Half-panel jackknife estimation of GMM models with fixed effects

Adrian Mehic

NEKN01: Master's thesis in Economics (15 ECTS)

Lund University
School of Economics and Management
Supervisor: Joakim Westerlund
20 August 2018

Abstract

In empirical economics, the generalized method of moments (GMM) is one of the most widely used methods for estimating models with fixed effects, such as the dynamic panel model. For the dynamic panel model, the two most widely used GMM-based estimators are the so-called *difference GMM* and *system GMM* estimators. However, it is typically the case that such estimators are asymptotically biased of order N^{-1}. To remedy this problem, this thesis extends the half-panel jackknife (HPJ) estimator of Dhaene and Jochmans (2015) to GMM models with fixed effects and $O(N^{-1})$ bias. In theory, this should reduce asymptotic bias from $O(N^{-1})$ to $O(N^{-2})$. The Monte Carlo results show that the HPJ gives satisfactory finite-sample bias improvements only for the difference GMM. For the system GMM, using the HPJ results in bias reductions only under very special circumstances.

Contents

1 Introduction

A panel data set contains observations of sample size N, collected during multiple time periods, T. If fixed effects are present, it may be the case that the maximum likelihood estimator (MLE) of the parameter of interest is inconsistent and biased of order $O(T^{-1})$, when the number of time periods T is fixed while the number of cross-sectional units N grows large. In recent years, researchers have developed methods for reducing this bias for at least one order of magnitude, that is, from $O(T^{-1})$ to $O(T^{-2})$. Most of these papers have covered the case without exogenous covariates − examples include analytical approaches by Lancaster (2002), Hahn and Kuersteiner (2004), and Arellano and Bonhomme (2009). Moreover, there are several non-analytical methods for bias correction, often based on the jackknife, a classical resampling technique from statistics. These include, for the MLE, the delete-one panel jackknife of Hahn and Newey (2004) and the split-panel jackknife (SPJ), which was proposed by Dhaene and Jochmans (2015). The latter approach is particularly appealing due to its relative simplicity, and it was extended to the case with $O(T^{-1} + N^{-1})$ bias by Fernández-Val and Weidner (2016), and to the case with with weakly exogenous covariates by Chudik et al. (2018).

However, for estimators based on the generalized method of moments (GMM) of Hansen (1982), it is often the case that under endogenity, the bias is instead $O(N^{-1})$. The most prominent example of a GMM-based model with $O(N^{-1})$ bias is the so-called *dynamic panel model*, which is characterized by the inclusion of the lagged dependent variable as a regressor, in addition to possible exogenous covariates. Due to the popularity of this model in empirical research, it is important that the bias is reduced. Consequently, bias minimization has been the topic of several papers in the dynamic panel literature. Choi et al. (2010) consider the recursive mean adjustment (RMA), while Lee et al. (2017) and Zhang and Zhou (2018) use the jackknife instrumental variables estimation (JIVE) of Angrist et al. (1999), which eliminates the bias by removing observations in the construction of the instrument matrix. Although this last method has proved to be successful in reducing finite-sample bias, it is not as computationally attractive as the SPJ.

The purpose of this thesis is thus to extend the SPJ approach to GMM. Specifically, I consider the simplest form of SPJ, namely the half-panel jackknife (HPJ). In theory, the HPJ should reduce the GMM bias from $O(N^{-1})$

to $O(N^{-2})$. The focus is on the dynamic panel model, and I will particularly consider the *difference GMM* estimator of Arellano and Bond (1991), as well as the *system GMM* of Blundell and Bond (1998), which are the two empirically most common GMM estimators of dynamic panel models.

The Monte Carlo results reported in the thesis shows that the HPJ considerably reduces both the finite-sample bias as well as the standard error of the difference GMM. However, the effect on the system GMM is minor. For large values of N, when the uncorrected estimates already have a very low bias, the HPJ actually increases finite-sample bias. For the so-called continuously updating system GMM, which is a variant of the GMM known for its low finite-sample bias, the HPJ leads to increased bias, and no significant reduction in the standard errors. Regardless of which of the GMM-based estimators is used, the jackknife approach results in the point estimate of the parameter of interest increasing in value vis-à-vis the uncorrected estimates.

The reminder of the thesis is organized as follows. Section 2 introduces the GMM estimation technique, as well the dynamic panel model and its corresponding GMM-based estimators. Section 3 describes the Monte Carlo design. Section 4 presents the results of the simulation study. The thesis concludes with Section 5.

Finally, a word on notation. Throughout this thesis, a sequence of random elements $\{X_N\}$ is said to be $O_p(1)$ 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, $X_N = O_p(Y_N)$ means that X_N/Y_N is bounded in probability. The notation $o_p(1)$ means that the sequence $\{X_N\}$ converges in probability to zero. I use $O(\cdot)$ to denote the order of the function. The symmetric matrix $\boldsymbol{B}$ is said to be *positive semi-definite* (p.s.d.) if the quadratic form $z'\boldsymbol{B}z$ is non-negative for all nonzero column vectors z. I use the abbreviation cdf for the cumulative density function. The notation $||\boldsymbol{A}||$ denotes the norm $\text{tr}(\boldsymbol{A}'\boldsymbol{A})^{1/2}$ for an arbitrary matrix $\boldsymbol{A}$, and $\boldsymbol{A} \otimes \boldsymbol{B}$ denotes the Kronecker product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. I use $\xrightarrow{a.s.}$ to denote almost sure convergence, $\xrightarrow{P}$ denotes convergence in probability, and $\xrightarrow{L}$ denotes convergence in law (in distribution). All other symbols are defined on site.

2 Theory

The structure of this section is as follows. Section 2.1 introduces the GMM estimation technique, while Section 2.2 discusses the issue of GMM bias. Section 2.3 is on the dynamic panel model. Section 2.4 introduces the HPJ for the GMM, and Section 2.5 discusses the bootstrap approach for calculation of standard errors used in the Monte Carlo analysis. Finally, Section 2.6 describes briefly the jackknife interpretation of the continuously updated estimator.

2.1 GMM estimation

Let $(\mathcal{D}, d)$ be a metric space, and let the panel data $\{z_{it}\} \equiv \{y_{it}, x_{it}\}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, be an i.i.d. random sample in $\mathbb{R}^s$. Assume that $\{z_{it}\}$ was generated by a stationary and ergodic stochastic process defined on the probability space $(\Omega, \mathcal{A}, P)$, where Ω is the sample space, $\mathcal{A}$ is a σ-algebra of random events, and P is a probability measure on $\mathcal{A}$. The sample is drawn from an unknown probability distribution $\mathcal{G}_0$. Our goal is to estimate the unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^k$, for which $\text{int}(\Theta) \neq \emptyset$, and Θ is assumed to be compact. The parameter θ_0 could be a scalar or a vector. Denote by $B(\theta, \xi)$ an open ball in Θ of radius ξ centered at θ, and by $L(\theta)$ a nonstochastic Borel measurable population objective function, for which $\sup_{\theta \in \Theta} L(\theta) < \infty$. We assume that there exists $\theta_0 \in \Theta : \forall \xi > 0, \inf_{\theta \notin B(\theta_0, \xi)} L(\theta) > L(\theta_0)$. Hence, θ_0 must uniquely minimize $L(\theta)$ over Θ. The sample equivalent of $L(\theta)$, which puts mass N^{-1} at each sample point, is denoted by $L_N(\theta)$, for which it holds that $\sup_{\theta \in \Theta} |L_N(\theta) - L(\theta)| \xrightarrow{P} 0$. Let α_i be a scalar individual effect, which is unobserved by the researcher. Fixed effects are standard in empirical economics, and a considerable share of the econometrics literature is, thus, focused on ameliorating fixed effects models. An example of a sample objective function is the average log-likelihood function,

$$L_N(\theta) = \frac{1}{NT} \sum_{i=1}^{T} \sum_{t=1}^{T} \log f_{it}(y_{it} | x_{it}; \theta, \hat{\alpha}_i(\theta)) \tag{1}$$

where $\hat{\alpha}_i(\theta) = \arg\sup_{\alpha_i} \frac{1}{T} \sum_{t=1}^{T} \log f_{it}(y_{it} | x_{it}; \hat{\alpha}_i, \theta)$. This is then used to construct the maximum likelihood estimator $\hat{\theta}_{ML}$ of θ_0,

$$\hat{\theta}_{ML} = \arg\sup_{\theta \in \Theta} L_N(\theta)$$

For fixed T, it holds that $\hat{\theta}_{ML}$ is inconsistent for θ_0, i.e. $\text{plim}_{N\to\infty}\hat{\theta}_{ML} \neq \theta_0$. More specifically, for T large, we can write $\text{plim}_{N\to\infty}(\hat{\phi}_{ML}-\theta_0) \approx -(1+\theta)/(T-1)$. This is known as the *incidental parameter problem* (Neyman and Scott 1948), because it is caused by the presence of the individual effects $\alpha_1, \ldots, \alpha_N$. The incidental parameters problem therefore leads to a bias, that is of order T^{-1}, or $O(T^{-1})$.

