot{\mathbf{Z}}_i = \begin{pmatrix} y_{i0}, x_{i1}, ..., x_{iT} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & y_{i0}, y_{i1}, x_{i1}, ..., x_{iT} & ... & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & y_{i0}, ..., y_{i,T-2}, x_{i1}, ..., x_{iT} \end{pmatrix},$$

$$\dot{\mathbf{q}}_i = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \qquad \dot{\mathbf{W}}_i = \begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{iT} \end{pmatrix}.$$

The one and two-step first-difference GMM estimators based on the above moment conditions are given by

$$\widehat{\boldsymbol{\delta}}^{dif}_{GMM1} = \left(\dot{\mathbf{S}}'_{ZW}\left(\dot{\mathbf{D}}_{1step}\right)^{-1}\dot{\mathbf{S}}_{ZW}\right)^{-1}\dot{\mathbf{S}}'_{ZW}\left(\dot{\mathbf{D}}_{1step}\right)^{-1}\dot{\mathbf{S}}_{Zq}, \tag{27}$$

$$\widehat{\boldsymbol{\delta}}^{dif}_{GMM2} = \left(\dot{\mathbf{S}}'_{ZW}\left(\dot{\mathbf{D}}_{2step}\right)^{-1}\dot{\mathbf{S}}_{ZW}\right)^{-1}\dot{\mathbf{S}}'_{ZW}\left(\dot{\mathbf{D}}_{2step}\right)^{-1}\dot{\mathbf{S}}_{Zq}, \tag{28}$$

where $\dot{\mathbf{S}}_{ZW} = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{Z}}'_i\dot{\mathbf{W}}_i$, $\dot{\mathbf{S}}_{Zq} = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{Z}}'_i\dot{\mathbf{q}}_i$, $\dot{\mathbf{D}}_{1step} = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{Z}}'_i\mathbf{H}\dot{\mathbf{Z}}_i$, $\dot{\mathbf{D}}_{2step} = \frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{Z}}'_i\widehat{\dot{\mathbf{u}}}_i\widehat{\dot{\mathbf{u}}}'_i\dot{\mathbf{Z}}_i$, $\widehat{\dot{\mathbf{u}}}_i = \dot{\mathbf{q}}_i - \dot{\mathbf{W}}_i\widehat{\boldsymbol{\delta}}^{dif}_{GMM1}$, and $\mathbf{H}$ is a matrix with 2's on the main diagonal, -1's on the first upper and lower sub-diagonals and 0's otherwise.

4.1.2 System GMM estimator

Although consistency of the first-difference GMM estimator is obtained under the no serial correlation assumption, Blundell and Bond (1998) demonstrated that it suffers from the so called weak instruments problem when γ is close to unity, and/or the variance ratio $\tau^2 = \Sigma_{i=1}^{N}var(\alpha_i)/\Sigma_{i=1}^{N}var(u_{it})$ is large. As a solution, these authors propose the system GMM estimator due to Arellano and Bover (1995) and show that it works well even if γ is close to unity. But as shown recently by Bun and Windmeijer (2010), the system GMM estimator continues to suffer from the weak instruments problem when the variance ratio, τ^2 is large. See also Appendix of Binder, Hsiao, and Pesaran (2005) where it is shown that the asymptotic variance of the GMM estimator is an increasing function of τ^2.

To introduce the moment conditions for the system GMM estimator, the following additional *homogeneity* assumptions are required:

$$E(y_{is}\alpha_i) = E(y_{it}\alpha_i), \qquad \text{for all } s \text{ and } t,$$
$$E(x_{is}\alpha_i) = E(x_{it}\alpha_i), \qquad \text{for all } s \text{ and } t.$$

Under these assumptions, we have the following moment conditions:

$$E\left[\Delta y_{is}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (s = 1, ..., t-1, t = 2, 3, ..., T), \tag{29}$$

$$E\left[\Delta x_{is}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (s, t = 2, 3, ..., T). \tag{30}$$

For the construction of the moment conditions for the system GMM estimator, given the moment conditions for the first-difference GMM estimator, some moment conditions in (29) and (30) are redundant. Hence, to implement the system GMM estimation, in addition to (25) and (26), we use the following moment conditions:

$$\begin{aligned} E\left[\Delta y_{i,t-1}\left(\alpha_i + u_{it}\right)\right] &= 0, \qquad (t = 2, 3, ..., T), & (31) \\ E\left[\Delta x_{it}\left(\alpha_i + u_{it}\right)\right] &= 0, \qquad (t = 2, 3, ..., T). & (32) \end{aligned}$$

The moment conditions (25), (26), (31) and (32) can be written compactly as

$$E\left[\ddot{\mathbf{Z}}_i' \ddot{\mathbf{u}}_i\right] = \mathbf{0},$$

where $\ddot{\mathbf{u}}_i = \ddot{\mathbf{q}}_i - \ddot{\mathbf{W}}_i \boldsymbol{\delta}$,

$$\begin{aligned} \ddot{\mathbf{Z}}_i &= \texttt{diag}\left(\dot{\mathbf{Z}}_i, \breve{\mathbf{Z}}_i\right), \quad \breve{\mathbf{Z}}_i = \begin{pmatrix} \Delta y_{i1}, \Delta x_{i2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Delta y_{i2}, \Delta x_{i3} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Delta y_{i,T-1}, \Delta x_{iT} \end{pmatrix}, \\ \ddot{\mathbf{q}}_i &= \begin{pmatrix} \dot{\mathbf{q}}_i \\ \breve{\mathbf{q}}_i \end{pmatrix}, \quad \breve{\mathbf{q}}_i = \begin{pmatrix} y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad \ddot{\mathbf{W}}_i = \begin{pmatrix} \dot{\mathbf{W}}_i \\ \breve{\mathbf{W}}_i \end{pmatrix}, \quad \breve{\mathbf{W}}_i = \begin{pmatrix} y_{i1} & x_{i2} \\ \vdots & \vdots \\ y_{i,T-1} & x_{iT} \end{pmatrix}. \end{aligned}$$

The one and two-step system GMM estimators based on the above conditions are given by

$$\begin{aligned} \widehat{\boldsymbol{\delta}}^{sys}_{GMM1} &= \left(\ddot{\mathbf{S}}_{ZW}' \left(\ddot{\mathbf{D}}_{1step}\right)^{-1} \ddot{\mathbf{S}}_{ZW}\right)^{-1} \ddot{\mathbf{S}}_{ZW}' \left(\ddot{\mathbf{D}}_{1step}\right)^{-1} \ddot{\mathbf{S}}_{Zq}, & (33) \\ \widehat{\boldsymbol{\delta}}^{sys}_{GMM2} &= \left(\ddot{\mathbf{S}}_{ZW}' \left(\ddot{\mathbf{D}}_{2step}\right)^{-1} \ddot{\mathbf{S}}_{ZW}\right)^{-1} \ddot{\mathbf{S}}_{ZW}' \left(\ddot{\mathbf{D}}_{2step}\right)^{-1} \ddot{\mathbf{S}}_{Zq}, & (34) \end{aligned}$$

where $\ddot{\mathbf{S}}_{ZW} = \frac{1}{N}\sum_{i=1}^N \ddot{\mathbf{Z}}_i' \ddot{\mathbf{W}}_i$, $\ddot{\mathbf{S}}_{Zq} = \frac{1}{N}\sum_{i=1}^N \ddot{\mathbf{Z}}_i' \ddot{\mathbf{q}}_i$ and $\ddot{\mathbf{D}}_{1step} = \texttt{diag}\left(\frac{1}{N}\sum_{i=1}^N \dot{\mathbf{Z}}_i' \mathbf{H} \dot{\mathbf{Z}}_i, \frac{1}{N}\sum_{i=1}^N \breve{\mathbf{Z}}_i' \breve{\mathbf{Z}}_i\right)$. The two-step system GMM estimator is obtained by replacing $\ddot{\mathbf{D}}_{1step}$ with $\ddot{\mathbf{D}}_{2step} = \frac{1}{N}\sum_{i=1}^N \ddot{\mathbf{Z}}_i' \widehat{\ddot{\mathbf{u}}}_i \widehat{\ddot{\mathbf{u}}}_i' \ddot{\mathbf{Z}}_i$, where $\widehat{\ddot{\mathbf{u}}}_i = \ddot{\mathbf{q}}_i - \ddot{\mathbf{W}}_i \widehat{\boldsymbol{\delta}}^{sys}_{GMM1}$.

Although additional moment conditions proposed by Ahn and Schmidt (1995) could be used, we mainly focus on the above two set of moment conditions since they are often used in applied research.

4.1.3 Continuous-updating GMM estimator

Since the two-step GMM estimators have undesirable finite sample bias property, (Newey and Smith, 2004), alternative estimation methods have been proposed in the literature. These include the empirical likelihood estimator, (Qin and Lawless, 1994), the exponential tilting estimator (Kitamura and Stutzer, 1997; Imbens, Spady, and Johnson, 1998) and the continuous updating (CU-) GMM estimator (Hansen, Heaton, and Yaron, 1996), where these are members of the generalized empirical likelihood estimator (Newey and Smith, 2004). Amongst these estimators, we focus on the CU-GMM estimator as an alternative to the two-step GMM estimator.

To define the CU-GMM estimator, we need some additional notation. Let $\check{\mathbf{Z}}_i$ denote $\dot{\mathbf{Z}}_i$ or $\ddot{\mathbf{Z}}_i$, and $\check{\mathbf{u}}_i$ denote $\dot{\mathbf{u}}_i$ or $\ddot{\mathbf{u}}_i$, and set

$$\mathbf{g}_i(\boldsymbol{\delta}) = \check{\mathbf{Z}}_i'\check{\mathbf{u}}_i, \quad \widehat{\mathbf{g}}_N(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\delta}), \quad \widehat{\boldsymbol{\Omega}}_N(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N} \left[\mathbf{g}_i(\boldsymbol{\delta}) - \widehat{\mathbf{g}}(\boldsymbol{\delta})\right]\left[\mathbf{g}_i(\boldsymbol{\delta}) - \widehat{\mathbf{g}}(\boldsymbol{\delta})\right]'.$$

Then, the CU-GMM estimator is defined as

$$\widehat{\boldsymbol{\delta}}_{GMM-CU} = \arg\min_{\delta} \widehat{\mathbf{g}}_N(\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}_N(\boldsymbol{\delta})^{-1}\widehat{\mathbf{g}}_N(\boldsymbol{\delta}). \tag{35}$$

Newey and Smith (2004) demonstrate that the CU-GMM estimator has a smaller finite sample bias than the two-step GMM estimator.

4.2 Inference using GMM estimators

4.2.1 Alternative standard errors

In the case of GMM estimators the choice of the covariance matrix is often as important as the choice of the estimator itself for inference. Although, it is clearly important that the estimator of the covariance matrix should be consistent, in practice it might not have favorable finite sample properties and could result in inaccurate inference. To address this problem, some modified standard errors have been proposed. For the two-step GMM estimators, Windmeijer (2005) proposes corrected standard errors for linear static panel data models which are applied to dynamic panel models by Bond and Windmeijer (2005). For the CU-GMM, while it is asymptotically equivalent to the two-step GMM estimator, it is more dispersed than the two-step GMM estimator in finite samples and inference based on conventional standard errors formula results in a large size distortions. To overcome this problem, Newey and Windmeijer (2009) propose an alternative estimator for the covariance matrix of CU-GMM estimator under many-weak moments asymptotics and demonstrate by simulation that the use of the modified standard errors improve the size property of the tests based on the CU-GMM estimators.[7]

[7] For the precise definition of many weak moments asymptotics, see Newey and Windmeijer (2009).

4.2.2 Weak instruments robust inference

As noted above, the first-difference and system GMM estimators could be subject to the weak instruments problem. In the presence of weak instruments, the estimators are biased and inference becomes inaccurate. As a remedy for this, some tests that have correct size regardless of the strength of instruments have been proposed in the literature. These include Stock and Wright (2000) and Kleibergen (2005). Stock and Wright (2000) propose a GMM version of the Anderson and Rubin(AR) test (Anderson and Rubin, 1949). Kleibergen (2005) proposes a Lagrange Multiplier (LM) test. This author also extends the conditional likelihood ratio (CLR) test of Moreira (2003) to the GMM case since the CLR test performs better than other tests in linear homoskedastic regression models.

We now introduce these tests. The GMM version of the $\mathcal{AR}$ statistic proposed by Stock and Wright (2000) is defined as

$$AR(\boldsymbol{\delta}) = 2N \cdot Q(\boldsymbol{\delta}), \tag{36}$$

where $Q(\boldsymbol{\delta}) = \widehat{\mathbf{g}}(\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\delta})/2$. Under the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, this statistic is asymptotically (as $N \to \infty$) distributed as χ^2_n, regardless of the strength of the instruments, where n is the dimension of $\mathbf{g}_i(\boldsymbol{\delta})$.

The LM statistic proposed by Kleibergen (2005) is

$$LM(\boldsymbol{\delta}) = N \cdot \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} \left[\widehat{\mathbf{D}}(\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1}\widehat{\mathbf{D}}(\boldsymbol{\delta})\right]^{-1} \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}, \tag{37}$$

where $\widehat{\mathbf{D}}(\boldsymbol{\delta}) = \left(\widehat{\mathbf{d}}_1(\boldsymbol{\delta}), \widehat{\mathbf{d}}_2(\boldsymbol{\delta})\right)$ with

$$\widehat{\mathbf{d}}_j(\boldsymbol{\delta}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \mathbf{g}_i(\boldsymbol{\delta})}{\partial \delta_j} - \left(\frac{1}{N}\sum_{i=1}^{N} \frac{\partial \mathbf{g}_i(\boldsymbol{\delta})}{\partial \delta_j}\mathbf{g}_i(\boldsymbol{\delta})'\right)\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\delta}), \qquad \text{for } j = 1, 2.$$

Under the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, this statistic follows χ^2_2, asymptotically.

The GMM version of the CLR statistic proposed by Kleibergen (2005) is given by

$$CLR(\boldsymbol{\delta}) = \frac{1}{2}\left[AR(\boldsymbol{\delta}) - \widehat{R}(\boldsymbol{\delta}) + \sqrt{\left(AR(\boldsymbol{\delta}) - \widehat{R}(\boldsymbol{\delta})\right)^2 + 4LM(\boldsymbol{\delta})\widehat{R}(\boldsymbol{\delta})}\right] \tag{38}$$

where $\widehat{R}(\boldsymbol{\delta})$ is a statistic which is large when instruments are strong and small when the instruments are weak, and is random only through $\widehat{\mathbf{D}}(\boldsymbol{\delta})$ asymptotically. In the simulation, following Newey and Windmeijer (2009), we use $\widehat{R}(\boldsymbol{\delta}) = N \cdot \lambda_{min}\left(\widehat{\mathbf{D}}(\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1}\widehat{\mathbf{D}}(\boldsymbol{\delta})\right)$ where $\lambda_{min}(\mathbf{A})$ denotes the smallest eigenvalue of $\mathbf{A}$. Under the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, this statistic asymptotically follows a nonstandard distribution whose critical values can be obtained by simulation[8].

These tests are derived under the standard asymptotics where the number of moment conditions is fixed. Recently, Newey and Windmeijer (2009) show that these results are valid even under many

[8] For the details of computation, see Kleibergen (2005) or Newey and Windmeijer (2009).

weak moments asymptotics.

5 Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to investigate the finite sample properties of the transformed QML approach and compare them to those of the various GMM estimators proposed in the literature and reviewed in the previous section.

5.1 ARX(1) model

We first consider a distributed lag model with one exogenous regressor, ARX(1), which is likely to be more relevant in practice than the pure AR(1) model which will be considered later.

5.1.1 Monte Carlo design

For each i, the time series processes $\{y_{it}\}$ are generated as

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + u_{it}, \quad \text{for } t = -m+1, -m+2, .., 0, 1, ..., T, \tag{39}$$

where $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$, with $\sigma_i^2 \sim \mathcal{U}[0.5, 1.5]$, so that $E(\sigma_i^2) = 1$. For the initial values, we set $y_{i,-m} = 0$ and note that for m sufficiently large,

$$y_{i0} \approx \left(\frac{1-\gamma^m}{1-\gamma}\right)\alpha_i + \beta\sum_{j=0}^{m-1}\gamma^j x_{i,-j} + \sum_{j=0}^{m-1}\gamma^j u_{i,-j}.$$

We discard the first $m = 50$ observations, and use the observations $t = 0$ through T for estimation and inference.[9] The regressor, x_{it}, is generated as

$$x_{it} = \mu_i + \zeta_{it}, \qquad \text{for } t = -m, -m+1, .., 0, 1, ..., T, \tag{40}$$

where $\mu_i \sim iid\mathcal{N}(0, 1)$

$$\zeta_{it} = \phi\zeta_{i,t-1} + \varepsilon_{it}, \qquad \text{for } t = -49-m, -48-m, ..., 0, 1, ..., T, \tag{41}$$

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{\varepsilon i}^2), \qquad \xi_{i,-m-50} = 0. \tag{42}$$

with $|\phi| < 1$. We also generate a set of heteroskedastic errors for the x_{it} process and generate $\sigma_{\varepsilon i}^2 \sim \mathcal{U}[0.5, 1.5]$, independently of σ_i^2, and ensure heterogeneous variance ratios $\sigma_i^2/\sigma_{\varepsilon i}^2$ across i. We discard the first 50 observations of ζ_{it} and use the remaining $T + 1 + m$ observations for generating x_{it} and y_{it}.

