A comparison of FDML and GMM for estimation of dynamic panel models with roots near unity

Adrian Mehic

NEKH01: Bachelor's thesis in Economics (15 ECTS)

Lund University
School of Economics and Management
Supervisor: Joakim Westerlund
23 August 2017

Abstract

This thesis compares the performance of the first-differenced maximum likelihood estimator (FDML) and the Blundell-Bond continuously-updating system GMM estimator of the autoregressive parameter in an AR(1) dynamic panel model without exogenous covariates, particularly focusing on the close-to-non-stationary case. This case is far from trivial, as a high degree of persistence is the norm rather than the exception in economic panels. The results of the Monte Carlo simulations show that the absolute mean and median biases of the FDML are higher for low values of N and T in the close-to-non-stationary case. However, the biases become negligible for both estimators as N and T increase. The power of the GMM is generally higher than that of the FDML, while, on the other hand, the GMM suffers from severe size distortions. This problem is magnified both when T increases, as well as when the process approaches non-stationarity. Finally, the GMM estimator is shown to display Cauchy properties when the process is very close to non-stationarity. This produces some peculiar bias results for certain combinations of N and T when using the GMM.

Contents

1 Introduction

Dynamic panel data models are of high importance in empirical econometrics. Their applications can be found in virtually all areas of economics: labor economics, macroeconomics, finance, development economics and in applied microeconometrics. The statistical techniques for parameter estimation in the dynamic panel model have traditionally been based on the generalized method of moments, GMM. Two GMM-based methods for dynamic panel data have been particularly successful in the empirical literature: the *difference GMM* estimator, which is due to Arellano and Bond (1991) and the *system GMM* of Blundell and Bond (1998). A likelihood-based estimator, known as the *first-differenced ML* (FDML), was developed by Hsiao et al. (2002). Recent research has shown that the FDML outperforms the GMM in terms of size, power and bias in most cases (cf. Hsiao and Zhang 2015). The FDML approach was extended to the case of cross-sectionally heteroscedastic errors by Hayakawa and Pesaran (2015), who demonstrated that the performance of the FDML does not deteriorate even when relaxing the homoscedasticty assumption.

However, there are other situations frequently encountered in empirical research that may cause problems for the econometrician. An example of this is the situation in which the data is close to non-stationary. The performance of GMM-based estimators in this case has been well covered in the literature. In these cases, the Arellano-Bond GMM estimator has been shown to perform poorly compared to the system GMM (cf. Blundell and Bond 1998; Madsen 2008; Hayakawa and Pesaran 2015). For instance, there is a considerable increase in bias for low N when using the difference GMM in lieu of the system GMM for close to non-stationary data. This is a situation with potentially serious ramifications, as a high degree of persistence is frequently encountered in economic panels. For the FDML, while it has been established that the performance of the estimator deteriorates when a unit root is present (Han and Philips 2013), its performance in the situation with almost non-stationary data has not been examined. Considering its robustness under heteroscedasticity, the FDML has recently emerged as a serious contender to the GMM in empirical economics research. However, failing to perform in the nearly non-stationary case would be considered a serious drawback, potentially limiting its usability in practical situations. The purpose of this thesis is, thus, to assess the finite-sample properties of the FDML and the system GMM estimator of Blundell and Bond, focusing particularly on the case of near-non-stationarity. The performance of the estimators is assessed by examining the mean and median biases, as well as the size and power of the two estimators.

The results of the Monte Carlo simulations show that the FDML has higher absolute bias than the system GMM in the nearly non-stationary case. The performance of the FDML is particularly weak when T is low. Moreover, the power of the FDML tends to be slightly lower than that of the GMM. However, for high values of N and T, the difference between the estimators is negligible. The results further show that GMM estimator suffers from severe size distortions, which are exacerbated as N and T increase. This problem for the system GMM was first noted by Hayakawa and Pesaran (2015), and is confirmed in this thesis. Finally, the GMM estimator is shown to display anomalies in the mean bias, which can be explained by the limiting distribution of the GMM estimator being Cauchy in the nearly non-stationary case.

The reminder of the thesis is organized as follows. Section 2 introduces the model and gives a brief outline of the two estimation methods considered together with some asymptotic results. Section 3 describes the Monte Carlo design. Section 4 presents the results of the simulation study. The thesis concludes with Section 5.

2 Theory

The dynamic panel data model considered in this thesis is based on the AR(1) time series model, and can be described by

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

for individuals $i = 1, \dots, N$ and time periods $t = 1, \dots, T$, where α_i are fixed effects with $\mathbb{E}[\alpha_i] = 0$, $\mathbb{V}[\alpha_i] = \sigma_\alpha^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \alpha_i^2 < \infty$, ϕ is the autoregressive parameter, and u_{it} is the idiosyncratic error term. It is assumed that the error terms u_{it} are independent and identically distributed, and that $\mathbb{E}[u_{it}] = 0$, $\mathbb{V}[u_{it}] = \sigma_u^2$, $\mathbb{E}[|u_{it}|^4] < \infty$ and $\mathbb{E}[u_{it}u_{is}] = 0$ for each $t \neq s$. Further, it is assumed that the initial observations $y_{i0} = \mathcal{O}_p(1)$ are observed, and that $\mathbb{E}[\alpha_i u_{it}] = 0$, as well as $\mathbb{E}[y_{i0}u_{it}] = 0$. Throughout this thesis, a sequence of random elements $\{X_N\}$ is said to be $\mathcal{O}_p(1)$ if it is bounded in probability (tight), i.e. if it for every $\varepsilon \in \mathbb{R}_+$ exists an integer $M < \infty$ such that $\mathbb{P}(||X_N|| \leq M) > 1 - \varepsilon$ for each $N \in \mathbb{N}$. Conversely, $X_N = \mathcal{O}_p(Y_N)$ means that X_N/Y_N is bounded in probability [1].

If $|\phi| = 1$, $\{y_{it}\}$ is a so-called *unit root* process, if $|\phi| < 1$, the process is called *stationary*, and if $|\phi| > 1$, the process is called *explosive*. The focus of this thesis is the case when the data are close to non-stationary; that is, when the value of ϕ is close to unity. Hence, this situation is also known as the local unit root (LUR) case.