Alternatively, if

$$\forall i : \mathbb{E}\left[g(z_i; \alpha_i, \theta)\right] \equiv \int g(z_i; \alpha_i, \theta)\, d\mathcal{G}_0 = 0 \tag{2}$$

denotes a set of moment equalities, where $g : \mathbb{R}^s \times \Theta \to \mathbb{R}^m$, $m \geq k$, is a known (up to θ) real-valued measurable map, with $z_i = (z_{i1}, \ldots, z_{iT})'$, and $\boldsymbol{W}_N(\theta) : \Theta \to \mathbb{R}^{m\times m}$ is a random weighting matrix, $L_N(\theta)$ could be the GMM objective function with continuously updating weighting matrix (Hansen et al. 1996),

$$L_N(\theta) = \bar{g}_N(\theta)'\boldsymbol{W}_N(\theta)^{+}\bar{g}_N(\theta) \tag{3}$$

where $\bar{g}_N(\theta) \equiv \bar{g}_N(z_i; \alpha_i, \theta) = \frac{1}{N}\sum_{i=1}^{N} g(z_i; \alpha_i, \theta)$. The notation $\boldsymbol{A}^{+}$ denotes the Moore-Penrose pseudoinverse of the arbitrary matrix $\boldsymbol{A}$. The weighting matrix is continously evaluated at the parameter values used for the moments, and hence, updated for each iteration. The GMM estimator $\hat{\theta}_{GMM}$ of θ_0 is given by

$$\hat{\theta}_{GMM} = \arg\inf_{\theta\in\Theta} L_N(\theta) \tag{4}$$

However, an empirically more common strategy for the weighting matrix is to use a two-step approach. Let $\boldsymbol{W}_N$ be a p.s.d. matrix, such as the identity matrix. Let also $\boldsymbol{\Lambda}_N(\theta) = N^{-1}\sum_{i=1}^{N}\left[g_i(\theta)g_i'(\theta)\right]$, where $g_i(\theta) = g(z_i; \alpha_i, \theta)$. Then, a preliminary estimate of θ_0 is given by $\dot{\theta} = \arg\inf_{\theta\in\Theta} \bar{g}_N(\theta)'\boldsymbol{W}_N^{-1}\bar{g}_N(\theta)$. The so-called *two-step GMM estimator* (Hansen 1982) is then constructed by using the objective function

$$L_N(\theta) = \bar{g}_N(\theta)'\boldsymbol{\Lambda}_N(\dot{\theta})^{-1}\bar{g}_N(\theta) \tag{5}$$

and using (4) to obtain the final $\hat{\theta}_{GMM}$. I shall refer to the first mentioned variant of the GMM estimator, which utilizes a parameter-dependent weighting matrix, as the *continuously updated estimator* (CUE). Note that the weighting matrix in the two-step estimator is not a function of the parameter θ, unlike $W_N(\theta)$. This also means that the first-order conditions (FOCs) associated with

(3) and (5) will not be the same, since θ appears in three terms in the CUE specification. However, Pakes and Pollard (1989) show that, when $\theta_0 \in \text{int}(\Theta)$, and given that $\boldsymbol{W}_N(\theta) \xrightarrow{P} \boldsymbol{W}$, $||\boldsymbol{W}_N(\theta_0)|| = O_p(1)$ and $\sup_{\theta\in\Theta}||\boldsymbol{W}_N(\theta)^{-1}|| = O_p(1)$, the CUE is consistent if the two-step estimator is consistent. Criteria for consistency of the GMM is discussed in some additional detail in the Appendix.

By Theorem 2.1 of Newey and Smith (2004), the CUE is a special case of the so-called *generalized empirical likelihood* (GEL) estimator (cf. Smith 1997). Let $\rho(\tau)$ be a scalar function, and $\mathcal{T}$ be an open interval containing zero. Let also $\Upsilon_N \equiv \Upsilon_N(\theta) = \{\upsilon : \upsilon' g_i(\theta), \in \mathcal{T}, i = 1, \ldots N\}$. The GEL estimator is the solution to the saddle point problem

$$\hat{\theta}_{GEL} = \arg \inf_{\theta\in\Theta} \sup_{\upsilon\in\Upsilon_N} \sum_{i=1}^{N} \rho\left[\upsilon' g_i(\theta)\right] \tag{6}$$

If $\rho(\tau) = -\frac{1}{2}\tau^2 - \tau$, we obtain the CUE [1]. The GEL representation of the CUE is of particular relevance when decomposing the bias of the GMM, as in Section 2.2 of this thesis.

2.2 Bias of the GMM

In models with endogenity, for instance in dynamic panel models with fixed effects, the GMM may be subject to bias. To discuss this in some more detail, we need some additional regularity assumptions, apart from those already established for consistency and asymptotic normality. Let ∇^j be a vector of the j:th partial derivative with respect to θ, and let be $\mathcal{Y}$ be a neighborhood around θ_0.

Assumption 1. *$\hat{\theta}_{GMM}$ is consistent and asymptotically normal.*

Assumption 2. $\exists\, (\boldsymbol{W}, \zeta(z)) : \boldsymbol{W}_N = \boldsymbol{W} + N^{-1}\sum_{i=1}^{N}\zeta(z_i)$*, where* $\mathbb{E}\left[\zeta(z_i))\right] = 0$ *and* $\mathbb{E}\left[|\zeta(z_i)|^6\right] < \infty$*, and* $\boldsymbol{W}$ *is p.s.d.*

Assumption 3. *Let* $\varphi(z)$ *with* $\mathbb{E}\left[\varphi(z_i)^6\right] < \infty$ *be such that on* $\mathcal{Y}$*,* $\exists$*:* $\nabla^j g(z,\theta)$ *satisfying* $\sup_{\theta\in\mathcal{Y}}||\nabla^4 g(z,\theta)|| \leq \varphi(z)\ \forall z$ *and for* $j \in [0,4]$*, and* $||\nabla^4 g(z,\theta) - \nabla^4 g(z,\theta_0)|| \leq \varphi(z)\, ||\theta - \theta_0||\, \forall\, \theta \in \mathcal{Y}$*.*

Let $\boldsymbol{\Gamma}$ be the Jacobian matrix, i.e. $\boldsymbol{\Gamma} = \mathbb{E}\left[\partial g_i(\theta_0)/\partial\theta\right]$, let $\boldsymbol{\Lambda} = \mathbb{E}\left[g_i(\theta_0)g_i(\theta_0)'\right]$.

[1] Another popular estimator, the *empirical likelihood* (EL) of Qin and Lawless (1994) and Imbens (1997), is a special case of the GEL with $\rho(\tau) = \log(1-\tau)$ and $\mathcal{T} = (-\infty, 1)$

Now, define the following matrices:

$$\boldsymbol{\Sigma} = (\boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma})^{-1} \qquad \boldsymbol{\Xi} = (\boldsymbol{\Sigma}'\boldsymbol{\Gamma}\boldsymbol{\Lambda})^{-1} \qquad \boldsymbol{V} = \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'\boldsymbol{\Lambda}^{-1}$$