[9] Hence, $T + 1$ is the actual length of the estimation sample.

In the simulations, we try the values $\gamma = 0.4, 0.9$, and $\phi = 0.5$. The slope coefficient, β, is set to ensure a reasonable degree of fit. But to deal with the error variance heterogeneity across the different equations in the panel we use the following average measure

$$R_y^2 = 1 - \frac{N^{-1}\sum_{i=1}^N Var(u_{it})}{N^{-1}\sum_{i=1}^N Var(y_{it}|c_i)},$$

where $Var(y_{it}|c_i)$ is the time-series variation of i^{th} unit. Since y_{it} is stationary and it is assumed to have started some time in the past we have

$$y_{it} = c_i + \beta\sum_{j=0}^{\infty}\gamma^j\zeta_{i,t-j} + \sum_{j=0}^{\infty}\gamma^j u_{i,t-j} = c_i + \beta w_{it} + \sum_{j=0}^{\infty}\gamma^j u_{i,t-j},$$

$c_i = (\alpha_i + \beta\mu_i)/(1-\gamma)$, and w_{it} is an AR(2) process, $w_{it} = \varphi_1 w_{i,t-1} + \varphi_2 w_{i,t-2} + \varepsilon_{it}$, with parameters $\varphi_1 = \gamma + \phi$, $\varphi_2 = -\phi\gamma$, and having the variance (Hamilton, 1994, p. 58)

$$Var(w_{it}) = \frac{(1+\phi\gamma)\,\sigma_{\varepsilon i}^2}{(1-\phi\gamma)\left[(1+\phi\gamma)^2 - (\gamma+\phi)^2\right]} = \frac{(1+\phi\gamma)\,\sigma_{\varepsilon i}^2}{(1-\gamma^2)\left(1-\phi^2\right)(1-\phi\gamma)}.$$

Hence

$$R_y^2 = 1 - \frac{N^{-1}\sum_{i=1}^N \sigma_i^2}{\frac{\beta^2(1+\phi\gamma)\left(N^{-1}\sum_{i=1}^N \sigma_{\varepsilon i}^2\right)}{(1-\gamma^2)\left(1-\phi^2\right)(1-\phi\gamma)} + \frac{N^{-1}\sum_{i=1}^N \sigma_i^2}{(1-\gamma^2)}} = \frac{\frac{\beta^2(1+\phi\gamma)\bar{\sigma}_{\varepsilon N}^2}{\left(1-\phi^2\right)(1-\phi\gamma)} + \gamma^2\bar{\sigma}_N^2}{\frac{\beta^2(1+\phi\gamma)\bar{\sigma}_{\varepsilon N}^2}{\left(1-\phi^2\right)(1-\phi\gamma)} + \bar{\sigma}_N^2},$$

$\bar{\sigma}_N^2 = N^{-1}\sum_{i=1}^N \sigma_i^2$, and $\bar{\sigma}_{\varepsilon N}^2 = N^{-1}\sum_{i=1}^N \sigma_{\varepsilon i}^2$. For N sufficiently large we now have (note that $\bar{\sigma}_N^2$ and $\bar{\sigma}_{\varepsilon N}^2 \to 1$ with $N \to \infty$)

$$R_y^2 = \frac{\frac{\beta^2(1+\phi\gamma)}{\left(1-\phi^2\right)(1-\phi\gamma)} + \gamma^2}{\frac{\beta^2(1+\phi\gamma)}{\left(1-\phi^2\right)(1-\phi\gamma)} + 1},$$

and

$$\beta^2 = \left(\frac{R_y^2 - \gamma^2}{1 - R_y^2}\right)\frac{\left(1-\phi^2\right)(1-\phi\gamma)}{(1+\phi\gamma)}.$$

We set β such that $R_y^2 = \gamma^2 + 0.1$. For $\gamma = 0.4$ and $\gamma = 0.9$ we have $R_y^2 = 0.26$ and $R_y^2 = 0.91$.

For the individual effects, we set

$$\alpha_i = \eta\left(\mu_i + \bar{u}_i + v_i\right),$$

where $\bar{u}_i = T^{-1}\sum_{t=1}^T u_{it}$, and $v_i \sim iid\mathcal{N}(0,1)$. Note that α_i is correlated with u_{it}. To set η we consider the variance ratio,

$$\tau^2 = \frac{N^{-1}\sum_{i=1}^N Var(\alpha_i)}{N^{-1}\sum_{i=1}^N Var(u_{it})} = \frac{\eta^2(T^{-1}\bar{\sigma}_N^2 + 2)}{\bar{\sigma}_N^2},$$

and use two values for τ, namely a low value of $\tau = 1$ often set in the Monte Carlo experiments conducted in the literature, and the high value of $\tau = 5$. The sample sizes considered are $N = 50, 150, 500$ and $T = 5, 10, 15$.

For the computation of the transformed ML estimators, we try two procedures. One is to maximize the log likelihood function directly, while the other is to use an iterative procedure suggested by Grassetti (2011). For the starting value of the nonlinear optimization, we use the minimum distance estimator of Hsiao, Pesaran, and Tahmiscioglu (2002) where ω is estimated by the one-step first-difference GMM estimator (27) in which $\dot{\mathbf{Z}}_i$ is replaced with

$$\dot{\mathbf{Z}}_i = \begin{pmatrix} y_{i0} & x_{i1} & 0 & 0 \\ y_{i1} & x_{i2} & y_{i0} & x_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ y_{i,T-2} & x_{i,T-1} & y_{i,T-3} & x_{i,T-2} \end{pmatrix}.$$

This GMM estimator is also used as the starting value for the iterative procedure.

For the GMM estimators, although there are many moment conditions for the first-difference GMM estimator as in (25) and (26), we consider three sets of moment conditions which only exploit a subset of instruments. The first set of moment conditions, denoted as "DIF1", consists of $E(y_{is}\Delta u_{it}) = 0$ for $s = 0, ..., t-2; t = 2, ..., T$ and $E(x_{is}\Delta u_{it}) = 0$ for $s = 1, ..., t; t = 2, ..., T$. In this case, the number of moment conditions are 24, 99, 224 for $T = 5, 10, 15$, respectively. The second set of moment conditions, denoted as "DIF2", consist of $E(y_{i,t-2-l}\Delta u_{it}) = 0$ with $l = 0$ for $t = 2$, $l = 0, 1$ for $t = 3, ..., T$ and $E(x_{i,t-l}\Delta u_{it}) = 0$ with $l = 0, 1$ for $t = 2$, $l = 0, 1, 2$ for $t = 3, ..., T$. In this case, the number of moment conditions are 18, 43, 68 for $T = 5, 10, 15$, respectively. The third set of moment conditions, denoted as "DIF3", consist of $\sum_{t=2}^{T} E(y_{i,t-2}\Delta u_{it}) = 0$, $\sum_{t=2}^{T-1} E(y_{i,t-2}\Delta u_{it}) = 0$, $\sum_{t=2}^{T} E(x_{it}\Delta u_{it}) = 0$, and $\sum_{t=2}^{T-1} E(x_{it}\Delta u_{it}) = 0$. The number of moment conditions for this case, often called the stacked instruments, are 4 for all T. Similarly, for the system GMM estimator, we add moment conditions (31) and (32) in addition to "DIF1" and "DIF2", which are denoted as "SYS1" and "SYS2", respectively. For "SYS1" we have 32, 117, 252 moment conditions for $T = 5, 10, 15$, respectively, while for "SYS2" we have 26, 61, 96 moment conditions for $T = 5, 10, 15$, respectively. Also, we add moment conditions $\sum_{t=2}^{T} E\left[\Delta y_{i,t-1}(\alpha_i + u_{it})\right] = 0$, $\sum_{t=2}^{T-1} E\left[\Delta y_{i,t-1}(\alpha_i + u_{it})\right] = 0$, $\sum_{t=2}^{T} E\left[\Delta x_{it}(\alpha_i + u_{it})\right] = 0$ and $\sum_{t=2}^{T-1} E\left[\Delta x_{it}(\alpha_i + u_{it})\right] = 0$ in addition to "DIF3", which is denoted as "SYS3". In this case, the number of moment conditions is 8 for any T.

In a number of cases where N is not sufficiently large relative to the number of moment conditions (for example, when $T = 15$ and $N = 50$) the inverse of the weighting matrix can not be computed. Such cases are denoted as "$-$" in the summary result tables.

For inference, we use the robust standard errors formula given in Theorem 2 for the transformed QML estimator. For the GMM estimators, in addition to the conventional standard errors, we also compute Windmeijer (2005)'s standard errors with finite sample correction for the two-step GMM estimators and Newey and Windmeijer (2009)'s alternative standard errors formula for the CU-GMM estimators. For the computation of optimal weighting matrix, a centered version is used except for the CU-GMM[10].

In addition to the Monte Carlo results for γ and β, we also report simulation results for the long-run coefficient defined by $\psi = \beta/(1-\gamma)$. We report median biases, median absolute errors (MAE), size and power for γ, β and ψ. The power is computed at $\gamma - 0.1$, $\beta - 0.1$ and $(\beta - 0.1)/(1-(\gamma-0.1))$, for selected null values of γ and β. All tests are carried out at the 5% significance level, and all experiments are replicated $1,000$ times.

5.1.2 Results for the ARX(1) model

To save space, we report the results of the GMM estimators which exploit moment conditions "DIF2" and "SYS2" with one-step estimation procedure only. The reason for selecting these moment conditions is that, in practice, these moment conditions are often used to mitigate the finite sample bias caused by using too many instruments. A complete set of results giving the remaining GMM estimators that make use of additional instruments are provided in a supplement available from the authors on request.

The small sample results for γ and β are summarized in Tables 1 to 4.[11] We first focus on the results of γ and then discuss the results for β. Table 1 (and A.2 in the supplement) provide the results of bias and MAE for the case of $\gamma = 0.4$, and shows that the transformed QMLE has a smaller bias than the GMM estimators in all cases with the exception of the CU-GMM estimator (see Table A.2). In terms of MAE the transformed QMLE outperforms the GMM estimators in all cases.

As for the effect of increasing the variance ratio, τ, on the various estimators, we first recall that the transformed QMLE is invariant to the choice of τ. In contrast, as to be expected the performance of the GMM estimators deteriorates (in some case substantially) as τ is increased from 1 to 5. This tendency is especially evident in the case of the system GMM estimators, and is in sharp contrast to the performance of the transformed QMLE which is robust to changes in τ. These observations also hold if we consider the experiments with $\gamma = 0.9$ (Table 2). Although the GMM estimators have smaller biases than the transformed likelihood estimator in a few cases, in terms of MAE, the transformed QMLE performs best in all cases (see also Table A.12 in the supplement).

We next consider size and power of the various tests, summarized in Tables 3 and 4 (A.3 and A.13 in the supplement). The results in these tables show that the empirical size of the transformed QMLE is close to the nominal size of 5% for all values of γ, T, N and τ. In contrast, for the GMM estimators, we find that the test sizes vary considerably depending on γ, T, N, τ, the estimation method (1step, 2step, CU), and whether corrections are applied to the standard errors. In the case of the GMM results without standard error corrections, most of the GMM methods are subject to substantial size distortions when N is small. For instance, when $\gamma = 0.4$, $N = 50$, $T = 5$, and $\tau = 1$, the size of the test based on the two-step procedure using moment conditions "DIF2" estimator is 34.2%. But the

[10] In the earlier version, we used centered weighting matirx. However, in this version, uncentered weighting matrix is used for the CU-GMM since it gave better performance than using centered weighting matrix.

[11] The corresponding tables in the supplement are labelled as Tables A.1 to A.9.

size distortion gets smaller as N increases. Increasing N to 500, reduces the size of this test to 7.7%. However, even with $N = 500$, the size distortion gets larger for two-step and CU-GMM estimators as T increases.

As to the effects of changes in τ on the estimators, we find that the system GMM estimators are significantly affected when τ is increased. When $\tau = 5$, all the system GMM estimators have large size distortions even when $T = 5$ and $N = 500$, where conventional asymptotics are expected to work well. This may be due to large finite sample biases caused by a large τ.

For the tests based on corrected GMM standard errors, Windmeijer (2005)'s correction seems to be quite useful, and in many cases it leads to accurate inference, although the corrections do not seem able to mitigate the size problem of the system GMM estimator when τ is large. The standard errors of Newey and Windmeijer (2009) are also helpful: they improve the size property in many cases.

Comparing power of the tests, we observe that the transformed likelihood estimator is in general more powerful than the GMM estimators. Specifically, the transformed likelihood estimators have higher power than the most efficient two-step system GMM estimator based on "SYS1" with Windmeijer's correction.

The above conclusions for size and power hold generally when we consider experiments with $\gamma = 0.9$ (Table 4 and A.13), except that the system GMM estimators now perform rather poorly even for a relatively large N. For example, when $\gamma = 0.9$, $T = 5$, $N = 500$ and $\tau = 1$, size distortions of the system GMM estimators are substantial, as compared to the case where $\gamma = 0.4$. Although it is known that the system GMM estimators break down when τ is large[12], the simulation results in Table 4 and A.13 reveal that they perform poorly even when τ is not so large ($\tau = 1$).

We next consider the small sample results for β (Tables 1 to 4, A.4 to A.6, and A.14 to A.16). The outcomes are similar to the results reported for γ. The transformed likelihood estimator tends to have smaller biases and MAEs than the GMM estimators in many cases, and there are almost no size distortions for all values of T, N and τ. The performance of the GMM estimators crucially depends on the values of T, N and τ. Unless N is large, the GMM estimators perform poorly and the system GMM estimators are subject to substantial size distortions when τ is large even for $N = 500$, although the magnitude of size distortions are somewhat smaller than those reported for γ.

The results for the long-run coefficient, $\psi = \beta/(1-\gamma)$, which are reported in the supplement (Tables A.7 to A.9 and A.17 to A.19), are very similar to those of γ and β. Although the GMM estimators outperform the transformed likelihood estimator in some cases, in terms of MAE, the transformed likelihood estimator performs best in almost all cases. As for inference, the transformed likelihood estimator has correct sizes for all values of T, N and τ when $\gamma = 0.4$. However, it shows some size distortions when $\gamma = 0.9$ and the sample size is small, say, when $T = 5$ and $N = 50$. However, size improves as T and/or N increase(s). When $T = 15$ and $N = 500$, there is essentially no size distortions. For the GMM estimators, it is observed that although the sizes are correct in some cases, say, the case with $T = 5$ and $N = 500$ when $\gamma = 0.4$, it is not the case when $\gamma = 0.9$; even for the case

[12] See Hayakawa (2007) and Bun and Windmeijer (2010).

of $T = 5$ and $N = 500$, there are size distortions and a large τ aggravates the size distortions.

Finally, we consider weak instruments robust tests, which are reported in Tables 5, A.10 and A.20. We find that test sizes are close to the nominal value only when $T = 5$ and $N = 500$. In other cases, especially when N is small and/or T is large, there are substantial size distortions. Although Newey and Windmeijer (2009) prove the validity of these tests under many weak moments asymptotics, they are essentially imposing $n^2/N \to 0$ or a stronger restriction where n is the number of moment conditions, which is unlikely to hold when N is small and/or T is large. Therefore, the weak instruments robust tests are less appealing, considering the very satisfactory size properties of the transformed likelihood estimator, the difficulty of carrying out inference on subset of the parameters using the weak instruments robust tests, and large size distortions observed for these tests when N is small.

In summary, for estimation of ARX panel data models the transformed likelihood estimator has several favorable properties over the GMM estimators in that the transformed likelihood estimator generally performs better than the GMM estimators in terms of biases, MAEs, size and power, and unlike GMM estimators, it is not affected by the variance ratio of individual effects to disturbances.