Using the standard fixed effects estimator to estimate (1) gives biased estimates of ϕ (Nickell 1981). This is because the fixed effects estimator uses demeaning, that is, subtracting the mean values on both sides of (1). However, using lagged values of the dependent variable as an explanatory variable generates correlation between the demeaned lagged dependent variable and the demeaned error term. Hence, estimates of ϕ will be biased. The bias is approximately equal to $-(1+\phi)/(1+T)$ as $N \to \infty$. Consequently, both for low values of T and in the LUR case, the bias can be sizable. Such a situation is particularly common in microeconometric models, when many individuals (large

[1] All stochastic objects in this thesis are assumed to be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The process $\{y_{ij}\}$ takes values in the measurable space $(S, \mathcal{S})$, where S is a Polish space and $\mathcal{S}$ is its Borel σ-field.

N) are sampled over a short time period (small T). The first to propose an unbiased estimator of (1) were Anderson and Hsiao (1981). This approach can be described as follows. In order to eliminate α_i, take the first difference of (1), which yields

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

for $t = 2, \ldots, T$, where $\Delta = (1 - L)$, and L is the lag operator. Now, Anderson and Hsiao suggest to use either $y_{i,t-2}$ or $\Delta y_{i,t-2}$ as instruments to remedy the bias problem. The corresponding estimators $\hat{\phi}_{AB1}$ and $\hat{\phi}_{AB2}$ are

$$\hat{\phi}_{AB1} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{it} y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{i,t-1} y_{i,t-2}} \tag{3}$$

and

$$\hat{\phi}_{AB2} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{it} \Delta y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{i,t-1} \Delta y_{i,t-2}} \tag{4}$$

However, the Anderson-Hsiao estimator has several issues that have led to its decline in practical use. Firstly, the estimator is asymptotically inefficient for all values of ϕ (Arellano and Bond 1991). Secondly, in the case of $\phi = 1$, the estimator is inconsistent for fixed T, and consistent but inefficient when $T \to \infty$ (Kruiniger 2008; Phillips 2015). Finally, when ϕ increases towards unity, the variance of the estimator explodes (Arellano and Bover 1995). This effect is present even when $\phi = 0.80$, which makes the AH estimator virtually useless in the LUR case (ibid.). Instead, this thesis will focus on GMM and FDML, which are introduced in sections 2.1 and 2.2. Section 2.3 presents the asymptotic properties of GMM and FDML.

2.1 GMM

The first estimation technique considered in this thesis is the *generalized method of moments* (GMM), which is due to Hansen (1982). The GMM is, as the name suggests, an extension of the standard method of moments technique. The difference between the GMM and the regular method of moments is that under GMM, it is possible to define more moment conditions than there are parameters to be estimated.

To simplify notation, let $\pi_{it} = \alpha_i + u_{it}$. Then, Arellano and Bond (1991) show that for $t = 3, \ldots, T$, the moment conditions

$$\mathbb{E}[y_{is} \Delta \pi_{is}] = 0 \tag{5}$$

can be utilized. If $u_{it} \sim MA(0)$, it holds that $s \geq 2$ and $s \geq 3$ if $u_{it} \sim MA(1)$. Now, define the instrument matrix $\boldsymbol{Z}_i$ as

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{i,T-2} \end{pmatrix} \tag{6}$$

and the vector of first-differenced errors as $\Delta\boldsymbol{\pi}_i = (\Delta\pi_{i3}, \dots, \Delta\pi_{iT})'$. Using this notation, the moment conditions can be written

$$\mathbb{E}[\boldsymbol{Z}_i'\Delta\boldsymbol{\pi}_i] = \boldsymbol{0} \tag{7}$$

for $i = 1, \dots, N$. Based on the moment conditions defined by (5), Arellano and Bond (1991) construct a GMM-based estimator of ϕ. This approach is similar to the Anderson-Hsiao technique. While Anderson and Hsiao utilize only $y_{i,t-2}$ or $\Delta y_{i,t-2}$, the Arellano-Bond estimator uses all available lags as instruments. The number of lags is highest for the time period closest to the final time T. However, it was shown by Blundell and Bond (1998) that the Arellano-Bond estimator significantly underestimates ϕ in the LUR case; the bias starts to increase already when $\phi = 0.80$. In order to remedy this problem, Arellano and Bover (1995) and Blundell and Bond (1998) introduce additional moment conditions, namely

$$\mathbb{E}[\pi_{it}\Delta y_{i,t-s}] = 0 \tag{8}$$

for $t = 3, \dots, T$ and $i = 1, \dots, N$. For $u_{it} \sim MA(0)$, it holds that $s = 1$ and if $u_{it} \sim MA(1)$, then $s = 2$. The joint moment conditions, that is, using (5) together with (8), can be compactly written in matrix form as

$$\mathbb{E}[\tilde{\boldsymbol{Z}}_i'\boldsymbol{\pi}_i^*] = \boldsymbol{0} \tag{9}$$

where $\tilde{\boldsymbol{Z}}_i = \mathrm{diag}(\boldsymbol{Z}_i, \Delta y_{i2}, \Delta y_{i3}, \dots, \Delta y_{i,T-1})$, $\boldsymbol{\pi}_i^* = (\Delta\boldsymbol{\pi}_i, \boldsymbol{\pi}_i)'$, where $\Delta\boldsymbol{\pi_i}$ is as defined previously and $\boldsymbol{\pi}_i = (\pi_{i3}, \dots, \pi_{iT})'$, for $i = 1, \dots, N$. Using these moment conditions, the Blundell-Bond estimator $\hat{\phi}_{GMM}$ of ϕ is the solution to the optimization problem

$$\hat{\phi}_{GMM} = \arg\min_{\phi\in\Phi} \left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\pi}_i^{*\prime}\tilde{\boldsymbol{Z}}_i\right) \boldsymbol{W}_N \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i'\boldsymbol{\pi}_i^*\right) \tag{10}$$

where Φ is the compact set of all possible parameters and $\boldsymbol{W}_N$ is a positive semi-definite weight matrix, for which it holds that $\boldsymbol{W}_N \xrightarrow{P} \boldsymbol{W}$ according to the weak law of large numbers. Here, $\xrightarrow{P}$, signifies convergence in probability. The closed-form expression for the solution to (10) is

