DISCUSSION PAPER SERIES

IZA DP No. 6583

Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models

Kazuhiko Hayakawa
M. Hashem Pesaran

May 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models

Kazuhiko Hayakawa
Hiroshima University

M. Hashem Pesaran
University of Cambridge, USC and IZA

Discussion Paper No. 6583
May 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6583
May 2012

ABSTRACT

Robust Standard Errors in Transformed Likelihood Estimation of Dynamic Panel Data Models*

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 crosssectionally heteroskedastic. This extension is not trivial due to the incidental parameters problem that arises, and its implications for estimation and inference. We approach the problem by working with a mis-specified homoskedastic model. It is shown 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 simulation, 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.

JEL Classification: C12, C13, C23

Keywords: dynamic panels, cross-sectional heteroskedasticity, Monte Carlo simulation, GMM estimation

Corresponding author:

M. Hashem Pesaran
Trinity College
University of Cambridge
CB2 1TQ Cambridge
United Kingdom
E-mail: mhp1@cam.ac.uk

* 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. Over the last few decades, a sizable literature has been developed on the estimation of dynamic panel data models. Early work includes the Instrumental Variables (IV) approach by Anderson and Hsiao (1981, 1982). 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) suggested to transform the dynamic model into first differences to eliminate the individual-specific effects, and then use a set of 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 one and/or the *variance ratio* of the individual effects to the idiosyncratic errors is large since in these cases the instruments are only weakly related to the lagged dependent variables.[1] Among others, the poor finite sample properties of GMM has been documented in Monte Carlo studies by Kiviet (2007). To deal with this problem, Arellano and Bover (1995) and Blundell and Bond (1998) proposed the use of extra moment conditions arising from the model in levels, 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 large.

The GMM estimators discussed so far have been widely adopted in the empirical literature, to investigate problems in areas such as labour economics, development economics, health economics, 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. Indeed, the incidental parameters issue and the initial values problem lead to a violation of the standard regularity conditions for the ML estimators of the structural parameters to be consistent. Hsiao et al. (2002) developed a transformed likelihood approach to overcome the incidental parameters problem. Binder et al. (2005) have extended this approach for estimating panel VAR (PVAR) models. Alvarez and Arellano (2004) have studies ML estimation of autoregressive panels in the presence of time-specific

[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.

heteroskedasticity (see also Bhargava and Sargan (1983)). Kruiniger (2008) considers ML estimation of a stationary/unit root AR(1) panel data models.

In this paper, we extend the analysis of Hsiao et al. (2002) to allow for cross-sectional heteroskedasticity. This extension is not trivial due to the incidental parameters problem that arises, and its implications for estimation and inference. To deal with the problem, we follow the GMM literature and ignore the error variance heterogeneity and work with a mis-specified homoskedastic model, and show that the transformed maximum likelihood estimator by Hsiao et al. (2002) continues to be consistent. We then derive a covariance matrix estimator which is robust to cross-sectional heteroskedasticity. Using Monte Carlo simulations, we investigate the finite sample performance of the transformed likelihood 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 likelihood estimator for cross-sectionally heteroskedastic errors. Section 4 reviews the GMM approach as applied to dynamic panels. 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 panel data model

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

for $i = 1, 2, ..., N$. It is supposed that these dynamic processes have started at time $t = -m$, (m being a finite positive constant) but we only observe the observations (y_{it}, x_{it}) over the period $t = 0, 1, 2, ..., T$. We assume that x_{it} is a scalar to simplify the notation. Extension to the case of multiple regressors is straightforward at the expense of notational complexity. We further assume that x_{it} is generated either by

$$x_{it} = \mu_i + \phi t + \sum_{j=0}^{\infty} a_j \varepsilon_{i,t-j}, \qquad \sum_{j=0}^{\infty} |a_j| < \infty \tag{2}$$

or

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

where μ_i can either be fixed constants, differing across i, or randomly distributed with a common mean, and ε_{it} are independently distributed over i and t with $E(\varepsilon_{it}) = 0$, and $var(\varepsilon_{it}) = \sigma^2_{\varepsilon i}$, with $0 < \sigma^2_{\varepsilon i} < K < \infty$.

We shall also consider the following assumptions:

Assumption 1 *(Initialization) Depending on whether the y_{it} process has reached stationarity, one of the following two assumptions holds:*
(i) $\mid \gamma \mid< 1$, 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 $y_{i,-m+1}$ with m finite, such that

$$E(\Delta y_{i,-m+1}|\Delta x_{i1}, \Delta x_{i2}, ..., \Delta x_{iT}) = b, \;\; for \;\; all \; i,$$

where b is a finite constant.

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

Assumption 3 *(shocks to regressors) ε_{it} in x_{it} are independently distributed over all i and t, with $E(\varepsilon_{it}) = 0$, and $E(\varepsilon_{it}^2) = \sigma_{\varepsilon i}^2$, and independent of u_{is} for all s and t.*

Assumption 4 *(constant variance ratio) $\sigma_{\varepsilon i}^2/\sigma_i^2 = c$, for $i = 1, 2, ..., N$, with $0 < c < K < \infty$.*

Remark 1 *Assumption 1.(ii) constrains the expected changes in the initial values to be the same across all individuals, but does not necessarily require that $\mid \gamma \mid< 1$. Assumptions 2, 3, and 4 allow for heteroskedastic disturbances in the equations for y_{it} and x_{it}, but to avoid the incidental parameter problem require their ratio to be constant over i. Also Assumption 3 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 et al. (2005). However, these further developments are beyond the scope of the present paper. See also the remarks in Section 6 .*

3 Transformed likelihood estimation

Take the first differences of (1) to eliminate the individual effects:

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \beta \Delta x_{it} + \Delta u_{it}, \tag{4}$$

which is well defined for $t = 2, 3, ..., T$, but not for $t = 1$, because the observations $y_{i,-1}$, $i = 1, 2, ..., N$, are not available. However, starting from $\Delta y_{i,-m+1}$, and by continuous substitution, we obtain

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

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

$$\eta_{i1} = E\left(\Delta y_{i1} | \Delta y_{i,-m+1}, \Delta x_{i1}, \Delta x_{i0}, ...\right) = \gamma^m \Delta y_{i,-m+1} + \beta \sum_{j=0}^{m-1} \gamma^j \Delta x_{i,1-j}, \tag{5}$$

is unknown, since the observations $\Delta x_{i,1-j}$, for $j = 1, 2, ..., m-1$, $i = 1, 2, ..., N$ are unavailable. 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 (1), where x_{it} follows either (2) or (3). Suppose that Assumptions 1, 2, 3, and 4 hold. Then Δy_{i1} can be expressed as:*

$$\Delta y_{i1} = b + \boldsymbol{\pi}' \Delta \mathbf{x}_i + v_{i1}, \tag{6}$$

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 $E(v_{i1}^2)/\sigma_i^2 = \omega$, with $0 < \omega < K < \infty$.

Note that Assumption 4 is used to show that $E(v_{i1}^2)/\sigma_i^2$ does not vary with i.

It is now possible to derive the likelihood function of the *transformed model* given by equations (6) and (4) for $t = 2, 3, ..., T$. Let $\Delta \mathbf{y}_i = (\Delta y_{i1}, \Delta y_{i2}, ..., \Delta y_{iT})'$,

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

and note that the transformed model can be rewritten as

$$\Delta \mathbf{y}_i = \Delta \mathbf{W}_i \boldsymbol{\varphi} + \mathbf{r}_i, \tag{7}$$

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:

$$E(\mathbf{r}_i \mathbf{r}_i') = \boldsymbol{\Omega}_i = \sigma_i^2 \begin{pmatrix} \omega & -1 & & & 0 \\ -1 & 2 & \ddots & & \\ & & \ddots & & \\ & & \ddots & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix} = \sigma_i^2 \boldsymbol{\Omega}, \tag{8}$$

where ω is a free parameter defined in Theorem 1.

The log-likelihood function of the transformed model (7) is given by

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

where $\boldsymbol{\psi}_N=\left(\boldsymbol{\varphi}',\omega,\sigma_1^2,\sigma_2^2,...\sigma_N^2\right)'$. Unfortunately, the maximum likelihood estimation based on $\ell(\boldsymbol{\psi}_N)$ encounters the incidental parameter 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, $\boldsymbol{\varphi}$ and ω, 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., $\sigma_i^2=\sigma^2$, $i=1,2,...,N$. Then under this mis-specification the pseudo log-likelihood function of the transformed model (7), is given by

$$\begin{aligned}\ell_p\left(\boldsymbol{\theta}\right) &= -\frac{NT}{2}\ln\left(2\pi\right)-\frac{NT}{2}\ln\left(\sigma^2\right)-\frac{N}{2}\ln\left[1+T\left(\omega-1\right)\right] \\ &\quad -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(\Delta\mathbf{y}_i-\Delta\mathbf{W}_i\boldsymbol{\varphi}\right)'\boldsymbol{\Omega}^{-1}\left(\Delta\mathbf{y}_i-\Delta\mathbf{W}_i\boldsymbol{\varphi}\right),\end{aligned} \tag{9}$$

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 in (9), and consider the pseudo-score vector

$$\frac{\partial\ell_p\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}=\begin{pmatrix}\frac{1}{\sigma^2}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\left(\Delta\mathbf{y}_i-\Delta\mathbf{W}_i\boldsymbol{\varphi}\right)\\ -\frac{NT}{2g}+\frac{1}{2\sigma^2g^2}\sum_{i=1}^{N}\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i\\ -\frac{NT}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{N}\mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i\end{pmatrix},$$

where $g=|\boldsymbol{\Omega}|=1+T(\omega-1)$, (see (40)), and

$$\boldsymbol{\Phi}=\begin{pmatrix}T^2 & T(T-1) & T(T-2) & \dots & T\\ T(T-1) & (T-1)^2 & (T-1)(T-2) & \dots & (T-1)\\ \vdots & \vdots & \vdots & \dots & \vdots\\ T & (T-1) & (T-2) & \dots & 1\end{pmatrix}. \tag{10}$$

Under heteroskedastic errors, the pseudo-true value of $\boldsymbol{\theta}$ denoted by $\boldsymbol{\theta}_* = (\boldsymbol{\varphi}'_*, \omega_*, \sigma_*^2)'$, is the solution of $E\left[\partial \ell_p\left(\boldsymbol{\theta}_*\right)/\partial\boldsymbol{\theta}\right] = \mathbf{0}$, namely

$$\sum_{i=1}^{N} E\left[\Delta\mathbf{W}_i'\boldsymbol{\Omega}_*^{-1}\left(\Delta\mathbf{y}_i - \Delta\mathbf{W}_i\boldsymbol{\varphi}_*\right)\right] = \mathbf{0}, \tag{11}$$

$$-\frac{NT}{2g_*} + \frac{1}{2\sigma_*^2 g_*^2}\sum_{i=1}^{N} E\left(\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i\right) = 0, \tag{12}$$

$$-\frac{NT}{2\sigma_*^2} + \frac{1}{2\sigma_*^4}\sum_{i=1}^{N} E\left(\mathbf{r}_i'\boldsymbol{\Omega}_*^{-1}\mathbf{r}_i\right) = 0, \tag{13}$$

where expectations are taken with respect to the true probability measure, and $g_* = 1 + T\left(\omega_* - 1\right)$.

Focusing first on (12) and (13), we have

$$\sum_{i=1}^{N} E\left(\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i\right) = \sum_{i=1}^{N}\sigma_i^2 tr\left(\boldsymbol{\Phi}\boldsymbol{\Omega}\right) = N\bar{\sigma}_N^2 Tg,$$

$$\sum_{i=1}^{N} E\left(\mathbf{r}_i'\boldsymbol{\Omega}_*^{-1}\mathbf{r}_i\right) = TN\bar{\sigma}_N^2 tr(\boldsymbol{\Omega}_*^{-1}\boldsymbol{\Omega})/T,$$

where $\bar{\sigma}_N^2 = N^{-1}\sum_{i=1}^{N}\sigma_i^2$ and (42) is used. Hence, using the above results in (12) and (13), we have

$$-\frac{NT}{2g_*} + \frac{1}{2\sigma_*^2 g_*^2}N\bar{\sigma}_N^2 tr\left(\boldsymbol{\Phi}\boldsymbol{\Omega}\right) = -\frac{NT}{2g_*} + \frac{1}{2\sigma_*^2 g_*^2}N\bar{\sigma}_N^2 Tg = 0,$$

$$-\frac{NT}{2\sigma_*^2} + \frac{1}{2\sigma_*^4}\sum_{i=1}^{N} E\left(\mathbf{r}_i'\boldsymbol{\Omega}_*^{-1}\mathbf{r}_i\right) = -\frac{NT}{2\sigma_*^2} + \frac{1}{2\sigma_*^4}TN\bar{\sigma}_N^2 tr(\boldsymbol{\Omega}_*^{-1}\boldsymbol{\Omega})/T = 0.$$

From the first equation, we have $\sigma_*^2/\bar{\sigma}_N^2 = g/g_* = \left[1 + T\left(\omega - 1\right)\right]/\left[1 + T\left(\omega_* - 1\right)\right]$. From the second equation, we have $\sigma_*^2/\bar{\sigma}_N^2 = tr(\boldsymbol{\Omega}_*^{-1}\boldsymbol{\Omega})/T$. Using these two, we have

$$\frac{1 + T\left(\omega - 1\right)}{1 + T\left(\omega_* - 1\right)} = \frac{1}{T}tr(\boldsymbol{\Omega}_*^{-1}\boldsymbol{\Omega}). \tag{14}$$

To solve this equation for ω_*, we first note that note that

$$tr(\boldsymbol{\Omega}_*^{-1}\boldsymbol{\Omega})/T = 1 + g_*^{-1}(\omega - \omega_*).$$

This result follows since all elements of $\boldsymbol{\Delta} = \boldsymbol{\Omega} - \boldsymbol{\Omega}_*$ are zero, except for the first element of $\boldsymbol{\Delta}$ which is given by $\omega - \omega_*$. Substituting this into (14), and after some algebra we have $(T-1)(\ \omega_* - \omega) = 0$, which yields $\omega_* = \omega$ for all $T > 1$. It also follows that $\sigma_*^2 = \lim_{N\to\infty}\bar{\sigma}_N^2$. Using the former result in (11), we have $\boldsymbol{\varphi}_* = \boldsymbol{\varphi}$. These results are stated formally in the following theorem.

Theorem 2 *Suppose that Assumptions 1, 2, 3, and 4 hold, and let* $\boldsymbol{\theta}_* = \left(\boldsymbol{\varphi}'_*, \omega_*, \sigma_*^2\right)'$ *be the pseudo* *true values of the ML estimator obtained by maximizing the pseudo log-likelihood function in (9). Then, we have*

$$\boldsymbol{\varphi}_* = \boldsymbol{\varphi}, \quad \omega_* = \omega, \quad \sigma_*^2 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_i^2.$$

This is one of the key results of this paper. This theorem shows that the first $(T+4)$ entries of $\boldsymbol{\theta}_*$ are identical to the first $(T+4)$ entries of $\boldsymbol{\psi}_N$. This indicates that the ML estimator of $\boldsymbol{\varphi}$ and ω obtained under mis-specified homoskedastic models will continue to be consistent, namely, the transformed ML estimator by Hsiao et al. (2002) is consistent even if cross-sectional heteroskedasticity is present.