5.2 AR(1) model

5.2.1 Monte Carlo design

The data generating process is the same as that in the previous section with $\beta = 0$. More specifically, y_{it} are generated as

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + u_{it}, \qquad (t = -m+1, ..., 1, ..., T; i = 1, ..., N), \tag{43}$$

with $y_{i,-m} = 0$ where $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i^2 \sim \mathcal{U}[0.5, 1.5]$, and

$$y_{i0} \approx \left(\frac{1-\gamma^m}{1-\gamma}\right)\alpha_i + \sum_{j=0}^{m-1} \gamma^j u_{i,-j}.$$

Individual effects are generated as

$$\alpha_i = \eta\left(\bar{u}_i + v_i\right),$$

where $v_i \sim iid\mathcal{N}(0,1)$, and η is set so that particular values of the variance ratio,

$$\tau^2 = \frac{N^{-1}\sum_{i=1}^N Var(\alpha_i)}{N^{-1}\sum_{i=1}^N Var(u_{it})} = \frac{\eta^2(T^{-1}\bar{\sigma}_N^2 + 1)}{\bar{\sigma}_N^2},$$

is achieved. Note that for N sufficiently large $\tau^2 \approx \eta^2(1 + 1/T)$.

For parameters and sample sizes, we consider $\gamma = 0.4, 0.9$, $T = 5, 10, 15, 20$ $N = 50, 150, 500$, and $\tau = 1, 5$.

Some comments on the computations are in order. For the starting value in the nonlinear optimization routine used to compute the transformed log-likelihood estimator, we use $(\widetilde{b}, \widetilde{\gamma}, \widetilde{\omega}, \widetilde{\sigma}^2)$ where

$\widetilde{b} = N^{-1}\sum_{i=1}^{N}\Delta y_{i1}$, $\widetilde{\gamma}$ is the one-step first-difference GMM estimator (27) where $\dot{\mathbf{W}}_i$ and $\dot{\mathbf{Z}}_i$ are replaced with[13]

$$\dot{\mathbf{W}}_i = \begin{pmatrix} \Delta y_{i1} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix}, \qquad \dot{\mathbf{Z}}_i = \begin{pmatrix} y_{i0} & 0 & 0 \\ y_{i1} & y_{i0} & 0 \\ y_{i2} & y_{i1} & y_{i0} \\ \vdots & \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} & y_{i,T-4} \end{pmatrix},$$

$\widetilde{\omega} = [(N-1)\widetilde{\sigma}_u^2]^{-1}\sum_{i=1}^{N}\left(\Delta y_{i1} - \widetilde{b}\right)^2$ and $\widetilde{\sigma}_u^2 = [2N(T-2)]^{-1}\sum_{i=1}^{N}\left(\Delta y_{it} - \widetilde{\gamma}\Delta y_{i,t-1}\right)^2$.

For the first-difference GMM estimators, we consider three sets of moment conditions. The first set of moment conditions, denoted as "DIF1", consists of $E(y_{is}\Delta u_{it}) = 0$ for $s = 0, ..., t-2; t = 2, ..., T$. In this case, the number of moment conditions are 10, 45, 105, 190 for $T = 5$, 10, 15, 20, respectively. The second set of moment conditions, denoted by "DIF2", consist of $E(y_{i,t-2-l}\Delta u_{it}) = 0$ with $l = 0$ for $t = 2$, and $l = 0, 1$ for $t = 3, ..., T$. In this case, the number of moment conditions are 7, 17, 27, 37 for $T = 5, 10, 15, 20$, respectively. The third set of moment conditions, denoted as "DIF3", consists of $\sum_{t=2}^{T} E(y_{i,t-2}\Delta u_{it}) = 0$, $\sum_{t=2}^{T-1} E(y_{i,t-2}\Delta u_{it}) = 0$ and $\sum_{t=2}^{T-2} E(y_{i,t-2}\Delta u_{it}) = 0$. In this case, the number of moment conditions are 3 for all T.

Similarly, for the system GMM estimator, we add moment conditions $E[\Delta y_{i,t-1}(\alpha_i + u_{it})] = 0$ for $t = 2, ..., T$ in addition to "DIF1" and "DIF2", which are denoted as "SYS1" and "SYS2", respectively. We also add moment conditions $\sum_{t=2}^{T} E\left[\Delta y_{i,t-1}(\alpha_i + u_{it})\right] = 0$, $\sum_{t=2}^{T-1} E\left[\Delta y_{i,t-1}(\alpha_i + u_{it})\right] = 0$, $\sum_{t=2}^{T} E\left[\Delta x_{it}(\alpha_i + u_{it})\right] = 0$ and $\sum_{t=2}^{T-1} E\left[\Delta x_{it}(\alpha_i + u_{it})\right] = 0$ in addition to "DIF3". For the moment conditions "SYS1", we have 14, 54, 119, 209 moment conditions for $T = 5, 10, 15, 20$, respectively, while for the moment conditions "SYS2", we have 11, 26, 41, 56 moment conditions for $T = 5$, 10, $15, 20$, respectively. The number of moment conditions for "SYS3" are 6 for all T. With regard to the inference, we use the robust standard errors formula given in Theorem 2 for the transformed log-likelihood estimator. For the GMM estimators, in addition to the conventional standard errors, we also compute Windmeijer (2005)'s standard errors for the two-step GMM estimators and Newey and Windmeijer (2009)'s standard errors for the CU-GMM estimators.

We report the median biases, median absolute errors (MAE), sizes ($\gamma = 0.4$ and 0.9) and powers (resp. $\gamma = 0.3$ and 0.8) with the nominal size set to 5%. As before, the number of replications is set to $1,000$.

5.2.2 Results

As in the case of ARX(1) experiments, to save space, we report the results of the transformed likelihood estimator and the GMM estimators exploiting moment conditions "DIF2" and "SYS2" with one-step estimation procedure. Complete set of results are provided in a supplement, which is available upon

[13] This type of estimator is considered in Bun and Kiviet (2006). Since the number of moment conditions are three, this estimator is always computable for any values of N and T considered in this paper. Also, since there are two more moments, we can expect that the first and second moments of the estimator to exist.

request. In the following, Tables 6 to 8 are given in the paper and Tables A.21 to A.28 are given in a supplement.

The biases and MAEs of the various estimators for the case of $\gamma = 0.4$ are summarized in Tables 6, A.22 and A.26. As can be seen from these tables, the transformed likelihood estimator performs best (in terms of MAE) in almost all cases, the exceptions being the CU-GMM estimators that show smaller biases in some experiments. As to be expected, the one- and two-step GMM estimators deteriorate as the variance ratio, τ, is increased from 1 to 5, and this tendency is especially evident for the system GMM estimator. For the case of $\gamma = 0.9$, we find that the system GMM estimators have smaller biases and MAEs than the transformed likelihood estimator in some cases. However, when $\tau = 5$, the transformed likelihood estimator outperforms the GMM estimators in all cases, both in terms of bias and MAE.

Consider now the size and power properties of the alternative procedures. The results for $\gamma = 0.4$ are summarized in Tables 7 and A.23. We first note that the transformed likelihood procedure shows almost correct sizes for all experiments. For the GMM estimators, although there are substantial size distortions when $N = 50$, the empirical sizes become close to the nominal value as N is increased. When $T = 5, 10$ and $N = 500$ and $\tau = 1$, the size distortions of the GMM estimators are small. However, when $\tau = 5$, there are severe size distortions for the system GMM estimator even when $N = 500$. For the effects of corrected standard errors, similar results to the ARX(1) case are obtained. Namely, Windmeijer (2005)'s correction is quite useful, and in many cases it leads to accurate inference although the corrections do result in severely under-sized tests in some cases. Also, this correction does not seem that helpful in mitigating the size problem of the system GMM estimator when τ is large. The standard errors of Newey and Windmeijer (2009) used for the CU-GMM estimators are also helpful: they improve the size property in many cases.

Size and power of the tests in the case of experiments with $\gamma = 0.9$ are summarized in Tables 7 and A.27, and show significant size distortions in many cases. The size distortion of the transformed likelihood gets reduced for relatively large sample sizes and its size declines to 8.0% when $\tau = 1$, $N = 500$ and $T = 20$. As to be expected, increasing the variance ratio, τ, to 5, does not change this result. A similar pattern can also be seen in the case of first-difference GMM estimators if we consider $\tau = 1$. But the size results are much less encouraging if we consider the system GMM estimators. Also, as to be expected, size distortions of GMM type estimators become much more pronounced when the variance ratio is increased to $\tau = 5$.

Finally, we consider the small sample performance of the weak instruments robust tests which are provided in Tables 8, A.24 and A.28. These results show that size distortions are reduced only when N is large ($N = 500$). In general, size distortions of these tests get worse as T, or the number of moment conditions, increases. In terms of power, the Lagrange multiplier test and conditional likelihood ratio test based on “SYS2" have almost the same power as the transformed likelihood estimator when $\gamma = 0.4$, $T = 5$, $N = 500$ and $\tau = 1$. For the case of $\gamma = 0.9$, the results are very similar to the case of $\gamma = 0.4$. Size distortions are small only when N is large. When N is small, there are substantial size distortions.

6 Concluding remarks

In this paper, we proposed the transformed likelihood approach to estimation and inference in dynamic panel data models with cross-sectionally heteroskedastic errors. It is shown that the transformed likelihood estimator by Hsiao, Pesaran, and Tahmiscioglu (2002) continues to be consistent and asymptotically normally distributed, but the covariance matrix of the transformed likelihood estimators must be adjusted to allow for the cross-sectional heteroskedasticity. By means of Monte Carlo simulations, we investigated the finite sample performance of the transformed likelihood estimator and compared it with a range of GMM estimators. Simulation results revealed that the transformed likelihood estimator for an ARX(1) model with a single exogenous regressor has very small bias and accurate size property, and in most cases outperformed GMM estimators, whose small sample properties vary considerably across parameter values (γ and β), the choice of moment conditions, and the value of the variance ratio, τ.

In this paper, x_{it} is assumed to be strictly exogenous. However, in practice, the regressors may be endogenous or weakly exogenous (c.f. Keane and Runkle, 1992). To allow for endogenous and weakly exogenous variables, one could consider extending the panel VAR approach advanced in Binder, Hsiao, and Pesaran (2005) to allow for cross-sectional heteroskedasticity. More specifically, consider the following bivariate model:

$$\begin{aligned} y_{it} &= \alpha_{yi} + \gamma y_{i,t-1} + \beta x_{it} + u_{it} \\ x_{it} &= \alpha_{xi} + \phi y_{i,t-1} + \rho x_{i,t-1} + v_{it} \end{aligned}$$

where $cov(u_{it}, v_{it}) = \theta$. In this model, x_{it} is strictly exogenous if $\phi = 0$ and $\theta = 0$, weakly exogenous if $\theta = 0$, and endogenous if $\theta \neq 0$. This model can be written as a PVAR(1) model as follows

$$\begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \begin{pmatrix} \alpha_{yi} + \beta\alpha_{xi} \\ \alpha_{xi} \end{pmatrix} + \begin{pmatrix} \gamma + \beta\phi & \beta\rho \\ \phi & \rho \end{pmatrix} \begin{pmatrix} y_{i,t-1} \\ x_{i,t-1} \end{pmatrix} + \begin{pmatrix} u_{it} + \beta v_{it} \\ v_{it} \end{pmatrix},$$

for $i = 1, 2, ..., N$. Let $\mathbf{A} = \{a_{ij}\}(i, j = 1, 2)$ be the coefficient matrix of $(y_{i,t-1}, x_{i,t-1})'$ in the above VAR model. Then, we have $\beta = a_{12}/a_{22}$, $\gamma = a_{11} - a_{12}a_{21}/a_{22}$, $\rho = a_{22}$ and $\phi = a_{21}$. Thus, if we estimate a PVAR model in (y_{it}, x_{it}), allowing for fixed effects and cross-sectional heteroskedasticity, we can recover the parameters of interest, γ and β, from the estimated coefficients of such a PVAR model. However, detailed analysis of such a model is beyond the scope of the present paper and is left to future research.

A Proofs

A.1 Preliminary results

In this appendix we provide some definitions and results useful for the derivations in the paper.

Lemma A1 *Let $\mathbf{\Omega}(\omega)$ be given by (10). Then the determinant and inverse of $\mathbf{\Omega}(\omega)$ are:*

$$\left|\mathbf{\Omega}(\omega)\right| = g(\omega) = 1 + T(\omega - 1), \tag{44}$$

$$\mathbf{\Omega}(\omega)^{-1} = g(\omega)^{-1} \left(\begin{array}{c|cccc} T & T-1 & \ldots & 2 & 1 \\ \hline T-1 & (T-1)\omega & \ldots & 2\omega & \omega \\ T-2 & & & & \\ & & & & \\ 2 & 2\omega & & 2\left[(T-2)\omega-(T-3)\right] & (T-2)\omega-(T-3) \\ 1 & \omega & \ldots & (T-2)\omega-(T-3) & (T-1)\omega-(T-2) \end{array} \right) \tag{45}$$

The generic (t,s)th element of the $(T-1)\times(T-1)$ lower block of $\mathbf{\Omega}(\omega)^{-1}$, denoted by $\widetilde{\mathbf{\Omega}}(\omega)$, can be calculated using the following formulas, for $t, s = 1, 2, ..., T-1$:

$$\left\{\widetilde{\mathbf{\Omega}}(\omega)\right\}_{ts} = \begin{cases} s(T-t)\omega - (s-1)(T-t), & (s \leq t) \\ t(T-s)\omega - (t-1)(T-s), & (s > t) \end{cases}. \tag{46}$$

Proof. See Hsiao, Pesaran, and Tahmiscioglu (2002). ■

Lemma A2 *Let $\mathbf{\Phi}$ be defined as*

$$\mathbf{\Phi} = \begin{pmatrix} T^2 & T(T-1) & T(T-2) & \ldots & T \\ T(T-1) & (T-1)^2 & (T-1)(T-2) & \ldots & (T-1) \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ T & (T-1) & (T-2) & \ldots & 1 \end{pmatrix} = \boldsymbol{\vartheta}\boldsymbol{\vartheta}' \tag{47}$$

where $\boldsymbol{\vartheta}' = (T, T-1, \ldots, 2, 1)$. Then, we have

$$\mathtt{tr}\left(\mathbf{\Phi}\mathbf{\Omega}(\omega)\right) = \mathtt{tr}\left(\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\mathbf{\Omega}(\omega)\right) = \boldsymbol{\vartheta}'\mathbf{\Omega}(\omega)\boldsymbol{\vartheta} = Tg(\omega), \tag{48}$$

where $g(\omega)$ is given by (44).