$$\hat{\phi}_{GMM} = \frac{\boldsymbol{q}_{-1}'\tilde{\boldsymbol{Z}}\boldsymbol{W}_N^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{q}}{\boldsymbol{q}_{-1}'\tilde{\boldsymbol{Z}}\boldsymbol{W}_N^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{q}_{-1}} \tag{11}$$

where $\boldsymbol{q} = (\boldsymbol{q}_1', \dots, \boldsymbol{q}_N')'$, $\boldsymbol{q}_i = (\Delta\boldsymbol{y_i}', \boldsymbol{y_i}')'$, where $\Delta\boldsymbol{y}_i = (\Delta y_{i1}, \Delta y_{i2}, \dots, \Delta y_{iT})'$ and $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$, and $\boldsymbol{q}_{-1}$ is the lagged version of $\boldsymbol{q}$. The weight matrix can be estimated by

$$\hat{\boldsymbol{W}}_N = \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i'\boldsymbol{H}\tilde{\boldsymbol{Z}}_i\right) \tag{12}$$

where $\boldsymbol{H}$ is the tridiagonal matrix

$$\boldsymbol{H} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix} \tag{13}$$

There are other ways of improving the finite-sample properties of the GMM estimator of ϕ, in addition to just adding more moment conditions. Instead of fixing the weight matrix in each stage of the estimation, the *continuously updating GMM* estimator alters the weight matrix as the value of $\hat{\phi}$ is changed during the minimization process. This type of GMM estimator is due to Hansen et al. (1996). Formally, the minimization problem can now be written

$$\hat{\phi}_{GMM-CU} = \arg\min_{\phi\in\Phi} \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\pi}_i^{*\prime} \tilde{\boldsymbol{Z}}_i \right) \boldsymbol{W}_N(\phi) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i^{\prime} \boldsymbol{\pi}_i^{*} \right) \tag{14}$$

Hence, the weight matrix is now a function of ϕ. It can be shown that introducing parameter-dependency on the weight matrix does not alter the asymptotic properties of the estimator, a result that holds for all GMM estimators (Pakes and Pollard 1989). Alternatively stated, the asymptotics of $\hat{\phi}_{GMM-CU}$ are the same as those of $\hat{\phi}_{GMM}$. However, even though the asymptotic properties of the two estimators are equivalent, the continuously-updating estimator has been shown to have a smaller finite-sample bias compared to the usual GMM estimator (Newey and Smith 2004).

2.2 FDML

An alternative approach to estimating ϕ is by using FDML, which is due to Hsiao et al. (2002). This estimator is denoted by $\hat{\phi}_{FDML}$. The FDML approach is the following. In order to eliminate α_i, take again the first difference of (1) to obtain

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \Delta u_{it}$$

which is the same as (2). For $t = 1$, the above expression is not well defined, since $\Delta y_{i1} = \phi\Delta y_{i0} + \Delta u_{i1}$ and Δy_{i0} is not observable. However, by continuous substitution,

$$\begin{aligned} \Delta y_{i1} &= \phi^m \Delta y_{i,-m+1} + \sum_{j=0}^{m-1} \phi^j \Delta u_{i,1-j} \\ &= \phi^m \Delta y_{i,-m+1} + \eta_{i1} \end{aligned} \tag{15}$$

Now, the analysis will differ slightly depending on whether the process is stationary or not. Assume first that $|\phi| < 1$ and $m \to \infty$, which means that the process has been going on for a very long time. Then, it holds for $t = 3, \dots, T$ and

$i = 1, \ldots, N$ that $\mathbb{E}[\Delta y_{i1}] = 0$, $\mathbb{V}[\Delta y_{i1}] = 2\sigma_u^2/(1+\phi)$, $\mathbb{C}[\eta_{i1}, \Delta u_{i2}] = -\sigma_u^2$, and $\mathbb{C}[\eta_{i1}, \Delta u_{it}] = 0$. Alternatively, if $|\phi| \geq 1$, the process has started from a finite point m that is behind the 0:th time point, so that $\mathbb{E}[\Delta y_{i1}] = b$, $\mathbb{V}[\Delta y_{i1}] = c\sigma_u^2$, $\mathbb{C}[\eta_{i1}, \Delta u_{i2}] = -\sigma_u^2$, and $\mathbb{C}[\eta_{i1}, \Delta u_{it}] = 0$, where $b, c \in \mathbb{R}_+$. However, both for unit root and for explosive processes, it is assumed that the increments are stationary. This enables the first difference of the process to be stationary. Let now $\Delta \boldsymbol{y}_i$ be as defined previously, and $\Delta \boldsymbol{u}_i^* = (\Delta y_{i1} - b^*, \Delta u_{i2}, \ldots, \Delta u_{iT})'$. Here, $b^* = 0$ if $|\phi| < 1$ and $b^* = b$ if $|\phi| \geq 1$. The covariance matrix of $\Delta \boldsymbol{u}_i^*$ can be expressed as

$$\mathbb{C}(\Delta \boldsymbol{u}_i^*) = \boldsymbol{\Omega} = \sigma_u^2 \boldsymbol{\Omega}^* \tag{16}$$

where $\boldsymbol{\Omega}^*$ is equal to

$$\boldsymbol{\Omega}^* = \begin{pmatrix} \omega & -1 & 0 & \ldots & 0 \\ -1 & 2 & -1 & \ldots & 0 \\ 0 & -1 & 2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix} \tag{17}$$

where $\omega = (1/\sigma_u^2)\mathbb{V}(\Delta y_{i1})$. This is equal to $2/(1+\phi)$ if $|\phi| < 1$, and c else. Now, in order to find the likelihood function of $\Delta \boldsymbol{y}_i$, note that $\Delta \boldsymbol{u}_i^*$ is a linear combination of $\Delta \boldsymbol{y}_i$, and that the Jacobian of this transformation is equal to unity. This means that the joint probability density functions (p.d.fs) of $\Delta \boldsymbol{u}_i^*$ and $\Delta \boldsymbol{y}_i$ are equal. Then, assuming that the u_{it}:s are independent normal, the joint p.d.f of $\Delta \boldsymbol{y}_i$ is equal to the likelihood function of (2), and is given by

$$\mathcal{L} = \prod_{i=1}^{N} (2\pi)^{-T/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{u}_i^{*\prime} \boldsymbol{\Omega}^{-1} \boldsymbol{u}_i^* \right\} \tag{18}$$