Let e_j be the j:th column of the identity matrix, and define $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}_i(\theta) = \partial g_i(\theta)/\partial\theta$, $\boldsymbol{M} = (\boldsymbol{\Gamma}'\boldsymbol{W}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}'\boldsymbol{W}^{-1}$ and $\bar{\boldsymbol{\Lambda}}_{\theta_j} = \mathbb{E}\left\{\partial\left[g_i(\theta_0)g_i(\theta_0)'\right]/\partial\theta_j\right\}$. Let also $a = (a_1, \ldots, a_m)'$, where each element is $a_j = \operatorname{tr}\left\{\boldsymbol{\Sigma}\mathbb{E}\left[\partial^2 g_{ij}(\theta_0)/\partial\theta\partial\theta'\right]\right\}/2$, and $g_{ij}(\theta)$ is the j:th element of $g_i(\theta)$. Then, we can decompose the GMM bias (Newey and Smith 2004; Anatolyev 2005) according to

$$\hat{\theta}_{GMM} - \theta_0 = B_1 + B_2 + B_3 + B_4 \tag{7}$$

where

$$B_1 = \boldsymbol{\Xi}\left[-a + \mathbb{E}(\boldsymbol{\Gamma}_i'\boldsymbol{\Xi}g_i)\right]/N$$

$$B_2 = -\boldsymbol{\Sigma}\mathbb{E}(\boldsymbol{\Gamma}_i'\boldsymbol{V}g_i)/N$$

$$B_3 = \boldsymbol{\Xi}\mathbb{E}(g_i g_i'\boldsymbol{V}g_i)/N$$

$$B_4 = -\boldsymbol{\Xi}\sum_{i=1}^{p}\bar{\boldsymbol{\Lambda}}_{\theta_j}(\boldsymbol{M} - \boldsymbol{\Xi})'e_j/N$$

Each of the four terms in (7) have their own interpretation. The term B_1 is the bias for a GMM estimator with the optimal covariance matrix (cf. Remark 2 of Theorem 2 of the Appendix), B_2 arises from the estimation of $\boldsymbol{\Gamma}$ and B_3 is the bias arising from the estimation of $\boldsymbol{\Lambda}$. B_2 will be zero if the $\boldsymbol{\Gamma}_i$'s are constant, whereas B_3 will be zero if the third moments are zero. However, both are typically nonzero under endogenity. Finally, the term B_4 arises from the choice of the first-step estimator, and is zero if $\boldsymbol{W}$ is a scalar multiple of $\boldsymbol{\Lambda}$.

It is important to note that the bias is now $O(N^{-1})$, and not $O(T^{-1})$ as in the ML case. This means that we cannot use the HPJ estimator of Dhaene and Jochmans (2015) for GMM-based models without adjustment. The infeasibility of the HPJ has not been addressed in previous literature, and is, thus, the motivation behind this thesis.

Turning to the CUE, recall that it is a special case of the GEL estimator with $\rho(\tau) = -\frac{1}{2}\tau^2 - \tau$. Now, let $\rho_j(\tau) = \partial\rho^j(\tau)/\partial\tau^j$ and $\rho_j = \rho_j(0)$. It is easy to verify that for the CUE, $\rho_1 = \rho_2 = -1$ and $\rho_3 = 0$. Given that $\rho(\tau)$ is four times differentiable, and the fourth derivative is Lipschitz continuous in a neighborhood of zero, it holds for the GEL that (Newey and Smith 2004),

$$\hat{\theta}_{GEL} - \theta_0 = B_1 + \left(1 + \frac{\rho_3}{2}\right) B_3 \tag{8}$$

provided that Assumptions 1 and 3 are satisfied. Hence,

$$\hat{\theta}_{CUE} - \theta_0 = B_1 + B_3 \tag{9}$$

Thus, using the CUE eliminates the bias arising from the Jacobian and from the preliminary estimator. [2] As a consequence of the reduction in asymptotic bias, the CUE is associated with lower finite-sample bias, which is of great importance in econometrics. The standard errors are typically also lower. However, due to the low bias and variance of the CUE, it tends to be oversized and thus unreasonable to use for testing purposes (cf. Hsiao et al. 2002).

2.3 GMM estimation of dynamic panel models

In this thesis, the focus is on one of the empirically most important panel models, the dynamic panel model. It can be described by

$$y_{it} = \alpha_i + \phi_0 y_{i,t-1} + u_{it} \tag{10}$$

I shall assume throughout this section that $0 < |\phi_0| < 1$, $\mathbb{E}[\alpha_i] = 0$, $\mathbb{E}[\alpha_i^2] = \sigma_\alpha^2$, and that $u_{it} \sim IID(0, \sigma_u^2)$ with $\mathbb{E}[|u_{it}|^8] < \infty$. It is further assumed that the initial observations $y_{i0} = O_p(1)$ are observed.

There are two classes of GMM estimators of dynamic panel models with fixed effects, the *difference GMM* of Arellano and Bond (1991), and the *system GMM*, which is due to Arellano and Bover (1995) and Blundell and Bond (1998). The main reason for the introduction of the GMM in the dynamic panel literature is the endogenity problem associated with the model. Since y_{it} depends on α_i, and α_i is the same for all time periods, the explanatory variable $y_{i,t-1}$ also depends on α_i. However, historically, the first remedies to the endogenity problem were not based on the GMM. Anderson and Hsiao (1981) suggested using lags of the dependent variable as instruments. Taking the first-difference to get rid of the fixed effects, the suggestion is to use $\Delta y_{i,t-2}$ as an instrument for $\Delta y_{i,t-1}$, which is uncorrelated with $\Delta u_{i,t}$ given that the errors are serially

[2] A slightly more formal description of the impact of CUE estimation on the Jacobian can be found in Donald and Newey (2000).

uncorrelated. However, as was first noted by Holtz-Eakin et al. (1988), this approach is inefficient, since not all information is used. For example, for $t = T$, the first-differenced model can be written

$$y_{i,T} - y_{i,T-1} = \phi(y_{i,T-1} - y_{i,T-2}) + u_{i,T} - u_{i,T-1} \tag{11}$$

for which the instruments $y_{i1}, y_{i2}, \ldots, y_{i,T-2}$ can be utilized. For $t = 3, \ldots, T$ and $s \geq 2$, the $(T-1)(T-2)/2$ moment conditions can be compactly written

$$\mathbb{E}[y_{i,t-s}\Delta u_{it}] = 0 \tag{12}$$

Now, define the instrument matrix $\boldsymbol{Z}_i$ as

$$\boldsymbol{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & y_{i1} & y_{i2} & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & y_{i1} & \ldots & y_{i,T-2} \end{pmatrix} \tag{13}$$

By construction of this matrix, the last row (corresponding to $t = T$), consists of zeros and all available instruments for the final time period. Let $\Delta u_i = (\Delta u_{i3}, \ldots, \Delta u_{iT})'$ be the vector of first-differenced errors. Using this notation, the moment conditions can be written

$$\mathbb{E}[\boldsymbol{Z}_i'\Delta u_i] = 0 \tag{14}$$

Because the term in brackets in (14) corresponds to $g(\cdot)$ in Section 2.1, Arellano and Bond (1991) construct a GMM-based estimator of ϕ_0 based on the above moment conditions as the solution to

$$\hat{\phi}_{AB} = \arg\inf_{\phi\in\Phi} \left(\frac{1}{N}\sum_{i=1}^{N}\Delta u_i'\boldsymbol{Z}_i\right)\boldsymbol{W}_N\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i'\Delta u_i\right) \tag{15}$$

where Φ is the compact set of all possible parameters. However, it was shown by Arellano and Bover (1995) that the Arellano-Bond estimator significantly underestimates ϕ_0 when the true value is close to unity, and that the bias starts to increase already when $\phi_0 = 0.8$. In order to remedy this problem, Arellano and Bover (1995) and Blundell and Bond (1998) introduce $(T-1)(T-2)/2$

additional moment conditions, namely

$$\mathbb{E}[\pi_{it}\Delta y_{i,t-s}] = 0 \tag{16}$$

for $t = 3, \dots, T$, where $\pi_{it} = \alpha_i + u_{it}$. The joint moment conditions, consisting of (14) and (16), can be compactly written in matrix form as

$$\mathbb{E}[\tilde{\boldsymbol{Z}}_i' \pi_i^*] = 0 \tag{17}$$

where $\tilde{\boldsymbol{Z}}_i = \text{diag}(z_i, \Delta y_{i2}, \Delta y_{i3}, \dots, \Delta y_{i,T-1})$, $\pi_i^* = (\Delta u_i, \pi_i)'$, where Δu_i is as defined previously and $\pi_i = (\pi_{i3}, \dots, \pi_{iT})'$, for $i = 3, \dots, N$. Using these moment conditions, the Blundell-Bond estimator $\hat{\phi}_{BB}$ of ϕ is the solution to the optimization problem

$$\hat{\phi}_{BB} = \arg\inf_{\phi \in \Phi} \left(\frac{1}{N} \sum_{i=1}^{N} \pi_i^{*\prime} \tilde{\boldsymbol{Z}}_i \right) \boldsymbol{W}_N \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i' \pi_i^* \right) \tag{18}$$