Proof. See Hsiao, Pesaran, and Tahmiscioglu (2002). ■

Lemma A3 *Let $\{x_i, i = 1, 2, ..., N\}$ and $\{z_i, i = 1, 2, ..., N\}$ be two sequences of independently distributed random variables, such that $x_i z_i$ are independently distributed across i, although x_i and z_i need* *not be independently distributed of each other. Then*

$$E\left[\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} z_i\right)\right] = \sum_{i=1}^{N} Cov(x_i, z_i) + \left[\sum_{i=1}^{N} E\left(x_i\right)\right]\left[\sum_{i=1}^{N} E\left(z_i\right)\right].$$

Lemma A4 *Consider the transformed model (9). Under Assumptions 1, 2, 3, 4, and 5, we have*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left[\Delta\mathbf{W}_i'\mathbf{\Omega}\left(\bar{\omega}_0\right)^{-1}\mathbf{r}_i\right] = \mathbf{0}, \tag{49}$$

where $\mathbf{\Omega}\left(\omega\right)$ *is given in (10),* $\bar{\omega}_0$ *is defined in (14). Further,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left(\Delta\mathbf{W}_i'\mathbf{\Phi}\mathbf{r}_i\right) = \begin{pmatrix} 0 & \mathbf{0}_{T\times 1} & \bar{\delta} & 0 \end{pmatrix}', \tag{50}$$

where $\mathbf{\Phi}$*, and* $\bar{\delta}$ *are given by (47) and (56), respectively.*

Proof. Let $\mathbf{p}_i = \mathbf{\Omega}\left(\bar{\omega}_0\right)^{-1}\mathbf{r}_i = (p_{i1}, ..., p_{iT})'$ and recall that $\mathbf{r}_i = (v_{i1}, \Delta u_{i2}, ..., \Delta u_{iT})'$. Hence, using (46) we have

$$\begin{aligned} p_{i1} &= Tv_{i1} + \sum_{s=2}^{T}(T-s+1)\Delta u_{is}, \\ p_{it} &= (T-t+1)v_{i1} + \sum_{s=2}^{t} h_{ts}\Delta u_{is} + \sum_{s=t+1}^{T} k_{ts}\Delta u_{is}, \qquad (t = 2, ..., T-1) \\ p_{iT} &= v_{i1} + \sum_{s=2}^{T} h_{Ts}\Delta u_{is} \end{aligned}$$

and

$$\begin{aligned} h_{ts}(\bar{\omega}_0) &= (T-t+1)\left[(s-1)\bar{\omega}_0 - (s-2)\right], \\ k_{ts}(\bar{\omega}_0) &= (T-s+1)\left[(t-1)\bar{\omega}_0 - (t-2)\right]. \end{aligned} \tag{51}$$

Also using (8) and since by Assumption 2, $\Delta\mathbf{x}_i = (\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT})'$ then it readily follows that

$$E\left[\Delta\mathbf{W}_i'\mathbf{\Omega}\left(\bar{\omega}_0\right)^{-1}\mathbf{r}_i\right] = E\left(\Delta\mathbf{W}_i'\mathbf{p}_i\right) = \left[0, \mathbf{0}_{T\times 1}, E\left(\Delta\widetilde{\mathbf{y}}_{i,-1}'\mathbf{p}_i\right), 0\right]',$$

where $\Delta\widetilde{\mathbf{y}}_{i,-1} = (0, \Delta y_{i1}, ..., \Delta y_{i,T-1})'$. Hence to establish (49) we need to prove that

$$\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left(\Delta\widetilde{\mathbf{y}}_{i,-1}'\mathbf{p}_i\right) = \mathbf{0}.$$

But, noting that $E(\Delta u_{is}\Delta y_{it}) = 0$ for $t < s-1$, we have

$$\begin{aligned}
E\left(\Delta\widetilde{\mathbf{y}}'_{i,-1}\mathbf{p}_i\right) &= \sum_{t=2}^{T} E\left(p_{it}\Delta y_{i,t-1}\right) = \sum_{t=2}^{T-1} E\left(p_{it}\Delta y_{i,t-1}\right) + E\left(p_{iT}\Delta y_{i,T-1}\right) \\
&= \sum_{t=2}^{T-1} E\left[(T-t+1)v_{i1}\Delta y_{i,t-1} + \sum_{s=2}^{t} h_{ts}(\bar{\omega}_0)\Delta u_{is}\Delta y_{i,t-1} + \sum_{s=t+1}^{T} k_{ts}(\bar{\omega}_0)\Delta u_{is}\Delta y_{i,t-1}\right] \\
&\quad + E\left(p_{iT}\Delta y_{i,T-1}\right) \\
&= \sum_{t=2}^{T}(T-t+1)E\left(v_{i1}\Delta y_{i,t-1}\right) + \sum_{t=2}^{T}\sum_{s=2}^{t} h_{ts}(\bar{\omega}_0)E\left(\Delta u_{is}\Delta y_{i,t-1}\right) \\
&= A_{1i} + A_{2i}.
\end{aligned}$$

Also[14]

$$E(v_{i1}\Delta y_{it}) = \begin{cases} \sigma_i^2\omega_{i0} & t=1 \\ \sigma_i^2\gamma^{t-2}(\gamma\omega_{i0}-1) & t=2,...,T \end{cases} \tag{52}$$

$$E(\Delta u_{is}\Delta y_{it}) = \begin{cases} -\sigma_i^2 & t=s-1 \\ \sigma_i^2(2-\gamma) & s=t \\ -\sigma_i^2(1-\gamma)^2\gamma^{t-s-1} & s<t \end{cases} \tag{53}$$

Using these results we now have

$$A_{1i} = \sigma_{i0}^2\left[(T-1)\omega_{i0} + (\gamma\omega_{i0}-1)\sum_{t=3}^{T}(T-t+1)\gamma^{t-3}\right], \tag{54}$$

[14] These results are obtained by noting that Δy_{it} can be written as follows

$$\begin{aligned}
\Delta y_{i1} &= b + \boldsymbol{\pi}'\Delta\mathbf{x}_i + v_{i1}, \\
\Delta y_{it} &= \gamma^{t-1}\Delta y_{i1} + \beta\left(\sum_{j=0}^{t-2}\gamma^j x_{i,t-j}\right) + \sum_{j=0}^{t-2}\gamma^j\Delta u_{i,t-j} \\
&= \gamma^{t-1}\left(b + \boldsymbol{\pi}'\Delta\mathbf{x}_i\right) + \gamma^{t-1}v_{i1} + \beta\left(\sum_{j=0}^{t-2}\gamma^j x_{i,t-j}\right) + \sum_{j=0}^{t-2}\gamma^j\Delta u_{i,t-j}, \qquad (t=2,...,T).
\end{aligned}$$

and (recalling that h_{ts} depends on $\bar{\omega}_0$)

$$\begin{aligned} A_{2i} &= h_{22}E(\Delta u_{i2}\Delta y_{i1}) \\ &\quad + h_{32}E(\Delta u_{i2}\Delta y_{i2}) + h_{33}E(\Delta u_{i3}\Delta y_{i2}) \\ &\quad + h_{42}E(\Delta u_{i2}\Delta y_{i3}) + h_{43}E(\Delta u_{i3}\Delta y_{i3}) + h_{44}E(\Delta u_{i4}\Delta y_{i3}) \\ &\quad + h_{52}E(\Delta u_{i2}\Delta y_{i4}) + h_{53}E(\Delta u_{i3}\Delta y_{i4}) + h_{54}E(\Delta u_{i4}\Delta y_{i4}) + h_{55}E(\Delta u_{i5}\Delta y_{i4}) \\ &\quad \vdots \\ &\quad + h_{T2}E(\Delta u_{i2}\Delta y_{i,T-1}) + h_{T3}E(\Delta u_{i3}\Delta y_{i,T-1}) + \cdots + h_{T,T-2}E(\Delta u_{i,T-2}\Delta y_{i,T-1}) + \\ &\qquad + h_{T,T-1}E(\Delta u_{i,T-1}\Delta y_{i,T-1}) + h_{TT}E(\Delta u_{iT}\Delta y_{i,T-1}) \\ &= \sigma_{i0}^2 \left[(-1)\sum_{s=2}^{T} h_{ss} + (2-\gamma)\sum_{s=2}^{T-1} h_{s+1,s} - (1-\gamma)^2 \sum_{t=4}^{T}\sum_{s=2}^{t-2} h_{ts}\gamma^{t-s-2} \right]. \end{aligned} \tag{55}$$

Then, by using (51), (54) and (55), and after some algebra, we obtain $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E\left(\Delta\widetilde{\mathbf{y}}'_{i,-1}\mathbf{p}_i\right) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}(A_{1i}+A_{2i}) = 0$.

To prove (50), first note that $E\left(\Delta\mathbf{W}'_i\mathbf{\Phi}\mathbf{r}_i\right)$ is a $(T+3)$ dimensional vector having all zeros, except for the $(T+2)$th entry, given by $E\left(\Delta\widetilde{\mathbf{y}}'_{i,-1}\mathbf{\Phi}\mathbf{r}_i\right)$. We have

$$\boldsymbol{\vartheta}'\mathbf{r}_i = \sum_{t=1}^{T}(T-t+1)v_{it} = Tv_{i1} + \sum_{t=2}^{T}(T-t+1)\Delta u_{it}, \qquad \boldsymbol{\vartheta}'\Delta\widetilde{\mathbf{y}}_{i,-1} = \sum_{s=1}^{T-1}(T-s)y_{is}.$$

Hence, using results (52)-(53), we have

$$\begin{aligned} \delta_i &= E\left(\boldsymbol{\vartheta}'\mathbf{r}_i\Delta\widetilde{\mathbf{y}}'_{i,-1}\boldsymbol{\vartheta}\right) = T\sum_{s=1}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) + \sum_{s=1}^{T-1}\sum_{t=2}^{T}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \\ &= T\sum_{s=1}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) + \sum_{s=1}^{T-1}\sum_{t=1}^{s+1}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \end{aligned}$$

which can be written as

$$\begin{aligned} \delta_i &= T(T-1)E(\Delta y_{i1}v_{i1}) + \sum_{s=2}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) + \sum_{s=1}^{T-1}\sum_{t=1}^{s-1}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \\ &\quad + \sum_{s=1}^{T-1}(T-s+1)(T-s)E\left(\Delta y_{is}\Delta u_{is}\right) + \sum_{s=1}^{T-1}(T-s)^2E\left(\Delta y_{is}\Delta u_{i,s+1}\right) \end{aligned}$$

Finally, we have

$$\bar{\delta} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\delta_i = T(T-1)\bar{\omega}_0 + (\gamma\bar{\omega}_0 - 1)\sum_{s=2}^{T-1}(T-s)\gamma^{s-2} \tag{56}$$

$$-(1-\gamma)^2\sum_{s=1}^{T-1}\sum_{t=1}^{s-1}(T-t+1)(T-s)\gamma^{s-t-1} + (2-\gamma)\sum_{s=1}^{T-1}(T-s+1)(T-s) - \sum_{s=1}^{T-1}(T-s)^2.$$

■

Lemma A5 *Let* $\mathbf{A}_N^* = -(1/N)\left(\partial^2\ell_p(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right)$, *where* $\ell_p(\boldsymbol{\theta})$ *is given by (12), and* $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_*, \omega_*, \sigma_*^2\right)' = \left(\boldsymbol{\varphi}_0, \bar{\omega}_0, \bar{\sigma}_0^2\right)'$ *is the vector of pseudo-true values. Then as* N *tends to infinity and for fixed* T, *we have*

$$p\lim_{N\to\infty}\mathbf{A}_N^* = \mathbf{A}^*,$$

where $\mathbf{A}^*$ *is a positive definite matrix.*

Proof. The elements of $\mathbf{A}_N^*$ are given by[15]

$$\mathbf{A}_{N,11}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = \frac{1}{\sigma_*^2}\frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i,$$

$$\mathbf{A}_{N,22}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\omega^2} = -\frac{T^2}{2g(\omega_*)^2} + \frac{T}{\sigma_*^2 g(\omega_*)^3}\frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i,$$

$$\mathbf{A}_{N,33}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial(\sigma^2)^2} = -\frac{T}{2(\sigma_*^2)^2} + \frac{1}{(\sigma_*^2)^3}\frac{1}{N}\sum_{i=1i}^{N}\mathbf{r}_i'\boldsymbol{\Omega}(\omega_*)^{-1}\mathbf{r}_i,$$

$$\mathbf{A}_{N,12}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\omega} = \frac{1}{\sigma_*^2 g(\omega_*)^2}\frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Phi}\mathbf{r}_i,$$

$$\mathbf{A}_{N,13}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\sigma^2} = \frac{1}{(\sigma_*^2)^2}\frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}(\omega_*)^{-1}\mathbf{r}_i,$$

$$\mathbf{A}_{N,23}^* = -\frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}_*)}{\partial\omega\partial\sigma^2} = \frac{1}{2(\sigma_*^2)^2 g(\omega_*)^2}\frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i.$$

[15] See also Hsiao, Pesaran, and Tahmiscioglu (2002).

Since $\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i$ and $\mathbf{r}_i'\mathbf{\Omega}(\omega_i)^{-1}\mathbf{r}_i$ are independent across i, with mean $T\sigma_i^2 g(\omega_i)$ and $T\sigma_i^2$, respectively, we have (recall from Assumption 5 that $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \sigma_i^2\omega_i = \bar{\sigma}_0^2\bar{\omega}_0$)

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E\left(\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i\right) &= T\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_i^2\left[1+T(\omega_i-1)\right] \qquad (57)\\ &= T\bar{\sigma}_0^2\left[1+T(\bar{\omega}_0-1)\right] = T\ \sigma_*^2 g(\omega_*)\\ \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\right) &= T\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \sigma_i^2 = T\bar{\sigma}_0^2 = T\sigma_*^2 \qquad (58)\end{aligned}$$

Then, using these and Lemma A4, the matrix $\mathbf{A}^*$ is given by

$$\mathbf{A}^* = \begin{pmatrix} p\lim_{N\to\infty}\frac{1}{N\sigma_*^2}\sum_{i=1}^N \Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i & p\lim_{N\to\infty}\frac{1}{Ng(\omega_*)^2\sigma_*^2}\sum_{i=1}^N \Delta\mathbf{W}_i'\mathbf{\Phi}\mathbf{r}_i & \mathbf{0} \\ p\lim_{N\to\infty}\frac{1}{Ng(\omega_*)^2\sigma_*^2}\sum_{i=1}^N \mathbf{r}_i'\mathbf{\Phi}\Delta\mathbf{W}_i & \frac{T^2}{2g(\omega_*)^2} & \frac{T}{2g(\omega_*)\sigma_*^2} \\ \mathbf{0} & \frac{T}{2g(\omega_*)\sigma_*^2} & \frac{T}{2(\sigma_*^2)^2} \end{pmatrix}. \qquad (59)$$

■

Lemma A6 *Let* $\mathbf{b}_N^* = \left(1/\sqrt{N}\right)\partial\ell_p(\boldsymbol{\theta}_*)/\partial\boldsymbol{\theta}$, *where* $\ell_p(\boldsymbol{\theta})$ *is given by (12), and* $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}_*,\omega_*,\sigma_*^2\right)' = \left(\boldsymbol{\varphi}_0,\bar{\omega}_0,\bar{\sigma}_0^2\right)'$ *is the vector of pseudo-true values. Then as* N *tends to infinity and for fixed* T, *we have*

$$\mathbf{b}_N^* \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\mathbf{B}^*\right). \qquad (60)$$

Proof. Note that $\mathbf{b}_N^*$ can be written as

$$\frac{1}{\sqrt{N}}\frac{\partial\ell_p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_*} = \frac{1}{\sqrt{N}}\begin{pmatrix}\frac{1}{\sigma_*^2}\sum_{i=1}^N \Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i \\ -\frac{NT}{2g(\omega_*)} + \frac{1}{2\sigma_*^2 g(\omega_*)^2}\sum_{i=1}^N \mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i \\ -\frac{NT}{2\sigma_*^2} + \frac{1}{2\sigma_*^4}\sum_{i=1}^N \mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\end{pmatrix} = \frac{1}{\sigma_*^2}\frac{1}{\sqrt{N}}\begin{pmatrix}\sum_{i=1}^N \Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i \\ \frac{1}{2g(\omega_*)^2}\sum_{i=1}^N \xi_i \\ \frac{1}{2\sigma_*^2}\sum_{i=1}^N \zeta_i\end{pmatrix}, \qquad (61)$$

where

$$\xi_i = \mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i - Tg(\omega_*)\sigma_*^2, \qquad \zeta_i = \mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i - T\sigma_*^2.$$

By Lemma A4, $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N E\left(\Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\right)$ has zero mean. Also, from (57) and (58), we have $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E(\xi_i) = 0$ and $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E(\zeta_i) = 0$. For the variance, using Lemma A3, we have

$$\begin{aligned}\mathbf{B}_{11}^* &= \lim_{N\to\infty}\frac{1}{\sigma_*^4}\frac{1}{N}E\left(\sum_{i=1}^N \Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\sum_{i=1}^N \mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right) \qquad (62)\\ &= \lim_{N\to\infty}\frac{1}{\sigma_*^4}\frac{1}{N}\sum_{i=1}^N E\left(\Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right).\end{aligned}$$

Again, using Lemma A3, (57) and (58) and recalling that $\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(\xi_i) = 0$, we have

$$\begin{aligned}
\mathbf{B}_{22}^{*} &= \lim_{N\to\infty} \frac{1}{4g(\omega_*)^4 \sigma_*^4} E\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i^2\right] = \frac{1}{4g(\omega_*)^4 \sigma_*^4}\lim_{N\to\infty} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i - Tg(\omega_*)\sigma_*^2\right)^2\right] \\
&= \lim_{N\to\infty} \frac{1}{4g(\omega_*)^4 \sigma_*^4}\frac{1}{N} E\left[\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i\right)^2 - 2Tg(\omega_*)\sigma_*^2\sum_{i=1}^{N}\left(\mathbf{r}_i'\ \mathbf{\Phi}\mathbf{r}_i\right) + NT^2 g(\omega_*)^2\sigma_*^4\right] \\
&= \frac{T^2}{4g(\omega_*)^4 \sigma_*^4}\lim_{N\to\infty} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i}{T}\right)^2 - g(\omega_*)^2\sigma_*^4\right]. \qquad (63)
\end{aligned}$$

Similarly

$$\begin{aligned}
\mathbf{B}_{33}^{*} &= \lim_{N\to\infty} \frac{1}{4N(\sigma_*^2)^4} E\left[\sum_{i=1}^{N}\zeta_i^2\right] = \lim_{N\to\infty}\frac{1}{4N(\sigma_*^2)^4}\left\{E\left[\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\right)^2\right] - NT^2\sigma_*^4\right\} \\
&= \frac{T^2}{4\sigma_*^8}\lim_{N\to\infty} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i}{T}\right)^2 - \sigma_*^4\right]. \qquad (64)
\end{aligned}$$