The corresponding log-likelihood is

$$\log \mathcal{L} = \frac{-NT}{2} \log(2\pi) - \frac{N}{2} \log|\boldsymbol{\Omega}| - \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{u}_i^{*\prime} \boldsymbol{\Omega}^{-1} \boldsymbol{u}_i^* \tag{19}$$

The two unknown elements of $\boldsymbol{\Omega}$ are σ_u^2 and ω. Proceeding from here, the FDML technique involves utilizing the Anderson-Hsiao estimator $\hat{\phi}_{AB2}$ to find an initial estimate of ϕ. Then, the variance σ_u^2 is estimated by

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta y_{it} - \hat{\phi}_{AB2} \Delta y_{i,t-1} \right)^2}{2N(T-2)} \tag{20}$$

In the stationary case, ω can be estimated either by $2(1 + \hat{\phi}_{AB2})^{-1}$ or by

$$\hat{\omega} = \frac{1}{N \hat{\sigma}_u^2} \sum_{i=1}^{N} \Delta y_{i1}^2 \tag{21}$$

In the non-stationary case, ω is estimated by

$$\tilde{\omega} = \frac{1}{(N-1)\hat{\sigma}_u^2} \sum_{i=1}^{N} (\Delta y_{i1} - \hat{b})^2 \tag{22}$$

where a consistent estimator of b is $\hat{b} = N^{-1} \sum_{i=1}^{N} \Delta y_{i1}$. Using these estimates, (19) is maximized numerically until convergence.

2.3 Asymptotic properties

This section considers the asymptotic properties of the GMM and FDML estimators by examining, in turn, the stationary, unit root and LUR cases. To simplify notation, the asymptotic results relating to the continuously updating GMM will be denoted just by $\hat{\phi}_{GMM}$. As described in Section 2.1, the asymptotic results are the same, regardless of whether the weight matrix is dependent on ϕ or not.

First, it is worth noting that under some fairly weak assumptions, it holds that both the FDML and GMM estimators are consistent, that is, $\hat{\phi}_{FDML} \xrightarrow{P} \phi$ and $\hat{\phi}_{GMM} \xrightarrow{P} \phi$. [2]. However, in addition to consistency, two other important properties of estimators are asymptotic unbiasedness and asymptotic efficiency. For the FDML in the context of dynamic panel models, the following theorem holds.

Theorem 1. *Given $|\phi| < 1$, for T large and N arbitrary, the limiting distribution of the FDML estimator $\hat{\phi}_{FDML}$ of ϕ is*

$$\sqrt{NT}(\hat{\phi}_{FDML} - \phi) \xrightarrow{L} \mathcal{N}(0, 1 - \phi^2) \tag{23}$$

Proof. See Kruiniger (2008). ■

In (23), $\xrightarrow{L}$, denotes convergence in law (in distribution). The implication of Theorem 1 is that the FDML estimator is asymptotically unbiased, and asymptotically normal. It can also be shown that the asymptotic variance in (23) is equal to the Cramér-Rao lower bound, which implies that the FDML is asymptotically efficient (Hahn and Kuersteiner 2002). However, when T is fixed and $N \to \infty$, the variance is higher than the Cramér-Rao bound, and hence, the

[2] This consistency result holds for all ML and GMM estimators. To see this, let $\boldsymbol{y}$ be a data vector with p.d.f denoted by $f(\boldsymbol{y})$ and $\boldsymbol{\theta}_0$ be an unknown parameter vector to be estimated. Then, the consistency conditions are (i) $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, which is compact, (ii) for each $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0, f(\boldsymbol{y}|\boldsymbol{\theta}) \neq f(\boldsymbol{y}|\boldsymbol{\theta}_0)$ (for the MLE) and $\boldsymbol{W}\mathbb{E}\left[g(\boldsymbol{y},\boldsymbol{\theta})\right] = \boldsymbol{0}$ for $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$ where $\boldsymbol{W}$ is p.s.d., and $\hat{\boldsymbol{W}} \xrightarrow{P} \boldsymbol{W}$ (for the GMM), (iii) for the MLE $\log f(\boldsymbol{y}|\boldsymbol{\theta})$, and for the GMM $g(\boldsymbol{y},\boldsymbol{\theta})$, is continuous at each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ almost surely, and (iv) $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \log f(\boldsymbol{y}|\boldsymbol{\theta})\right] < \infty$ for the MLE, and $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} ||g(\boldsymbol{y},\boldsymbol{\theta})||\right] < \infty$ for the GMM. Then, it holds that $\hat{\boldsymbol{\theta}}_{ML} \xrightarrow{P} \hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_{GMM} \xrightarrow{P} \hat{\boldsymbol{\theta}}_0$. See Newey and McFadden (1994) for further details.

estimator is inefficient (Kruninger 2008).[3] However, in this case, the estimator is still consistent, asymptotically unbiased and asymptotically normal.

For the GMM estimator $\hat{\phi}_{GMM}$, Álvarez and Arellano (2003) show that, given $|\phi| < 1$ and $(\log T)^2/N \to 0$, as $(N,T) \to \infty$,

$$\mathbb{E}\left[\sqrt{NT}(\hat{\phi}_{GMM} - \phi)\right] = -(1+\phi)\sqrt{\zeta} + \mathcal{O}_p\left(\frac{\log T}{\sqrt{NT}}\right) + o_p(1) \tag{24}$$

In equation (24), $\zeta = T/N$ [4]. Using this result, it is possible to derive the following theorem regarding the asymptotic properties of $\hat{\phi}_{GMM}$ in the stationary case.