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

Theorem 3 *Suppose that Assumptions 1, 2, 3 and 4 hold and let* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\varphi}}', \widehat{\omega}, \widehat{\sigma}^2\right)'$ *be the ML estimator obtained by maximizing the pseudo log-likelihood function in (9). Then as N tends to infinity,* $\widehat{\boldsymbol{\theta}}$ *is asymptotically normal with*

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

where $\boldsymbol{\theta}_* = (\boldsymbol{\varphi}', \omega, \sigma_*^2)'$,

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

To obtain consistent estimators of $\mathbf{A}^*$ and $\mathbf{B}^*$, robust to unknown heteroskedasticity, let

$$\widehat{\mathbf{r}}_i = \Delta\mathbf{y}_i - \Delta\mathbf{W}_i\widehat{\boldsymbol{\varphi}}.$$

Further, let

$$\widetilde{\sigma}_{NT}^2 = (TN)^{-1}\sum_{i=1}^{N}\widehat{\mathbf{r}}_i'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{r}}_i,$$

be an estimator of $N^{-1}\sum_{i=1}^{N}\sigma_i^2$. Then a consistent estimator of $\mathbf{A}^*$, denoted as $\widehat{\mathbf{A}}^*$, is given by

$$\widehat{\mathbf{A}}^* = \begin{pmatrix} \frac{1}{N\widetilde{\sigma}_{NT}^2}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\widehat{\boldsymbol{\Omega}}^{-1}\Delta\mathbf{W}_i & \frac{1}{g^2N\widetilde{\sigma}_{NT}^2}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\boldsymbol{\Phi}\widehat{\mathbf{r}}_i & \mathbf{0} \\ \frac{1}{g^2N\widetilde{\sigma}_{NT}^2}\sum_{i=1}^{N}\widehat{\mathbf{r}}_i'\boldsymbol{\Phi}\Delta\mathbf{W}_i & \frac{T^2}{2g^2} & \frac{T}{2g\widetilde{\sigma}_{NT}^2} \\ \mathbf{0} & \frac{T}{2g\widetilde{\sigma}_{NT}^2} & \frac{T}{2\left(\widetilde{\sigma}_{NT}^2\right)^2} \end{pmatrix}.$$

To obtain a consistent estimator of $\mathbf{B}^*$, denoted by $\widehat{\mathbf{B}}^*$, we also need to assume that the fourth moment of $(v_{i1} - u_{i1})/\sigma_i$ is homogeneous across i. In particular,

Assumption 5 *(kurtosis condition) Assume that* $E(\eta_{i1}^4) = \kappa = \gamma_2 + 3$ *for* $i = 1, 2, ..., N$, *where* $\eta_{i1} = (v_{i1} - u_{i1})/[\sigma_i(\omega-1)^{1/2}]$, *and* γ_2 *is the Pearson's measure of kurtosis.*

This assumption is used in combination with Assumption 2 to consistently estimate $N^{-1}\sum_{i=1}^{N}\sigma_i^4$ by $\widetilde{\sigma}_{NT}^4$ defined in the Appendix by (66). Then the elements of $\widehat{\mathbf{B}}^*$are given by:

$$\widehat{\mathbf{B}}_{11}^* = \frac{1}{N\left(\widetilde{\sigma}_{NT}^2\right)^2}\sum_{i=1}^{N}\Delta\mathbf{W}_i'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{r}}_i\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\Delta\mathbf{W}_i,$$

$$\widehat{\mathbf{B}}_{22}^* = \frac{T^2}{4\widehat{g}^4\left(\widetilde{\sigma}_{NT}^2\right)^2}\left\{N^{-1}\sum_{i=1}^{N}\left(\frac{\widehat{\mathbf{r}}_i'\mathbf{\Phi}\widehat{\mathbf{r}}_i}{T}\right)^2 - \widehat{g}^2\widetilde{\sigma}_{NT}^4\right\},$$

$$\widehat{\mathbf{B}}_{33}^* = \frac{T^2}{4\left(\widetilde{\sigma}_{NT}^2\right)^4}\left\{N^{-1}\sum_{i=1}^{N}\left(\frac{\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{r}}_i}{T}\right)^2 - \widetilde{\sigma}_{NT}^4\right\},$$

$$\widehat{\mathbf{B}}_{21}^* = \frac{1}{2N\widehat{g}^2\left(\widetilde{\sigma}_{NT}^2\right)^2}\sum_{i=1}^{N}\left(\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\Delta\mathbf{W}_i\right)\left(\widehat{\mathbf{r}}_i'\mathbf{\Phi}\widehat{\mathbf{r}}_i\right),$$

$$\widehat{\mathbf{B}}_{31}^* = \frac{1}{2N\left(\widetilde{\sigma}_{NT}^2\right)^3}\sum_{i=1}^{N}\left(\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\Delta\mathbf{W}_i\right)\left(\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{r}}_i\right),$$

$$\widehat{\mathbf{B}}_{32}^* = \frac{T^2}{4\widehat{g}^2\left(\widetilde{\sigma}_{NT}^2\right)^3}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{\mathbf{r}}_i'\mathbf{\Phi}\widehat{\mathbf{r}}_i}{T}\frac{\widehat{\mathbf{r}}_i'\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{r}}_i}{T} - \widehat{g}\widetilde{\sigma}_{NT}^4\right].$$

4 GMM approach

In this section, we review the GMM approach as a basis for the simulation studies in the next section. In the GMM approach, it is assumed that α_i and u_{it} have an error components structure, in which[2]

$$E(\alpha_i) = 0, \quad E(u_{it}) = 0, \quad E(\alpha_i u_{it}) = 0, \qquad \text{for } i = 1,..,N;\ \text{and } t = 1,2,...,T, \tag{16}$$

and the errors are uncorrelated with the initial values

$$E(y_{i0}u_{it}) = 0, \qquad \text{for } i = 1,2,..,N,\ \text{and } t = 1,2,...,T. \tag{17}$$

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

$$E(u_{it}u_{is}) = 0, \qquad \text{for } i = 1,2,..,N,\ \text{and } t \neq s = 1,2,...,T. \tag{18}$$

[2]Note that no restrictions are placed on $E(\alpha_i u_{it})$ under the transformed likelihood approach

4.1 Estimation

4.1.1 The first-difference GMM estimator

Under (16)-(18), and focusing on the equation in first differences, (4), Arellano and Bond (1991) suggest the following $T(T-1)/2$ moment conditions:

$$E\left[y_{is}\Delta u_{it}\right]=0, \qquad (s=0,1,...,t-2, t=2,3,...,T). \tag{19}$$

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

$$E\left[x_{is}\Delta u_{it}\right]=0, \qquad (s,t=2,...,T). \tag{20}$$

The moment conditions (19) and (20) 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

$$\begin{aligned}
\dot{\mathbf{Z}}_i &= \begin{pmatrix} y_{i0},x_{i1},...,x_{iT} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & y_{i0},y_{i1},x_{i1},...,x_{iT} & ... & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & y_{i0},...,y_{i,T-2},x_{i1},...,x_{iT} \end{pmatrix}, \\
\dot{\mathbf{q}}_i &= \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \qquad \dot{\mathbf{W}}_i=\begin{pmatrix} \Delta y_{i1} & \Delta x_{i2} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{iT} \end{pmatrix}.
\end{aligned}$$

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

$$\widehat{\boldsymbol{\delta}}^{dif}_{GMM1} = \left(\dot{\mathbf{S}}_{ZW}'\left(\dot{\mathbf{D}}_{1step}\right)^{-1}\dot{\mathbf{S}}_{ZW}\right)^{-1}\dot{\mathbf{S}}_{ZW}'\left(\dot{\mathbf{D}}_{1step}\right)^{-1}\dot{\mathbf{S}}_{Zq}, \tag{21}$$

$$\widehat{\boldsymbol{\delta}}^{dif}_{GMM2} = \left(\dot{\mathbf{S}}_{ZW}'\left(\dot{\mathbf{D}}_{2step}\right)^{-1}\dot{\mathbf{S}}_{ZW}\right)^{-1}\dot{\mathbf{S}}_{ZW}'\left(\dot{\mathbf{D}}_{2step}\right)^{-1}\dot{\mathbf{S}}_{Zq}, \tag{22}$$

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 sub-diagonal and 0's otherwise.

4.1.2 System GMM estimator

Although consistency of the first-difference GMM estimator is obtained under a mild assumption of no serial correlation, Blundell and Bond (1998) demonstrated that it suffers from the so called weak instruments problem when γ is close to one and/or the variance ratio $var(\alpha_i)/var(u_{it})$ 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 $var(\alpha_i)/var(u_{it})$ is large.

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

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

Under these assumptions, we have the following moment conditions:

$$E\left[\Delta y_{is}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (s = 1, ..., t-1, t = 2, 3, ..., T), \tag{23}$$

$$E\left[\Delta x_{is}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (s, t = 2, 3, ..., T). \tag{24}$$

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 (23) and (24) are redundant. Hence, to implement the system GMM estimation, in addition to (19) and (20), we use the following moment conditions:

$$E\left[\Delta y_{i,t-1}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (t = 2, 3, ..., T), \tag{25}$$

$$E\left[\Delta x_{it}\left(\alpha_i + u_{it}\right)\right] = 0, \qquad (t = 2, 3, ..., T). \tag{26}$$

The moment conditions (19), (20), (25) and (26) can be written compactly as

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

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

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

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

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

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

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

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 $\ddot{\mathbf{D}}_{1step} = diag\left(\frac{1}{N}\sum_{i=1}^{N}\dot{\mathbf{Z}}'_i\mathbf{H}\dot{\mathbf{Z}}_i, \frac{1}{N}\sum_{i=1}^{N}\breve{\mathbf{Z}}'_i\breve{\mathbf{Z}}_i\right)$. 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 $\widehat{\ddot{\mathbf{u}}}_i = \ddot{\mathbf{q}}_i - \ddot{\mathbf{W}}_i\widehat{\boldsymbol{\delta}}^{sys}_{GMM1}$.

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 mainly 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 $\ddot{\mathbf{u}}_i$. Also, let m be the number of columns of $\check{\mathbf{Z}}_i$, i.e., the number of instruments, and set

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

Then, the CU-GMM estimator is defined as

$$\widehat{\boldsymbol{\delta}}_{GMM-CU} = \arg\min_{\delta} Q(\boldsymbol{\delta}), \tag{29}$$

$$Q(\boldsymbol{\delta}) = \widehat{\mathbf{g}}(\boldsymbol{\delta})'\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\delta})/2. \tag{30}$$

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

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 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 distortion. To overcome this problem, Newey and Windmeijer (2009) propose an alternative estimator for the covariance matrix of CU-GMM estimators 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.[3]

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

$$AR(\boldsymbol{\delta}) = 2N \cdot Q(\boldsymbol{\delta}). \tag{31}$$

Under the null hypothesis $H_0 : \boldsymbol{\delta} = \boldsymbol{\delta}_0$, this statistic is asymptotically (as $N \to \infty$) distributed as χ^2_m, regardless of the strength of the instruments, where m is the dimension of δ.

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

The LM statistic proposed by Kleibergen (2005) is

$$LM(\boldsymbol{\delta}) = N \cdot \frac{\partial Q(\boldsymbol{\delta})'}{\partial \boldsymbol{\delta}} \left[\widehat{\mathbf{D}}(\boldsymbol{\delta})' \widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})^{-1} \widehat{\mathbf{D}}(\boldsymbol{\delta}) \right]^{-1} \frac{\partial Q(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}, \tag{32}$$

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

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

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

$$CLR(\boldsymbol{\delta}) = \frac{1}{2} \left[AR(\boldsymbol{\delta}) - \widehat{R}(\boldsymbol{\delta}) + \sqrt{\left(AR(\boldsymbol{\delta}) - \widehat{R}(\boldsymbol{\delta}) \right)^2 + 4LM(\boldsymbol{\delta}) \widehat{R}(\boldsymbol{\delta})} \right] \tag{33}$$

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 which can be obtained by simulation[4].

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

5 Monte Carlo simulations

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

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

5.1.1 Monte Carlo design

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

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + \beta x_{it} + u_{it}, \quad \text{for } t = -m+1, -m+2, .., 0, 1, ..., T, \tag{34}$$

with the initial value $y_{i,-m} = \alpha_i + \beta x_{i,-m} + u_{i,-m}$, and $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$, with $\sigma_i^2 \sim \mathcal{U}[0.5, 1.5]$, so that $E(\sigma_i^2) = 1$. We discard the first m observations, and use the observations $t = 0$ through T for estimation and inference.[5] The regressor, x_{it}, is generated as

$$x_{it} = \mu_i + gt + \zeta_{it}, \qquad \text{for } t = -m, -m+1, .., 0, 1, ..., T, \tag{35}$$

where

$$\zeta_{it} = \phi \zeta_{i,t-1} + \varepsilon_{it}, \qquad \text{for } t = -49-m, -48-m, ..., 0, 1, ..., T, \tag{36}$$

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{\varepsilon i}^2), \qquad \xi_{i,-m-50} = 0. \tag{37}$$

with $|\phi| < 1$. We discard the first 50 observations of ζ_{it} and use the remaining $T+1+m$ observations for generating x_{it} and y_{it}.

In the simulations, we try the values $\gamma = 0.4, 0.9$, $\beta = 0.5$, $\phi = 0.5$, and $g = 0.01$. The error variances $\sigma_{\varepsilon i}^2$ are set so that to ensure a reasonable fit, namely[6]

$$\sigma_{\varepsilon i}^2 = \frac{\sigma_i^2 R_{\Delta y}^2 (1+\phi)(1-\phi\gamma)}{\beta^2 \left(1 - R_{\Delta y}^2\right)},$$

with $R_{\Delta y}^2 = 0.4$. The sample sizes considered are $N = 50, 150, 500$ and $T = 5, 10, 15$. For the individual effects, we set $\alpha_i = \tau \left(\frac{q_i - 1}{\sqrt{2}}\right)$, where $q_i \sim \chi_1^2$. For the value of τ, which is the variance ratio, $\tau = var(\alpha_i)/var(u_{it})$, we consider the values of $\tau = 1$ often used in the literature, and the high value of $\tau = 5$. Further, we assume that both y_{it} and x_{it} depend linearly on the same individual effects, by taking $\mu_i = \eta\alpha_i$ where the value of η is computed by (69) in the Appendix A.5 with $R_y^2 = 0.4$.[7]

In computing the transformed ML estimators we use the minimum distance estimator of Hsiao et al. (2002) as starting values for the nonlinear optimization where ω is estimated by the one-step

[5] Hence, $T + 1$ is the actual length of the estimation sample.
[6] For the derivation of $R_{\Delta y}^2$, see Appendix A.5.
[7] Since (69) is a quadratic equation, we have two solutions. In the simulations, we used the positive solution.

first-difference GMM estimator (21) in which $\dot{\mathbf{Z}}_i$ is replaced with

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

For the GMM estimators, although there are many moment conditions for the first-difference GMM estimator as in (19) and (20), we consider two 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. Similarly, for the system GMM estimator, we add moment conditions (25) and (26) 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.

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 $-$ in the summary result tables.

For inference, we use the robust standard errors formula given in Theorem 2 for the transformed likelihood 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.

In addition to the MC results for γ and β, we also report simulation results for the long-run coefficient defined by $\delta = \beta/(1-\gamma)$. We report median biases, median absolute errors (MAE), size and power for γ, β and δ. 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" 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 are summarized in Tables 1 to 4. Table 1 provides the results for the case of $\gamma = 0.4$, and shows that the transformed likelihood estimator has a smaller bias than the GMM estimators in all cases with the exception of the CU-GMM estimator (the last panel of Table 1). In terms of MAE the transformed likelihood estimator 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 likelihood estimator must be invariant to the choice of τ, although the estimates reported in Table 1 do show some effects, albeit small. The observed impact of changes in τ on the performance of the transformed likelihood estimator is solely due to computational issues and reflects the dependence of the choice of initial values on τ in computation of the transformed ML estimators. One would expect that these initial value effects to disappear as N is increased, and this is seen to be the case from the results summarized in Table 1. In contrast, 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 likelihood estimator which are 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 likelihood estimator performs best in all cases.

We next consider size and power of the various tests, summarized in Tables 3 and 4. Table 3 shows that the empirical size of the transformed likelihood estimator is close to the nominal size of 5% for all values of T, N and τ.

For the GMM estimators, we find that the test sizes vary considerably depending on 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 $N = 50$, $T = 5$, and $\tau = 1$, the size of the test based on DIF2(2step) estimator is 30.4%. But the size distortion gets smaller as N increases. Increasing N to 500, reduces the size of this test to 6.6%. 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.

Amongst 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 not always helpful, and although they improve the size property in some cases, they have either little effects or tend to worsen the test sizes in other cases.

Comparing power of the tests, we observe that the transformed likelihood estimator is in general more powerful than the GMM estimators. For example when $N = 150$, the transformed likelihood estimators have higher power than "SYS2(2step$_W$)" which is the most efficient amongst the reported GMM estimators.

The above conclusions hold generally when we consider experiments with $\gamma = 0.9$ (Table 4), 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[8], the simulation results in Table 4 reveal that they perform poorly even when τ is not so large ($\tau = 1$).

The small sample results for β (Tables 5 to 8), 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, $\delta = \beta/(1 - \gamma)$, are reported in a supplementary appendix, and 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 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 9 and 10. 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 $m^2/N \to 0$ or a stronger restriction where m 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

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

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

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + u_{it}, \qquad \text{for } t = 1, ..., T \text{ and } i = 1, ..., N, \tag{38}$$

$$y_{i0} = \frac{\alpha_i}{1-\gamma} + u_{i0}\sqrt{\frac{1}{1-\gamma^2}}, \tag{39}$$

where $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i^2 \sim \mathcal{U}[0.5, 1.5]$. Note that y_{it} are covariance stationary. Individual effects are generated as $\alpha_i = \tau(q_i - 1)/\sqrt{2}$ where $q_i \sim \chi_1^2$.

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 (21) where $\dot{\mathbf{W}}_i$ and $\dot{\mathbf{Z}}_i$ are replaced with[9]

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

$\widetilde{\omega} = [(N-1)\widetilde{\sigma}_u^2]\sum_{i=1}^{N}\left(\Delta y_{i1} - \widetilde{b}\right)^2$ 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 two 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 for $T =$ 5, 10, 15, 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 for $T = 5, 10, 15$, respectively.