The off-diagonal elements of $\mathbf{B}^*$ are (using Lemma A3 and noting that $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left(\Delta\mathbf{W}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\right) = 0$ and $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E(\xi_i) = 0$):

$$\begin{aligned}
\mathbf{B}_{21}^{*} &= \lim_{N\to\infty} \frac{1}{2\sigma_*^4 g(\omega_*)^2} E\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right] \\
&= \lim_{N\to\infty} \frac{1}{2\sigma_*^4 g(\omega_*)^2} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right)\left(\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i - Tg(\omega_*)\sigma_*^2\right)\right] \\
&= \lim_{N\to\infty} \frac{1}{2\sigma_*^4 g(\omega_*)^2} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right)\left(\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i\right)\right], \qquad (65) \\
\mathbf{B}_{31}^{*} &= \lim_{N\to\infty} \frac{1}{2\sigma_*^6} E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\Delta\mathbf{W}_i\right)\left(\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i\right)\right]. \qquad (66)
\end{aligned}$$

Similarly, using Lemma A3, (57) and (58), we have

$$
\begin{aligned}
\mathbf{B}_{32}^{*} &= \lim_{N\to\infty} \frac{1}{4\sigma_*^6 g(\omega_*)^2} E\left(\frac{1}{N}\sum_{i=1}^{N} \xi_i \zeta_i\right) \\
&= \lim_{N\to\infty} \frac{T^2}{4\sigma_*^6 g(\omega)^2} E\left[\frac{1}{N}\sum_{i=1}^{N} \left(\frac{\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i}{T} - g(\omega_*)\sigma_*^2\right)\left(\frac{\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i}{T} - \sigma_*^2\right)\right] \\
&= \lim_{N\to\infty} \frac{T^2}{4\sigma_*^6 g(\omega_*)^2} E\left[\frac{1}{N}\sum_{i=1}^{N} \left(\frac{\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i}{T}\frac{\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i}{T} - g(\omega_*)\sigma_*^2 \frac{\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i}{T} - \sigma_*^2 \frac{\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i}{T} + g(\omega_*)\sigma_*^4\right)\right] \\
&= \lim_{N\to\infty} \frac{T^2}{4\sigma_*^6 g(\omega_*)^2} E\left[\frac{1}{N}\sum_{i=1}^{N} \left(\frac{\mathbf{r}_i'\mathbf{\Phi}\mathbf{r}_i}{T}\frac{\mathbf{r}_i'\mathbf{\Omega}(\omega_*)^{-1}\mathbf{r}_i}{T} - g(\omega_*)\sigma_*^4\right)\right]. \qquad (67)
\end{aligned}
$$

For fixed T, the elements inside the sum operator in expressions (62)-(67) are finite for all i. Hence, (60) is established by applying the central limit theorem for independent and heterogeneous random variables (White, 2001). ■

A.2 Proof of Theorem 1

First note that equation (6) can be rewritten as

$$\eta_{i1} = E(\eta_{i1}|\Delta\mathbf{x}_i) + [\eta_{i1} - E(\eta_{i1}|\Delta\mathbf{x}_i)] = E(\eta_{i1}|\Delta\mathbf{x}_i) + \varsigma_{i1}, \qquad (68)$$

where $\varsigma_{i1} = \eta_{i1} - E(\eta_{i1}|\Delta\mathbf{x}_i)$. Also, we have

$$E(\eta_{i1}|\Delta\mathbf{x}_i) = \gamma^m E(\Delta y_{i,-m+1}|\Delta\mathbf{x}_i) + \beta\Delta x_{i1} + \beta\sum_{j=1}^{m-1}\gamma^j E(\Delta x_{i,1-j}|\Delta\mathbf{x}_i). \qquad (69)$$

Using either (3) or (4) we have

$$\Delta x_{it} = \phi + \sum_{j=0}^{\infty} \widetilde{d}_j \varepsilon_{i,t-j}, \qquad (70)$$

with $\widetilde{d}_j = d_j$ under (4), $\widetilde{d}_j = a_j - a_{j-1}$ under (3), and $\widetilde{d}_0 = a_0$. Hence, it is easily seen that under (70)

$$E(\Delta x_{i,1-j}|\Delta\mathbf{x}_i) = b_j + \boldsymbol{\pi}_j'\Delta\mathbf{x}_i, \quad (j = 1, ..., m-1) \qquad (71)$$

where b_j and $\boldsymbol{\pi}_j$ do not depend on i. Using Assumption 3 and (71) in (69), we have

$$\begin{aligned} E(\eta_{i1}|\Delta\mathbf{x}_i) &= \gamma^m\left(b_m+\boldsymbol{\pi}_m'\Delta\mathbf{x}_i\right)+\beta\Delta x_{i1}+\beta\sum_{j=1}^{m-1}\gamma^j\left(b_j+\boldsymbol{\pi}_j'\Delta\mathbf{x}_i\right) \\ &= \left(\gamma^m b_m+\beta\sum_{j=1}^{m-1}\gamma^j b_j\right)+\left(\boldsymbol{\pi}_m+\beta\mathbf{e}_1+\beta\sum_{j=1}^{m-1}\gamma^j\boldsymbol{\pi}_j\right)'\Delta\mathbf{x}_i \\ &= b+\boldsymbol{\pi}'\Delta\mathbf{x}_i \end{aligned} \tag{72}$$

where $\mathbf{e}_1=(1,0,...,0)'$, b is a constant, and $\boldsymbol{\pi}$ is a T-dimensional vector of parameters. Then, using (5), (68) and (72), Δy_{i1} can be written as

$$\begin{aligned} \Delta y_{i1} &= \eta_{i1}+\sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j}=E(\eta_i|\Delta\mathbf{x}_i)+\varsigma_{i1}+\sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j} \\ &= b+\boldsymbol{\pi}'\Delta\mathbf{x}_i+v_{i1}, \end{aligned} \tag{73}$$

where $v_{i1}=\varsigma_{i1}+\sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j}$, In the above equation, v_{i1} has zero mean and variance $E\left(v_{i1}^2\right)=\omega_i\sigma_i^2$. ■

A.3 Proof of Theorem 2

To simplify the derivation and better understand the model, we consider an alternative expression of the model proposed by Grassetti (2011). By pre-multiplying (9) by the $T\times T$ accumulation matrix,

$$\mathbf{L}_T=\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

to obtain

$$\begin{pmatrix} y_{i1}-y_{i0} \\ y_{i2}-y_{i0} \\ \vdots \\ y_{iT}-y_{i0} \end{pmatrix}=\begin{pmatrix} 1 & \Delta\mathbf{x}_i' & 0 & 0 \\ 1 & \Delta\mathbf{x}_i' & y_{i1}-y_{i0} & x_{i2}-x_{i1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta\mathbf{x}_i' & y_{i,T-1}-y_{i0} & x_{iT}-x_{i1} \end{pmatrix}\boldsymbol{\varphi}+\begin{pmatrix} \xi_i+u_{i1} \\ \xi_i+u_{i2} \\ \vdots \\ \xi_i+u_{iT} \end{pmatrix},$$

which can be written more compactly as

$$\dot{\mathbf{y}}_i=\dot{\mathbf{W}}_i\boldsymbol{\varphi}+\dot{\mathbf{r}}_i, \tag{74}$$

where

$$v_{i1} = u_{i1} + (v_{i1} - u_{i1}) = u_{i1} + \xi_i, \quad \text{and } \dot{\mathbf{r}}_i = \boldsymbol{\iota}_T \xi_i + \mathbf{u}_i. \tag{75}$$

Since $\mathbf{L}_T$ does not depend on any parameters, then the likelihood functions for (9) and (74) are identical, also noting that the Jacobian of the transformation , given by $|\mathbf{L}_T| = 1$. Hence, the ML estimators based on the transformed ML estimator for (9) and (74) will be identical.

The tth row of (74) can be written as

$$(y_{it} - y_{i0}) = b + \Delta \mathbf{x}_i' \boldsymbol{\pi} + (y_{i,t-1} - y_{i0})\gamma + (x_{it} - x_{i1})\beta + \xi_i + u_{it.} \tag{76}$$

Also, from the definition of ξ_i,

$$\xi_i = v_{i1} - u_{i1} = \left(\varsigma_{i1} + \sum_{j=0}^{m-1} \gamma^j \Delta u_{i,1-j} \right) - u_{i1} = \varsigma_{i1} - (1-\gamma) u_{i0} - (1-\gamma) \sum_{j=1}^{m-2} \gamma^j u_{i,-j} - \gamma^{m-1} u_{i,-m+1}$$

where $\varsigma_{i1} = \eta_{i1} - E\left(\eta_{i1} | \Delta \mathbf{x}_i\right)$, $\eta_{i1} = E\left(\Delta y_{i1} | \Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...\right)$. Note that $var(\xi_i) = \sigma^2_{\xi i} = \sigma_i^2 \left(\omega_i - 1\right)$. Using Assumption 5, we have $\bar{\sigma}^2_{\xi 0} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \sigma^2_{\xi i0} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \left(\omega_i - 1\right) \sigma_i^2 = \bar{\sigma}_0^2 (\bar{\omega}_0 - 1)$. Although (74) looks like the standard random effect model, it is not the case since the regressor $(y_{i,t-1} - y_{i0})$ and new individual effects ξ_i are correlated.

For some σ^2 and $\sigma_\xi^2 = \sigma^2 \left(\omega - 1\right)$, define

$$\begin{aligned} \mathbf{V}_T &= E\left(\dot{\mathbf{r}}_i \dot{\mathbf{r}}_i'\right) = \sigma^2 \mathbf{I}_T + \sigma_\xi^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T', \ \mathbf{V}_T^{-1} = \frac{1}{\sigma^2} \left[\mathbf{I}_T - (1-\psi) \frac{1}{T} \boldsymbol{\iota}_T \boldsymbol{\iota}_T' \right], \\ \mathbf{Q}_T &= \mathbf{I}_T - \frac{1}{T} \boldsymbol{\iota}_T \boldsymbol{\iota}_T', \quad \psi = \frac{\sigma^2}{\sigma^2 + T\sigma_\xi^2} = \frac{1}{1 + T\left(\omega - 1\right)}, \qquad 1 - \psi = \frac{T\left(\omega - 1\right)}{1 + T\left(\omega - 1\right)}. \end{aligned} \tag{77}$$

Then, by using $|\mathbf{V}_T| = \sigma^{2(T-1)}(\sigma^2 + T\sigma_\xi^2) = \sigma^{2T}\left[1 + T\left(\omega - 1\right)\right]$, the alternative expression for the pseudo log-likelihood function under homoskedasticity can be written as

$$\begin{aligned} \ell_{RE}\left(\boldsymbol{\theta}\right) &= -\frac{NT}{2} \ln\left(2\pi\right) - \frac{N}{2} \ln|\mathbf{V}_T| - \frac{1}{2} \sum_{i=1}^N \dot{\mathbf{r}}_i' \mathbf{V}_T^{-1} \dot{\mathbf{r}}_i \\ &\propto -\frac{NT}{2} \ln \sigma^2 - \frac{N}{2} \ln\left[1 + T\left(\omega - 1\right)\right] - \frac{1}{2\sigma^2} \sum_{i=1}^N \left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i \boldsymbol{\varphi}\right)' \mathbf{Q}_T \left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i \boldsymbol{\varphi}\right) \\ &\quad - \frac{1}{2\sigma^2 T\left[1 + T\left(\omega - 1\right)\right]} \sum_{i=1}^N \left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i \boldsymbol{\varphi}\right)' \boldsymbol{\iota}_T \boldsymbol{\iota}_T' \left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i \boldsymbol{\varphi}\right) \end{aligned}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\varphi}', \omega, \sigma^2\right)'$. Under heteroskedastic errors, the pseudo-true value of $\boldsymbol{\theta}$ denoted by $\boldsymbol{\theta}_* =$

$(\boldsymbol{\varphi}_*', \omega_*, \sigma_*^2)'$, is the solution of $\lim_{N\rightarrow\infty} N^{-1}E\left[\partial\ell_{RE}\left(\boldsymbol{\theta}_*\right)/\partial\boldsymbol{\theta}\right] = \mathbf{0}$, and can be written as

$$\begin{aligned}
\boldsymbol{\varphi}_* &= \left[\lim_{N\rightarrow\infty} N^{-1}\sum_{i=1}^{N} E\left(\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{W}}_i\right)\right]^{-1}\lim_{N\rightarrow\infty} N^{-1}\sum_{i=1}^{N} E\left(\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{y}}_i\right), \quad (78)\\
1+T\left(\omega_*-1\right) &= \frac{1}{\sigma_*^2}\lim_{N\rightarrow\infty}\frac{1}{NT}\sum_{i=1}^{N} E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right],\\
\sigma_*^2 &= \lim_{N\rightarrow\infty}\frac{1}{N(T-1)}\sum_{i=1}^{N} E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\mathbf{Q}_T\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right], \quad (79)
\end{aligned}$$

where $\mathbf{V}_{T*} = \sigma_*^2\mathbf{I}_T + \sigma_*^2\left(\omega_*-1\right)\boldsymbol{\iota}_T\boldsymbol{\iota}_T'$. Substituting σ_*^2 into the expression of $\sigma_{\xi*}^2$, we have

$$1+T\left(\omega_*-1\right) = \frac{\lim_{N\rightarrow\infty}\frac{1}{NT}\sum_{i=1}^{N} E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right]}{\lim_{N\rightarrow\infty}\frac{1}{N(T-1)}\sum_{i=1}^{N} E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\mathbf{Q}_T\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right]}. \quad (80)$$

The expectations in the above first order conditions are taken with respect to the true heteroskedatic model. To derive these expectations we first note that

$$\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_* = \dot{\mathbf{r}}_i-\dot{\mathbf{W}}_i\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right),$$

and obatin

$$\begin{aligned}
\frac{1}{T}E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right] &= \sigma_i^2\left[1+T(\omega_i-1)\right]-2\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'E\left(T^{-1}\dot{\mathbf{W}}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\dot{\mathbf{r}}_i\right)\\
&\quad+\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'E\left(T^{-1}\dot{\mathbf{W}}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\dot{\mathbf{W}}_i\right)\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{T-1}E\left[\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\mathbf{Q}_T\left(\dot{\mathbf{y}}_i-\dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right] &= \sigma_i^2-2\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'E\left(\frac{1}{T-1}\dot{\mathbf{W}}_i'\mathbf{Q}_T\dot{\mathbf{r}}_i\right)\\
&\quad+\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'E\left(\frac{1}{T-1}\dot{\mathbf{W}}_i'\mathbf{Q}_T\dot{\mathbf{W}}_i\right)\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right).
\end{aligned}$$

Using the above results in (80) we obtain

$$\begin{aligned}
\omega_*-\bar{\omega}_0 &= -\frac{1+T\left(\omega_*-1\right)}{\bar{\sigma}_0^2\left(T-1\right)}\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'\lim_{N\rightarrow\infty} N^{-1}\sum_{i=1}^{N} E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T-h_a\,\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{W}}_i\right]\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)\\
&\quad+\frac{2\left[1+T\left(\omega_*-1\right)\right]}{\bar{\sigma}_0^2\left(T-1\right)}\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'\lim_{N\rightarrow\infty} N^{-1}\sum_{i=1}^{N} E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T-h_a\,\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{r}}_i\right] \quad (81)
\end{aligned}$$

where $h_a = \omega_* / [1 + T(\omega_* - 1)]$. Similarly, using the first order condition (78) we also have

$$\begin{aligned} \left[\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E\left(T^{-1}\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{W}}_i\right)\right](\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0) &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left(T^{-1}\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{r}}_i\right) \\ &= \frac{1}{\sigma_*^2}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T - h_b\,\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{r}}_i\right] \end{aligned} \tag{82}$$

where $h_b = (\omega_* - 1)/[1 + T(\omega_* - 1)]$. Since the regressors are assumed to be exogenous then (recall also that $\dot{\mathbf{r}}_i = \boldsymbol{\iota}_T\xi_i + \mathbf{u}_i$)

$$E\left[\dot{\mathbf{W}}_i'\left(\mathbf{I}_T - \frac{\omega_* - 1}{1 + T(\omega_* - 1)}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{r}}_i\right] = \mathbf{e}_3\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} E\left[\dot{\mathbf{y}}_{i,-1}'\left(\mathbf{I}_T - h_b\,\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)(\boldsymbol{\iota}_T\xi_i + \mathbf{u}_i)\right], \tag{83}$$

where $\dot{\mathbf{y}}_{i,-1} = (0, y_{i1} - y_{i0}, ..., y_{i,T-1} - y_{i0})'$, and $\mathbf{e}_3 = (0, \mathbf{0}_{T\times 1}', 1, 0)'$.