Theorem 2. *Given $|\phi| < 1$ and $(logT)^2/N \to 0$, as $(N,T) \to \infty$, the GMM estimator $\hat{\phi}_{GMM}$ of ϕ is asymptotically biased. The limiting distribution of $\hat{\phi}_{GMM}$ is*

$$\sqrt{NT}\left[\hat{\phi}_{GMM} - \left(\phi - \frac{1}{N}(1+\phi)\right)\right] \xrightarrow{L} \mathcal{N}(0, 1-\phi^2) \tag{25}$$

Proof. See Álvarez and Arellano (2003). An alternative proof is given in the supplement of Hsiao and Zhou (2017). ∎

Thus, the asymptotic bias is $\mathcal{O}_p(\sqrt{T/N})$, which implies that as $N/T \to \infty$, the asymptotic bias disappears. Moreover, it follows directly from (25) that as $N/T \to \infty$, the asymptotic variance of the GMM estimator attains the Cramér-Rao lower bound.

It was recently shown (Hsiao and Zhou 2017) that using only one lag as instrument will yield asymptotically unbiased estimators. However, in this case, the asymptotic variance is equal to $(1+\phi)\left[(1+\frac{\sigma_\alpha^2}{\sigma_u^2}) + \phi(\frac{\sigma_\alpha^2}{\sigma_u^2} - 1)\right] > (1-\phi^2)$. This implies that the asymptotic variance will be particularly inflated when ϕ is close to unity.

Since this thesis deals with the situation with close to non-stationary data, it is natural to describe the asymptotic results for $\hat{\phi}_{FDML}$ and $\hat{\phi}_{GMM}$ when ϕ is exactly equal to one and when ϕ is local to one. These are dramatically different from the stationary case, as manifested by Theorems 3, 4 and 5.

Theorem 3. *For $|\phi| = 1$, the limiting distribution of the FDML estimator $\hat{\phi}_{FDML}$ of ϕ is*

$$T\sqrt{N}(\hat{\phi}_{FDML} - 1) \xrightarrow{L} \mathcal{N}(0, 8) \tag{26}$$

as $N, T \to \infty$ jointly.

[3] The reason that the asymptotic variance does not attain the Cramér-Rao lower bound in the latter case is that with T fixed, the individual effects α_i are not ancillary for ϕ and σ_u^2 (Kruninger 2008). Andersen (1970) has derived conditions under which certain ML-based estimators reach the Cramér-Rao bound. For a general discussion about ancillary statistics, see e.g. Hogg, McKean and Craig (2014, pp. 426-433).

[4] The notation $o_p(1)$ means that the sequence $\{X_N\}$ converges in probability to zero. Similarly, $X_N = o_p(Y_N)$ means that X_N/Y_N converges in probability to zero. Since convergence in probability implies convergence in law, if a sequence is $o_p(1)$, it is also $\mathcal{O}_p(1)$ by Prokhorov's theorem.

Proof. See Kruiniger (2008) or Han and Phillips (2013). ■

The rate of convergence here is $\mathcal{O}_p(T\sqrt{N})$, which is faster than the rate of convergence in the stationary case, which is $\mathcal{O}_p(\sqrt{NT})$ according to Theorem 1. Hence, the larger limiting variance in (26) is compensated by a faster convergence in distribution. The GMM equivalent of Theorem 4 is

Theorem 4. *For $|\phi| = 1$, the limiting distribution of the GMM estimator $\hat{\phi}_{GMM}$ of ϕ is*

$$\sqrt{T}(\hat{\phi}_{GMM} - 1) \xrightarrow{L} 2\,\mathcal{C} \tag{27}$$

as $N, T \to \infty$ jointly.

Proof. See Phillips (2014). ■

In Theorem 4, $\mathcal{C}$ denotes a standard Cauchy variate, viz. a Cauchy distributed random variable with location parameter equal to zero and scale parameter equal to unity. It would be outright impossible to compare asymptotic means and variances for (26) and (27), since the moments for the Cauchy distribution are undefined [5]. However, the median of such a variate exists and is equal to zero.

Finally, consider the LUR case. Regrettably, there exists a considerable research gap regarding the asymptotic results for the FDML in the LUR case. However, for GMM, it is possible to formulate the following asymptotic results.

Theorem 5. *For $\phi = 1 + \frac{c}{T^\gamma}$, given $c < 0$ fixed, the following asymptotic results regarding $\hat{\phi}_{GMM}$ hold.*
(i) For $\gamma = 1/2$ and $N \to \infty$ followed by $T \to \infty$, or $T \to \infty$ followed by $N \to \infty$,

$$\sqrt{NT}(\hat{\phi}_{GMM} - \phi) \xrightarrow{L} \mathcal{N}(0, 4) \tag{28}$$

(ii) For $\gamma = 1$ and $N \to \infty$ followed by $T \to \infty$,

$$\sqrt{NT}(\hat{\phi}_{GMM} - \phi) \xrightarrow{L} \mathcal{N}\left(0, -8c\frac{1 - 2c - e^{2c}}{(1 + 2c - e^{2c})^2}\right) \tag{29}$$

(iii) For $\gamma = 1$ and $T \to \infty$ followed by $N \to \infty$,

$$\sqrt{NT}(\hat{\phi}_{GMM} - \phi) \xrightarrow{L} \mathcal{N}\left(0, -8c\frac{1 - 2c - e^{2c}}{(e^{2c} - 2c - 1)^2}\right) \tag{30}$$

[5] Let X be standard Cauchy, with p.d.f $f(x) = \frac{1}{\pi(1+x^2)}$ for $x \in \mathbb{R}$. The definition of expected value is $\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{a} xf(x)\,dx + \int_{a}^{\infty} xf(x)\,dx$, for some finite $a \in \mathbb{R}$. It suffices that only one of these integrals is finite for the expected value to exist. However, $\int_a^\infty xf(x)\,dx = \int_a^\infty x\frac{1}{\pi(1+x^2)}\,dx = \frac{1}{\pi}\lim_{w\to\infty}\int_a^w \frac{x}{(1+x^2)}\,dx = \frac{1}{2\pi}\left[\log(1+x^2)\right]_a^w$, which diverges, and similarly for $\int_{-\infty}^{a} xf(x)\,dx$. Then, by the Lyapunov inequality, it follows that if $\mathbb{E}[X]$ diverges, so does $\mathbb{E}[X^2]$. Finally, since the variance of X is $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, the variance does not exist either.