An asymptotically optimal weighting matrix can be estimated by

$$\boldsymbol{W}_N = \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i' \widehat{\Delta u_i} \widehat{\Delta u_i}' \tilde{\boldsymbol{Z}}_i \right)^{-1} \tag{19}$$

where $\widehat{\Delta u_i}$ are the first-differenced residuals obtained from the one-step estimator. If the u_{it}'s are homoscedastic, the efficient GMM estimator can be obtained in one-step, using instead the weighting matrix

$$\boldsymbol{W}_{1N} = \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i' \boldsymbol{H} \tilde{\boldsymbol{Z}}_i \right)^{-1} \tag{20}$$

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

$$\boldsymbol{H} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix} \tag{21}$$

There are other ways of improving the finite-sample properties of the GMM estimator of ϕ, in addition to just adding more moment conditions. As described in Section 2.1, instead of fixing the weighting matrix in each stage of the estimation, the CUE alters the weighting matrix as the value of $\hat{\phi}$ is changed during the minimization process. Formally, the minimization problem with Blundell-Bond moment conditions can now be written

$$\hat{\phi}_{CUE} = \arg \inf_{\phi \in \Phi} \left(\frac{1}{N} \sum_{i=1}^{N} \pi_i^{*\prime} \tilde{\boldsymbol{Z}}_i \right) \boldsymbol{W}_N(\phi) \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{\boldsymbol{Z}}_i^{\prime} \pi_i^{*} \right) \tag{22}$$

Hence, the weighting matrix is now a function of ϕ. However, since the weighting matrix is updated in each iteration, there cannot exist a closed-form solution to the minimization problem in (22). The following theorem establishes the asymptotic properties of two-step Arellano-Bond and Blundell-Bond estimators of ϕ_0.

Theorem 1. *As N and $T \to \infty$, the following asymptotic results regarding the two-step Arellano-Bond and Blundell-Bond estimators of ϕ_0 hold.*

(i) $\hat{\phi}_{AB} \xrightarrow{P} \phi_0$ *if (log T^2)/$N \to 0$*

(ii) $\hat{\phi}_{BB} \xrightarrow{P} \phi_0$ *if (log T^2)/$N \to 0$*

(iii) $\sqrt{NT}\left[\hat{\phi}_{AB} - \phi_0 + \frac{(1+\phi_0)}{N}\right] \xrightarrow{L} \mathcal{N}\left(0, 1-\phi_0^2\right)$ *if (log T^2)/$N \to 0$ and $T/N \to c\,(0 \leq c < \infty)$*

(iv) $\sqrt{NT}\left[\hat{\phi}_{BB} - \phi_0 + \frac{1}{N}\frac{(1+\phi_0)(r+1+\phi_0)}{2(r+1)}\right] \xrightarrow{L} \mathcal{N}\left(0, 1-\phi_0^2\right)$ *if (log T^2)/$N \to 0$ and $T/N \to c\,(0 \leq c < \infty)$*

Proof. See Álvarez and Arellano (2003) for (a) and (c), and Hayakawa (2015) for (b) and (d). ■

Although it is clear from the discussion in Section 2.1 that $\hat{\phi}_{CUE} \xrightarrow{P} \phi_0$, the closed-form expression for asymptotic bias of the CUE is unknown to the literature.

As with any estimator, efficiency is an important property. Hahn and Kuersteiner (2002) show that for the more general case when y_{it} in (10) is a vector, the error terms u_{it} are i.i.d., and $\hat{\phi}$ is an arbitrary estimator of $\boldsymbol{\phi}_0$, the Cramér-Rao lower bound on the variance of vec$(\hat{\phi} - \phi_0)$ is $\boldsymbol{\Omega} \otimes \boldsymbol{\Psi}^{-1}$, where $\boldsymbol{\Omega} = \mathbb{E}(s_{it}s_{it}^{\prime})^{-1}$ and $s_{it} = \partial \log f_{it}(y_{it}|x_{it}; \theta, \alpha_i(\omega)/\partial\theta$, and $\boldsymbol{\Psi} = \boldsymbol{\Omega} + \boldsymbol{\phi}_0\boldsymbol{\Omega}\boldsymbol{\phi}_0^{\prime} + \boldsymbol{\phi}_0^2\boldsymbol{\Omega}(\boldsymbol{\phi}_0^{\prime})^2 + \ldots$.

Under the assumption that the elements of the sequence $\{u_{it}\}$ are multivariate Gaussian, it follows that the limiting distribution of $\sqrt{NT}\text{vec}(\hat{\phi} - \phi_0)$ cannot be more concentrated than $\mathcal{N}(0, \mathbf{\Omega} \otimes \mathbf{\Psi}^{-1})$, which for the univariate case corresponds to $\mathcal{N}(0, 1 - \phi_0^2)$. Hence, for an unbiased estimator, $1 - \phi_0^2$ is the lowest possible asymptotic variance.

From parts (*iii*) and (*iv*) of Theorem 1, it is clear that both the Arellano-Bond and Blundell-Bond estimators are biased. The source of the GMM bias in the dynamic panel case is the fact that the GMM is an instrumental variable (IV) estimator, in which instruments are used to eliminate correlations between the regressors and the errors. Since this technique involves taking cross-sectional averages, the resulting correlations between sample moments are $O(N^{-1})$. If T is fixed, there is no asymptotic bias, because the scale factor is proportional to $\sqrt{N}$. However, if T and N increase simultaneously, so that $N/T \to c \neq 0$ (as in Theorem 1), the scale factor becomes to $\sqrt{NT}$, which is asymptotically equivalent to $\sqrt{c}N$. This leads to an asymptotic bias of order $\sqrt{c}$.

Considering that the difference and system GMM are asymptotically biased, there have been a number of studies on the finite-sample performance of these estimators. Research by Blundell et al. (1998), and Hayakawa and Pesaran (2015) show that the two-step system GMM is to be preferred vis-à-vis the two-step difference estimator, as the finite-sample bias tends to be lower. However, whereas the finite-sample bias of the system GMM is lower, the size tends to be higher than the difference GMM, which makes it unsuitable for testing purposes. Hayakawa and Pesaran (2015) find that the continuously updated BB GMM has almost zero bias even for low values of N and T. However, this leads to severe oversizing, a problem that increases with T and decreases with N. Mehic (2017) considers highly persistent panels, and finds that the oversizing problem for the CUE is exacerbated as ϕ increases towards unity, although the power is also increasing with ϕ.

It is well-known that using many instruments leads to size distortions, and it is a feature present in many types of IV models (Staiger and Stock 1997; Bun and Windmeijer 2010). In the dynamic panel setting, this translates into severe oversizing in the system GMM, which utilizes more moment conditions. Using an adjusted covariance matrix developed by Newey and Windmeijer (2009) mitigates the size distorsion slightly, although the problem is still pronounced when N is low (cf. Hayakawa and Pesaran 2012). A further note on the weak instrument problem in the dynamic panel case and how it can be remedied is given in Roodman (2009). Recent research has established that using only one lag of the dependent variable as instrument eliminates the asymptotic bias for the difference estimator (Hsiao and Zhou 2017). Nevertheless, there is a considerable knowledge gap on the weak instruments problem for dynamic panels, especially for the system GMM. Thus, the Monte Carlo approach of this thesis utilizes all available instruments.

2.4 The SPJ

This section introduces the SPJ in detail. To the best of this author's knowledge, this paper is the first to consider the SPJ for models with $O(N^{-1})$ bias. Let $\hat{\theta}$ be an estimate of an unknown parameter θ_0. To be consistent with the existing literature on the subject, let $\hat{\theta} = \theta_N(p)$ be the estimator associated with the entire sample N, and p be an $N \times 1$ vector with mean μ such that $p - \mu = O_p(N^{-1/2})$. The Taylor approximation of order k associated with $\theta_N(p)$ can be written

$$\theta_{N,k} = \sum_{s=0}^{N} \frac{1}{s!} \left\{ \left[(p-\mu)' \frac{\partial}{\partial p} \right]^s \theta_N(p) \right\}_{p=\mu} \tag{23}$$

Then, by Theorems A1 and A2 of Sargan (1976), it holds that

$$\mathbb{E}\left(|\theta_N(p)^j|\right) = E\left(|\theta_{N,k}(p)^j|\right) + O(N^{-\gamma k}) \tag{24}$$

for $\gamma > 0$ and suitably large k. Hence, we can approximate j:th moment of $\theta_N(p)$ by the j:th moment of $\theta_{N,k}(p)$, a property used in the jackknife resampling technique.