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.

[9] 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.

For the moment conditions "SYS1", we have 14, 54, 119 moment conditions for $T = 5, 10$, 15, respectively, while for the moment conditions "SYS2", we have 11, 26, 41 moment conditions for $T = 5$, 10, 15, respectively. 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". Complete set of results are provided in a supplement, which is available upon request.

The biases and MAEs of the various estimators for the case of $\gamma = 0.4$ are summarized in Table 11. As can be seen from this table, 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 $\gamma = 0.9$ (Table 12), 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 Table 13. 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 not always helpful: although they improve the size property in some cases, they have almost no effects in some cases or worsen the test sizes in other cases.

Size and power of the tests in the case of experiments with $\gamma = 0.9$ are summarized in Table 14, 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 7.7% when $\tau = 1$, $N > 150$ and $T > 15$. 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 GMM-DIF 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 a supplement, to save space. 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, although "LM(SYS2)" and "CLR(SYS2)" tests have almost the same power as the transformed likelihood estimator when $\gamma = 0.4$, $T = 5$, $N = 500$ and $\tau = 1$, their powers decline when $\tau = 5$, unlike the transformed likelihood estimator which is invariant to changes in τ. 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 et al. (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 et al. (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 $\mathbf{\Omega}$ be given by (8). Then the determinant and inverse of $\mathbf{\Omega}$ are:*

$$|\mathbf{\Omega}| = g = 1 + T(\omega - 1), \tag{40}$$

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

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

$$\left\{\widetilde{\mathbf{\Omega}}\right\}_{ts} = \begin{cases} s(T-t)\omega - (s-1)(T-t), & (s \leq t) \\ t(T-s)\omega - (t-1)(T-s), & (s > t) \end{cases}. \tag{41}$$

Proof. See Hsiao et al. (2002). ■

Lemma A2 *Let $\mathbf{\Phi}$ be defined in (10). We have*

$$\mathbf{\Phi} = \boldsymbol{\vartheta}\boldsymbol{\vartheta}',$$

where $\boldsymbol{\vartheta}' = (T, T-1, \ldots, 2, 1)$ *and*

$$tr\left(\mathbf{\Phi\Omega}\right) = tr\left(\boldsymbol{\vartheta\vartheta}'\mathbf{\Omega}\right) = \boldsymbol{\vartheta}'\mathbf{\Omega}\boldsymbol{\vartheta} = Tg, \tag{42}$$

where g *is given by (40).*

Proof. See Hsiao et al. (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 (7). Under Assumptions 1 and 2 we have*

$$E\left(\Delta\mathbf{W}_i'\mathbf{\Omega}^{-1}\mathbf{r}_i\right) = \mathbf{0}, \qquad (i = 1, 2, ..., N), \tag{43}$$

where $\mathbf{\Omega}$ *is given in (8). Further,*

$$E\left(\mathbf{r}_i'\mathbf{\Phi}\Delta\mathbf{W}_i\right) = \begin{pmatrix} 0 & \mathbf{0} & \varpi & 0 \end{pmatrix}, \qquad (i = 1, 2, ..., N), \tag{44}$$

where $\mathbf{\Phi}$ *is given by (10), and* $\varpi \neq 0$.

Proof. Let $\Delta\widetilde{\mathbf{y}}_{i,-1} = (0, \Delta y_{i1}, ..., \Delta y_{i,T-1})'$ and note that, for (43) to hold, it is only needed to prove that $E\left(\Delta\widetilde{\mathbf{y}}_{i,-1}'\mathbf{\Omega}^{-1}\mathbf{r}_i\right) = \mathbf{0}$. To show this, let $\mathbf{p}_i = \mathbf{\Omega}^{-1}\mathbf{r}_i = (p_{i1}, ..., p_{iT})'$ where by (41)

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

and

$$\begin{aligned}
h_{ts} &= (T-t+1)\left[(s-1)\omega - (s-2)\right], \\
k_{ts} &= (T-s+1)\left[(t-1)\omega - (t-2)\right].
\end{aligned} \tag{45}$$

Then, we have

$$\begin{aligned} E\left[\Delta\widetilde{\mathbf{y}}_{i,-1}'\mathbf{\Omega}^{-1}\mathbf{r}_i\right] &= \sum_{t=2}^{T} E\left[p_{it}\Delta y_{i,t-1}\right] = \sum_{t=2}^{T-1} E\left[p_{it}\Delta y_{i,t-1}\right] + E\left[p_{iT}\Delta y_{i,T-1}\right] \\ &= \sum_{t=2}^{T-1} E\left[(T-t+1)v_{i1}\Delta y_{i,t-1} + \sum_{s=2}^{t} h_{ts}\Delta u_{is}\Delta y_{i,t-1} + \sum_{s=t+1}^{T} k_s \Delta u_{is}\Delta y_{i,t-1}\right] \\ &\quad + E\left(p_{iT}\Delta y_{i,T-1}\right) \\ &= \sum_{t=2}^{T}(T-t+1)E\left(v_{i1}\Delta y_{i,t-1}\right) + \sum_{t=2}^{T}\sum_{s=2}^{t} h_{ts}E\left(\Delta u_{is}\Delta y_{i,t-1}\right) \\ &= A_1 + A_2. \end{aligned}$$

where we used $E(\Delta u_{is}\Delta y_{it}) = 0$ for $t < s-1$. To derive A_1 and A_2, we use the followings[10]:

$$\sigma_i^{-2}E(v_{i1}\Delta y_{it}) = \begin{cases} \omega & t=1 \\ \gamma^{t-2}(\gamma\omega - 1) & t = 2, ..., T \end{cases} \tag{46}$$

$$\sigma_i^{-2}E(\Delta u_{is}\Delta y_{it}) = \begin{cases} -1 & t = s-1 \\ (2-\gamma) & s = t \\ -(1-\gamma)^2\gamma^{t-s-1} & s < t \end{cases} \tag{47}$$

Using (46) and (47), we have

$$A_1 = \sigma_i^2\left[(T-1)\omega + (\gamma\omega - 1)\sum_{t=3}^{T}(T-t+1)\gamma^{t-3}\right], \tag{48}$$

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

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

$$
\begin{aligned}
A_2 \;=\; & \; h_{22}E(\Delta u_{i2}\Delta y_{i1}) \\
& + h_{32}E(\Delta u_{i2}\Delta y_{i2}) + h_{33}E(\Delta u_{i3}\Delta y_{i2}) \\
& + h_{42}E(\Delta u_{i2}\Delta y_{i3}) + h_{43}E(\Delta u_{i3}\Delta y_{i3}) + h_{44}E(\Delta u_{i4}\Delta y_{i3}) \\
& + h_{52}E(\Delta u_{i2}\Delta y_{i4}) + h_{53}E(\Delta u_{i3}\Delta y_{i4}) + h_{54}E(\Delta u_{i4}\Delta y_{i4}) + h_{55}E(\Delta u_{i5}\Delta y_{i4}) \\
& \qquad \vdots \\
& + h_{T2}E(\Delta u_{i2}\Delta y_{i,T-1}) + h_{T3}E(\Delta u_{i3}\Delta y_{i,T-1}) + \cdots + h_{T,T-2}E(\Delta u_{i,T-2}\Delta y_{i,T-1}) + \\
& \qquad + h_{T,T-1}E(\Delta u_{i,T-1}\Delta y_{i,T-1}) + h_{TT}E(\Delta u_{iT}\Delta y_{i,T-1}) \\
=\; & \sigma_i^2 \left[(-1)\sum_{s=2}^{T} h_{ss} + (2-\gamma)\sum_{s=2}^{T-1} h_{s+1,s} - (1-\gamma)^2 \sum_{t=4}^{T}\sum_{s=2}^{t-2} h_{ts}\gamma^{t-s-2} \right]. \qquad (49)
\end{aligned}
$$

Then, by using (45), (48) and (49), and after some algebra, we obtain $E\left[\Delta\widetilde{\mathbf{y}}_{i,-1}'\boldsymbol{\Omega}^{-1}\mathbf{r}_i\right] = A_1 + A_2 = 0$.

To prove (44), first note that $E\left(\Delta\mathbf{W}_i'\boldsymbol{\Phi}\mathbf{r}_i\right)$ is a $(T+3)$ dimensional vector having all zeros, except for the $(T+2)$th entry, given by $E\left(\Delta\widetilde{\mathbf{y}}_{i,-1}'\boldsymbol{\Phi}\mathbf{r}_i\right)$. 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 (46)-(47), we have

$$
\begin{aligned}
\sigma_i^{-2}E\left(\boldsymbol{\vartheta}'\mathbf{r}_i\Delta\widetilde{\mathbf{y}}_{i,-1}'\boldsymbol{\vartheta}\right) &= \varpi = T\sum_{s=1}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) + \sum_{s=1}^{T-1}\sum_{t=2}^{T}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \\
&= T\sum_{s=1}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) + \sum_{s=1}^{T-1}\sum_{t=1}^{s+1}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \\
&= T(T-1)E(\Delta y_{i1}v_{i1}) + \sum_{s=2}^{T-1}(T-s)E(\Delta y_{is}v_{i1}) \\
&\quad + \sum_{s=1}^{T-1}\sum_{t=1}^{s-1}(T-t+1)(T-s)E\left(\Delta y_{is}\Delta u_{it}\right) \\
&\quad + \sum_{s=1}^{T-1}(T-s+1)(T-s)E\left(\Delta y_{is}\Delta u_{is}\right) + \sum_{s=1}^{T-1}(T-s)^2E\left(\Delta y_{is}\Delta u_{i,s+1}\right) \\
&= T(T-1)\omega + (\gamma\omega - 1)\sum_{s=2}^{T-1}(T-s)\gamma^{s-2} \\
&\quad - (1-\gamma)^2\sum_{s=1}^{T-1}\sum_{t=1}^{s-1}(T-t+1)(T-s)\gamma^{s-t-1} \\
&\quad + (2-\gamma)\sum_{s=1}^{T-1}(T-s+1)(T-s) - \sum_{s=1}^{T-1}(T-s)^2. \qquad (50)
\end{aligned}
$$

which in general is different from zero. ■

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 (9), and* $\boldsymbol{\theta}_* = (\boldsymbol{\varphi}', \omega, \sigma_*^2)'$ *is the vector of pseudo-true values. Then as* N *tends to infinity and for fixed* T, *we have*

$$\underset{N\to\infty}{\text{plim}}\ \mathbf{A}_N^* = \mathbf{A}^*,$$

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

Proof. The elements of $\mathbf{A}_N^*$ are given by[11]

$$\begin{aligned}
\mathbf{A}_{N,11}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} = \frac{1}{\sigma_*^2}\frac{1}{N}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\Delta\mathbf{W}_i, \\
\mathbf{A}_{N,22}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial\omega^2} = -\frac{T^2}{2g^2} + \frac{T}{\sigma_*^2 g^3 N}\sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i, \\
\mathbf{A}_{N,33}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial(\sigma^2)^2} = -\frac{T}{2\left(\sigma_*^2\right)^2} + \frac{1}{\left(\sigma^{2*}\right)^3 N}\sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i, \\
\mathbf{A}_{N,12}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\omega} = \frac{1}{\sigma_*^2 g^2 N}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Phi}\mathbf{r}_i, \\
\mathbf{A}_{N,13}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial\boldsymbol{\varphi}\partial\sigma^2} = \frac{1}{\left(\sigma_*^2\right)^2 N}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i, \\
\mathbf{A}_{N,23}^* &= -\frac{1}{N}\frac{\partial^2 \ell_p(\boldsymbol{\theta}_*)}{\partial\omega\partial\sigma^2} = \frac{1}{2\left(\sigma_*^2\right)^2 g^2 N}\sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i.
\end{aligned}$$

Given that, under Assumptions 2 and 3 $\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i$, are independent across i, and, by Lemma A4, have zero mean, and have finite variance for fixed T, by applying the law of large numbers for heterogeneous observations (White, 2001), we have

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i = \mathbf{0}.$$

Further, $\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i$ and $\mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i$ are independent across i, with mean $T\sigma_i^2 g$ and $T\sigma_i^2$, respectively, and finite variances for fixed T, so that

$$\underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i = T\sigma_*^2 g, \qquad \underset{N\to\infty}{\text{plim}}\ \frac{1}{N}\sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i = T\sigma_*^2.$$

[11]See also Hsiao et al. (2002).

Hence, the matrix $\mathbf{A}^*$ is given by

$$\mathbf{A}^* = \begin{pmatrix} \operatorname*{plim}_{N\to\infty} \frac{1}{N\sigma_*^2} \sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\Delta\mathbf{W}_i & \operatorname*{plim}_{N\to\infty} \frac{1}{Ng^2\sigma_*^2} \sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Phi}\mathbf{r}_i & \mathbf{0} \\ \operatorname*{plim}_{N\to\infty} \frac{1}{Ng^2\sigma_*^2} \sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Phi}\Delta\mathbf{W}_i & \frac{T^2}{2g^2} & \frac{T}{2g\sigma*^2} \\ \mathbf{0} & \frac{T}{2g\sigma_*^2} & \frac{T}{2\left(\sigma_*^2\right)^2} \end{pmatrix}. \tag{51}$$

■

Lemma A6 *Let* $\mathbf{b}_N^* = \left(1/\sqrt{N}\right)\partial\ell_p\left(\boldsymbol{\theta}_*\right)/\partial\boldsymbol{\theta}$*, where* $\ell_p\left(\boldsymbol{\theta}\right)$ *is given by (9), and* $\boldsymbol{\theta}_* = (\boldsymbol{\varphi}', \omega, \sigma_*^2)'$ *is the vector of pseudo-true values. Then as* N *tends to infinity and for fixed* T*, we have*

$$\mathbf{b}_N^* \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{B}^*\right). \tag{52}$$

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

$$\frac{1}{\sqrt{N}}\frac{\partial\ell_p\left(\boldsymbol{\theta}_*\right)}{\partial\boldsymbol{\theta}} = \frac{1}{\sigma_*^2}\frac{1}{\sqrt{N}} \begin{pmatrix} \sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i \\ \frac{1}{2g^2}\sum_{i=1}^N \xi_i \\ \frac{1}{2\sigma_*^2}\sum_{i=1}^N \zeta_i \end{pmatrix}, \tag{53}$$

where ξ_i and ζ_i are given by

$$\xi_i = \mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i - Tg\sigma_i^2, \qquad \zeta_i = \mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i - T\sigma_i^2. \tag{54}$$

By Lemma A4, $\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i$ has zero mean for all i. It is easily seen that ξ_i and ζ_i have also zero mean. Then, using Lemma A3, we have

$$\mathbf{B}_{11}^* = \frac{1}{N\left(\sigma_*^2\right)^2} E\left(\sum_{i=1}^N \Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i \sum_{i=1}^N \mathbf{r}_i'\boldsymbol{\Omega}^{-1}\Delta\mathbf{W}_i\right) = \frac{1}{N\left(\sigma_*^2\right)^2}\sum_{i=1}^N E\left(\Delta\mathbf{W}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i\mathbf{r}_i'\boldsymbol{\Omega}^{-1}\Delta\mathbf{W}_i\right).$$

Again, using Lemma A3, and recalling that $E(\xi_i) = 0$, we have

$$\begin{aligned} \mathbf{B}_{22}^* &= \frac{1}{4Ng^4\left(\sigma_*^2\right)^2} E\left[\sum_{i=1}^N \xi_i^2\right] = \frac{1}{4Ng^4\left(\sigma_*^2\right)^2} E\left[\sum_{i=1}^N \left(\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i - Tg\sigma_i^2\right)^2\right] \\ &= \frac{1}{4Ng^4\left(\sigma_*^2\right)^2} E\left[\sum_{i=1}^N \left(\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i\right)^2 - 2Tg\sum_{i=1}^N \sigma_i^2\left(\mathbf{r}_i'\ \boldsymbol{\Phi}\mathbf{r}_i\right) + T^2g^2\sum_{i=1}^N \sigma_i^4\right] \\ &= \frac{T^2}{4g^4\left(\sigma_*^2\right)^2} E\left[N^{-1}\sum_{i=1}^N \left\{\left(\frac{\mathbf{r}_i'\boldsymbol{\Phi}\mathbf{r}_i}{T}\right)^2 - g^2\sigma_i^4\right\}\right]. \end{aligned} \tag{55}$$

Similarly

$$\begin{aligned}
\mathbf{B}_{33}^{*} &= \frac{1}{4N\left(\sigma_{*}^{2}\right)^{4}}E\left[\sum_{i=1}^{N}\zeta_{i}^{2}\right]=\frac{1}{4N\left(\sigma_{*}^{2}\right)^{4}}\left\{E\left[\sum_{i=1}^{N}\left(\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}\right)^{2}\right]-T^{2}\sum_{i=1}^{N}\sigma_{i}^{4}\right\} \\
&= \frac{T^{2}}{4\left(\sigma_{*}^{2}\right)^{4}}E\left[N^{-1}\sum_{i=1}^{N}\left\{\left(\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}\right)^{2}-\sigma_{i}^{4}\right\}\right]. \qquad (56)
\end{aligned}$$