To evaluate the expectations in the above formulas, we first derive some preliminary results. From the model (76), for $t = 2, ..., T$, we have

$$\begin{aligned} \dot{y}_{it} &= y_{it} - y_{i0} = \left(1 + \gamma_0 + ... + \gamma_0^{t-2}\right)\left(b_0 + \boldsymbol{\pi}_0'\Delta\mathbf{x}_i\right) + \gamma_0^{t-1}(y_{i1} - y_{i0}) \\ &\quad + \beta_0\left(\sum_{j=0}^{t-2}\gamma_0^j(x_{i,t-j} - x_{i1})\right) + \left(1 + \gamma_0 + ... + \gamma_0^{t-2}\right)\xi_i + \sum_{j=0}^{t-2}\gamma_0^j u_{i,t-j} \\ &= \left(\frac{1 - \gamma_0^t}{1 - \gamma_0}\right)\left(b_0 + \boldsymbol{\pi}_0'\Delta\mathbf{x}_i\right) + \beta_0\left(\sum_{j=0}^{t-2}\gamma_0^j(x_{i,t-j} - x_{i1})\right) + \left(\frac{1 - \gamma_0^t}{1 - \gamma_0}\right)\xi_i + \sum_{j=0}^{t-1}\gamma_0^j u_{i,t-j}. \end{aligned}$$

Then, for $s = 1, ..., T$

$$E[(y_{it} - y_{i0})u_{is}] = E\left[\left(u_{it} + \gamma_0 u_{i,t-1} + ... + \gamma_0^{t-1}u_{i1}\right)u_{is}\right] = \begin{cases} \sigma_{i0}^2\gamma_0^{t-s} & t \geq s \\ 0 & t < s \end{cases}. \tag{84}$$

Also, we have

$$E[\xi_i(y_{it} - y_{i0})] = \left(\frac{1 - \gamma_0^t}{1 - \gamma_0}\right)\sigma_{\xi i0}^2 = \left(\frac{1 - \gamma_0^t}{1 - \gamma_0}\right)\sigma_{i0}^2(\omega_{i0} - 1) \tag{85}$$

Then, using (84) and (85), we have

$$E(\dot{\mathbf{y}}_{i,-1}'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\mathbf{u}_i) = \boldsymbol{\iota}_T' E\left(\mathbf{u}_i\dot{\mathbf{y}}_{i,-1}'\right)\boldsymbol{\iota}_T = \left(\frac{\sigma_{i0}^2}{1 - \gamma_0}\right)\left[T - \frac{1 - \gamma_0^T}{1 - \gamma_0}\right] = T\phi_0\sigma_{i0}^2, \tag{86}$$

$$E\left(\xi_i\dot{\mathbf{y}}_{i,-1}'\boldsymbol{\iota}_T\right) = \sum_{t=1}^{T-1} E[\xi_i(y_{it} - y_{i0})] = \left(\frac{\sigma_{\xi i0}^2}{1 - \gamma_0}\right)\sum_{t=1}^{T-1}(1 - \gamma_0^t) = T\phi_0\sigma_{i0}^2(\omega_{i0} - 1) \tag{87}$$

where

$$\phi_0 = \frac{1}{1 - \gamma_0}\left(1 - \frac{1}{T}\frac{1 - \gamma_0^T}{1 - \gamma_0}\right). \tag{88}$$

Using the above results it now readily follows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\dot{\mathbf{y}}_{i,-1}'\left(\mathbf{I}_T-h\ \boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\left(\boldsymbol{\iota}_T\xi_i+\mathbf{u}_i\right)\right]=T\phi_0\bar{\sigma}_0^2\left[\left(1-hT\right)\left(\bar{\omega}_0-1\right)-h\right]. \quad (89)$$

Using this result with $h=h_b=(\omega_*-1)/\left[1+T\left(\omega_*-1\right)\right]$ in (83) and then in (82) yields

$$\begin{aligned}\left[\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(T^{-1}\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{W}}_i\right)\right](\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0) &= \mathbf{e}_3\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(T^{-1}\dot{\mathbf{y}}_{i,-1}'\mathbf{V}_{T*}^{-1}\dot{\mathbf{r}}_i\right)\\ &= \mathbf{e}_3\frac{-\phi_0\bar{\sigma}_0^2}{\sigma_*^2}\frac{(\omega_*-\bar{\omega}_0)}{1+T\left(\omega_*-1\right)}. \quad (90)\end{aligned}$$

Similarly,

$$\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T-h_a\ \boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{r}}_i\right]=-\mathbf{e}_3\frac{\phi_0\bar{\sigma}_0^2\left[\left(T-1\right)\left(\bar{\omega}_0-1\right)+\omega_*\right]}{1+T\left(\omega_*-1\right)},$$

and hence using (81) we have

$$\begin{aligned}\omega_*-\bar{\omega}_0 &= -\frac{1+T\left(\omega_*-1\right)}{\bar{\sigma}_0^2\left(T-1\right)}\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T-h_a\ \boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{W}}_i\right]\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)\\ &\quad -\frac{2\phi_0\left[\left(T-1\right)\left(\bar{\omega}_0-1\right)+\omega_*\right]}{\left(T-1\right)}\left(\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0\right)'\mathbf{e}_3 \quad (91)\end{aligned}$$

Furthermore, we note that the following limits exist

$$\mathbf{A}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(T^{-1}\dot{\mathbf{W}}_i'\dot{\mathbf{W}}_i\right),\ \mathbf{B}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(T^{-1}\dot{\mathbf{W}}_i'\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\dot{\mathbf{W}}_i\right), \quad (92)$$

and

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(T^{-1}\dot{\mathbf{W}}_i'\mathbf{V}_{T*}^{-1}\dot{\mathbf{W}}_i\right) &= \frac{1}{\sigma_*^2}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[T^{-1}\dot{\mathbf{W}}_i'\left(\mathbf{I}_T-h_b\,\boldsymbol{\iota}_T\boldsymbol{\iota}_T'\right)\dot{\mathbf{W}}_i\right]\\ &= \frac{1}{\sigma_*^2}\left[\mathbf{A}-\frac{\omega_*-1}{1+T\left(\omega_*-1\right)}\mathbf{B}\right],\end{aligned}$$

which is a positive definite matrix. Using these result in (90) and (91) we have

$$\boldsymbol{\varphi}_*-\boldsymbol{\varphi}_0=-\phi_0\bar{\sigma}_0^2\frac{(\omega_*-\bar{\omega}_0)}{1+T\left(\omega_*-1\right)}\left[\mathbf{A}-\frac{\omega_*-1}{1+T\left(\omega_*-1\right)}\mathbf{B}\right]^{-1}\mathbf{e}_3, \quad (93)$$

and

$$(\omega_* - \bar{\omega}_0) = -\frac{1+T(\omega_*-1)}{(T-1)\bar{\sigma}_0^2}(\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0)'\left[\mathbf{A}-\frac{\omega_*}{1+T(\omega_*-1)}\mathbf{B}\right](\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0) - \frac{2\phi_0[(T-1)(\bar{\omega}_0-1)+\omega_*]}{(T-1)}(\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0)'\mathbf{e}_3. \tag{94}$$

Substituting $\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0$ from (93) in the above and after some algebra we have

$$\left\{1 - \frac{2\kappa_2\phi_0^2\bar{\sigma}_0^2[(T-1)(\bar{\omega}_0-1)+\omega_*]}{(T-1)[1+T(\omega_*-1)]}\right\}(\omega_* - \bar{\omega}_0) = \frac{-\kappa_1\phi_0^2\bar{\sigma}_0^2}{(T-1)[1+T(\omega_*-1)]}(\omega_* - \bar{\omega}_0)^2. \tag{95}$$

where

$$\kappa_1 = \mathbf{e}_3'\left[\mathbf{A}-\frac{\omega_*-1}{1+T(\omega_*-1)}\mathbf{B}\right]^{-1}\left[\mathbf{A}-\frac{\omega_*}{1+T(\omega_*-1)}\mathbf{B}\right]\left[\mathbf{A}-\frac{\omega_*-1}{1+T(\omega_*-1)}\mathbf{B}\right]^{-1}\mathbf{e}_3,$$

$$\kappa_2 = \mathbf{e}_3'\left[\mathbf{A}-\frac{\omega_*-1}{1+T(\omega_*-1)}\mathbf{B}\right]^{-1}\mathbf{e}_3.$$

Also, using (79)

$$\sigma_*^2 = \lim_{N\to\infty}\frac{1}{N(T-1)}\sum_{i=1}^{N}E\left[\left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)'\mathbf{Q}_T\left(\dot{\mathbf{y}}_i - \dot{\mathbf{W}}_i\boldsymbol{\varphi}_*\right)\right]$$
$$= \bar{\sigma}_0^2 + \frac{2\phi_0\bar{\sigma}_0^2}{T-1}(\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0)'\mathbf{e}_3 + \left(\frac{T}{T-1}\right)(\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0)'\mathbf{C}(\boldsymbol{\varphi}_* - \boldsymbol{\varphi}_0) \tag{96}$$

where $\mathbf{C} = \mathbf{A} - T^{-1}\mathbf{B}$.

It is clear that for a finite T all the terms in (95) are finite and as required $\omega_* = \bar{\omega}_0$ is a solution of the first order equations. Using this result in (95) and (96) it also follows that $\boldsymbol{\varphi}_* = \boldsymbol{\varphi}_0$ and $\sigma_*^2 = \bar{\sigma}_0^2$. However, for a finite T this solution is not unique and (95) has another solution given implicitly by

$$\omega_* = \bar{\omega}_0 - \frac{(T-1)[1+T(\omega_*-1)] - 2\phi_0^2\bar{\sigma}_0^2[(T-1)(\bar{\omega}_0-1)+\omega_*]\kappa_2}{\phi_0^2\bar{\sigma}_0^2\kappa_1}. \tag{97}$$

Under this solution $\boldsymbol{\varphi}_* \neq \boldsymbol{\varphi}_0$.

A.4 Proof of Theorem 3

First, take a Taylor series expansion of $\left(1/\sqrt{N}\right)\partial\ell_p(\widehat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}$ around $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_*$, yielding

$$\mathbf{0} = \frac{1}{\sqrt{N}}\frac{\partial\ell_p\left(\widehat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\theta}} = \frac{1}{\sqrt{N}}\frac{\partial\ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\theta}} + \frac{1}{N}\frac{\partial^2\ell_p(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right) + \boldsymbol{\delta}_N,$$

where $\boldsymbol{\delta}_N$ is an approximation error which, given the consistency of $\widehat{\boldsymbol{\theta}}$, goes to zero as N tend to infinity. Rearranging, we have

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*\right)=\left[-\frac{1}{N}\frac{\partial^2 \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]^{-1}\frac{1}{\sqrt{N}}\frac{\partial \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}}+o_p\left(1\right).$$

As $N\to\infty$ and for fixed T, we have

$$\mathbf{A}_N^*=-\frac{1}{N}\frac{\partial^2 \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\overset{p}{\to}\mathbf{A}^*,$$

where, by Lemma A5, $\mathbf{A}^*$ is a symmetric and positive definite matrix (see expression (59)). Then by the Slutsky's theorem

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*\right)=\mathbf{A}^*\frac{1}{\sqrt{N}}\frac{\partial \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}}+o_p\left(1\right).$$

Further, by Lemma A6, as $N\to\infty$ and for a fixed T we have

$$\mathbf{b}_N^*=\frac{1}{\sqrt{N}}\frac{\partial \ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}}\overset{d}{\to}\mathcal{N}\left(\mathbf{0},\mathbf{B}^*\right),$$

where the elements of $\mathbf{B}^*$ are given in expressions (62)-(67). Hence, result (15) follows, and $\widehat{\boldsymbol{\theta}}$ is asymptotically normally distributed for a fixed T, and as N tends to infinity.■

References

Ahn, S. C., and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68(1), 5–27.

Alvarez, J., and M. Arellano (2004): "Robust Likelihood Estimation of Dynamic Panel Data Models," mimeo.

Anderson, T. W., and C. Hsiao (1981): "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76(375), 598–606.

——— (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18(1), 47–82.

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20(1), 46–63.

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(2), 277–297.

Arellano, M., and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68(1), 29–51.

Bhargava, A., and J. D. Sargan (1983): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51(6), 1635–1659.

Binder, M., C. Hsiao, and M. H. Pesaran (2005): "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," *Econometric Theory*, 21(4), 795–837.

Blundell, R., and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87(1), 115–143.

Blundell, R., S. Bond, and F. Windmeijer (2000): "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator," in *Nonstationary Panels, Panel Cointegration and Dynamic Panels*, ed. by B. H. Baltagi, vol. 15 of *Advances in Econometrics*, pp. 53–91. JAI Press, Amsterdam.

Bond, S., and F. Windmeijer (2005): "Reliable Inference For GMM Estimators? Finite Sample Properties of Alternative Test Procedures in Linear Panel Data Models," *Econometric Reviews*, 24(1), 1–37.

Breusch, T. S. (1987): "Maximum Likelihood Estimation of Random Effects Models," *Journal of Econometrics*, 36(3), 383–389.

Bun, M. J. G., and J. F. Kiviet (2006): "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," *Journal of Econometrics*, 132(2), 409–444.

Bun, M. J. G., and F. Windmeijer (2010): "The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models," *Econometrics Journal*, 13(1), 95–126.

Grassetti, L. (2011): "A Note on Transformed Likelihood Approach in Linear Dynamic Panel Models," *Statistical Methods & Applications*, 20(2), 221–240.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press.

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14(3), 262–80.

Hayakawa, K. (2007): "Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models," *Economics Letters*, 95(1), 32–38.

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–1395.

Hsiao, C., M. H. Pesaran, and K. A. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109(1), 107–150.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66(2), 333–357.

Keane, M. P., and D. E. Runkle (1992): "On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business and Economic Statistics*, 10(1), 1–9.

Kent, J. T. (1982): "Robust Properties of Likelihood Ratio Tests," *Biometrika*, 69(1), 19–27.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica: Journal of the Econometric Society*, 65(4), 861–874.

Kiviet, J. F. (2007): "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models," in *The Refinement of Econometric Estimation and Test Procedures*, ed. by G. D. A. Phillips, and E. Tzavalis, pp. 282–318. Cambridge University Press.

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73(4), 1103–1123.

Kruiniger, H. (2008): "Maximum Likelihood Estimation and Inference Methods for the Covariance Stationary Panel AR(1)/Unit Root Model," *Journal of Econometrics*, 144(2), 447–464.

Maddala, G. S. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39(2), 341–358.

Moreira, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

Newey, W. K., and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72(1), 219–255.

Newey, W. K., and F. Windmeijer (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77(3), 687–719.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22(1), 300–325.

STOCK, J. H., AND J. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68(5), 1055–1096.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.

——— (2001): *Asymptotic Theory for Econometricians*. Academic Press.

WINDMEIJER, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126(1), 25–51.