The jackknife was introduced in statistics by Quenouille (1949). Originally used to approximate serial correlations, its current use is primarily for bias-reduction. Split the N observations into m samples of length l, so that $N = m \times l$. In its most basic form (cf. Quenouille 1956), the jackknife estimator $\hat{\theta}_{Jack}$ of θ_0 is given by

$$\hat{\theta}_{Jack} = \frac{m}{m-1}\hat{\theta} - \frac{1}{m^2 - m}\sum_{i=1}^{m} \hat{\theta}_{li} \tag{25}$$

where $\hat{\theta}_{li}$ is the estimate of θ_0 formed by using the i:th subsample. If for the vector p the expansion

$$\mathbb{E}(p) = \mu + \frac{q_1}{N} + O(N^{-2}) \tag{26}$$

holds for some constant q_1, a similar expansion yields

$$\mathbb{E}(\hat{\theta}) = \theta_0 + \frac{r_1}{N} + O(N^{-2}) \tag{27}$$

and

$$\mathbb{E}(\hat{\theta}_{li}) = \theta_0 + \frac{r_1}{N/m} + O\left((N/m)^{-2}\right) \tag{28}$$

for some constant r_1. Taking expectations in (25) and substituting (27) and (28) into the resulting equation, yields

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_{Jack}) &= \mathbb{E}\left[\frac{m}{m-1}\hat{\theta} - \frac{1}{m^2-m}\sum_{i=1}^{m}\hat{\theta}_{li}\right] \\
&= \frac{m}{m-1}\left(\theta_0 + \frac{r_1}{N}\right) - \frac{1}{m^2-m}\sum_{i=1}^{m}\left(\theta_0 + \frac{r_1}{N/m}\right) + O(N^{-2}) \\
&= \theta_0 + O(N^{-2})
\end{aligned} \tag{29}$$

Thus, if the bias is $O(N^{-1})$, the jackknife reduces the bias to $O(N^{-2})$. Note also that the bias reduction is invariant to the choice of m.

The simplest jackknife approach, the HPJ, sets $m = 2$. This means splitting the sample vector into two parts. Consider again the panel case where we want to estimate the unknown parameter ϕ_0, where ϕ_0 could be the AR(1) autoregressive coefficient described by (12). Slightly abusing notation, we may partition the panel $\{1, \ldots, N\}$ into two half-panels, $S_1 = \{1, \ldots, N/2\}$ and $S_2 = \{N/2+1, \ldots, N\}$, given that N is even[3]. Let $\bar{\phi}_{1/2} \equiv \frac{1}{2}(\hat{\phi}_{S1} + \hat{\phi}_{S2})$. Then, the HPJ estimator of ϕ_0 can be written

$$\hat{\phi}^{HPJ} \equiv 2\hat{\phi}_{GMM} - \bar{\phi}_{1/2} \tag{30}$$

Since the bias reduction is invariant with respect to the choice of the subsample length m, we can deduce that the HPJ reduces the asymptotic bias from $O(N^{-1})$ to $O(N^{-2})$.

This approach is an extension of the SPJ for $O(T^{-1})$ bias MLE described by Dhaene and Jochmans (2015) to the case when the bias is $O(N^{-1})$. However, whereas it is straightforward to split the panel into two parts when the bias is $O(T^{-1})$, as the time series observations are ordered sequentially, there is no natural ordering when the panel is split with respect to the cross-sectional units.

[3] For the sake of simplicity, I will not discuss the case when N is odd.

Hence, whereas it in the original case only exists one possible way to split the panel, there are $C = \binom{N}{N/2} = \frac{N!}{[(N/2)!]^2}$ combinations in the case considered in this thesis. For computational ease, Fernández-Val and Weidner (2016) suggest to instead average over J randomly chosen cross sectional partitions, where $0 < J << C$, in order to obtain an estimate of $\bar{\phi}_{1/2}$ [4].

2.5 Calculation of standard errors

To calculate the standard errors associated with the jackknifed estimates $\hat{\phi}_{1/2}$, note that

$$\mathrm{Var}(\hat{\phi}_{1/2}) = 4\,\mathrm{Var}(\hat{\theta}_{GMM}) + \mathrm{Var}(\bar{\phi}_{1/2}) - 4\,\mathrm{Cov}(\hat{\theta}_{GMM}, \bar{\phi}_{1/2}). \tag{31}$$

Whereas the first two terms on the right-hand side of this equation are known, $\mathrm{Cov}(\hat{\theta}_{GMM}, \bar{\phi}_{1/2})$ is unknown. This problem can remedied by using the bootstrap approach of Kapetanios (2008), which is the method applied by Dhaene and Jochmans (2015). Again suppressing the " T" subscript for notational convenience, let $z \equiv (z_1, \ldots, z_N)$ denote the original panel and $\dot{z} \equiv (z_{d1}, \ldots, z_{dN})$ denote the bootstrap panel, where $\{d_1, \ldots, d_N\}$ are i.i.d. draws with replacement from $\{1, \ldots, N\}$. Then, the variance of the bootstrap distribution can be shown to be a consistent estimate of $\mathrm{Var}(\hat{\phi}_{1/2})$, and its associated coverage rates to be asymptotically equal to the desired significance level α.

2.6 Jackknife interpretation of the CUE

One final note is that by construction of the CUE, it can be seen as a jackknife estimator in itself. This was shown by to Donald and Newey (2000). To see that the CUE can be written as a jackknife estimator, we consider again the general expression for the CUE described by (5), namely that the estimator is the solution to

$$\arg\inf_{\theta \in \Theta} \bar{g}_N(\theta)' \boldsymbol{W}_N(\theta)^{+} \bar{g}_N(\theta) \tag{32}$$

The FOC associated with the above optimization problem can be shown to be

$$\tilde{\boldsymbol{\Gamma}}' \boldsymbol{W}_N^{-1} \hat{g} - \hat{g}' \boldsymbol{W}_N^{-1} \boldsymbol{\Lambda}_N \boldsymbol{W}_N^{-1} \hat{g} = 0 \tag{33}$$

[4]Hence, we obtain J estimates of $\bar{\phi}_{1/2}$. The average of these J estimates is then used in (30).

where $\tilde{\mathbf{\Gamma}} = \partial g_N(\hat{\theta})/\partial\theta$, $\mathbf{\Lambda}_N = \mathbf{\Lambda}_N(\hat{\theta})$, $\boldsymbol{W}_N = \boldsymbol{W}_N(\hat{\theta})$, and $\hat{g} = g_N(\hat{\theta})$. Now, let $\tilde{\mathbf{\Gamma}}_i$ be the i:th element of $\tilde{\mathbf{\Gamma}}$ and $\hat{\boldsymbol{B}} = \boldsymbol{W}_N^{-1}\mathbf{\Lambda}_N$, where $\mathbf{\Lambda}_N = N^{-1}\sum_{i=1}^N \left[g_i(\hat{\theta})g_i'(\hat{\theta})\right]$, and $\hat{\boldsymbol{U}}_i = \tilde{\mathbf{\Gamma}}_i - \hat{\boldsymbol{B}}\hat{g}_i$, where $\hat{g}_i$ is the i:th element of $\hat{g}$. By using the property $\sum_{i=1}^N \hat{\boldsymbol{U}}_i\hat{g}_i'/N = 0$, we may write (33) as

$$
\begin{aligned}
0 &= (\tilde{\mathbf{\Gamma}} - \hat{\boldsymbol{B}}'\hat{g})'\boldsymbol{W}_N^{-1}\hat{g} \\
&= \left(\frac{1}{N}\sum_{i=1}^N \hat{\boldsymbol{U}}_i\right)' \boldsymbol{W}_N^{-1}\hat{g} \\
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \hat{\boldsymbol{U}}_j'\boldsymbol{W}_N^{-1}\hat{g}_i \\
&= \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}^N \hat{\boldsymbol{U}}_j'\boldsymbol{W}_N^{-1}\hat{g}_i \\
&= \frac{1}{N}\sum_{i=1}^N \left(\frac{1}{N}\sum_{j\neq i}^N \hat{\boldsymbol{U}}_j\boldsymbol{W}_N^{-1}\right)\hat{g}_i \qquad (34)
\end{aligned}
$$

Equation (34) shows that the CUE is indeed a jackknife estimator of the 'delete one observation' type.