The off-diagonal elements of $\mathbf{B}^{*}$ are (using Lemma A3 and noting that $E\left(\Delta\mathbf{W}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}\right)=0$ and $E\left(\xi_{i}\right)=0$):

$$\begin{aligned}
\mathbf{B}_{21}^{*} &= \frac{1}{2N\left(\sigma_{*}^{2}\right)^{2}g^{2}}E\left[\sum_{i=1}^{N}\xi_{i}\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\Delta\mathbf{W}_{i}\right]=\frac{1}{2N\left(\sigma_{*}^{2}\right)^{2}g^{2}}E\left[\sum_{i=1}^{N}\left(\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\Delta\mathbf{W}_{i}\right)\left(\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}-Tg\sigma_{i}^{2}\right)\right] \\
&= \frac{1}{2N\left(\sigma_{*}^{2}\right)^{2}g^{2}}E\left[\sum_{i=1}^{N}\left(\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\Delta\mathbf{W}_{i}\right)\left(\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}\right)\right], \qquad (57) \\
\mathbf{B}_{31}^{*} &= \frac{1}{2N\left(\sigma_{*}^{2}\right)^{3}}E\left[\sum_{i=1}^{N}\left(\Delta\mathbf{W}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}\right)\left(\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}\right)\right], \qquad (58)
\end{aligned}$$

Similarly, using Lemma A3 we have

$$\begin{aligned}
\mathbf{B}_{32}^{*} &= \frac{1}{4N\left(\sigma_{*}^{2}\right)^{3}g^{2}}E\left(\sum_{i=1}^{N}\xi_{i}\zeta_{i}\right)=\frac{T^{2}}{4N\left(\sigma_{*}^{2}\right)^{3}g^{2}}E\left[\sum_{i=1}^{N}\left(\frac{\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}}{T}-g\sigma_{i}^{2}\right)\left(\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}-\sigma_{i}^{2}\right)\right] \\
&= \frac{T^{2}}{4N\left(\sigma_{*}^{2}\right)^{3}g^{2}}E\left[\sum_{i=1}^{N}\left(\frac{\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}}{T}\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}-g\sigma_{i}^{2}\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}-\sigma_{i}^{2}\frac{\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}}{T}+g\sigma_{i}^{4}\right)\right] \\
&= \frac{T^{2}}{4N\left(\sigma_{*}^{2}\right)^{3}g^{2}}\left[\sum_{i=1}^{N}E\left(\frac{\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}}{T}\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}\right)-g\sum_{i=1}^{N}\sigma_{i}^{4}-g\sum_{i=1}^{N}\sigma_{i}^{4}+g\sum_{i=1}^{N}\sigma_{i}^{4}\right] \\
&= \frac{T^{2}}{4N\left(\sigma_{*}^{2}\right)^{3}g^{2}}E\left[\sum_{i=1}^{N}\left(\frac{\mathbf{r}_{i}'\mathbf{\Phi}\mathbf{r}_{i}}{T}\frac{\mathbf{r}_{i}'\mathbf{\Omega}^{-1}\mathbf{r}_{i}}{T}-g\sigma_{i}^{4}\right)\right]. \qquad (59)
\end{aligned}$$

For fixed T, and under Assumption 2, the elements inside the sum operator in expressions (55)-(59) are finite for all i. Hence, (52) 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, under Assumption 1, equation (5) can be rewritten as

$$\eta_{i1}=b+\beta\Delta x_{i1}+\beta\sum_{j=1}^{m-1}\gamma^{j}E\left(\Delta x_{i,1-j}|\Delta\mathbf{x}_{i}\right)+\varsigma_{i1}, \qquad (60)$$

where $\varsigma_{i1} = \eta_{i1} - E(\eta_{i1}|\Delta\mathbf{x}_i)$, and b is zero under Assumption 1.(i) and is equal to $\widetilde{b}$ otherwise. Using either (2) or (3) we have

$$\Delta x_{it} = \phi + \sum_{j=0}^{\infty} \widetilde{d}_j \varepsilon_{i,t-j}, \tag{61}$$

with $\widetilde{d}_j = d_j$ under (3), $\widetilde{d}_j = a_j - a_{j-1}$ under (2), and $\widetilde{d}_0 = a_0$. Hence, it is easily seen that under (61)

$$E(\Delta x_{i,1-j}|\Delta\mathbf{x}_i) = b_j + \boldsymbol{\pi}_j'\Delta\mathbf{x}_i, \tag{62}$$

where b_j and $\boldsymbol{\pi}_j$ do not depend on i. Using (62) in (6) and (60), the marginal distribution of Δy_{i1} conditional on $\Delta\mathbf{x}_i$ can be written as

$$\Delta y_{i1} = b + \beta\Delta x_{i1} + \beta\sum_{j=1}^{m-1}\gamma^j\left(b_j + \boldsymbol{\pi}_j'\Delta\mathbf{x}_i\right) + \varsigma_{i1} + \sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j},$$

or, more compactly,

$$\Delta y_{i1} = b + \boldsymbol{\pi}'\Delta\mathbf{x}_i + v_{i1}, \tag{63}$$

where $v_{i1} = \varsigma_{i1} + \sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j}$, b is a constant, and $\boldsymbol{\pi}$ is a T-dimensional vector of parameters. Note that $b = 0$ under Assumption 1.(i) and if $\phi = 0$, while it is a nonzero constant otherwise. In the above equation, v_{i1} has zero mean and its variance satisfies

$$\begin{aligned}
\omega &= \frac{1}{\sigma_i^2}E\left(v_{i1}^2\right) = \frac{1}{\sigma_i^2}E\left[\left(\beta\sum_{j=1}^{m-1}\gamma^j\left[\Delta x_{i,1-j} - E(\Delta x_{i,1-j}|\Delta\mathbf{x}_i)\right] + \sum_{j=0}^{m-1}\gamma^j\Delta u_{i,1-j}\right)^2\right] \\
&= \frac{\beta^2}{\sigma_i^2}\sum_{j,\ell=0}^{m-1}\gamma^{j+\ell}E\left\{\left[\Delta x_{i,1-j} - E(\Delta x_{i,1-j}|\mathbf{x}_i)\right]\left[\Delta x_{i,1-\ell} - E(\Delta x_{i,1-\ell}|\mathbf{x}_i)\right]\right\} \\
&\quad + \frac{1}{\sigma_i^2}\sum_{j,\ell=0}^{m-1}\gamma^{j+\ell}E\left(\Delta u_{i,1-j}\Delta u_{i,1-\ell}\right) \\
&= \frac{1}{\sigma_i^2}\left[\beta^2\sigma_{\varepsilon i}^2\sum_{j,\ell=0}^{m-1}\gamma^{j+\ell}\varpi_{j\ell} + 2\sigma_i^2\sum_{j=0}^{m-1}\gamma^{2j} - \sigma_i^2\sum_{j=1}^{m-1}\gamma^{2j-1} - \sigma_i^2\sum_{j=0}^{m-2}\gamma^{2j+1}\right] \\
&= \beta^2\frac{\sigma_{\varepsilon i}^2}{\sigma_i^2}\sum_{j,\ell=0}^{m-1}\gamma^{j+\ell}\varpi_{j\ell} + \left[2\sum_{j=0}^{m-1}\gamma^{2j} - \sum_{j=1}^{m-1}\gamma^{2j-1} - \sum_{j=0}^{m-2}\gamma^{2j+1}\right].
\end{aligned} \tag{64}$$

where $\varpi_{j\ell} = \frac{1}{\sigma_{\varepsilon i}^2}E\left\{\left[\Delta x_{i,1-j} - E(\Delta x_{i,1-j}|\mathbf{x}_i)\right]\left[\Delta x_{i,1-\ell} - E(\Delta x_{i,1-\ell}|\mathbf{x}_i)\right]\right\}$ is given by

$$\varpi_{j\ell} = \frac{1}{\sigma_{\varepsilon i}^2}\sum_{h,k=0}^{\infty}\widetilde{d}_h\widetilde{d}_k E\left[\left(\varepsilon_{i,1-j-h} - \boldsymbol{\pi}_h'\boldsymbol{\varepsilon}_{i,-h-j}\right)\left(\varepsilon_{i,1-\ell-k} - \boldsymbol{\pi}_k'\boldsymbol{\varepsilon}_{i,-k-\ell}\right)\right]$$

where $\boldsymbol{\varepsilon}_{i,-h-j} = \left(\varepsilon_{i,1-h-j}, \varepsilon_{i,2-h-j}, ..., \varepsilon_{i,T-h-j}\right)'$, and is easily seen to be finite and constant across i, for fixed T. It follows that v_{i1}/σ_i has a constant variance under Assumption 4. We also have that $\frac{1}{\sigma_i^2}E\left(v_{i1}^2\right) > 1$, and $E\left(v_{i1}\Delta u_{i2}\right) = -\sigma_i^2$, $E\left(v_{i1}\Delta u_{it}\right) = 0$ for $t = 3, 4, ..., T$. Finally, note that under Assumption 1.(i), the term in the square bracket in (64) reduces to

$$2\sum_{j=0}^{\infty}\gamma^{2j} - \sum_{j=1}^{\infty}\gamma^{2j-1} - \sum_{j=0}^{\infty}\gamma^{2j+1} = \frac{2}{1+\gamma}.$$

■

A.3 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

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

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

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

where, by Lemma A5, $\mathbf{A}^*$ is a symmetric and positive definite matrix (see expression (51)). 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

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

where the elements of $\mathbf{B}^*$ are given in expressions (55)-(59). Hence, result (15) follows, and $\widehat{\boldsymbol{\theta}}$ is asymptotically normally distributed for a fixed T, and as N tends to infinity.■

A.4 Estimation of $N^{-1}\sum_{i=1}^{N}\sigma_i^4$

To obtain an estimator of $N^{-1}\sum_{i=1}^{N}\sigma_i^4$, we first note that $\mathbf{r}_i$ can be written as

$$\mathbf{r}_i = \begin{pmatrix} 1 & 1 & 0 & 0 & ... & 0 \\ 0 & -1 & 1 & 0 & ... & 0 \\ 0 & 0 & -1 & ... & & \\ ... & & & & 1 & 0 \\ 0 & 0 & 0 & ... & -1 & 1 \end{pmatrix} \begin{pmatrix} \vartheta_i \\ u_{i1} \\ u_{i2} \\ \\ u_{iT} \end{pmatrix} = \underset{T\times(T+1)}{\mathbf{H}}\underset{(T+1)\times 1}{\boldsymbol{\varsigma}_i},$$

where $\vartheta_i = \varsigma_{i1} - u_{i0} + \sum_{j=1}^{m-1}\gamma^j \Delta u_{i,1-j} = v_{i1} - u_{i1}$ (see equation (63)) is independent of $u_{i1}, u_{i2}, ..., u_{iT}$. Clearly, the elements of $\boldsymbol{\varsigma}_i$ are independent of each other. Noting that $\frac{1}{\sigma_i^2}E\left(\vartheta_i^2\right) = \frac{1}{\sigma_i^2}E\left(v_{i1}^2\right) + \frac{1}{\sigma_i^2}E\left(u_{i1}^2\right) - 2\frac{1}{\sigma_i^2}E\left(v_{i1}u_{i1}\right) = \omega - 1 > 0$, the random vector $\boldsymbol{\varsigma}_i$ has variance

$$E\left(\boldsymbol{\varsigma}_i\boldsymbol{\varsigma}_i'\right) = \underset{(T+1)\times(T+1)}{\boldsymbol{\Omega}_{\boldsymbol{\varsigma}_i}} = \sigma_i^2 \begin{pmatrix} (\omega-1) & 0 & 0 & & 0 \\ 0 & 1 & & & \\ & & 1 & & \\ & & & ... & \\ 0 & & & & 1 \end{pmatrix} = \sigma_i^2\boldsymbol{\Omega}_{\boldsymbol{\varsigma}},$$

so that

$$E\left(\mathbf{r}_i\mathbf{r}_i'\right) = \sigma_i^2\boldsymbol{\Omega} = E\left(\mathbf{H}\boldsymbol{\varsigma}_i\boldsymbol{\varsigma}_i'\mathbf{H}'\right) = \mathbf{H}\boldsymbol{\Omega}_{\boldsymbol{\varsigma}_i}\mathbf{H}' = \sigma_i^2\mathbf{H}\boldsymbol{\Omega}_{\boldsymbol{\varsigma}}\mathbf{H}'.$$

Let $\boldsymbol{\eta}_i = \boldsymbol{\Omega}_{\boldsymbol{\varsigma}_i}^{-1/2}\boldsymbol{\varsigma}_i = \frac{1}{\sigma_i}\left(\frac{\vartheta_i}{(\omega-1)^{0.5}}, u_{i1}, u_{i2}, ..., u_{iT}\right)' = (\eta_{i1}, ..., \eta_{iT}, \eta_{i,T+1})'$ and note that $E\left(\eta_{it}\right) = 0$, $E(\eta_{it}^2) = 1$ for $i = 1, 2, ..., N$, for $i = 1, 2, ..., N$, $t = 1, 2, ..., T+1$. Also under Assumptions 2 and 5, we have $E(\eta_{it}^4) = \kappa = \gamma_2 + 3$ for $t = 1, ..., T+1$, where γ_2 is the Pearson's measure of kurtosis. Then using results on moments of quadratic forms for independent random variables under non-normality, we have

$$E\left[\left(\frac{1}{\sigma_i^2}\mathbf{r}_i'\boldsymbol{\Omega}^{-1}\mathbf{r}_i\right)^2\right] = \frac{1}{\sigma_i^4}E\left[\left(\boldsymbol{\varsigma}_i'\mathbf{H}'\boldsymbol{\Omega}^{-1}\mathbf{H}\boldsymbol{\varsigma}_i\right)^2\right] = E\left[\left(\boldsymbol{\eta}_i'\boldsymbol{\Omega}_{\boldsymbol{\varsigma}}^{1/2}\mathbf{H}'\boldsymbol{\Omega}^{-1}\mathbf{H}\boldsymbol{\Omega}_{\boldsymbol{\varsigma}}^{1/2}\boldsymbol{\eta}_i\right)^2\right] = E\left[\left(\boldsymbol{\eta}_i'\mathbf{G}\boldsymbol{\eta}_i\right)^2\right]$$

where $\mathbf{G}$ is a $(T+1)\times(T+1)$ matrix $\mathbf{G} = \boldsymbol{\Omega}_{\boldsymbol{\varsigma}}^{1/2}\mathbf{H}'\boldsymbol{\Omega}^{-1}\mathbf{H}\boldsymbol{\Omega}_{\boldsymbol{\varsigma}}^{1/2}$. Then using[12]

$$E\left[\boldsymbol{\eta}_i'\mathbf{G}\boldsymbol{\eta}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right] = \gamma_2(\mathbf{I}_{T+1}\odot\mathbf{G}) + tr(\mathbf{G})\mathbf{I}_{T+1} + 2\mathbf{G} \tag{65}$$

[12] See Ullah (2004, p. 187).

and $tr(\mathbf{G}) = tr(\mathbf{G}^2) = T$, we have

$$\begin{aligned} E\left[\left(\boldsymbol{\eta}_i'\mathbf{G}\boldsymbol{\eta}_i\right)^2\right] &= E\left[\left(\boldsymbol{\eta}_i'\mathbf{G}\boldsymbol{\eta}_i\right) tr\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\mathbf{G}\right)\right] = tr\left[E\left(\boldsymbol{\eta}_i'\mathbf{G}\boldsymbol{\eta}_i\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right)\mathbf{G}\right] \\ &= \gamma_2 tr\left[(\mathbf{I}_{T+1}\odot\mathbf{G})\mathbf{G}\right] + [tr(\mathbf{G})]^2 + 2tr(\mathbf{G}^2) \\ &= \gamma_2 \sum_{t=1}^{T+1} g_{tt}^2 + T\left(T+2\right), \end{aligned}$$

where g_{tt} are the diagonal elements of $\mathbf{G}$. On the basis of the above result, we consider the following estimator of $N^{-1}\sum_{i=1}^N \sigma_i^4$:

$$\widetilde{\sigma}_{NT}^4 = \widehat{\frac{1}{N}\sum_{i=1}^N \sigma_i^4} = \frac{1}{N\left[\widehat{\gamma}_2\sum_{t=1}^{T+1}\widehat{g}_{tt}^2 + T\left(T+2\right)\right]}\sum_{i=1}^N\left(\widehat{\mathbf{r}}_i'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{r}}_i\right)^2. \tag{66}$$

where $\widehat{g}_{tt}^2$ are the diagonal elements of $\widehat{\mathbf{G}} = \widehat{\boldsymbol{\Omega}}_{\boldsymbol{\varsigma}}^{1/2}\mathbf{H}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{H}\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\varsigma}}^{1/2}$. In the case of normal errors, $\kappa = 3$ and $\gamma_2 = 0$, so that the above expression reduces to:

$$\widetilde{\sigma}_{NT}^4 = \frac{1}{NT\left(T+2\right)}\sum_{i=1}^N\left(\widehat{\mathbf{r}}_i'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{r}}_i\right)^2.$$