(iv) For $\gamma > 1$ and $N \to \infty$ followed by $T \to \infty$,

$$\sqrt{NT^{3-2\gamma}}(\hat{\phi}_{GMM} - \phi) \xrightarrow{L} \mathcal{N}\left(0, \frac{8}{c^2}\right) \tag{31}$$

(v) For $\gamma > 1$ and $T \to \infty$ followed by $N \to \infty$,

$$\sqrt{T}(\hat{\phi}_{GMM} - \phi) \xrightarrow{L} 2\,\mathcal{C} \tag{32}$$

Proof. See Phillips (2014). ■

From Theorem 5, it is clear that the asymptotic properties depend on both how close ϕ is to unity, as well as the order of convergence. Specifically, (v) shows that normality can break down even when ϕ is lower than one. Note that the distance between ϕ and unity is related to the value of γ. Hence, $\gamma > 1$ in (iv) and (v) implies that in these two cases, ϕ is closer to unity than in cases (i) to (iii). Hence, ϕ must be sufficiently close to one in order for the limiting distribution to become Cauchy. Furthermore, T must tend to infinity before N; otherwise, the limiting distribution will be normal with limiting variance as in (31).

To conclude this section, a few heuristic comments on the strengths and weaknesses of the two estimators. The main issue with GMM-based estimators is the rapid growth of the moment conditions. From equation (8), it is straightforward to see that the increase in moment conditions is $\mathcal{O}(T^2)$; that is, the number of moment conditions increases quadratically with T. Hence, by increasing T and keeping N fixed, there will eventually be more moment conditions than observations. This situation may lead to efficiency problems (Satchachai and Schmidt 2008). On the other hand, the main advantage of the GMM vis-à-vis ML-based methods is that it is not necessary to assume normality of the errors. For example, in finance, normality is usually too strong of an assumption, due to the heavy-tailedness of financial returns.

While the asymptotic properties of the two estimators differ quite substantially, particularly in the unit root case, the infinite sample-results are not of immense practical importance. For instance, an econometrician examining cross-country differences in inflation rates is limited to $N \approx 200$, which is the approximate number of countries of the world. Additionally, more often that than not, data from developing countries will be scarcely available; hence, N is usually far less than 200. Similarly, most macroeconomic indicators are collected annually or quarterly at best, so T is usually relatively small as well. This further implies that it is, instead, the finite-sample properties that are crucial for the practical usability of the two estimators.

3 Monte Carlo setup

This section will briefly describe the Monte Carlo procedure. Again, the model of interest is

$$y_{it} = \alpha_i + \phi y_{i,t-1} + u_{it}$$

where $u_{it} \sim \mathcal{N}(0, 1)$. Individual effects are generated according to $\alpha_i = (\lambda - 1)/\sqrt{2}$, where $\lambda \sim \chi^2(1)$, so that $\mathbb{E}[\alpha_i] = 0$ and $\mathbb{V}[\alpha_i] = 1$. The autoregressive parameter is varied according to $\phi \in \{0.90, 0.95, 0.99\}$, while the values considered for N and T are $N \in \{50, 150, 500\}$ and $T \in \{5, 10, 20, 30, 50\}$, respectively. The number of Monte Carlo replications is set to 1,000. Due to the poor performance of the Arelllano-Bond GMM when ϕ is near unity, only the FDML and Blundell-Bond continuously-updating GMM will be used as estimation methods.

The experiment design is similar to that of Hayakawa and Pesaran (2015), with two major differences. Firstly, in this thesis, ϕ takes values very close to unity, while the Hayakawa and Pesaran paper only considers $\phi = 0.4$ and $\phi = 0.9$. Secondly, the previously mentioned paper weakens the homoscedasticity assumption and allows for cross-sectional heteroscedasticity, i.e. $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$, where the error variances are $\sigma_i^2 \sim \mathcal{U}(0.5, 1.5)$ [6]. However, since the purpose of this thesis is to investigate the effects of near non-stationarity rather than of heteroscedasticty on the bias, size and power, the framework is simplified slightly.

The software used in the Monte Carlo simulations is Matlab, version R2016a, together with an external add-in for FDML and GMM for dynamic panel data (Hayakawa 2017).

4 Results

Tables 1-3 in the Appendix report the mean and median bias, as well as the size and power of the FDML and GMM estimators. The tables correspond to $\phi = 0.90$, $\phi = 0.95$, and $\phi = 0.99$, respectively. For the power calculations, the value of ϕ under the alternative hypothesis is always 0.1 units lower than the value under the null hypothesis. The reason that the numerical difference between the null and alternative hypotheses is fixed, is because given fixed sample size, the power is larger when the difference between the two values of ϕ is high.

Consider first the bias results as T increases and N is fixed. The absolute values of the biases, both the mean and median, are decreasing as T increases, which is expected. The same results hold when N increases and T is fixed. Hence, the absolute biases are decreasing both with N and T. However, generally, in order to reduce bias, it is more important to increase T rather than N. This holds for all values of ϕ, and is particularly true for the FDML. That the bias decreases faster with T for the FDML than the GMM is also noted by Hayakwa and Pesaran (2015).

The absolute biases are generally larger when ϕ is closer to unity; this applies for both estimators. Also, the absolute biases are greater for the FDML in most cases. However, for $T = 20$, $T = 30$ and $T = 50$, the FDML generally outperforms the continuously-updating GMM when $\phi = 0.90$ and $\phi = 0.95$.

Diverting attention to the case with $\phi = 0.99$, the GMM estimator is better

[6] This implies that $\mathbb{E}[\sigma_i^2] = 1$ and $\mathbb{V}[\sigma_i^2] = \frac{1}{12}$.

for virtually every combination of N and T; the exceptions being the mean bias for the combination of $T = 50$ and $N = 500$, as well as the median biases for the $T = 30$ when $N = 150$ and 500. The FDML performs relatively poorly when ϕ is this close to unity, especially when $T < 20$. The GMM, on the other hand, performs much better. Table 3 shows that the mean and median biases are close to zero for the GMM when $\phi = 0.99$. Considering that the Blundell-Bond version of the GMM is more or less tailor-made for the situation with close-to-non-stationary data, the bias results do not come as a major surprise. The mean and median biases are generally negative for the FDML, while for GMM, they tend to be negative for $\phi = 0.90$ and $\phi = 0.95$ and positive for $\phi = 0.99$.