3 Monte Carlo setup

This section will briefly describe the Monte Carlo (MC) procedure. The model of interest is

$$y_{it} = (1-\phi_0)\alpha_i + \phi_0 y_{i,t-1} + u_{it} \qquad (35)$$

where $u_{it} \sim \mathcal{N}(0, \sigma_i^2)$. I consider $\sigma_i^2 \sim \mathcal{U}(0.5, 1.5)$, so that $\mathbb{E}[\sigma_i^2] = 1$ and $\mathbb{V}[\sigma_i^2] = 1/12$, as well as $\sigma_i^2 \sim \mathcal{U}(1,3)$, so that $\mathbb{E}[\sigma_i^2] = 2$ and $\mathbb{V}[\sigma_i^2] = 1/3$. 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 $\mathbb{V}[\alpha_i] = 1$. For the initial observations y_{i0}, I set $y_{i0} = \alpha_i/(1-\phi_0) + u_{i0}$, where $u_{i0} \sim \mathcal{N}[0, 1/(1-\phi_0^2)]$. The autoregressive parameter is varied according to $\phi_0 \in \{0.50, 0.80\}$. The number of MC replications is set to 500, while the number of cross-sectional partitions, J, is set to 50. When calculating the standard errors for the jackknifed estimators, the number of bootstrap replications is set to 25.

The parameterization in (35), where the individual effects are multiplied by $(1-\phi_0)$, is a standard approach in the literature when dealing with almost non-stationary data (cf. Han and Phillips 2013; Bun et al. 2017). Without this correction, the individual effects would have too much of an impact on the MC results when the true value ϕ_0 is close to unity, e.g. when $\phi_0 = 0.80$.

4 Results

N	T	ϕ_0	$\hat{\phi}_{AB}$	$\hat{\phi}_{AB}^{HPJ}$	$\hat{\phi}_{BB}$	$\hat{\phi}_{BB}^{HPJ}$	$\hat{\phi}_{CUE}$	$\hat{\phi}_{CUE}^{HPJ}$
				Parameter estimates				
20	5	0.50	0.3526	0.4651	0.4727	0.5012	0.5029	0.5451
20	5	0.80	0.6414	0.7675	0.7469	0.7836	0.8805	1.0535
50	5	0.50	0.4557	0.5243	0.5045	0.5237	0.5169	0.5374
50	5	0.80	0.7495	0.8170	0.7860	0.8225	0.8168	0.8649
100	5	0.50	0.4680	0.5074	0.5010	0.5139	0.5083	0.5147
100	5	0.80	0.7664	0.8004	0.7938	0.8181	0.8106	0.8346
100	10	0.50	0.4758	0.5033	0.4994	0.5147	0.5024	0.5124
100	10	0.80	0.7755	0.8067	0.7963	0.8264	0.8102	0.8378
				Bias				
20	5	0.50	-0.1474	-0.0349	-0.0273	0.0012	0.0029	0.0451
20	5	0.80	-0.1586	-0.0325	-0.0531	-0.0164	0.0805	0.2535
50	5	0.50	-0.0443	0.0243	0.0045	0.0237	0.0169	0.0374
50	5	0.80	-0.0505	0.0170	-0.0140	0.0225	0.0168	0.0649
100	5	0.50	-0.0320	0.0074	0.0010	0.0139	0.0083	0.0147
100	5	0.80	-0.0336	0.0004	-0.0062	0.0181	0.0106	0.0346
100	10	0.50	-0.0242	0.0033	-0.0004	0.0147	0.0024	0.0124
100	10	0.80	-0.0245	0.0067	-0.0037	0.0264	0.0102	0.0378
				Standard errors				
20	5	0.50	0.1759	0.1184	0.0861	0.0438	0.0886	0.0937
20	5	0.80	0.1856	0.1240	0.0824	0.0383	0.0928	0.0877
50	5	0.50	0.1460	0.1254	0.0904	0.0690	0.0897	0.0691
50	5	0.80	0.1434	0.1270	0.0928	0.0674	0.0948	0.0702
100	5	0.50	0.1134	0.1047	0.0741	0.0647	0.0740	0.0643
100	5	0.80	0.1083	0.1022	0.0822	0.0671	0.0822	0.0678
100	10	0.50	0.0362	0.0243	0.0245	0.0140	0.0244	0.0140
100	10	0.80	0.0323	0.0220	0.0234	0.0129	0.0230	0.0130

Table 1: Parameter and standard errors estimates with $\sigma_i^2 \sim \mathcal{U}(0.5, 1.5)$.

Tables 1 and 2 below present the results. Table 1 corresponds to the case with unit variance, whereas the results in Table 2 correspond to $\sigma_i^2 \sim \mathcal{U}(1, 3)$. Consider first bias correction using the two-step AB estimator, the results of which are given in the columns $\hat{\phi}_{AB}$ and $\hat{\phi}_{AB}^{HPJ}$, respectively, in Tables 1 and 2. Without correction, the AB estimator is severely biased. However, it is clear that,

N	T	ϕ_0	$\hat{\phi}_{AB}$	$\hat{\phi}_{AB}^{HPJ}$	$\hat{\phi}_{BB}$	$\hat{\phi}_{BB}^{HPJ}$	$\hat{\phi}_{CUE}$	$\hat{\phi}_{CUE}^{HPJ}$
Parameter estimates								
20	5	0.50	0.3646	0.4768	0.4613	0.5013	0.4846	0.5470
20	5	0.80	0.5717	0.7171	0.7247	0.7671	0.9429	1.1832
50	5	0.50	0.4599	0.5251	0.4974	0.5216	0.5113	0.5387
50	5	0.80	0.7202	0.8120	0.7826	0.8294	0.8153	0.8777
100	5	0.50	0.4711	0.5083	0.4985	0.5156	0.5083	0.5209
100	5	0.80	0.7491	0.8001	0.7902	0.8189	0.8089	0.8394
100	10	0.50	0.4782	0.5035	0.4985	0.5169	0.5017	0.5162
100	10	0.80	0.7665	0.8041	0.7926	0.8254	0.8024	0.8292
Bias								
20	5	0.50	-0.1354	-0.0232	-0.0387	-0.0387	-0.0154	0.0470
20	5	0.80	-0.2283	-0.0829	-0.0753	-0.0329	0.1429	0.3832
50	5	0.50	-0.0401	0.0251	-0.0026	0.0216	0.0113	0.0387
50	5	0.80	-0.0798	0.0120	-0.0174	0.0294	0.0153	0.0777
100	5	0.50	-0.0289	0.0083	-0.0015	0.0156	0.0083	0.0209
100	5	0.80	-0.0509	0.0001	-0.0098	0.0189	0.0089	0.0394
100	10	0.50	-0.0218	0.0035	-0.0015	0.0169	0.0017	0.0162
100	10	0.80	-0.0335	0.0041	-0.0074	0.0254	0.0024	0.0292
Standard errors								
20	5	0.50	0.1768	0.1198	0.0863	0.0452	0.0874	0.1378
20	5	0.80	0.2209	0.1424	0.0872	0.0448	0.0945	0.0844
50	5	0.50	0.1423	0.1328	0.0876	0.0655	0.0870	0.0674
50	5	0.80	0.1779	0.1695	0.0944	0.0654	0.0940	0.0676
100	5	0.50	0.1103	0.1035	0.0707	0.0594	0.0706	0.0596
100	5	0.80	0.1357	0.1231	0.0785	0.0630	0.0776	0.0626
100	10	0.50	0.0348	0.0234	0.0245	0.0140	0.0243	0.0139
100	10	0.80	0.0372	0.0249	0.0236	0.0133	0.0230	0.0127

Table 2: Parameter and standard errors estimates with $\sigma_i^2 \sim \mathcal{U}(1, 3)$.

regardless of the value of σ_i^2, the HPJ provides significant bias reduction. When N is large, the corrected estimates are very close to the true values. However, for $N = 20$, the bias is still relatively large in absolute terms.

Turning to the two-step BB estimator, we see that the nonadjusted version, although almost always underestimating the true parameter value, is considerably closer to the true value compared to the two-step AB. For $N \geq 50$, we see that, both for $\sigma_i^2 \sim \mathcal{U}(0.5, 1.5)$ and $\sigma_i^2 \sim \mathcal{U}(1, 3)$, the absolute bias is lower with the non-adjusted version. Hence, for the two-step BB, the HPJ only performs under very special circumstances - namely the case when the sample size is very low.