To obtain an estimator of γ_2 (i.e., the kurtosis of η_{it}) in the more general case of non-normal errors, we can exploit information on r_{it}. In particular, note that, for $i = 1, 2, ..., N$ and $t = 1$, under the Assumption 5,

$$\begin{aligned} E(r_{i1}^4) &= E\left[\left(\vartheta_i + u_{i1}\right)^4\right] = \sigma_i^4\left[1 + \left(\omega - 1\right)^2\right]\kappa + 6\sigma_i^4\left(\omega - 1\right) \\ &= \sigma_i^4\left\{\left[1 + \left(\omega - 1\right)^2\right]\gamma_2 + 3\left[1 + \left(\omega - 1\right)^2\right] + 6\left(\omega - 1\right)\right\}, \end{aligned}$$

while for $t = 2, ..., T$, under Assumption 2

$$E(r_{it}^4) = E\left[\left(u_{it} - u_{i,t-1}\right)^4\right] = \sigma_i^4\left(2\kappa + 6\right) = \sigma_i^4\left(2\gamma_2 + 12\right).$$

Then

$$\begin{aligned} E\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T r_{it}^4\right) &= \frac{1}{NT}\sum_{i=1}^N \sigma_i^4\left\{\left[1 + \left(\omega - 1\right)^2\right]\gamma_2 + 3\left[1 + \left(\omega - 1\right)^2\right] + 6\left(\omega - 1\right)\right. \\ &\qquad \left. + 2\left(T - 1\right)\gamma_2 + 12\left(T - 1\right)\right\} \\ &= \frac{1}{NT}\sum_{i=1}^N \sigma_i^4\left\{\left[\left(\omega - 1\right)^2 + 2T - 1\right]\gamma_2 + 3\left(\omega - 1\right)^2 + 6\left(\omega - 1\right) + 12T - 9\right\} \\ &= \frac{1}{NT}\sum_{i=1}^N \sigma_i^4\left\{\left[\left(\omega - 1\right)^2 + 2T - 1\right]\gamma_2 + 3\omega^2 + 12\left(T - 1\right)\right\}. \end{aligned}$$

Hence

$$\gamma_2 = \left[(\omega-1)^2 + 2T - 1\right]^{-1} \left\{ T \frac{E\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} r_{it}^4\right)}{N^{-1}\sum_{i=1}^{N}\sigma_i^4} - \left[3\omega^2 + 12\left(T-1\right)\right] \right\},$$

and γ_2 can be consistently estimated by

$$\widehat{\gamma}_2 = \left[(\widehat{\omega}-1)^2 + 2T - 1\right]^{-1} \left\{ \frac{\left[\widehat{\gamma}_2 \sum_{t=1}^{T+1} \widehat{g}_{tt}^2 + T\left(T+2\right)\right] \sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{r}_{it}^4}{\sum_{i=1}^{N}\left(\widehat{\mathbf{r}}_i' \widehat{\mathbf{\Omega}}^{-1} \widehat{\mathbf{r}}_i\right)^2} - \left[3\widehat{\omega}^2 + 12\left(T-1\right)\right] \right\}$$

or

$$\widehat{\gamma}_2 = \frac{T\left(T+2\right)\widehat{q} - 3\widehat{\omega}^2 - 12\left(T-1\right)}{(\widehat{\omega}-1)^2 + 2T - 1 - \widehat{q}\sum_{t=1}^{T+1}\widehat{g}_{tt}^2}, \tag{67}$$

where

$$\widehat{q} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{r}_{it}^4}{\sum_{i=1}^{N}\left(\widehat{\mathbf{r}}_i' \widehat{\mathbf{\Omega}}^{-1} \widehat{\mathbf{r}}_i\right)^2}.$$

A.5 Derivation of R_y^2 and $R_{\Delta y}^2$

We derive R_y^2 for models (34) and (35) where homoskedasticity, $\sigma_i^2 = \sigma^2$ and $\sigma_{\varepsilon i}^2 = \sigma_\varepsilon^2$ for all i is assumed for simplicity. We also let $var(\alpha_i) = \sigma_\alpha^2$, $var(\mu_i) = \sigma_\mu^2$, and $cov(\alpha_i, \mu_i) = \sigma_{\alpha\mu}$. We assume that the process has been going for a long time (i.e., $m \to \infty$) as follows:

$$y_{it} = \frac{\alpha_i}{1-\gamma} + \beta \sum_{j=0}^{\infty} \gamma^j x_{i,t-j} + \sum_{j=0}^{\infty} \gamma^j u_{i,t-j},$$

and, in first differences,

$$\Delta y_{it} = \beta \sum_{j=0}^{\infty} \Delta x_{i,t-j} + \sum_{j=0}^{\infty} \Delta u_{i,t-j}.$$

The population value of R_y^2 is given by

$$R_y^2 = 1 - \frac{Var\left(y_{it}|x_{it}, x_{i,t-1}, ...\right)}{Var\left(y_{it}\right)}.$$

We have

$$\begin{aligned} Var\left(y_{it}|x_{it}, x_{i,t-1}, ...\right) &= Var\left(\frac{\alpha_i}{1-\gamma} + \sum_{j=0}^{\infty}\gamma^j u_{i,t-j}\right) = \frac{\sigma_\alpha^2}{(1-\gamma)^2} + Var\left(\sum_{j=0}^{\infty}\gamma^j u_{i,t-j}\right) \\ &= \frac{\sigma_\alpha^2}{(1-\gamma)^2} + \frac{\sigma^2}{1-\gamma^2} = \frac{1}{1-\gamma^2}\left[\sigma^2 + \frac{(1+\gamma)\sigma_\alpha^2}{1-\gamma}\right]. \end{aligned}$$

Further,

$$
\begin{aligned}
Var\left(y_{it}\right) &= \beta^2 Var\left(\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}\right) + Var\left(\frac{\alpha_i}{1-\gamma}+\sum_{j=0}^{\infty}\gamma^j u_{i,t-j}\right) + 2Cov\left(\beta\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}, \frac{\alpha_i}{1-\gamma}\right) \\
&= \beta^2 Var\left(\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}\right) + \left[\frac{\sigma_\alpha^2}{(1-\gamma)^2}+\frac{\sigma^2}{1-\gamma^2}\right] + 2Cov\left(\beta\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}, \frac{\alpha_i}{1-\gamma}\right).
\end{aligned}
$$

Using (35)-(37),

$$
Var\left(\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}\right) = Var\left[\sum_{j=0}^{\infty}\gamma^j\left(\mu_i + g\left(t-j\right)+\zeta_{i,t-j}\right)\right] = \frac{\sigma_\mu^2}{1-\gamma^2} + Var\left(\sum_{j=0}^{\infty}\gamma^j\zeta_{i,t-j}\right).
$$

Let

$$
w_{it} = \sum_{j=0}^{\infty}\gamma^j\zeta_{i,t-j} = \frac{1}{\left(1-\gamma L\right)\left(1-\phi L\right)}\varepsilon_{it} = \frac{1}{\left(1-\left(\gamma+\phi\right)L+\phi\gamma L^2\right)}\varepsilon_{it},
$$

Note that w_{it} 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 variance (Hamilton, 1994, p. 58)

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

It follows that

$$
Var\left(\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}\right) = \frac{\sigma_\mu^2}{1-\gamma^2} + \frac{\left(1+\phi\gamma\right)\sigma_\varepsilon^2}{\left(1-\gamma^2\right)\left(1-\phi^2\right)\left(1-\phi\gamma\right)}.
$$

Further,

$$
Cov\left(\beta\sum_{j=0}^{\infty}\gamma^j x_{i,t-j}, \frac{\alpha_i}{1-\gamma}\right) = \frac{\beta}{1-\gamma}E\left(\alpha_i\sum_{j=0}^{\infty}\gamma^j\mu_i\right) = \frac{\beta\sigma_{\alpha\mu}}{(1-\gamma)^2},
$$

and

$$
\begin{aligned}
Var\left(y_{it}\right) &= \beta^2\left[\frac{\sigma_\mu^2}{1-\gamma^2}+\frac{(1+\phi\gamma)\sigma_\varepsilon^2}{(1-\gamma^2)(1-\phi^2)(1-\phi\gamma)}\right] + \left[\frac{\sigma_\alpha^2}{(1-\gamma)^2}+\frac{\sigma^2}{1-\gamma^2}\right] + \frac{2\beta\sigma_{\alpha\mu}}{(1-\gamma)^2} \\
&= \frac{1}{1-\gamma^2}\left[\beta^2\sigma_\mu^2 + \sigma^2 + \frac{\beta^2(1+\phi\gamma)\sigma_\varepsilon^2}{(1-\phi^2)(1-\phi\gamma)} + \frac{(1+\gamma)(\sigma_\alpha^2+2\beta\sigma_{\alpha\mu})}{1-\gamma}\right].
\end{aligned}
$$

Using the above results, R_y^2 is given by

$$
R_y^2 = 1 - \frac{\sigma^2+\frac{(1+\gamma)\sigma_\alpha^2}{1-\gamma}}{\beta^2\sigma_\mu^2+\sigma^2+\frac{\beta^2(1+\phi\gamma)\sigma_\varepsilon^2}{(1-\phi^2)(1-\phi\gamma)}+\frac{(1+\gamma)(\sigma_\alpha^2+2\beta\sigma_{\alpha\mu})}{1-\gamma}}. \tag{68}
$$

Then, using $\sigma_\alpha^2 = \tau^2$, $\sigma_\mu^2 = \eta^2\tau^2$ and $\sigma_{\alpha\mu} = \eta\tau^2$, we have

$$R_y^2 = 1 - \frac{\sigma^2 + \frac{(1+\gamma)\tau^2}{1-\gamma}}{\beta^2\eta^2\tau^2 + \sigma^2 + \frac{\beta^2(1+\phi\gamma)\sigma_\varepsilon^2}{(1-\phi^2)(1-\phi\gamma)} + \frac{(1+\gamma)(\tau^2+2\beta\eta\tau^2)}{1-\gamma}},$$

or

$$\beta^2\tau^2\eta^2 + \frac{2\beta\tau^2(1+\gamma)}{1-\gamma}\eta + \sigma^2 + \frac{\beta^2(1+\phi\gamma)\sigma_\varepsilon^2}{(1-\phi^2)(1-\phi\gamma)} + \frac{(1+\gamma)\tau^2}{1-\gamma} - \frac{\sigma^2 + \left(\frac{1+\gamma}{1-\gamma}\right)\tau^2}{1-R_y^2} = 0. \qquad (69)$$

Note that (69) is a quadratic equations of η.

We now derive $R_{\Delta y}^2$. We have

$$\begin{aligned} Var\left(\Delta y_{it}|\Delta x_{it}, \Delta x_{i,t-1}, ...\right) &= Var\left(\sum_{j=0}^{\infty}\gamma^j \Delta u_{i,t-j}\right) = \frac{2\sigma^2}{1+\gamma}, \\ Var\left(\Delta y_{it}\right) &= \beta^2 Var\left(\sum_{j=0}^{\infty}\gamma^j \Delta x_{i,t-j}\right) + \frac{2\sigma^2}{1+\gamma}. \end{aligned}$$

Using result D.11 in Hsiao et al. (2002), where $\theta = 0$,

$$Var\left(\sum_{j=0}^{\infty}\gamma^j \Delta x_{i,t-j}\right) = \frac{2\sigma_\varepsilon^2}{(1+\gamma)(1+\phi)(1-\phi\gamma)},$$

and it follows that

$$R_{\Delta y}^2 = \frac{\beta^2\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma^2(1+\phi)(1-\phi\gamma)}. \qquad (70)$$

References

Ahn, S. C. and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 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, 598–606.

——— (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 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, 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, 277–297.

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

Bhargava, A. and J. D. Sargan (1983): "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 51, 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, 795–837.

Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 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, Amsterdam: JAI Press, vol. 15 of *Advances in Econometrics*, 53–91.

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, 1–37.

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, 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, 95–126.

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, 262–80.

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

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371–95.

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, 107–150.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66, 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, 1–9.

Kent, J. T. (1982): “Robust Properties of Likelihood Ratio Tests,” *Biometrika*, 69, 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, 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, Cambridge University Press, 282–318.

Kleibergen, F. (2005): “Testing Parameters in GMM Without Assuming that They Are Identified,” *Econometrica*, 73, 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, 447–464.

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

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

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

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

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

Stock, J. H. and J. Wright (2000): “GMM with Weak Identification,” *Econometrica*, 68, 1055–1096.

Ullah, A. (2004): *Finite Sample Econometrics*, Oxford University Press.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 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, 25–51.

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

$\gamma = 0.4$	median bias(×100)			MAE(×100)			median bias(×100)			MAE(×100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
	Transformed likelihood estimator											
50	-0.539	-0.183	-0.162	4.128	2.354	1.624	-0.367	-0.183	-0.153	4.259	2.354	1.625
150	-0.100	0.007	-0.119	2.456	1.336	1.101	-0.100	0.007	-0.119	2.456	1.336	1.101
500	0.014	-0.048	-0.050	1.272	0.729	0.554	0.014	-0.048	-0.050	1.272	0.729	0.554
	One-step first-difference GMM estimator based on "DIF2"											
50	-3.079	-2.202	–	5.239	3.479	–	-4.013	-3.054	–	6.118	3.947	–
150	-1.271	-0.742	-0.730	3.106	1.818	1.430	-1.378	-1.099	-1.030	3.425	2.126	1.702
500	-0.174	-0.231	-0.270	1.569	1.011	0.789	-0.214	-0.267	-0.317	1.798	1.194	0.948
	Two-step first-difference GMM estimator based on "DIF2"											
50	-2.812	-0.874	–	6.366	5.935	–	-3.579	-2.032	–	7.068	6.826	–
150	-0.867	-0.514	-0.486	3.183	2.130	1.811	-1.257	-0.819	-0.824	3.611	2.444	2.024
500	-0.196	-0.190	-0.257	1.609	1.057	0.863	-0.296	-0.267	-0.335	1.768	1.188	1.012
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	0.599	1.365	–	7.576	8.408	–	0.420	1.477	–	8.899	9.454	–
150	0.291	0.464	0.473	3.267	2.238	2.076	0.312	0.329	0.248	3.753	2.738	2.309
500	0.161	0.081	0.030	1.611	1.016	0.876	0.134	0.085	0.006	1.795	1.213	1.008
	One-step system GMM estimator based on "SYS2"											
50	1.218	–	–	4.647	–	–	29.721	–	–	29.721	–	–
150	0.545	0.814	0.766	2.809	1.851	1.512	19.854	20.312	19.851	19.854	20.312	19.851
500	0.366	0.275	0.156	1.527	0.961	0.750	9.322	9.227	9.077	9.322	9.227	9.077
	Two-step system GMM estimator based on "SYS2"											
50	1.331	–	–	5.815	–	–	28.124	–	–	28.124	–	–
150	0.440	0.490	0.553	2.760	2.133	2.037	12.942	14.353	14.263	12.942	14.353	14.263
500	0.226	0.171	0.091	1.311	0.998	0.848	2.654	3.156	3.073	2.833	3.156	3.073
	Continuous-updating system GMM estimator based on "SYS2"											
50	0.779	–	–	8.205	–	–	4.799	–	–	9.930	–	–
150	0.055	0.004	0.067	2.963	2.382	2.536	0.272	0.123	0.125	3.073	2.375	2.552
500	0.066	0.029	-0.056	1.316	0.982	0.871	0.095	0.046	-0.016	1.414	1.007	0.845

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, two-step and continuous-updating first-difference GMM estimators are computed by (21), (22) and (29) 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, two-step and continuous-updating system GMM estimators are computed by (27), (28) and (29) 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(×100) and MAE(×100) of γ ($\gamma = 0.9, \beta = 0.5$) for ARX(1) model

$\gamma = 0.9$	median bias(×100)			MAE(×100)			median bias(×100)			MAE(×100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
	Transformed likelihood estimator											
50	-0.344	-0.193	-0.113	5.419	2.286	1.349	-0.557	-0.217	-0.104	5.673	2.273	1.343
150	-0.086	-0.124	-0.122	3.281	1.280	0.843	-0.060	-0.102	-0.117	3.269	1.284	0.847
500	0.059	-0.021	0.011	1.535	0.716	0.436	0.056	-0.011	0.017	1.534	0.715	0.441
	One-step first-difference GMM estimator based on "DIF2"											
50	-4.990	-3.756	–	6.704	4.224	–	-5.207	-3.944	–	6.680	4.307	–
150	-1.746	-1.318	-1.268	3.642	2.045	1.671	-1.783	-1.380	-1.322	3.665	2.078	1.757
500	-0.293	-0.358	-0.408	1.789	1.069	0.875	-0.259	-0.321	-0.429	1.767	1.074	0.908
	Two-step first-difference GMM estimator based on "DIF2"											
50	-4.452	-2.831	–	7.516	6.164	–	-4.745	-3.172	–	7.214	6.542	–
150	-1.860	-1.178	-1.198	3.928	2.294	1.918	-1.973	-1.271	-1.290	3.945	2.243	2.016
500	-0.353	-0.409	-0.411	1.793	1.113	0.928	-0.344	-0.388	-0.371	1.719	1.115	0.968
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	0.086	-0.687	–	8.339	8.673	–	0.048	-1.226	–	8.313	8.744	–
150	0.023	0.155	-0.039	3.811	2.291	2.051	-0.028	0.153	-0.114	3.989	2.371	2.021
500	0.174	-0.017	-0.028	1.909	1.095	0.897	0.266	0.021	0.011	1.904	1.154	0.986
	One-step system GMM estimator based on "SYS2"											
50	4.841	–	–	4.931	–	–	7.238	–	–	7.238	–	–
150	3.672	3.598	3.519	3.732	3.598	3.519	7.068	7.094	7.054	7.068	7.094	7.054
500	1.983	1.830	1.723	2.139	1.854	1.723	6.476	6.459	6.459	6.476	6.459	6.459
	Two-step system GMM estimator based on "SYS2"											
50	5.158	–	–	5.380	–	–	7.285	–	–	7.285	–	–
150	3.804	3.664	3.408	4.007	3.685	3.415	7.190	7.146	7.148	7.190	7.146	7.148
500	1.873	1.678	1.473	2.184	1.733	1.497	6.560	6.484	6.459	6.560	6.484	6.459
	Continuous-updating system GMM estimator based on "SYS2"											
50	4.297	–	–	6.973	–	–	7.054	–	–	7.525	–	–
150	0.833	0.556	0.769	4.273	2.936	3.085	5.779	4.929	5.819	6.528	5.710	6.050
500	0.195	0.045	-0.019	1.586	1.037	0.852	0.991	0.160	0.093	2.473	1.173	0.922

Note: See notes to Table 1.