Table 1: Median bias($\times$100) and MAE($\times$100) of γ and β ($\gamma = 0.4, \beta = 0.5$) for ARX(1) model

$\gamma = 0.4$

$\gamma = 0.4$	median bias($\times$100)			MAE($\times$100)			median bias($\times$100)			MAE($\times$100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	-0.253	-0.081	-0.078	7.477	3.294	2.655	-0.253	-0.081	-0.078	7.477	3.294	2.655
150	-0.184	0.063	-0.039	3.830	2.153	1.631	-0.184	0.063	-0.039	3.830	2.153	1.631
500	0.042	-0.054	-0.107	2.073	1.192	0.827	0.042	-0.054	-0.107	2.073	1.192	0.827
One-step first-difference GMM estimator based on "DIF2"												
50	-10.112	-5.709	–	11.943	6.308	–	-19.590	-14.454	–	20.144	14.546	–
150	-4.124	-1.949	-1.962	6.335	3.149	2.733	-9.768	-6.161	-5.505	11.082	6.491	5.669
500	-1.107	-0.648	-0.572	3.260	1.642	1.189	-3.204	-2.047	-1.782	5.366	3.048	2.283
One-step system GMM estimator based on "SYS2"												
50	7.035	–	–	9.044	–	–	46.392	–	–	46.392	–	–
150	3.955	2.767	2.331	5.525	3.425	2.810	35.877	35.838	34.855	35.877	35.838	34.855
500	2.369	0.941	0.673	3.042	1.657	1.219	23.131	20.218	18.993	23.131	20.218	18.993

$\beta = 0.5$

$\beta = 0.5$	median bias($\times$100)			MAE($\times$100)			median bias($\times$100)			MAE($\times$100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	0.040	0.144	0.028	5.340	3.025	2.389	0.040	0.144	0.028	5.340	3.025	2.389
150	-0.024	-0.115	0.219	3.073	1.988	1.454	-0.024	-0.115	0.219	3.073	1.988	1.454
500	-0.056	-0.066	0.040	1.488	0.951	0.760	-0.056	-0.066	0.040	1.488	0.951	0.760
One-step first-difference GMM estimator based on "DIF2"												
50	-0.771	-0.237	–	5.609	3.860	–	-1.007	-0.872	–	5.488	3.802	–
150	-0.276	-0.072	-0.009	3.644	2.369	1.895	-0.449	-0.384	-0.308	3.577	2.418	1.901
500	-0.120	0.009	-0.014	1.818	1.162	0.941	-0.181	-0.120	-0.131	1.806	1.150	0.948
One-step system GMM estimator based on "SYS2"												
50	2.031	–	–	6.408	–	–	4.064	–	–	7.800	–	–
150	1.011	0.986	1.056	3.775	2.667	2.013	3.751	3.780	4.187	5.286	4.359	4.253
500	0.523	0.350	0.379	1.954	1.271	1.105	2.549	2.368	2.277	3.330	2.472	2.327

Note: "DIF2" denotes Arellano and Bond type moment conditions: $E(y_{i,t-2-l}\Delta u_{it}) = 0$ with $l = 0$ for $t = 2$, $l = 0, 1$ for $t = 3, ..., T$ and $E(x_{i,t-l}\Delta u_{it}) = 0$ with $l = 0, 1$ for $t = 2$, $l = 0, 1, 2$ for $t = 3, ..., T$. One-step first-difference GMM estimator is computed by (27) with a suitable modification of $\dot{\mathbf{Z}}_i$. "SYS2" denotes Blundell and Bond type moment conditions: $E[\Delta y_{i,t-1}(\alpha_i + u_{it})] = 0$ and $E[\Delta x_{it}(\alpha_i + u_{it})] = 0$ for $t = 2, ..., T$ in addition to the ones used in "DIF2". One-step system GMM estimator is computed by (33) with a suitable modification of $\ddot{\mathbf{Z}}_i$. The numbers of moment conditions of "DIF2" and "SYS2" are 18 and 26 when $T = 5$, 43 and 61 when $T = 10$ and 68 and 96 when $T = 15$. "$-$" denotes the cases where the GMM estimators are not computed since the number of moment conditions exceeds the sample size.

Table 2: Median bias($\times 100$) and MAE($\times 100$) of γ and β ($\gamma = 0.9, \beta = 0.5$) for ARX(1) model

$\gamma = 0.9$

$\gamma = 0.9$	median bias($\times 100$)			MAE($\times 100$)			median bias($\times 100$)			MAE($\times 100$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	-0.061	-0.128	0.108	7.251	3.003	1.646	-0.076	-0.126	0.084	7.117	2.970	1.637
150	-0.030	0.017	-0.092	4.284	1.728	1.027	-0.017	0.037	-0.090	4.237	1.732	1.024
500	0.115	-0.009	-0.020	2.091	0.842	0.520	0.091	-0.023	-0.022	2.091	0.846	0.515
One-step first-difference GMM estimator based on "DIF2"												
50	-10.001	-6.630	–	10.747	6.855	–	-11.931	-7.690	–	12.431	7.816	–
150	-3.894	-2.761	-2.402	5.916	3.267	2.605	-4.955	-3.239	-2.657	6.481	3.696	2.774
500	-1.536	-0.767	-0.714	2.909	1.462	1.075	-1.831	-0.890	-0.867	3.007	1.678	1.188
One-step system GMM estimator based on "SYS2"												
50	5.682	–	–	5.686	–	–	9.155	–	–	9.155	–	–
150	5.343	4.594	4.197	5.343	4.594	4.197	8.992	8.827	8.765	8.992	8.827	8.765
500	4.625	3.349	2.875	4.625	3.349	2.875	8.717	8.260	8.122	8.717	8.260	8.122

$\beta = 0.5$

$\beta = 0.5$	median bias($\times 100$)			MAE($\times 100$)			median bias($\times 100$)			MAE($\times 100$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	0.157	0.229	0.118	5.203	3.085	2.368	0.101	0.246	0.107	5.210	3.085	2.353
150	0.043	-0.104	0.177	3.122	1.985	1.482	0.039	-0.108	0.181	3.125	1.984	1.481
500	-0.007	-0.020	0.042	1.520	0.912	0.771	-0.002	-0.019	0.041	1.506	0.910	0.771
One-step first-difference GMM estimator based on "DIF2"												
50	-2.939	-2.221	–	6.203	4.424	–	-3.399	-2.628	–	6.400	4.721	–
150	-1.259	-1.079	-0.788	3.918	2.674	2.111	-1.434	-1.274	-0.983	3.971	2.691	2.143
500	-0.421	-0.233	-0.223	1.931	1.310	0.984	-0.487	-0.306	-0.302	1.935	1.302	1.039
One-step system GMM estimator based on "SYS2"												
50	1.766	–	–	5.821	–	–	2.145	–	–	6.054	–	–
150	2.598	2.013	1.978	4.026	2.841	2.406	3.743	3.276	3.229	4.692	3.593	3.349
500	2.755	1.996	1.572	2.870	2.039	1.648	4.285	3.963	3.723	4.290	3.963	3.723

Note: See notes to Table 1.

Table 3: Size(%) and power(%) of γ and β ($\gamma = 0.4, \beta = 0.5$) for ARX(1) model

$\gamma = 0.4$

	size ($H_0 : \gamma = 0.4$)			power ($H_1 : \gamma = 0.3$)			size ($H_0 : \gamma = 0.4$)			power ($H_1 : \gamma = 0.3$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	9.1	6.4	5.5	28.1	51.2	75.6	9.1	6.4	5.5	28.1	51.2	75.6
150	7.3	5.2	5.8	46.8	91.4	99.9	7.3	5.2	5.8	46.8	91.4	99.9
500	7.7	5.3	5.8	86.5	100.0	100.0	7.7	5.3	5.8	86.5	100.0	100.0
One-step first-difference GMM estimator based on "DIF2"												
50	14.4	13.2	–	36.7	59.3	–	21.1	29.3	–	41.1	66.8	–
150	7.8	8.6	8.1	41.0	79.4	95.5	12.9	16.4	18.9	36.4	65.2	85.4
500	6.1	5.7	5.3	68.1	99.3	100.0	8.1	8.9	8.5	43.2	86.9	98.3
One-step system GMM estimator based on "SYS2"												
50	16.1	–	–	7.4	–	–	97.7	–	–	93.6	–	–
150	10.5	12.2	12.5	16.5	47.3	71.4	93.7	99.7	100.0	81.9	97.5	99.4
500	11.1	8.7	5.9	53.2	97.8	100.0	85.6	97.8	99.8	54.5	64.1	71.2

$\beta = 0.5$

	size ($H_0 : \beta = 0.5$)			power ($H_1 : \beta = 0.4$)			size ($H_0 : \beta = 0.5$)			power ($H_1 : \beta = 0.4$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	5.7	6.6	5.9	30.2	58.9	80.4	5.7	6.6	5.9	30.2	58.9	80.4
150	6.0	6.7	5.3	62.9	95.4	99.9	6.0	6.7	5.3	62.9	95.4	99.9
500	4.9	4.0	5.1	99.1	100.0	100.0	4.9	4.0	5.1	99.1	100.0	100.0
One-step first-difference GMM estimator based on "DIF2"												
50	5.8	5.5	–	27.5	44.9	–	6.7	6.0	–	29.8	50.1	–
150	5.1	7.3	6.1	52.2	83.2	94.0	4.9	7.8	5.6	53.2	85.3	95.5
500	5.5	3.8	4.8	95.9	100.0	100.0	5.5	4.0	5.2	95.8	100.0	100.0
One-step system GMM estimator based on "SYS2"												
50	7.5	–	–	15.8	–	–	7.0	–	–	9.2	–	–
150	6.1	8.4	9.7	35.4	69.6	84.7	7.7	15.4	23.0	14.3	27.4	38.4
500	6.1	4.7	5.4	90.5	100.0	100.0	10.4	16.5	23.4	44.8	86.9	98.1

Note: For the definition of "DIF2" and "SYS2", see notes to Table 1.

Table 4: Size(%) and power(%) of γ and β ($\gamma = 0.9, \beta = 0.5$) for ARX(1) model

$\gamma = 0.9$

	size ($H_0 : \gamma = 0.9$)			power ($H_1 : \gamma = 0.8$)			size ($H_0 : \gamma = 0.9$)			power ($H_1 : \gamma = 0.8$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	5.9	5.6	5.2	26.3	62.1	90.0	5.7	5.4	5.2	26.7	62.4	90.1
150	4.8	5.9	5.2	40.3	93.1	99.9	4.9	5.8	4.9	40.2	93.4	99.9
500	5.5	4.1	5.1	83.1	100.0	100.0	5.2	4.3	5.2	83.5	100.0	100.0
One-step first-difference GMM estimator based on "DIF2"												
50	17.9	22.2	–	46.6	81.1	–	18.5	25.1	–	47.3	81.3	–
150	9.9	11.2	13.9	51.2	90.6	99.4	10.0	13.4	15.2	48.5	88.4	99.2
500	5.8	5.9	6.5	82.0	100.0	100.0	6.6	6.7	6.7	77.9	99.9	100.0
One-step system GMM estimator based on "SYS2"												
50	58.3	–	–	37.9	–	–	99.8	–	–	16.5	–	–
150	62.4	79.1	88.0	57.8	94.5	99.7	100.0	100.0	100.0	31.9	68.0	88.1
500	79.4	78.0	80.9	94.7	100.0	100.0	100.0	100.0	100.0	79.3	99.2	100.0

$\beta = 0.5$

	size ($H_0 : \beta = 0.5$)			power ($H_1 : \beta = 0.4$)			size ($H_0 : \beta = 0.5$)			power ($H_1 : \beta = 0.4$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Transformed likelihood estimator												
50	5.3	6.6	6.0	27.5	57.2	80.5	5.2	6.6	5.9	27.5	57.3	80.5
150	5.3	6.3	5.5	57.6	94.6	99.8	5.4	6.3	5.5	57.6	94.6	99.8
500	4.9	4.6	4.9	98.4	100.0	100.0	4.9	4.6	4.9	98.4	100.0	100.0
One-step first-difference GMM estimator based on "DIF2"												
50	9.2	8.7	–	35.2	54.9	–	9.4	9.1	–	36.3	57.0	–
150	6.3	7.7	6.4	53.9	83.8	95.1	5.6	8.0	7.0	53.9	84.5	95.5
500	5.4	4.7	4.6	94.8	100.0	100.0	6.0	4.5	5.6	94.6	100.0	100.0
One-step system GMM estimator based on "SYS2"												
50	6.4	–	–	17.3	–	–	7.2	–	–	16.1	–	–
150	9.3	10.6	12.8	31.2	64.2	82.8	11.7	17.4	23.4	23.1	47.5	67.6
500	18.2	19.0	18.1	77.9	99.3	100.0	32.1	56.2	69.2	53.5	89.9	98.9

Note: See notes to Table 3.

Table 5: Size(%) and power(%) of weak instruments robust tests for ARX(1) model

$\boldsymbol{\theta}$=(0.4,0.5)'

	size ($H_0 : \boldsymbol{\theta} = (0.4, 0.5)'$)			power ($H_1 : \boldsymbol{\theta} = (0.3, 0.4)'$)			size ($H_0 : \boldsymbol{\theta} = (0.4, 0.5)'$)			power ($H_1 : \boldsymbol{\theta} = (0.3, 0.4)'$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Anderson and Rubin test based on moment conditions "DIF2"												
50	49.9	100.0	–	56.4	100.0	–	49.7	100.0	–	55.1	100.0	–
150	10.7	53.3	95.1	30.5	85.2	99.5	11.4	53.8	95.0	25.8	78.3	98.8
500	8.9	15.1	22.9	72.6	98.5	100.0	8.3	14.2	22.5	62.8	91.2	99.3
Anderson and Rubin test based on moment conditions "SYS2"												
50	84.2	–	–	89.5	–	–	85.6	–	–	88.4	–	–
150	23.7	88.3	99.9	47.5	98.0	100.0	25.0	89.1	100.0	48.3	98.6	100.0
500	11.8	22.3	50.0	79.2	99.6	100.0	13.6	23.2	49.0	80.8	99.6	100.0
Lagrange Multiplier test based on moment conditions "DIF2"												
50	33.7	77.8	–	40.5	82.7	–	35.9	81.9	–	43.2	82.9	–
150	7.4	26.3	70.1	12.2	29.3	86.2	8.4	28.8	73.1	8.1	29.3	84.4
500	6.7	7.1	8.4	30.8	88.8	98.7	6.6	8.6	8.7	8.2	16.4	40.7
Lagrange Multiplier test based on moment conditions "SYS2"												
50	40.7	–	–	41.9	–	–	40.5	–	–	43.4	–	–
150	11.9	28.8	52.4	31.1	40.0	74.5	10.7	26.4	48.4	29.2	34.2	57.1
500	7.9	10.3	11.4	67.1	98.0	99.8	6.6	11.4	11.3	58.9	96.7	98.7
Conditional likelihood ratio test based on moment conditions "DIF2"												
50	50.9	78.0	–	56.0	82.9	–	51.5	82.0	–	55.9	83.2	–
150	9.0	30.0	80.8	15.1	38.8	90.6	11.9	40.8	86.9	13.8	47.3	92.0
500	6.4	7.2	8.1	31.4	89.3	98.8	6.7	8.6	8.8	9.8	19.2	42.9
Conditional likelihood ratio test based on moment conditions "SYS2"												
50	44.8	–	–	45.2	–	–	41.0	–	–	44.3	–	–
150	12.6	35.5	52.9	33.4	44.5	75.1	11.6	27.1	48.6	31.0	35.4	57.4
500	8.1	10.2	11.6	67.4	98.1	99.8	6.8	11.9	11.6	60.4	96.8	98.8

$\boldsymbol{\theta}$=(0.9,0.5)'

	size ($H_0 : \boldsymbol{\theta} = (0.9, 0.5)'$)			power ($H_1 : \boldsymbol{\theta} = (0.8, 0.4)'$)			size ($H_0 : \boldsymbol{\theta} = (0.9, 0.5)'$)			power ($H_1 : \boldsymbol{\theta} = (0.8, 0.4)'$)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
Anderson and Rubin test based on moment conditions "DIF2"												
50	48.0	100.0	–	53.4	99.9	–	48.2	100.0	–	53.3	100.0	–
150	11.8	54.7	95.2	25.3	79.1	99.2	10.9	54.8	94.5	24.1	77.3	99.2
500	9.2	13.6	23.5	59.9	93.5	99.8	8.5	13.4	23.3	56.4	90.7	99.4
Anderson and Rubin test based on moment conditions "SYS2"												
50	87.1	–	–	88.9	–	–	87.5	–	–	89.4	–	–
150	31.1	89.8	100.0	50.1	98.8	100.0	42.7	92.4	100.0	51.6	99.0	100.0
500	31.8	34.5	53.4	89.5	99.9	100.0	68.1	56.2	66.3	91.7	100.0	100.0
Lagrange Multiplier test based on moment conditions "DIF2"												
50	36.6	81.1	–	42.6	85.2	–	37.5	77.9	–	42.4	84.4	–
150	7.7	23.7	67.9	7.7	22.0	90.0	7.7	23.3	66.4	7.9	22.2	89.2
500	4.6	6.2	7.7	16.1	69.3	92.4	4.0	6.6	7.6	9.6	50.0	79.6
Lagrange Multiplier test based on moment conditions "SYS2"												
50	47.4	–	–	53.3	–	–	42.0	–	–	46.8	–	–
150	15.6	41.7	64.4	19.6	70.8	88.9	20.7	28.5	52.5	22.6	61.7	78.7
500	16.9	8.9	12.0	62.7	41.5	44.6	9.7	11.9	15.4	68.3	39.9	73.6
Conditional likelihood ratio test based on moment conditions "DIF2"												
50	46.4	81.0	–	52.7	85.1	–	47.1	78.0	–	52.9	84.4	–
150	8.0	24.9	71.9	9.0	28.5	91.7	8.3	25.7	71.4	8.8	28.4	91.2
500	4.7	6.3	7.5	16.9	69.7	92.5	4.0	6.5	7.9	10.1	51.0	80.0
Conditional likelihood ratio test based on moment conditions "SYS2"												
50	47.5	–	–	53.4	–	–	41.8	–	–	46.6	–	–
150	16.0	42.0	64.6	19.8	71.2	89.0	20.7	28.4	52.5	22.2	61.4	78.9
500	17.6	8.7	12.2	62.8	42.0	44.7	9.9	11.7	15.3	68.3	39.8	73.8

For the definition of "DIF2" and "SYS2", see notes to Table 1. "Anderson and Rubin test" denotes Anderson and Rubin test for GMM (Stock and Wright 2000)(eq. (36)). "Lagrange multiplier test" denotes Kleibergen's(2005) LM test (eq. (37)). "Conditional likelihood ratio test" denotes the conditional likelihood ratio test of Moreira (2003)(extended by Kleibergen(2005)) (eq.(38)). "–" denotes the cases where the GMM estimators are not computed since the number of moment conditions exceeds the sample size.