However, there is a noteworthy peculiarity for the GMM mean bias results when $\phi = 0.99$. Firstly, for $N = 150$ and $N = 500$, the absolute values of the mean biases decrease, and then increase again, as T increases. Secondly, for high values of T ($T = 30$ and $T = 50$), the absolute mean biases do not decrease with N, which is remarkable given that the sample size more than triples from 150 to 500. This instability of the mean noted in the GMM case, but not for the FDML, is an indication of the presence of Cauchy properties, which are predicted by theory given that either of Theorem 4 or Theorem 5(v) holds. Considering the GMM median bias instead, the median bias will asymptotically tend to zero, and the rate of convergence is $\mathcal{O}_p(\sqrt{T})$ according to theory. However, in the $\phi = 0.99$ case, it is only as $N = 500$ that the median bias tends zero quickly as T increases. Moreover, already with $T = 30$, the bias results of the FDML are on par with those of the GMM.

Another property of the normal distribution is that the mean and median are asymptotically equal. For the FDML, the mean and median absolute biases are almost identical when N and T are simultaneously equal to 500 and 50, with a slightly larger difference for $\phi = 0.99$. However, for the GMM, this is never true. Instead, there are considerable differences between the mean and median biases, again indicating the presence of Cauchy properties in the GMM estimator.

While the performance of the GMM is superior to the FDML in terms of bias, the size is considerably higher than 5% for the GMM, irrespective of the value of ϕ. Interestingly, the size of the GMM estimator is increasing with T, a peculiarity also noted in Hayakawa and Pesaran (2015). Additionally, the size of the GMM estimator is increasing with the autoregressive parameter ϕ, an effect not observed in the likelihood estimator. For example, when $\phi = 0.99$, the size is above 96% even for $N = 150$ and $T = 30$. The augmentation of the size distortion for high values of ϕ is noted in Hsiao and Zhang (2015), although in that paper, the highest value considered for ϕ is 0.8. This thesis shows that the size problem for the GMM is further exacerbated when ϕ is very close to one. Regarding the power, it is considerably lower for the FDML, especially for small values of T. Additionally, the power of the FDML is deteriorating as ϕ approaches unity. However, when $T = 30$ and 50, the power of both estimators is close to 100%, regardless of N.

A final comment on the GMM is that there are no results when $N = 50$ and T is 20, 30 and 50. This is because there are too many moment conditions relative to the number of observations in these cases. Although this can be remedied by not using the full set of moment conditions, this will cause an increase in the asymptotic variance, particularly in the LUR case, as shown in Section 2.3.

5 Concluding remarks

The purpose of this thesis has been to investigate the finite-sample performances of the FDML and continuously-updating system GMM in the dynamic panel context, with a particular focus on the situation with close to non-stationary data. Because of the wide range of applications of the dynamic panel model in applied econometrics, the finite-sample properties of the different estimators are of great importance.

The main finding of this thesis is the relatively large increase in bias of the FDML as the value of the autoregressive parameter approaches unity. Additionally, the results of the Monte Carlo simulations show that the absolute bias of the system GMM is lower for most combinations of N and T. However, as expected, the biases are decreasing rapidly as N and T increase. The size and power of the GMM estimator are both considerably higher than those of the FDML. Notably, the size of the GMM estimator is increasing with both the value of the autoregressive parameter ϕ and with the number of time periods T. This peculiarity is not seen in the likelihood-based estimator. The power of both estimators is shown to be close to 100% with N and T sufficiently large, although the power of the FDML deteriorates slightly in the case of $\phi = 0.99$, even in the large-sample case. Another important result of the thesis concerns the Cauchy properties of the GMM, which are manifested by instability in the values for the mean bias. These peculiarities become evident when $\phi = 0.99$.

As with any study, there are a number of limitations. Particularly, increasing the number of Monte Carlo replications will certainly lower the bias and increase the power. However, even with a lower number of replications, the relative difference between the estimators can be inferred. Likewise, there is virtually an infinite number of combinations of N and T that can be explored. However, this will be saved for future research.

This thesis gives rise to two additional research questions. Firstly, whether the FDML can be modified, so as to improve in cases with near-non-stationarity. Secondly, if the results hold when extending the model to include more independent variables than just the lagged dependent variable. This situation is even more interesting in applied research, as researchers are often interested in the effect of exogenous covariates on the dependent variable.

References

[1] Álvarez, J. and Arellano, M. (2003). The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica* 71 (4), 1121-1159.

[2] Andersen, E.B. (1970). Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society, Series B* 32 (2), 283-301.

[3] Anderson, T.W. and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76 (375), 598-606.

[4] Arellano, M. and Bond, S.R. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58 (2), 277-297.

[5] Arellano, M. and Bover, O. (1995). Another Look at the Instrumental Variables Estimation of Error Components Models. *Journal of Econometrics* 68 (1), 29-52.

[6] Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87 (1), 115-143.

[7] Hahn, J. and Kuersteiner, G. (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large. *Econometrica* 70 (4), 1639-1657.

[8] Han, C. and Phillips, P.C.B. (2013). First difference maximum likelihood and dynamic panel estimation. *Journal of Econometrics* 175 (1), 35-45.

[9] Hansen, L.P. (1982). Large Sample Properties of Generalized Methods of Moments Estimators. *Econometrica* 50 (4), 1029-1054.

[10] Hansen, L.P., Heaton, J. and Yaron, A. (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* 14 (3), 262-280.

[11] Hayakwa, K. *Home Page* [online]. Available at http://home.hiroshima-u.ac.jp/kazuhaya/cv-e.html. Published 4 April 2017. Accessed 9 August 2017.

[12] Hayakwa, K. and Pesaran, M.H. (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics* 188 (1), 111-134.

[13] Hogg, R.V., McKean, J. and Craig, A.T. (2014). *Introduction to Mathematical Statistics*. Seventh Edition. Harlow, Pearson.

[14] Hsiao, C., Pesaran, M.H. and Tahmiscioglu, A.K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109 (1), 107-150.

[15] Hsiao, C. and Zhang, J. (2015). IV, GMM or likelihood approach to estimate dynamic panel models when either N or T or both are large. *Journal of Econometrics* 187 (1), 312-322.