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

	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	6.4	5.7	4.6	42.4	83.7	97.2	6.4	5.7	4.6	42.8	83.7	97.6
150	4.1	5.2	4.1	78.6	99.9	100.0	4.1	5.2	4.1	78.6	99.9	100.0
500	3.8	6.3	5.2	99.9	100.0	100.0	3.8	6.3	5.2	99.9	100.0	100.0
	One-step first-difference GMM estimator based on "DIF2"											
50	8.6	8.9	–	45.3	78.7	–	10.4	9.8	–	42.6	75.2	–
150	6.6	5.8	5.6	72.8	98.2	100.0	6.9	5.9	7.4	65.2	95.7	99.9
500	4.4	4.6	5.6	99.3	100.0	100.0	4.7	5.9	7.2	97.6	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2"											
50	30.4	75.9	–	62.7	88.5	–	30.8	77.7	–	61.5	88.3	–
150	13.1	20.9	32.7	77.0	98.2	100.0	13.8	20.9	33.6	70.3	96.6	99.6
500	6.6	9.2	11.6	99.3	100.0	100.0	7.0	9.1	13.8	97.6	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors											
50	8.4	0.7	–	26.3	2.5	–	7.3	0.7	–	25.4	2.8	–
150	5.7	5.1	2.9	65.6	91.9	96.2	6.6	5.4	3.9	57.6	84.3	91.2
500	4.9	5.3	6.1	98.5	100.0	100.0	5.4	5.4	6.2	96.9	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	36.7	83.1	–	53.1	88.4	–	40.1	82.4	–	53.1	86.4	–
150	11.3	25.5	40.3	68.6	95.9	99.5	12.4	25.5	39.2	60.3	90.6	97.4
500	7.4	9.4	11.9	98.3	100.0	100.0	6.9	8.7	13.2	96.3	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors											
50	45.3	33.6	–	62.3	40.9	–	45.1	37.3	–	56.6	42.3	–
150	12.0	42.3	73.2	67.2	97.6	99.9	12.5	41.4	68.3	58.0	94.5	98.7
500	6.9	10.6	17.0	98.1	100.0	100.0	6.7	10.1	17.8	96.0	100.0	100.0
	One-step system GMM estimator based on "SYS2"											
50	8.7	–	–	24.2	–	–	89.6	–	–	72.1	–	–
150	6.2	6.2	6.5	62.6	92.5	99.3	78.0	97.1	99.6	41.1	57.2	66.4
500	4.3	4.8	6.3	99.1	100.0	100.0	53.8	87.4	97.5	9.1	12.3	13.2
	Two-step system GMM estimator based on "SYS2"											
50	45.8	–	–	64.4	–	–	96.5	–	–	89.3	–	–
150	16.3	33.7	52.5	80.2	98.1	99.8	80.8	97.1	98.9	59.2	75.5	84.7
500	7.5	11.2	15.2	99.8	100.0	100.0	39.6	65.8	78.9	83.6	95.5	99.2
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors											
50	3.5	–	–	10.4	–	–	38.6	–	–	22.7	–	–
150	4.8	3.6	0.5	62.6	77.5	55.7	42.9	71.1	58.3	15.6	21.5	18.9
500	5.1	5.6	4.7	99.7	100.0	100.0	19.6	40.3	47.8	67.4	84.1	94.0
	Continuous-updating system GMM estimator based on "SYS2"											
50	58.6	–	–	71.1	–	–	75.7	–	–	80.6	–	–
150	17.9	38.5	62.6	80.7	97.5	99.0	28.1	50.9	71.8	84.5	98.3	99.2
500	7.7	10.8	16.1	99.9	100.0	100.0	11.4	14.6	21.4	99.9	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors											
50	53.9	–	–	63.4	–	–	36.0	–	–	43.5	–	–
150	19.4	62.5	42.8	81.7	98.9	91.3	18.8	33.5	14.4	78.5	92.4	75.5
500	8.1	14.4	23.6	99.9	100.0	100.0	7.7	12.8	23.0	99.5	100.0	100.0

Note: For the definition of "DIF2" and "SYS2", see notes to Table 1. "NW" denotes Newey and Windmeijer's(2009) standard errors.

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

	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	5.1	5.4	5.6	33.2	79.5	98.1	4.4	5.5	5.7	32.9	80.0	98.3
150	4.4	6.0	5.5	56.8	99.6	100.0	4.5	6.0	5.7	57.1	99.6	100.0
500	4.9	5.5	5.4	95.9	100.0	100.0	4.9	5.3	5.6	95.9	100.0	100.0
	One-step first-difference GMM estimator based on "DIF2"											
50	12.2	12.7	–	50.3	86.3	–	11.7	13.5	–	49.8	85.9	–
150	8.3	8.0	9.9	67.7	97.8	100.0	8.3	7.6	10.9	67.6	97.1	100.0
500	5.7	7.1	8.4	96.6	100.0	100.0	5.5	7.6	8.2	96.6	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2"											
50	32.9	77.4	–	65.6	91.9	–	30.9	77.2	–	65.5	91.8	–
150	14.5	22.3	34.5	73.0	98.2	99.9	14.9	23.4	35.9	72.1	98.1	99.9
500	7.2	10.1	14.5	96.9	100.0	100.0	7.4	9.7	15.1	96.7	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors											
50	7.3	1.2	–	31.0	3.7	–	7.5	1.2	–	29.2	3.8	–
150	7.9	5.3	4.1	61.9	92.5	96.7	6.9	6.0	5.2	61.5	92.0	95.6
500	5.7	6.6	7.8	96.3	100.0	100.0	5.9	6.5	8.9	96.4	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	38.0	83.3	–	53.4	88.5	–	37.5	83.0	–	51.6	88.3	–
150	13.2	22.7	37.2	61.3	94.5	99.0	13.9	23.9	36.7	61.4	93.6	98.2
500	7.3	9.1	13.6	95.8	100.0	100.0	7.2	9.4	14.5	95.8	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors											
50	42.9	34.5	–	57.5	46.8	–	42.2	33.7	–	55.5	46.8	–
150	13.2	38.5	70.4	60.1	96.3	99.7	13.2	38.9	66.8	60.3	96.1	99.4
500	6.7	10.7	18.0	95.6	100.0	100.0	7.0	11.1	18.3	95.5	100.0	100.0
	One-step system GMM estimator based on "SYS2"											
50	40.4	–	–	42.5	–	–	99.6	–	–	92.2	–	–
150	30.5	53.7	66.5	72.1	98.1	99.9	99.4	100.0	100.0	95.1	100.0	100.0
500	21.6	34.8	41.9	99.4	100.0	100.0	98.7	100.0	100.0	99.2	100.0	100.0
	Two-step system GMM estimator based on "SYS2"											
50	78.7	–	–	81.1	–	–	100.0	–	–	98.5	–	–
150	60.3	80.3	89.0	88.2	99.5	100.0	99.8	100.0	100.0	97.6	100.0	100.0
500	35.2	48.7	56.4	99.7	100.0	100.0	99.8	100.0	100.0	99.7	100.0	100.0
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors											
50	18.7	–	–	10.9	–	–	73.7	–	–	42.4	–	–
150	25.9	21.0	8.9	41.8	61.0	37.7	78.7	91.1	91.6	54.7	66.4	66.0
500	15.7	17.6	19.9	94.1	100.0	100.0	69.5	95.3	99.5	45.2	51.5	69.2
	Continuous-updating system GMM estimator based on "SYS2"											
50	81.8	–	–	82.2	–	–	97.9	–	–	94.2	–	–
150	57.6	68.4	81.2	89.2	96.8	98.2	89.5	93.7	96.4	90.9	96.4	97.5
500	23.5	25.3	28.1	99.6	100.0	100.0	66.8	67.5	70.4	95.8	99.9	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors											
50	55.4	–	–	57.0	–	–	59.0	–	–	48.6	–	–
150	39.3	43.5	39.6	78.1	89.5	86.5	59.2	49.1	51.3	65.3	50.0	44.5
500	13.2	14.2	22.6	97.3	99.8	100.0	29.5	11.7	9.8	89.3	98.7	99.6

Note: See notes to Table 3.

Table 5: Median bias(×100) and MAE(×100) of β ($\gamma = 0.4, \beta = 0.5$) for ARX(1) model

$\beta = 0.5$	median bias(×100)			MAE(×100)			median bias(×100)			MAE(×100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
	Transformed likelihood estimator											
50	0.082	-0.066	-0.026	3.187	1.865	1.478	0.052	-0.066	-0.026	3.141	1.865	1.478
150	-0.026	-0.004	0.063	1.803	1.021	0.838	-0.026	-0.004	0.063	1.803	1.021	0.838
500	0.000	-0.002	0.031	0.878	0.567	0.462	0.000	-0.002	0.031	0.878	0.567	0.462
	One-step first-difference GMM estimator based on "DIF2"											
50	-0.034	0.105	–	3.460	2.167	–	-0.043	-0.280	–	3.613	2.163	–
150	-0.116	0.148	0.140	2.050	1.379	1.029	-0.187	0.010	-0.012	2.091	1.410	1.070
500	-0.081	-0.011	0.050	1.011	0.714	0.556	-0.044	-0.092	-0.022	1.069	0.766	0.592
	Two-step first-difference GMM estimator based on "DIF2"											
50	-0.188	0.045	–	3.937	4.220	–	-0.088	-0.625	–	4.124	4.627	–
150	-0.100	-0.062	-0.028	2.225	1.531	1.382	-0.137	-0.143	-0.150	2.248	1.649	1.390
500	-0.137	-0.069	-0.015	1.050	0.714	0.601	-0.082	-0.139	-0.036	1.075	0.777	0.638
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	-0.238	0.090	–	4.936	6.625	–	0.525	0.139	–	5.135	6.629	–
150	-0.099	-0.019	-0.061	2.247	1.663	1.545	0.068	-0.021	-0.053	2.366	1.801	1.589
500	-0.132	-0.065	-0.019	1.065	0.724	0.615	-0.046	-0.072	0.005	1.083	0.803	0.648
	One-step system GMM estimator based on "SYS2"											
50	0.446	–	–	3.640	–	–	6.885	–	–	7.149	–	–
150	0.117	0.216	0.269	2.119	1.385	1.075	4.682	5.077	5.058	4.824	5.077	5.058
500	0.062	0.029	0.079	1.146	0.735	0.603	2.249	2.403	2.410	2.698	2.408	2.416
	Two-step system GMM estimator based on "SYS2"											
50	0.243	–	–	4.381	–	–	4.933	–	–	6.683	–	–
150	0.103	-0.193	-0.115	2.128	1.543	1.447	2.557	2.223	2.254	3.372	2.524	2.543
500	-0.089	-0.065	-0.009	1.015	0.732	0.600	0.420	0.376	0.395	1.189	0.869	0.684
	Continuous-updating system GMM estimator based on "SYS2"											
50	0.009	–	–	5.858	–	–	1.713	–	–	6.721	–	–
150	-0.058	-0.203	-0.234	2.250	1.677	1.832	0.031	-0.165	-0.091	2.365	1.706	1.861
500	-0.121	-0.103	-0.038	1.031	0.759	0.611	-0.119	-0.071	-0.031	1.033	0.774	0.597

Note: See notes to Table 1.

Table 6: Median bias(×100) and MAE(×100) of β ($\gamma = 0.9, \beta = 0.5$) for ARX(1) model

$\beta = 0.5$	median bias(×100)			MAE(×100)			median bias(×100)			MAE(×100)		
	$\tau = 1$						$\tau = 5$					
N/T	5	10	15	5	10	15	5	10	15	5	10	15
	Transformed likelihood estimator											
50	-0.019	-0.013	0.013	3.939	2.163	1.715	0.078	-0.015	0.016	3.838	2.169	1.718
150	-0.046	0.016	0.008	2.241	1.185	0.954	-0.038	0.017	0.007	2.231	1.183	0.953
500	0.014	0.026	0.028	1.125	0.670	0.559	0.010	0.025	0.028	1.122	0.671	0.559
	One-step first-difference GMM estimator based on "DIF2"											
50	-1.365	-1.263	–	4.769	2.869	–	-1.177	-1.278	–	4.522	2.856	–
150	-0.533	-0.428	-0.294	2.704	1.783	1.345	-0.528	-0.392	-0.406	2.790	1.801	1.331
500	-0.174	-0.214	-0.084	1.393	0.999	0.752	-0.162	-0.268	-0.120	1.389	0.959	0.743
	Two-step first-difference GMM estimator based on "DIF2"											
50	-0.763	-0.743	–	5.639	5.598	–	-0.832	-0.786	–	5.272	5.683	–
150	-0.468	-0.374	-0.395	2.931	2.002	1.732	-0.383	-0.336	-0.493	2.903	2.110	1.756
500	-0.187	-0.206	-0.074	1.410	1.013	0.771	-0.143	-0.221	-0.094	1.388	0.993	0.799
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	0.899	-0.019	–	6.866	8.450	–	0.686	-0.031	–	6.839	8.998	–
150	-0.030	-0.004	0.033	3.039	2.140	2.005	0.109	0.118	-0.040	3.105	2.299	2.024
500	-0.038	-0.056	0.063	1.413	1.005	0.778	0.005	-0.065	0.026	1.397	0.981	0.806
	One-step system GMM estimator based on "SYS2"											
50	1.375	–	–	4.260	–	–	2.511	–	–	4.484	–	–
150	1.056	1.227	1.179	2.444	1.835	1.521	2.211	2.552	2.486	2.737	2.591	2.495
500	0.558	0.647	0.668	1.356	0.993	0.874	2.226	2.383	2.422	2.242	2.383	2.422
	Two-step system GMM estimator based on "SYS2"											
50	1.561	–	–	5.070	–	–	2.321	–	–	5.182	–	–
150	0.708	0.535	0.453	2.425	1.880	1.805	1.216	1.175	0.986	2.694	2.039	1.804
500	0.315	0.292	0.280	1.219	0.938	0.749	1.482	1.452	1.376	1.704	1.487	1.398
	Continuous-updating system GMM estimator based on "SYS2"											
50	1.414	–	–	6.836	–	–	1.961	–	–	6.725	–	–
150	0.037	0.020	0.190	3.008	2.192	2.154	1.095	0.801	0.848	3.161	2.333	2.390
500	-0.067	-0.066	-0.021	1.312	0.933	0.704	0.448	0.120	0.161	1.565	1.054	0.767

Note: See notes to Table 1.