For the CUE, the precision of the non-adjusted estimator $\hat{\phi}_{CUE}$ is, with a few exceptions, excellent. Since using HPJ increases the estimate of ϕ_0, the HPJ CUE estimator actually increases bias. Another interesting point is that, except for in a minority of cases considered, there is no improvement in terms of bias when using the unadjusted CUE as opposed to the unadjusted two-step estimator. Since the CUE is a jackknife-type estimator in itself (as demonstrated in Section 2.6), it appears that the "extra" jackknifing deteriorates the performance of the estimator by increasing the estimate of ϕ_0.

Two conclusions hold regardless of which of the three estimators is considered. First, neither the unadjusted nor the HPJ estimates improve as T increases. This is clear as T is increased from 5 to 10 with $N = 100$, and is consistent with the discussion in Section 2.4. Second, the general results, namely that the HPJ decreases bias only for the difference estimator, and in very special cases also the the system two-step estimator, hold regardless of the value of ϕ_0 and σ_i^2.

Consequently, the half-panel jackknife approach gives in satisfactory finite-sample bias reduction only when using the Arellano-Bond two-step GMM, and for very low combinations of N and T with the Blundell-Bond two-step estimator. In the other cases, as well as in all cases when using the CU Blundell-Bond GMM, the HPJ approach is not effective in reducing finite-sample bias.

However, using HPJ reduces the standard errors compared to the non-adjusted estimators, particularly for the two-step estimators. It should be noted that the reduction in standard errors was even greater than the one seen in the JIVE estimator by Zhang and Zhou (2018). In sum, the reduction in both bias and standard errors makes the HPJ very attractive for the AB two-step estimator and for the BB two-step in the small-sample case. However, for the BB two-step with moderate N, as well as for all cases with the CUE, the performance of the HPJ is not particularly impressive, despite the reduction in standard errors. Also, regardless of which estimator is being corrected, the advantages with the HPJ must be weighed against the additional computational effort required to construct this estimator.

5 Concluding remarks

The purpose of this thesis has been to study the performance of the half-panel jackknife (HPJ) in reducing the $O(N^{-1})$ asymptotic bias associated with GMM estimators with fixed effects, focusing on the AR(1) dynamic panel model without exogenous covariates. Considering the additional computational effort associated with constructing the HPJ, it is of importance that the bias reduction is significant.

The results show that for the Arellano-Bond two-step estimator, the bias reduction is considerable. Together with a large reduction in the standard errors, the HPJ is a potentially valuable alternative to other bias-reducing methods proposed in the literature, such as the JIVE and the RMA. For the two-step Blundell-Bond estimator, the bias reduction is negligible except for very low values of N, although the results showed a considerable reduction in standard errors for all combinations of N and T. However, for the CUE estimator, the HPJ approach actually increases the finite-sample bias, and the associated standard errors are generally close to those of the non-corrected estimator.

As with any study, there are a number of limitations. By construction of the jackknifed estimators, there is a certain degree of randomness which is likely to be effected by the number of cross-sectional partitions. Furthermore, it is not clear, at least not to this author, why the jackknife exclusively leads to the estimates of the autoregressive parameter increasing vis-à-vis the non-jackknifed estimates.

Two questions could be of interest for future research. Firstly, it could be valuable to examine the effect on bias of the numbers of instruments used. As mentioned in Section 2.3, research indicates that at least for the difference GMM, there is some finite-sample bias reduction when only one lag of the dependent variable is used as instrument. Additionally, it would be useful to extend the HPJ GMM to include exogenous covariates, which has hitherto only been done for the MLE.

References

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

Anatolyev, S. (2005). GMM, GEL, Serial Correlation, and Asymptotic Bias. *Econometrica* 73 (3), 983-1002.

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

Andrews, D.W.K. (1994). Empirical process methods in econometrics. In: Engle, R.F. and McFadden, D.L. (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: North Holland.

Andrews, D.W.K. (2002). Generalized Method of Moments Estimation When a Parameter Is on a Boundary. *Journal of Business and Economic Statistics* 20 (4), 530-544.

Angrist, J.D., G.W. Imbens and A.B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14 (1), 57-67.

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

Arellano, M. and S. Bonhomme (2009). Robust Priors in Nonlinear Panel Data Models. *Econometrica* 77 (2), 489-536.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68 (1), 29-52.

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

Blundell, R., S. Bond and F. Windmeijer (1998). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator, in Badi H. Baltagi, Thomas B. Fomby, R. Carter Hill (eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels* (Advances in Econometrics, 15). Bingley: Emerald Group Publishing, 53-91.

Bun, M.J.G., M.A Carree and A. Juodis (2017). On Maximum Likelihood Estimation of Dynamic Panel Data Models. *Oxford Bulletin of Economics and Statistics* 79 (4), 463-494.

Bun, M.J.G. and F. Windmeijer (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal* 13 (1), 95-126.

Chen, X., O. Linton and I. Van Keilegom (2003). Estimation of Semiparametric Models when the Criterion Function Is Not Smooth. *Econometrica* 71 (5), 1591-1608.

Choi, C.-Y., N.C. Mark and D. Sul (2010). Bias Reduction in Dynamic Panel Data Models by Common Recursive Mean Adjustment. *Oxford Bulletin of Economics and Statistics* 72 (5), 567-599.

Chudik, A., M.H. Pesaran and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics*, forthcoming.

Dhaene, G. and K. Jochmans (2015). Split-panel Jackknife Estimation of Fixed-effect Models. *Review of Economic Studies* 82 (3), 991-1030.

Donald, S.G. and W.K. Newey (2000). A jackknife interpretation of the continuous updating estimator. *Economics Letters* 67 (3), 239-243.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics* 192 (1), 291-312.

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

Hahn, J. and G. Kuersteiner (2004). Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. *Econometric Theory* 27 (6), 1152-1191.

Hahn, J. and W.K. Newey (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* 72 (4), 1295-1319.

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

Hansen, L.P. (1982). Large Sample Properties of Generalized Methods of

Moments Estimators. *Econometrica* 50 (4), 1029-1054.

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

Hayakawa, K. (2015). The Asymptotic Properties of the System GMM Estimator in Dynamic Panel Data Models When Both N and T are Large. *Econometric Theory* 31 (3), 647-667.

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

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

Hsiao, C. and Q. Zhou (2017). First difference or forward demeaning: Implications for the method of moments estimators. *Econometric Reviews* 36 (6-9), 883-897.

Imbens, G.W. (1997). One-Step Estimators for Over-Identified Generalized Method of Moments Models. *Review of Economic Studies* 64 (3), 359-383.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal* 11 (2), 377-395.

Koltchinskii, V.I. (1981). On the central limit theorem for empirical measures. *Theory of Probability and Mathematical Statistics* 24, 71–82.

Lancaster, T. (2002). Orthogonal Parameters and Panel Data. *Review of Economic Studies* 69 (3), 647-666.

Lee, N., H.R. Moon and Q. Zhou (2017). Many IVs estimation of dynamic panel regression models with measurement error. *Journal of Econometrics* 200 (2), 251-259.

Mehic, A. (2017). FDML versus GMM for dynamic panel models with roots near unity. *Working paper, Lund University.*

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

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

NEYMAN, J. AND E.L. SCOTT (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16 (1), 1-32.

ROODMAN, D. (2009). A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* 71 (1), 135-158.

OSSIANDER, M. (1987). A Central Limit Theorem Under Metric Entropy with L_2 Bracketing. *Annals of Probability* 15 (3), 897-919.

PAKES, A. AND D. POLLARD (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica* 57 (5), 1027-1057.

PESARAN, M.H. (2015). *Time Series and Panel Data Econometrics.* Oxford: Oxford University Press.

POLLARD, D. (1982). A central limit theorem for empirical processes. *Journal of the Australian Mathematical Society* 33 (2), 235-248.

QIN, J. AND J. LAWLESS (1994). Empirical Likelihood and General Estimating Equations. *Annals of Statistics* 22 (1), 300-325.

QUENOUILLE, M.H. (1949). Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society, Series B* 11 (1), 68-84.

QUENOUILLE, M.H. (1956). Notes on Bias in Estimation. *Biometrika* 43 (3/4), 353-360.