[16] Hsiao, C. and Zhou, Q. (2017). First difference or forward demeaning: Implications for the method of moments estimators. *Econometric Reviews*. Forthcoming.

[17] Kruiniger, H. (2008). Maximum likelihood estimation and inference methods for the covariance stationary panel AR(1)/unit root model. *Journal of Econometrics* 144 (2), 447-464.

[18] Madsen, E. (2008). GMM Estimators and Unit Root Tests in the AR(1) Panel Data Model. *Working paper, University of Copenhagen*.

[19] Newey, W.K. and McFadden, D.L. (1994). Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D.L. (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam, North Holland.

[20] Newey, W.K. and Smith, R.J. (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72 (1), 219-255.

[21] Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica* 49 (6), 1417-1426.

[22] Pakes, A. and Pollard, D. (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica* 57 (5), 1027-1057.

[23] Phillips, P.C.B. (2014). Dynamic Panel GMM with Roots Near Unity. *Working paper, Yale University*.

[24] Phillips, P.C.B. (2015). Dynamic Panel Anderson-Hsiao Estimation With Roots Near Unity. *Econometric Theory*, First View, 1-24.

[25] Satchachai, P. and Schmidt, P. (2008). GMM with more moment conditions than observations. *Economics Letters* 99 (2), 252-255.

Result tables

	FDML						*GMM*				
N/T	5	10	20	30	50		5	10	20	30	50
					Mean bias ($\times$ 100)						
50	-7.333	-1.501	-0.037	0.025	0.090		0.739	-1.496	–	–	–
150	-4.692	0.120	0.225	-0.006	-0.056		-2.088	-1.087	-0.709	-0.659	-0.050
500	-2.345	0.537	0.052	0.048	-0.034		-0.942	-0.439	-0.114	-0.173	-0.123
					Median bias ($\times$ 100)						
50	-4.397	-0.826	-0.288	-0.135	-0.0090		3.323	0.889	–	–	–
150	-2.993	0.077	-0.017	-0.010	0.423		-2.933	-0.071	-0.407	-0.369	1.342
500	-1.512	0.184	-0.041	0.023	-0.035		0.196	-0.053	-0.240	-0.056	-0.058
					Size						
50	15.0	16.6	17.7	9.6	4.8		54.5	80.0	–	–	–
150	14.5	20.1	10.2	6.1	5.8		32.6	40.7	57.0	71.2	97.7
500	13.5	18.8	6.3	5.3	5.0		16.5	18.4	22.8	22.8	33.3
					Power (H_1: $\phi = 0.80$)						
50	26.6	43.1	75.6	98.8	100.0		60.4	89.3	–	–	–
150	32.1	51.2	98.5	99.9	100.0		32.3	86.0	97.2	97.5	98.7
500	38.7	79.2	99.1	100.0	100.0		83.3	98.1	100.0	100.0	100.0

Table 1: Mean bias, median bias, size and power for $\phi = 0.90$.

N/T	*FDML* 5	10	20	30	50	*GMM* 5	10	20	30	50
					Mean bias ($\times$ 100)					
50	-9.397	-3.618	-0.593	0.018	0.042	2.526	-0.551	–	–	–
150	-6.623	-2.045	-0.075	0.129	0.034	-0.148	-1.419	-1.165	-0.842	0.355
500	-4.297	-0.785	0.172	0.074	0.053	-1.053	-0.876	-0.314	0.192	-0.141
					Median bias ($\times$ 100)					
50	-5.272	-1.964	-0.184	-0.055	-0.071	3.286	3.543	–	–	–
150	-3.379	-1.419	-0.261	-0.003	0.079	-3.679	-0.010	-0.301	-0.369	1.268
500	-2.458	0.424	-0.033	-0.047	0.056	-0.310	-0.111	-0.113	-0.076	-0.023
					Size					
50	14.9	15.9	17.9	19.3	12.5	68.6	87.1	–	–	–
150	14.4	14.2	19.5	16.8	7.5	51.3	63.4	76.4	80.2	98.6
500	18.5	13.9	17.7	9.4	3.6	29.4	36.8	38.4	41.0	50.1
					Power (H_1: $\phi = 0.85$)					
50	23.5	37.0	68.0	95.3	99.6	89.0	94.8	–	–	–
150	31.2	46.8	96.3	99.4	99.9	86.1	91.5	96.3	97.5	99.6
500	35.7	64.2	98.9	99.8	100.0	87.0	95.2	100.0	100.0	100.0

Table 2: Mean bias, median bias, size and power for $\phi = 0.95$.

N/T	*FDML* 5	10	20	30	50	*GMM* 5	10	20	30	50
					Mean bias ($\times$ 100)					
50	-12.133	-5.181	-2.394	-1.382	-0.711	1.429	0.066	–	–	–
150	-8.169	-3.529	-1.550	-0.850	-0.361	1.411	0.082	0.090	0.064	-0.262
500	-5.734	-2.599	-0.923	-0.465	-0.183	0.479	0.059	0.027	-0.315	-0.224
					Median bias ($\times$ 100)					
50	-8.922	-3.323	-1.660	-0.798	-0.384	0.915	0.925	–	–	–
150	-4.920	-2.098	-1.037	-0.265	-0.207	0.973	0.790	0.809	0.785	0.908
500	-3.219	-1.659	-0.547	-0.270	-0.165	0.861	0.636	0.349	0.149	0.082
					Size					
50	18.1	15.2	12.5	13.1	14.5	77.6	92.0	–	–	–
150	15.5	13.2	13.3	13.0	12.5	71.4	87.9	93.0	96.2	99.1
500	13.5	15.1	13.5	13.6	14.7	67.6	80.5	82.9	86.6	89.4
					Power (H_1: $\phi = 0.89$)					
50	15.0	33.7	57.8	90.0	97.2	97.8	99.0	–	–	–
150	26.9	41.7	89.2	95.2	98.1	97.8	98.7	99.6	99.6	99.3
500	36.1	48.2	95.1	97.5	99.8	98.5	98.1	100.0	100.0	100.0

Table 3: Mean bias, median bias, size and power for $\phi = 0.99$.