Table 7: Size(%) and power(%) of β ($\gamma = 0.4, \beta = 0.5$) for ARX(1) model

	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	7.8	5.9	6.3	63.9	96.3	99.7	8.3	5.9	6.0	63.9	96.3	99.7
150	5.9	5.2	5.3	97.3	100.0	100.0	5.9	5.2	5.3	97.3	100.0	100.0
500	4.6	6.5	5.3	100.0	100.0	100.0	4.6	6.5	5.3	100.0	100.0	100.0
	One-step first-difference GMM estimator based on "DIF2"											
50	7.4	5.3	–	55.5	87.8	–	8.2	6.4	–	54.7	88.0	–
150	6.5	6.4	6.6	92.6	99.9	100.0	6.0	6.1	6.9	92.5	99.9	100.0
500	4.6	5.0	4.7	100.0	100.0	100.0	5.2	5.7	4.5	99.9	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2"											
50	26.7	75.7	–	72.3	93.5	–	27.3	78.3	–	71.2	93.5	–
150	13.5	20.4	32.0	94.2	100.0	100.0	13.8	20.7	32.9	92.6	99.9	100.0
500	6.7	8.9	9.6	100.0	100.0	100.0	6.6	8.8	10.7	100.0	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors											
50	5.9	1.1	–	39.4	5.0	–	6.7	0.9	–	38.3	3.4	–
150	7.1	5.1	3.2	89.0	99.5	99.8	7.5	6.0	3.9	87.9	98.8	99.3
500	5.5	5.1	4.6	100.0	100.0	100.0	5.5	5.7	4.7	100.0	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	35.3	83.4	–	68.6	90.9	–	38.2	83.8	–	65.3	91.8	–
150	15.6	24.3	38.7	92.4	100.0	100.0	15.4	24.4	39.0	90.7	99.6	100.0
500	6.6	9.5	10.4	100.0	100.0	100.0	6.9	9.2	11.3	100.0	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors											
50	45.7	40.2	–	72.6	56.5	–	44.5	43.7	–	69.3	60.3	–
150	14.9	40.3	74.0	92.1	100.0	100.0	14.9	40.1	71.2	89.7	99.9	100.0
500	6.5	10.6	15.0	100.0	100.0	100.0	6.4	10.2	14.8	100.0	100.0	100.0
	One-step system GMM estimator based on "SYS2"											
50	7.2	–	–	47.9	–	–	25.8	–	–	11.5	–	–
150	5.5	5.7	5.8	89.6	99.8	100.0	21.0	48.7	67.1	22.7	46.0	65.5
500	4.5	5.2	5.7	100.0	100.0	100.0	13.0	28.7	48.4	72.9	98.3	100.0
	Two-step system GMM estimator based on "SYS2"											
50	42.0	–	–	76.1	–	–	57.9	–	–	56.3	–	–
150	15.3	27.1	49.1	95.9	99.9	100.0	31.2	49.0	70.2	73.6	93.5	97.8
500	7.5	10.6	14.1	100.0	100.0	100.0	10.8	15.8	22.7	99.8	100.0	100.0
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors											
50	3.0	–	–	24.3	–	–	3.0	–	–	4.5	–	–
150	7.1	3.9	0.7	87.8	98.0	88.7	10.1	11.2	4.1	47.9	62.5	37.2
500	5.0	5.3	5.4	100.0	100.0	100.0	6.4	7.4	7.0	99.7	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2"											
50	55.3	–	–	77.6	–	–	61.6	–	–	75.4	–	–
150	18.3	34.6	60.4	94.9	99.8	100.0	21.5	36.7	64.4	93.4	99.8	99.9
500	8.1	11.0	15.2	100.0	100.0	100.0	9.3	11.0	15.3	100.0	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors											
50	58.1	–	–	74.5	–	–	42.4	–	–	53.5	–	–
150	21.7	62.6	42.8	94.4	100.0	98.5	21.8	50.3	30.4	93.6	99.7	94.8
500	7.8	13.9	24.6	100.0	100.0	100.0	8.7	13.6	23.1	100.0	100.0	100.0

Note: See notes to Table 3.

Table 8: Size(%) and power(%) of β ($\gamma = 0.9, \beta = 0.5$) for ARX(1) model

	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	7.9	5.8	5.8	47.9	87.8	98.1	7.8	5.8	5.8	48.5	87.8	98.1
150	5.9	5.2	5.6	86.7	100.0	100.0	6.1	5.3	5.7	86.6	100.0	100.0
500	4.1	5.4	5.7	100.0	100.0	100.0	4.1	5.4	5.7	100.0	100.0	100.0
	One-step first-difference GMM estimator based on "DIF2"											
50	8.0	7.2	–	46.5	76.9	–	7.9	7.5	–	46.4	77.2	–
150	6.5	7.3	7.1	79.4	98.6	99.9	6.4	6.9	7.3	78.6	98.2	99.9
500	5.4	5.2	4.7	99.8	100.0	100.0	5.4	5.2	5.2	99.8	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2"											
50	29.9	76.9	–	61.9	91.6	–	29.7	78.7	–	62.1	90.7	–
150	14.1	19.9	32.2	82.5	98.7	99.7	14.4	20.3	31.8	82.4	98.4	99.7
500	6.4	7.5	10.3	99.9	100.0	100.0	6.4	8.4	12.1	99.8	100.0	100.0
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors											
50	7.1	0.7	–	31.3	3.2	–	6.8	0.9	–	30.1	3.0	–
150	8.3	5.8	3.2	72.6	94.1	96.7	8.0	5.7	3.0	73.3	93.3	95.4
500	5.8	5.4	5.2	99.7	100.0	100.0	5.6	5.3	5.6	99.7	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2"											
50	39.2	84.9	–	55.4	89.2	–	38.6	85.2	–	55.5	87.4	–
150	15.2	23.3	38.5	77.3	96.9	98.9	15.5	24.5	38.9	77.3	96.7	98.9
500	6.5	8.8	10.3	99.8	100.0	100.0	6.6	9.1	12.2	99.8	100.0	100.0
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors											
50	48.3	42.0	–	59.9	53.9	–	46.0	38.5	–	62.5	51.5	–
150	14.7	39.6	70.1	76.6	98.3	99.4	15.0	39.0	69.3	77.0	98.5	99.4
500	6.4	9.9	14.5	99.7	100.0	100.0	6.5	10.2	16.4	99.7	100.0	100.0
	One-step system GMM estimator based on "SYS2"											
50	8.5	–	–	35.8	–	–	9.8	–	–	30.0	–	–
150	6.7	9.0	9.8	76.7	97.9	100.0	11.1	20.0	28.5	64.1	93.5	99.5
500	4.8	8.7	10.1	100.0	100.0	100.0	20.9	46.9	67.2	98.7	100.0	100.0
	Two-step system GMM estimator based on "SYS2"											
50	43.0	–	–	65.8	–	–	45.6	–	–	63.5	–	–
150	16.9	32.7	53.4	87.6	99.6	99.7	20.5	35.9	54.3	81.6	99.0	99.6
500	6.9	11.2	15.1	100.0	100.0	100.0	17.3	32.1	45.2	99.8	100.0	100.0
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors											
50	2.7	–	–	13.1	–	–	2.3	–	–	8.5	–	–
150	5.3	3.1	0.6	71.8	88.9	69.1	5.3	3.6	1.6	52.1	66.8	42.5
500	4.6	5.9	4.8	100.0	100.0	100.0	7.9	12.8	15.0	95.6	99.8	100.0
	Continuous-updating system GMM estimator based on "SYS2"											
50	57.3	–	–	68.3	–	–	57.2	–	–	68.7	–	–
150	22.0	38.3	58.6	87.3	99.3	99.2	26.3	42.4	63.4	80.3	96.5	98.5
500	7.5	11.2	15.2	100.0	100.0	100.0	14.0	15.4	19.6	99.6	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors											
50	52.8	–	–	62.5	–	–	38.3	–	–	48.5	–	–
150	22.0	58.9	37.7	85.7	98.8	94.9	21.5	39.5	23.6	73.3	85.8	74.4
500	6.4	13.0	22.4	99.8	100.0	100.0	10.5	13.7	19.6	99.2	100.0	100.0

Note: See notes to Table 3.

Table 9: Size(%) and power(%) of weak instruments robust tests ($\boldsymbol{\theta} = (0.4, 0.5)'$) for ARX(1) model

	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 "DIF2"											
50	51.7	100.0	–	61.8	100.0	–	51.6	100.0	–	56.7	100.0	–
150	14.2	65.1	99.2	43.2	98.8	100.0	13.8	67.0	99.1	28.6	90.5	100.0
500	6.6	14.9	34.5	88.4	100.0	100.0	7.0	14.2	34.1	56.1	98.0	99.9
	Anderson and Rubin test based on "SYS2"											
50	84.5	–	–	93.2	–	–	86.7	–	–	89.3	–	–
150	24.5	94.6	100.0	74.3	100.0	100.0	24.6	94.9	100.0	51.0	99.9	100.0
500	9.4	26.9	60.8	99.2	100.0	100.0	8.6	26.2	60.8	80.6	100.0	100.0
	Lagrange multiplier test based on "DIF2"											
50	33.2	98.9	–	46.9	99.6	–	33.2	99.2	–	41.5	99.2	–
150	8.9	29.2	64.5	54.2	72.2	99.8	8.9	27.6	67.3	29.1	53.2	97.8
500	6.4	8.7	11.4	98.8	100.0	100.0	5.7	9.1	12.5	83.6	100.0	100.0
	Lagrange multiplier test based on "SYS2"											
50	54.1	–	–	69.8	–	–	54.9	–	–	57.5	–	–
150	11.7	42.2	78.2	75.9	95.5	100.0	12.8	42.9	79.3	39.4	78.8	98.6
500	5.4	12.1	15.8	100.0	100.0	100.0	5.7	11.5	16.2	93.0	99.1	87.9
	Conditional likelihood ratio test based on "DIF2"											
50	44.3	98.9	–	57.8	99.6	–	44.4	99.2	–	50.8	99.2	–
150	9.4	33.4	68.2	55.5	82.6	99.9	9.5	34.3	71.5	31.9	67.2	98.1
500	6.3	8.6	11.3	98.8	100.0	100.0	5.9	8.8	13.1	84.0	100.0	100.0
	Conditional likelihood ratio test based on "SYS2"											
50	57.6	–	–	72.7	–	–	57.2	–	–	59.6	–	–
150	12.0	48.5	78.4	78.4	96.9	100.0	13.9	45.3	79.4	41.5	80.1	98.7
500	5.3	12.1	15.8	100.0	100.0	100.0	5.7	11.0	16.5	93.2	99.3	90.3

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. (31)). "Lagrange multiplier test" denotes Kleibergen's(2005) LM test (eq. (32)). "Conditional likelihood ratio test" denotes the conditional likelihood ratio test of Moreira (2003)(extended by Kleibergen(2005)) (eq.(33)). "–" denotes the cases where the GMM estimators are not computed since the number of moment conditions exceeds the sample size.

Table 10: Size(%) and power(%) of weak instruments robust tests ($\boldsymbol{\theta} = (0.9, 0.5)'$) for ARX(1) model

	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 "DIF2"											
50	50.6	100.0	–	49.6	100.0	–	50.0	100.0	–	49.6	100.0	–
150	14.6	69.3	99.1	15.5	72.7	99.5	15.5	68.6	99.5	15.6	70.5	99.4
500	7.2	14.7	35.1	10.7	31.1	73.8	7.6	14.1	34.2	8.7	18.0	42.0
	Anderson and Rubin test based on "SYS2"											
50	84.5	–	–	90.3	–	–	87.7	–	–	90.0	–	–
150	25.6	94.6	100.0	72.7	100.0	100.0	25.8	94.4	100.0	43.9	98.6	100.0
500	10.3	25.0	60.5	99.1	100.0	100.0	11.6	26.5	62.1	59.8	94.7	99.8
	Lagrange Multiplier test based on "DIF2"											
50	40.1	99.3	–	49.3	99.1	–	42.3	99.1	–	48.5	99.3	–
150	9.5	37.5	78.2	10.2	48.4	94.9	10.3	36.8	77.9	11.3	53.1	91.5
500	6.0	9.5	11.0	10.4	42.1	73.9	5.4	9.0	13.7	8.7	16.0	24.1
	Lagrange Multiplier test based on "SYS2"											
50	58.3	–	–	72.4	–	–	56.9	–	–	65.8	–	–
150	14.7	46.4	79.1	59.7	95.9	99.6	17.4	44.2	77.6	45.1	71.6	92.6
500	5.8	13.3	18.1	98.6	100.0	99.0	7.5	11.9	16.5	78.7	94.7	97.8
	Conditional likelihood ratio test based on "DIF2"											
50	49.1	99.3	–	54.8	99.1	–	49.2	99.1	–	54.3	99.3	–
150	14.0	51.3	82.8	15.7	66.1	96.2	16.1	54.0	83.1	16.8	68.3	93.2
500	6.4	10.5	12.1	12.7	46.2	80.2	7.5	11.7	18.9	10.3	22.2	38.2
	Conditional likelihood ratio test based on "SYS2"											
50	61.0	–	–	75.4	–	–	58.6	–	–	68.7	–	–
150	15.8	52.5	79.3	62.0	96.7	99.5	17.8	45.1	77.8	46.0	75.5	92.5
500	5.8	13.3	18.1	98.7	100.0	99.6	7.8	12.3	17.0	78.9	95.0	97.9

Note: See notes to Table 9.

Table 11: Median bias(×100) and MAE(×100) of γ ($\gamma = 0.4$) for AR(1) model

$\gamma = 0.4$	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	-0.619	-0.139	-0.155	0.079	7.375	3.875	2.939	2.254	-0.619	-0.139	-0.155	0.079	7.374	3.875	2.939	2.254
150	-0.361	-0.165	0.051	-0.090	4.414	2.048	1.687	1.285	-0.360	-0.165	0.051	-0.090	4.414	2.049	1.687	1.285
500	0.094	-0.011	-0.006	-0.011	2.264	1.184	0.898	0.720	0.093	-0.011	-0.006	-0.011	2.264	1.184	0.898	0.720
	One-step first-difference GMM estimator based on "DIF2"															
50	-6.284	-2.997	-2.518	-1.689	11.586	5.762	4.566	3.614	-17.257	-11.002	-7.401	-5.357	22.769	12.744	8.435	6.761
150	-2.408	-1.354	-0.787	-0.744	6.588	3.149	2.273	2.019	-6.799	-5.226	-3.398	-2.255	12.391	7.630	5.063	3.563
500	-0.755	-0.238	-0.342	-0.321	3.342	1.831	1.309	1.100	-2.357	-1.225	-1.291	-0.892	6.764	3.570	2.599	2.024
	Two-step first-difference GMM estimator based on "DIF2"															
50	-5.976	-2.194	-1.629	-0.519	12.144	6.741	5.397	5.806	-17.765	-10.405	-5.854	-3.818	24.138	13.280	9.538	8.743
150	-2.021	-1.079	-0.664	-0.356	6.666	3.424	2.566	2.278	-7.078	-3.966	-2.404	-1.330	12.764	7.138	4.936	3.472
500	-0.742	-0.174	-0.216	-0.162	3.414	1.772	1.227	1.151	-2.438	-1.508	-1.021	-0.560	6.958	3.783	2.417	2.018
	Continuous-updating first-difference GMM estimator based on "DIF2"															
50	0.247	1.741	1.643	2.456	12.857	7.552	6.338	7.872	-0.639	0.889	1.136	2.014	28.519	14.958	10.879	10.859
150	0.271	0.232	0.475	0.740	7.001	3.280	2.647	2.365	-0.814	0.844	0.880	0.844	13.092	7.270	4.719	3.455
500	-0.011	0.223	0.102	0.172	3.387	1.824	1.241	1.185	0.434	0.141	0.066	0.253	6.880	3.598	2.355	2.017
	One-step system GMM estimator based on "SYS2"															
50	3.114	2.511	2.683	–	8.459	5.355	4.309	–	41.534	40.665	40.834	–	41.567	40.665	40.834	–
150	0.527	0.825	1.009	0.933	4.820	2.782	2.464	2.039	27.032	26.997	27.378	26.370	27.048	26.997	27.378	26.370
500	0.499	0.393	0.223	0.249	2.615	1.703	1.256	1.055	13.733	13.585	13.040	12.626	14.145	13.585	13.040	12.626
	Two-step system GMM estimator based on "SYS2"															
50	2.054	1.938	1.458	–	8.152	6.670	6.420	–	38.281	36.411	36.955	–	38.346	36.411	36.955	–
150	0.138	0.184	0.217	0.267	4.518	2.820	2.573	2.256	20.238	19.530	20.456	19.588	20.238	19.530	20.456	19.588
500	0.172	0.135	0.079	-0.016	2.248	1.576	1.166	1.034	5.447	5.700	5.595	5.521	6.003	5.770	5.595	5.521
	Continuous-updating system GMM estimator based on "SYS2"															
50	0.171	0.012	0.074	–	9.562	8.114	10.749	–	4.030	1.588	6.424	–	11.641	9.260	17.122	–
150	-0.237	-0.307	-0.214	-0.252	4.650	2.953	2.682	2.438	0.022	-0.393	-0.299	-0.143	5.023	2.908	2.681	2.381
500	-0.014	0.025	-0.125	-0.156	2.224	1.554	1.168	1.055	-0.054	0.023	-0.103	-0.121	2.362	1.564	1.208	1.048