SARGAN, J.D. (1976). Econometric Estimators and the Edgeworth Approximation. *Econometrica* 44 (3), 421-448.

SMITH, R.J. (1997). Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation. *Economic Journal* 107 (441), 503-519.

STAIGER, D. AND J.H. STOCK (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica* 65 (3), 557-586.

ZHANG, Y. AND Q. ZHOU (2018). Bias correction for Arellano-Bond type GMM estimation of dynamic panel models. *Working paper, Louisiana State University.*

Appendix: Consistency and asymptotic normality of the GMM

The large sample theory in this appendix is based on Pakes and Pollard (1989). A similar outline for the semiparametric case is given in Chen et al. (2003). Since the results in this section are not exclusively valid for GMM estimators of panel models, consider the more general case where the data $z = (z_1, \dots, z_N)'$ are an i.i.d. sample from $\mathbb{R}^s$, and are subject to the same regularity conditions described previously for the panel data $\{z_{it}\}$.

Theorem 2. *Under the following conditions,*

(i) $||g_N(\hat{\theta}_{GMM})|| \leq o_p(1) + \inf_{\theta \in \Theta} ||g_N(\hat{\theta})||$

(ii) $\inf_{||\theta - \theta_0 > \delta||} ||g(\theta)|| = o_p(1), \quad \forall \delta > 0$

(iii) $\sup_{\theta \in \Theta} \frac{||g_N(\theta) - g(\theta)||}{1 + ||g_N(\theta)|| + ||g(\theta)||} = o_p(1)$

$\hat{\theta}_{GMM} \xrightarrow{P} \theta_0$*, where* θ_0 *is the unique* $\theta_0 \in \Theta$ *for which* $g(\theta_0) = 0$.

Proof. Omitted. ■

Let $\{\delta_N\} = o_p(1)$ be a sequence of positive numbers. Theorem 2 below gives the conditions for asymptotic normality.

Theorem 3. *Let* $\hat{\theta}_{GMM} \xrightarrow{P} \theta_0$*, where* θ_0 *is the unique* $\theta_0 \in \Theta$ *for which* $g(\theta_0) = 0$*. Then, if the following conditions are satisfied,*

(i) $\theta_0 \in int(\Theta)$

(ii) $||g_N(\hat{\theta}_{GMM})|| \leq o_p(N^{-1/2}) + \inf_{\theta \in \Theta} ||g_N(\hat{\theta})||$

(iii) $\sup_{||\theta - \theta_N|| < \delta_N} \frac{||g_N(\theta) - g(\theta) - g_N(\theta_0)||}{N^{-1/2} + ||g_N(\theta)|| + ||g(\theta)||} = o_p(1)$

(iv) $\sqrt{N} g_N(\theta_0) \xrightarrow{L} \mathcal{N}(0, \mathbf{\Lambda})$.

(v) $g()$ *is differentiable at* $\theta_0 \in \Theta$*, and the matrix* $\mathbf{\Gamma}$ *exists for* $\theta \in \Theta$*, is of full column rank, and is continuous at* $\theta = \theta_0$.

$\sqrt{N}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{L} \mathcal{N}(0, \mathbf{\Phi})$*, where* $\mathbf{\Phi} = (\mathbf{\Gamma}'\boldsymbol{W}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}'\boldsymbol{W}\mathbf{\Lambda}\boldsymbol{W}\mathbf{\Gamma}(\mathbf{\Gamma}'\boldsymbol{W}\mathbf{\Gamma})$.

Proof. Omitted. ■

Remark 1 (Identification assumption)**.** *Note that for consistency, it is not required that $\theta_0 \in int(\Theta)$, i.e. θ_0 can be anywhere in Θ. However, for asymptotic normality, we impose this condition. For a discussion on estimation when the true parameter is on a boundary of the parameter space, see Andrews (2002).*

Remark 2 (Optimal weighting matrix)**.** *For the two-step GMM, the choice $\boldsymbol{W} = \boldsymbol{\Lambda}^{-1}$ minimizes the asymptotic variance. It is straightforward to show that, for this particular case, $\boldsymbol{\Phi}$ reduces to $(\boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma})^{-1}$. For a proof that this is actually the smallest possible asymptotic variance, see e.g. Pesaran (2015, p. 232).*

Note that for condition (iii) of Theorem 2 to be satisfied, it must hold that $\mathcal{G}$ is $P-$Donsker, where $\mathcal{G} = \{g(\theta) : \theta \in \Theta\}$, while condition (iii) of Theorem 1 will be satisfied if $\mathcal{G}$ is $P-$Glivenko-Cantelli. To understand the difference between these two classes, introduce the following concepts. If again $z_1, \ldots, z_N$ denotes a random sample from a distribution P on $(\mathcal{D}, d)$, the *empirical distribution* $\mathbb{P}_N$ is defined as $\mathbb{P}_N = N^{-1}\sum_{i=1}^{N}\delta_{z_i}$, where δ_{z_i} is the degenerate distribution, which has cdf $F_\delta(z_i, k_0) = 1$ if $z_i \geq k_0$, and 0 else. Given a Borel measurable function $f : \mathcal{D} \to \mathbb{R}$, let $\mathbb{P}_N f = N^{-1}\sum_{i=1}^{N} f(z_i)$ and $Pf = \int f dP$, be the expectation under the empirical measure and expectation under P. Let $f \in \mathcal{F}$, where $\mathcal{F}$ is a class of measurable functions. If $\mathcal{F}$ satisfies $\sup_{f\in\mathcal{F}}|\mathbb{P}_N f - Pf| \xrightarrow{a.s.} 0$, then $\mathcal{F}$ is said to be $P-$*Glivenko-Cantelli*. On the other hand, if $\mathbb{G}_N = \sqrt{N}(\mathbb{P}_N - P) \xrightarrow{L} \mathbb{G}$, where $\mathbb{G}$ is a tight Gaussian process, $\mathcal{F}$ is said to be $P-$*Donsker*. Hence, asymptotic normality depends crucially on the $P-$Donsker property.

In turn, whether $\mathcal{G}$ is $P-$Donsker or not is related to its covering and bracketing numbers. To see this, introduce some additional notation. First, for $r \in [1, \infty)$, let $L_r(P)$ be the space of measurable real-valued functions $f : \mathcal{D} \to \mathbb{R}$. Then, the *covering number* $N(\varepsilon, \mathcal{F}, L_r(P))$ is the minimal number of balls of radius ε in $L_r(P)$ needed to cover $\mathcal{F}$ [5]. The *bracketing number* $N_{[]}(\varepsilon, \mathcal{F}, L_r(P))$ is the minimal number of ε-brackets needed to cover $\mathcal{F}$, where an ε-bracket with respect to the metric d in $L_r(P)$ is a pair of functions $l, u \in L_r(P)$ with $l(X) \leq u(X)$ and $d(l, u) \leq \varepsilon$.

Then, if $\int_0^\infty \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}d\varepsilon < \infty$, the class of measurable functions $\mathcal{F}$ is $P-$Donsker (Ossiander 1987). It also possible to use the covering number to establish the $P-$Donsker property. Denote by F the envelope of $\mathcal{F}$ (an envelope for $\mathcal{F}$ is a function F such that $|f| \leq F$ for all $f \in \mathcal{F}$), for which it holds that

[5] Let B be a (closed) ball in $\mathbb{R}^k$. A collection $\mathcal{B}$ of closed balls in $\mathbb{R}^k$ is a *cover* of the set $A \subset \mathbb{R}^k$ if $A \subset \bigcup_{B\in\mathcal{B}} B$.

$\int F^2 < \infty$, and by $\mathcal{Q}$ the class of all finitely discrete probability measures on $(\mathcal{D}, d)$. If $\int_0^\infty \sup_{Q \in \mathcal{Q}} \sqrt{\log N(\varepsilon ||F||_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon < \infty$, then $\mathcal{F}$ is $P-$Donsker (Koltchinskii 1981; Pollard 1982). Since $\Theta \subset \mathbb{R}^k$ is closed and bounded, the covering and bracketing numbers of Θ are known; Andrews (1994) shows that the GMM estimator $\hat{\theta}_{GMM}$ of θ_0 indeed satisfies the Ossiander condition [6]. Hence, $\mathcal{G} = \{g(\theta) : \theta \in \Theta\}$ is $P-$Donsker.

[6]Theorem 3 of Chen et al. (2003) gives the same conclusion for the semiparametric case.