Table 6: Median bias(×100) and MAE(×100) for AR(1) model

$\gamma = 0.4$

$\gamma = 0.4$	median bias(×100)				MAE(×100)				median bias(×100)				MAE(×100)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Transformed likelihood estimator																
50	-0.360	0.286	-0.107	-0.135	7.148	3.666	2.791	2.286	0.121	0.301	-0.091	-0.135	7.522	3.713	2.804	2.286
150	-0.268	-0.133	-0.017	-0.071	4.340	1.992	1.454	1.304	-0.088	-0.133	-0.017	-0.071	4.472	1.992	1.454	1.304
500	0.020	0.002	0.059	-0.045	2.283	1.240	0.876	0.697	0.033	0.002	0.059	-0.045	2.298	1.240	0.876	0.697
One-step first-difference GMM estimator based on "DIF2"																
50	-7.230	-2.561	-2.709	-1.777	11.683	6.174	4.711	3.574	-16.953	-11.360	-8.042	-5.812	21.862	12.651	9.452	6.945
150	-2.548	-1.288	-0.716	-0.540	6.455	3.323	2.433	1.918	-7.222	-5.031	-3.083	-2.090	12.886	7.155	4.838	3.691
500	-0.729	-0.318	-0.122	-0.213	3.730	1.741	1.283	1.082	-2.076	-1.726	-1.052	-0.853	7.010	3.852	2.609	1.975
One-step system GMM estimator based on "SYS2"																
50	2.241	3.216	2.477	–	8.479	5.620	4.282	–	40.852	41.051	40.739	–	40.852	41.051	40.739	–
150	1.047	0.862	1.047	1.170	4.948	2.817	2.439	2.063	27.590	28.270	26.976	26.619	27.621	28.270	26.976	26.619
500	0.458	0.414	0.390	0.281	2.819	1.654	1.188	1.037	13.344	13.208	13.080	12.126	13.785	13.208	13.080	12.126

$\gamma = 0.9$

$\gamma = 0.9$	median bias(×100)				MAE(×100)				median bias(×100)				MAE(×100)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Transformed likelihood estimator																
50	0.472	1.472	0.423	0.325	9.900	6.412	4.317	3.070	-0.341	1.430	0.202	0.162	9.900	6.418	4.064	2.893
150	2.332	1.118	0.481	0.109	8.745	4.377	2.578	1.580	1.299	0.460	0.078	0.073	8.522	4.037	2.325	1.519
500	4.258	1.031	0.218	0.087	6.831	2.663	1.321	0.849	2.172	0.579	0.279	0.179	6.327	2.480	1.342	0.871
One-step first-difference GMM estimator based on "DIF2"																
50	-59.725	-33.679	-21.641	-15.368	59.822	33.679	21.641	15.368	-66.448	-46.397	-37.081	-31.207	66.448	46.397	37.081	31.207
150	-41.482	-20.518	-10.119	-6.810	42.419	20.681	10.561	7.150	-59.917	-45.943	-34.220	-27.238	60.026	45.943	34.220	27.238
500	-18.534	-7.117	-3.859	-2.205	22.327	8.246	4.769	3.115	-46.246	-32.406	-23.978	-19.485	46.509	32.406	24.014	19.485
One-step system GMM estimator based on "SYS2"																
50	7.194	7.362	7.190	–	7.534	7.362	7.190	–	9.824	9.878	9.872	–	9.824	9.878	9.872	–
150	5.901	5.860	5.758	5.705	6.568	5.906	5.758	5.705	9.778	9.739	9.734	9.759	9.778	9.739	9.734	9.759
500	3.184	3.743	3.449	3.189	4.041	3.796	3.453	3.193	9.343	9.474	9.404	9.404	9.343	9.474	9.404	9.404

Note: "DIF2" denotes Arellano and Bond type moment conditions $E(y_{i,t-2-l}\Delta u_{it}) = 0$ with $l = 0$ for $t = 2$, $l = 0, 1$ for $t = 3, ..., T$. One-step first-difference GMM estimator is computed by (27) with a suitable modification of $\dot{\mathbf{Z}}_i$ and $\dot{\mathbf{W}}_i$. "SYS2" denotes Blundell and Bond type moment conditions $E[\Delta y_{i,t-1}(\alpha_i + u_{it})] = 0$ for $t = 2, ..., T$ in addition to the ones used in "DIF2". One-step system GMM estimator is computed by (33), (34) and (35) with a suitable modification of $\ddot{\mathbf{Z}}_i$ and $\ddot{\mathbf{W}}_i$. The numbers of moment conditions of "DIF2" and "SYS2" are 7 and 11 when $T = 5$, 17 and 26 when $T = 10$, 27 and 41 when $T = 15$ and 37 and 56 when $T = 20$.

Table 7: Size(%) and power(%) of γ ($\gamma = 0.4$) for AR(1) model

$\gamma = 0.4$

	size ($H_0 : \gamma = 0.4$)				power ($H_1 : \gamma = 0.3$)				size ($H_0 : \gamma = 0.4$)				power ($H_1 : \gamma = 0.3$)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Transformed likelihood estimator																
50	7.8	6.9	7.2	6.7	24.1	47.7	70.1	85.7	10.5	7.4	7.3	6.7	26.1	48.1	70.1	85.7
150	5.0	4.7	4.2	5.8	42.4	90.3	99.6	99.8	7.9	4.8	4.2	5.8	43.7	90.4	99.6	99.8
500	5.1	5.3	5.0	4.7	81.4	100.0	100.0	100.0	5.5	5.3	5.0	4.7	81.5	100.0	100.0	100.0
One-step first-difference GMM estimator based on "DIF2"																
50	7.9	8.9	7.9	7.0	21.8	36.9	54.4	67.0	15.5	13.6	13.3	13.2	26.0	35.7	44.1	55.2
150	5.5	6.1	5.9	6.9	27.9	67.3	86.9	96.0	8.2	9.8	7.4	9.2	22.9	37.1	54.2	73.1
500	6.1	5.8	4.8	4.1	53.0	97.2	100.0	100.0	6.3	5.6	5.2	4.5	23.5	55.7	86.6	98.3
One-step system GMM estimator based on "SYS2"																
50	9.5	11.5	8.8	–	11.0	21.0	29.0	–	76.6	93.6	97.8	–	64.0	83.1	89.7	–
150	6.3	5.4	5.6	7.9	24.4	57.1	76.8	88.4	59.9	86.5	94.5	98.4	42.1	60.5	68.7	74.0
500	5.9	6.2	4.8	5.5	64.4	97.6	99.9	100.0	34.8	65.6	84.6	92.5	12.6	15.8	17.1	15.6

$\gamma = 0.9$

	size ($H_0 : \gamma = 0.9$)				power ($H_1 : \gamma = 0.8$)				size ($H_0 : \gamma = 0.9$)				power ($H_1 : \gamma = 0.8$)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Transformed likelihood estimator																
50	15.0	22.0	19.9	21.3	23.9	32.9	49.6	68.2	14.6	22.3	19.1	19.9	24.2	33.3	51.4	70.2
150	20.9	20.8	17.6	12.2	25.3	45.1	71.1	87.6	20.0	19.1	14.7	11.5	27.4	47.8	75.0	89.2
500	25.7	18.3	9.9	8.0	32.4	65.7	88.4	94.1	23.3	16.4	10.8	9.4	37.0	70.1	87.1	90.7
One-step first-difference GMM estimator based on "DIF2"																
50	34.5	31.5	26.2	23.7	45.1	53.5	58.1	67.4	38.1	41.1	40.3	41.1	47.8	60.9	63.5	70.0
150	25.4	17.7	10.0	9.9	36.2	43.2	52.5	73.4	31.6	37.3	33.3	32.8	42.2	55.8	61.0	69.7
500	13.3	7.9	6.0	5.6	25.1	40.1	72.7	91.8	26.3	28.3	25.7	22.6	37.0	51.1	60.6	66.2
One-step system GMM estimator based on "SYS2"																
50	30.6	54.1	63.9	–	1.6	5.2	11.9	–	95.5	99.8	100.0	–	0.0	0.6	1.8	–
150	28.6	46.1	58.3	66.9	3.7	17.4	34.2	50.8	95.3	99.9	100.0	100.0	0.2	1.6	2.6	3.0
500	20.4	33.0	40.0	45.2	22.3	72.2	93.5	98.5	94.0	99.8	100.0	100.0	1.7	6.8	12.3	18.7

Note: For the definition of "DIF2" and "SYS2", see notes to Table 6.

Table 8: Size(%) and power(%) of weak instruments robust tests for AR(1) model

$\theta = 0.4$

	size ($H_0 : \theta = 0.4'$)				power ($H_1 : \theta = 0.3$)				size ($H_0 : \theta = 0.4'$)				power ($H_1 : \theta = 0.3$)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Anderson and Rubin test based on moment conditions "DIF2"																
50	12.8	45.1	87.1	99.1	15.1	50.9	90.5	99.5	13.1	45.3	87.8	99.8	12.9	48.4	89.0	99.7
150	6.3	13.5	22.1	40.5	12.5	32.2	53.2	76.9	7.2	13.1	21.5	42.6	8.2	18.6	32.3	60.8
500	4.7	7.0	7.8	9.2	24.9	66.9	90.6	97.2	5.0	6.0	8.3	9.6	9.0	16.6	39.9	65.9
Anderson and Rubin test based on moment conditions "SYS2"																
50	23.7	86.2	100.0	–	29.5	89.1	100.0	–	26.0	84.9	100.0	–	29.6	87.8	100.0	–
150	9.1	23.8	49.1	81.3	24.9	54.7	80.9	96.5	9.2	23.3	46.6	81.6	22.7	52.3	79.1	95.8
500	5.8	7.5	13.5	17.5	48.0	81.4	95.7	98.8	5.8	7.3	13.8	18.1	42.4	78.8	94.5	98.7
Lagrange Multiplier test based on moment conditions "DIF2"																
50	11.8	28.3	45.2	61.7	11.7	32.3	68.8	76.0	12.4	29.7	45.3	62.4	11.5	30.0	56.2	74.0
150	5.3	8.7	12.7	17.0	19.5	53.8	74.8	80.4	5.8	8.5	12.5	18.5	8.9	17.1	33.2	51.2
500	6.2	6.3	5.9	6.3	46.3	96.6	99.9	100.0	5.7	6.8	5.3	6.5	15.3	45.9	82.1	97.3
Lagrange Multiplier test based on moment conditions "SYS2"																
50	16.1	37.2	72.5	–	20.5	41.9	76.7	–	17.0	39.1	70.7	–	21.4	40.8	74.1	–
150	7.1	11.1	14.4	23.7	41.2	73.0	82.0	59.9	7.3	12.3	16.5	23.5	35.9	58.7	54.9	34.5
500	4.8	7.6	6.5	8.0	82.4	99.5	100.0	100.0	5.2	7.9	6.9	7.9	75.5	99.2	100.0	100.0
Conditional likelihood ratio test based on moment conditions "DIF2"																
50	14.9	40.0	48.9	61.8	16.1	44.1	72.7	76.3	14.9	43.4	48.5	62.7	14.6	42.6	60.7	74.2
150	5.4	8.6	12.9	18.0	19.6	54.1	75.3	82.1	6.9	9.2	13.3	19.8	8.8	18.3	35.2	54.0
500	6.0	6.5	6.1	6.2	46.4	96.6	99.9	100.0	5.8	6.6	5.7	6.4	15.5	46.1	81.7	97.5
Conditional likelihood ratio test based on moment conditions "SYS2"																
50	19.5	39.6	72.3	–	25.5	43.4	76.5	–	18.1	39.3	70.5	–	22.2	41.2	74.0	–
150	7.1	11.5	15.9	26.3	41.3	73.6	84.0	63.3	7.5	12.9	17.3	23.5	35.8	58.8	55.2	34.8
500	4.9	7.6	6.6	8.4	82.1	99.5	100.0	100.0	5.2	8.0	6.8	7.7	75.8	99.2	100.0	100.0

$\theta = 0.9$

	size ($H_0 : \theta = 0.9'$)				power ($H_1 : \theta = 0.8$)				size ($H_0 : \theta = 0.9'$)				power ($H_1 : \theta = 0.8$)			
	$\tau = 1$								$\tau = 5$							
N/T	5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
Anderson and Rubin test based on moment conditions "DIF2"																
50	11.5	45.1	86.6	99.6	11.5	46.4	87.4	99.9	12.3	45.4	86.2	100.0	11.8	45.7	86.9	99.8
150	6.8	12.6	21.3	41.8	7.3	15.5	27.6	56.4	6.9	13.1	22.2	41.0	7.1	12.8	22.0	42.1
500	5.7	5.5	7.6	9.1	6.1	8.8	22.3	41.4	5.6	5.4	9.0	9.3	5.6	5.6	9.3	11.0
Anderson and Rubin test based on moment conditions "SYS2"																
50	23.8	85.8	100.0	–	26.6	88.6	100.0	–	23.8	85.7	99.9	–	26.4	88.5	100.0	–
150	9.3	22.2	49.0	81.4	20.5	48.2	79.7	96.4	9.1	23.1	48.3	80.1	19.4	47.8	79.6	96.7
500	5.6	6.8	15.2	19.0	45.2	78.7	94.2	98.9	5.2	6.8	14.2	17.7	45.1	79.3	94.0	98.8
Lagrange Multiplier test based on moment conditions "DIF2"																
50	15.4	38.4	57.0	68.0	15.6	44.2	68.5	77.4	14.8	40.9	54.9	68.1	15.2	46.7	62.8	73.4
150	6.8	10.2	13.3	19.8	6.2	10.3	18.9	29.7	7.3	12.7	19.8	30.6	7.8	12.0	23.7	33.8
500	5.3	6.1	6.0	6.6	7.6	17.8	50.4	82.4	7.0	7.0	7.2	7.2	7.2	7.0	5.9	10.4
Lagrange Multiplier test based on moment conditions "SYS2"																
50	16.8	35.6	74.0	–	17.1	42.9	75.8	–	17.1	37.1	72.8	–	17.6	41.6	74.9	–
150	7.8	10.7	14.9	23.7	28.8	48.4	41.6	31.6	7.6	11.9	14.7	23.1	28.9	32.9	29.9	31.0
500	5.5	6.4	6.1	8.1	76.7	99.2	100.0	100.0	5.9	7.1	6.6	7.8	74.8	88.1	84.7	82.7
Conditional likelihood ratio test based on moment conditions "DIF2"																
50	12.6	44.2	60.1	68.1	12.1	45.3	71.4	77.7	12.9	44.3	57.8	68.1	13.4	46.3	65.0	74.2
150	7.3	12.2	16.8	25.0	7.7	14.8	23.5	37.7	7.7	13.5	21.1	36.8	7.6	13.2	22.6	38.6
500	5.6	6.0	6.0	6.8	7.7	17.5	50.6	82.6	6.1	5.5	8.6	9.4	5.4	6.4	8.9	12.1
Conditional likelihood ratio test based on moment conditions "SYS2"																
50	17.1	36.0	74.3	–	17.8	43.2	75.8	–	16.8	37.1	72.7	–	17.6	41.6	74.9	–
150	8.2	11.2	15.1	24.0	29.2	48.6	42.5	32.2	7.5	11.7	15.0	23.3	29.2	33.2	29.9	30.8
500	5.6	6.4	5.9	8.2	76.6	99.2	100.0	100.0	5.9	7.0	6.5	7.8	74.8	88.4	84.9	82.7

For the definition of "DIF2" and "SYS2", see notes to Table 6. "Anderson and Rubin test" denotes Anderson and Rubin test for GMM (Stock and Wright 2000)(eq. (36)). "Lagrange multiplier test" denotes Kleibergen's(2005) LM test (eq. (37)). "Conditional likelihood ratio test" denotes the conditional likelihood ratio test of Moreira (2003)(extended by Kleibergen(2005)) (eq.(38)). "–" denotes the cases where the GMM estimators are not computed since the number of moment conditions exceeds the sample size.