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, two-step and continuous-updating first-difference GMM estimators are computed by (21), (22) and (29) 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, two-step and continuous-updating system GMM estimators are computed by (27), (28) and (29) 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 12: Median bias(×100) and MAE(×100) of γ ($\gamma = 0.9$) for AR(1) model

$\gamma = 0.9$	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	-4.488	-0.657	-0.274	-0.218	8.311	4.768	3.350	2.552	-4.506	-0.653	-0.275	-0.220	8.275	4.767	3.352	2.549
150	-1.246	-0.433	-0.168	-0.095	5.868	3.470	2.127	1.530	-1.264	-0.435	-0.168	-0.094	5.875	3.467	2.124	1.529
500	-0.247	0.046	0.058	0.044	4.714	2.340	1.251	0.831	-0.241	0.039	0.057	0.044	4.719	2.339	1.250	0.830
	One-step first-difference GMM estimator based on "DIF2"															
50	-59.932	-33.381	-21.769	-14.764	60.190	33.426	21.769	14.776	-65.887	-47.540	-36.902	-30.315	66.670	47.540	36.902	30.315
150	-39.709	-18.212	-10.582	-6.200	41.184	18.479	10.903	6.441	-60.174	-43.735	-33.156	-26.312	60.435	43.892	33.156	26.312
500	-21.088	-7.396	-3.903	-2.362	23.081	8.448	4.696	3.135	-46.847	-34.896	-23.210	-19.520	47.260	34.896	23.210	19.520
	Two-step first-difference GMM estimator based on "DIF2"															
50	-65.532	-37.165	-23.820	-16.218	66.340	37.321	23.905	17.157	-75.593	-59.001	-47.123	-38.379	76.025	59.001	47.123	38.448
150	-42.734	-20.413	-10.636	-6.009	43.817	20.792	11.194	6.514	-66.991	-53.417	-44.130	-35.875	67.567	53.417	44.130	35.875
500	-22.014	-7.326	-3.646	-2.180	24.012	8.859	4.510	3.077	-49.659	-41.006	-31.070	-26.926	49.842	41.006	31.070	26.926
	Continuous-updating first-difference GMM estimator based on "DIF2"															
50	-35.511	-9.495	-5.778	-5.482	83.591	42.818	28.368	26.071	-60.452	-43.158	-37.088	-24.588	102.765	86.448	79.294	68.952
150	-10.682	-1.212	-0.538	-0.296	41.350	14.780	7.070	4.561	-27.787	-14.913	-3.598	-1.313	79.975	68.023	57.868	50.278
500	-1.061	-0.090	-0.351	-0.073	19.699	7.013	3.684	2.626	-9.838	-0.486	1.390	1.500	51.871	35.635	20.615	15.745
	One-step system GMM estimator based on "SYS2"															
50	7.300	6.999	7.145	–	7.675	7.070	7.154	–	9.893	9.891	9.873	–	9.893	9.891	9.873	–
150	5.916	5.584	5.729	5.619	6.354	5.607	5.741	5.619	9.820	9.733	9.749	9.769	9.820	9.733	9.749	9.769
500	3.487	3.639	3.366	3.327	4.180	3.737	3.371	3.330	9.350	9.398	9.390	9.437	9.350	9.398	9.390	9.437
	Two-step system GMM estimator based on "SYS2"															
50	6.365	6.155	6.449	–	7.632	6.524	6.670	–	9.833	9.800	9.816	–	9.833	9.800	9.816	–
150	4.605	4.121	4.134	4.219	5.946	4.589	4.282	4.239	9.653	9.555	9.632	9.593	9.659	9.555	9.632	9.593
500	2.026	1.943	1.656	1.730	3.513	2.527	1.926	1.948	8.820	8.876	8.988	8.982	8.820	8.876	8.988	8.982
	Continuous-updating system GMM estimator based on "SYS2"															
50	4.604	2.964	2.649	–	11.663	9.554	9.915	–	9.572	9.134	9.309	–	11.582	10.931	11.466	–
150	1.693	0.436	0.032	-0.493	6.771	4.115	3.461	3.250	8.912	6.034	3.156	1.788	10.592	7.543	5.926	4.806
500	0.266	0.158	0.009	-0.065	3.802	2.157	1.541	1.288	6.418	3.092	0.873	0.349	7.507	4.607	2.401	1.773

Note: See notes to Table 11.

Table 13: Size(%) and power(%) of γ ($\gamma = 0.4$) for AR(1) model

	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	5.2	7.2	7.0	5.9	20.7	49.8	69.1	85.3	5.1	7.2	7.0	5.9	20.6	49.8	69.1	85.3
150	6.1	4.9	5.0	5.6	42.2	90.0	99.3	100.0	6.1	4.9	5.0	5.6	42.2	90.0	99.3	100.0
500	4.8	4.9	5.3	5.0	83.6	100.0	100.0	100.0	4.8	4.9	5.3	5.0	83.6	100.0	100.0	100.0
	One-step first-difference GMM estimator based on "DIF2"															
50	9.7	8.1	9.6	6.1	21.8	38.0	56.8	66.9	15.9	14.8	14.1	11.4	25.6	34.7	41.9	53.8
150	5.3	6.0	6.0	6.5	28.3	65.6	88.8	95.6	9.3	7.3	9.5	7.4	20.6	36.3	55.8	74.8
500	5.6	5.1	5.2	5.1	55.0	97.1	100.0	100.0	5.1	5.4	5.9	7.1	24.2	53.6	88.3	98.2
	Two-step first-difference GMM estimator based on "DIF2"															
50	17.7	29.7	43.5	65.6	31.6	57.7	75.7	84.3	25.9	38.2	50.7	71.4	37.1	57.7	71.3	81.4
150	8.6	10.4	12.7	18.5	32.8	72.8	89.7	95.3	11.7	15.1	19.3	21.4	24.2	43.7	68.2	83.3
500	5.9	7.0	6.7	7.7	56.6	96.9	100.0	100.0	5.9	7.9	7.0	10.6	27.4	59.2	92.1	98.8
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors															
50	7.8	5.1	3.0	1.2	16.4	22.2	16.8	5.4	11.7	7.1	2.9	1.1	19.5	17.2	8.6	3.0
150	5.8	5.2	4.8	5.2	27.0	60.3	77.9	86.7	8.2	6.4	6.9	5.9	18.1	29.0	46.1	60.7
500	5.4	5.3	4.6	4.0	54.2	96.5	100.0	100.0	5.3	5.7	5.4	6.3	24.0	54.0	88.0	97.9
	Continuous-updating first-difference GMM estimator based on "DIF2"															
50	20.6	35.0	51.1	75.0	25.5	46.4	66.5	78.8	30.8	42.5	55.5	74.3	34.0	43.5	59.6	77.8
150	8.9	11.1	15.5	20.5	26.1	62.1	82.8	90.8	12.4	16.3	17.9	22.5	18.2	30.1	50.8	71.0
500	6.0	7.3	6.7	7.3	51.9	96.0	100.0	100.0	6.2	7.4	6.5	10.2	21.2	48.8	85.7	97.7
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors															
50	22.1	52.8	58.6	38.1	26.2	62.0	70.6	43.8	24.8	49.5	54.0	37.5	28.6	50.5	57.3	40.5
150	7.1	12.6	21.3	35.2	24.9	65.3	87.9	95.5	9.7	16.5	22.6	35.8	15.5	29.8	57.4	81.1
500	5.7	7.3	7.6	9.2	50.3	96.2	100.0	100.0	5.2	7.1	6.9	11.1	18.8	47.4	85.4	97.9
	One-step system GMM estimator based on "SYS2"															
50	9.9	9.4	9.3	–	12.4	20.5	28.2	–	76.9	92.3	97.9	–	65.3	82.0	89.8	–
150	5.8	5.0	6.5	6.3	25.7	55.5	78.2	88.9	56.9	82.8	94.3	98.7	38.7	57.1	68.9	73.0
500	5.4	6.6	5.2	5.1	65.9	96.5	100.0	100.0	37.5	66.7	83.7	94.0	15.0	18.5	16.4	16.8
	Two-step system GMM estimator based on "SYS2"															
50	25.5	52.3	76.0	–	38.0	61.2	85.5	–	90.8	98.3	100.0	–	84.5	94.6	98.3	–
150	12.1	14.3	23.3	28.0	47.7	80.7	91.0	96.9	74.1	92.3	97.5	99.5	64.9	74.5	83.7	85.9
500	7.0	9.2	9.9	11.7	85.8	99.6	100.0	100.0	50.1	68.1	81.3	85.8	59.6	70.2	78.2	82.9
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors															
50	7.2	3.2	2.1	–	12.8	9.6	3.5	–	65.0	53.4	11.2	–	53.4	38.2	8.0	–
150	6.4	3.6	4.6	4.3	34.6	59.4	72.8	76.0	46.2	69.3	81.8	85.6	33.2	42.8	53.2	49.7
500	5.6	6.5	4.8	5.3	82.8	99.3	100.0	100.0	24.4	43.2	54.0	62.6	31.4	39.7	43.7	50.8
	Continuous-updating system GMM estimator based on "SYS2"															
50	32.7	58.4	85.8	–	43.9	72.0	89.5	–	64.1	80.6	95.4	–	71.5	85.7	95.5	–
150	12.3	16.8	25.8	34.7	49.7	83.8	93.4	96.4	31.2	37.2	47.2	54.4	62.3	91.4	97.1	98.0
500	7.4	8.6	9.1	11.9	86.7	99.8	100.0	100.0	14.1	15.9	17.2	20.0	86.2	99.8	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors															
50	33.9	50.3	32.1	–	44.4	62.5	40.9	–	36.0	22.0	23.7	–	43.8	27.6	24.0	–
150	12.3	19.6	40.9	56.9	47.4	85.5	96.5	98.5	13.3	16.7	24.3	24.1	46.8	81.3	90.9	90.5
500	6.6	8.5	10.0	15.4	84.9	99.8	100.0	100.0	5.7	8.9	9.8	14.2	81.4	99.5	100.0	100.0

Note: For the definition of "DIF2" and "SYS2", see notes to Table 11. "NW" denotes Newey and Windmeijer's(2009) standard errors.

Table 14: Size(%) and power(%) of γ ($\gamma = 0.9$) for AR(1) model

	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	13.5	15.9	13.1	10.2	28.6	44.0	60.1	77.1	13.5	15.9	13.1	10.3	28.6	44.0	60.1	77.1
150	16.7	12.7	7.7	6.7	31.6	58.4	81.9	95.3	16.8	12.6	7.7	6.6	31.6	58.3	81.9	95.2
500	17.0	7.7	5.1	4.5	44.6	76.2	95.3	100.0	17.2	7.7	5.1	4.5	44.7	76.2	95.2	100.0
	One-step first-difference GMM estimator based on "DIF2"															
50	33.8	30.8	27.5	24.1	45.0	53.3	57.9	66.3	37.5	43.5	39.8	38.5	47.4	58.8	64.3	71.1
150	22.9	16.2	11.7	9.1	32.5	39.2	56.9	72.4	30.8	36.5	32.6	29.9	41.7	54.8	61.4	68.6
500	14.7	7.8	7.0	6.7	27.9	41.2	72.1	92.0	27.3	30.1	27.0	25.8	39.5	51.7	56.7	67.5
	Two-step first-difference GMM estimator based on "DIF2"															
50	59.0	69.6	74.1	80.9	66.1	81.6	87.7	91.3	64.5	80.9	87.2	90.2	71.8	88.4	93.6	96.0
150	39.1	42.8	41.0	35.7	48.8	62.7	76.9	85.8	52.1	69.1	77.0	79.0	60.2	80.8	88.8	93.0
500	21.6	16.1	14.7	14.6	33.8	49.0	77.3	94.0	39.0	56.2	57.4	63.3	47.9	71.4	79.6	84.7
	Two-step first-difference GMM estimator based on "DIF2" with Windmeijer standard errors															
50	29.1	20.6	8.6	5.0	34.7	29.5	14.7	7.0	33.3	30.2	17.4	6.7	39.5	38.1	23.2	9.0
150	20.7	15.3	12.6	7.5	27.7	31.4	45.5	57.3	28.1	30.9	33.4	29.3	35.6	42.6	48.7	48.6
500	13.6	10.0	9.5	8.6	24.4	37.7	70.8	91.1	23.9	28.1	32.0	34.1	30.5	43.3	52.0	62.1
	Continuous-updating first-difference GMM estimator based on "DIF2"															
50	50.7	61.9	70.1	81.8	54.0	65.9	75.6	85.8	58.1	78.2	88.4	92.2	60.9	79.4	88.6	93.1
150	30.8	31.9	25.7	23.2	35.3	37.0	47.0	63.6	42.4	63.3	70.4	73.7	47.2	65.9	72.9	74.5
500	14.4	9.3	5.9	7.7	19.3	25.3	56.5	84.2	28.9	40.6	39.6	43.1	32.9	44.6	44.6	49.7
	Continuous-updating first-difference GMM estimator based on "DIF2" with NW standard errors															
50	37.1	45.5	46.3	39.9	41.6	49.8	53.7	45.2	43.5	52.7	52.0	48.1	46.7	55.1	55.0	51.1
150	20.1	21.4	22.3	33.0	25.1	28.1	46.2	73.1	29.3	34.8	32.8	40.4	32.0	38.9	35.2	42.8
500	8.7	5.9	6.0	8.6	13.7	19.7	54.5	85.3	17.4	15.6	13.9	17.5	21.0	20.4	19.5	23.3
	One-step system GMM estimator based on "SYS2"															
50	31.6	48.1	65.2	–	1.1	6.2	11.7	–	96.1	100.0	100.0	–	0.2	0.8	1.9	–
150	27.7	42.6	57.5	65.2	3.8	19.8	37.2	52.1	96.3	99.7	100.0	100.0	0.7	1.6	3.1	3.7
500	19.3	30.9	39.1	49.9	26.6	73.9	94.3	98.9	93.7	99.8	100.0	100.0	2.3	8.3	12.3	18.2
	Two-step system GMM estimator based on "SYS2"															
50	56.9	77.1	90.8	–	46.1	69.1	85.8	–	98.5	100.0	100.0	–	43.6	61.6	82.5	–
150	44.1	56.1	69.6	78.2	42.2	74.1	86.8	92.6	98.4	100.0	100.0	100.0	39.7	51.9	52.5	63.0
500	26.5	36.8	39.3	46.6	75.0	97.3	99.8	100.0	97.0	100.0	100.0	100.0	52.5	63.7	68.8	72.4
	Two-step system GMM estimator based on "SYS2" with Windmeijer standard errors															
50	17.9	12.7	4.9	–	7.3	7.3	4.3	–	78.5	81.6	40.0	–	4.9	5.4	2.5	–
150	17.6	20.4	23.8	24.4	9.7	24.1	36.1	38.7	86.0	98.9	99.9	99.9	3.6	8.8	13.9	16.3
500	10.9	15.4	16.2	16.0	43.4	87.0	97.1	99.9	86.5	99.7	99.8	100.0	5.3	17.7	26.1	34.0
	Continuous-updating system GMM estimator based on "SYS2"															
50	69.8	85.8	94.3	–	63.4	81.6	92.2	–	97.4	98.2	99.3	–	73.8	89.5	95.8	–
150	49.4	51.9	60.1	67.2	58.4	89.7	96.0	97.7	94.3	95.9	93.3	95.8	75.1	96.2	99.3	99.5
500	29.2	27.3	27.7	27.7	82.4	99.9	100.0	100.0	90.1	87.3	83.4	79.3	93.3	100.0	100.0	100.0
	Continuous-updating system GMM estimator based on "SYS2" with NW standard errors															
50	43.5	30.3	21.3	–	35.2	28.3	19.8	–	46.8	27.4	21.2	–	10.0	2.5	1.6	–
150	30.3	27.1	28.8	26.2	38.2	70.8	81.8	81.0	63.0	48.0	34.4	27.4	22.5	25.7	21.1	12.4
500	17.5	14.6	13.3	14.8	66.5	96.6	99.3	99.9	57.6	44.9	29.8	19.7	44.1	79.7	89.8	93.6

Note: See notes to Table 13.