econometrics

MDPI

Article

Subset-Continuous-Updating GMM Estimators for Dynamic Panel Data Models

Richard A. Ashley [1] and Xiaojin Sun [2,*]

[1] Department of Economics, Virginia Tech, Blacksburg, VA 24060, USA; ashleyr@vt.edu
[2] Department of Economics and Finance, University of Texas at El Paso, El Paso, TX 79968, USA
* Correspondence: xsun3@utep.edu; Tel.: +1-915-747-7783

Academic Editors: In Choi, Ryo Okui, Marc S. Paolella and Kerry Patterson
Received: 25 May 2016; Accepted: 25 November 2016; Published: 30 November 2016

Abstract: The two-step GMM estimators of Arellano and Bond (1991) and Blundell and Bond (1998) for dynamic panel data models have been widely used in empirical work; however, neither of them performs well in small samples with weak instruments. The continuous-updating GMM estimator proposed by Hansen, Heaton, and Yaron (1996) is in principle able to reduce the small-sample bias, but it involves high-dimensional optimizations when the number of regressors is large. This paper proposes a computationally feasible variation on these standard two-step GMM estimators by applying the idea of continuous-updating to the autoregressive parameter only, given the fact that the absolute value of the autoregressive parameter is less than unity as a necessary requirement for the data-generating process to be stationary. We show that our subset-continuous-updating method does not alter the asymptotic distribution of the two-step GMM estimators, and it therefore retains consistency. Our simulation results indicate that the subset-continuous-updating GMM estimators outperform their standard two-step counterparts in finite samples in terms of the estimation accuracy on the autoregressive parameter and the size of the Sargan-Hansen test.

Keywords: dynamic panel data models; Arellano-Bond GMM estimator; Blundell-Bond GMM estimator; subset-continuous-updating GMM estimators

JEL Classification: C18; C23

1. Introduction

In recent decades, dynamic panel data models with unobserved individual-specific heterogeneity have been widely used to investigate the dynamics of economic activities. Several estimators have been suggested for estimating the model parameters. A standard estimation procedure is to first-difference the model, so as to eliminate the unobserved heterogeneity, and then base GMM estimation on the moment conditions implied where endogenous differences of the variables are instrumented by their lagged levels. This is the well known Arellano-Bond estimator, or first-difference (DIF) GMM estimator (see Arellano and Bond [1]). The DIF GMM estimator was found to be inefficient since it does not make use of all available moment conditions (see Ahn and Schmidt [2]); it also has very poor finite sample properties in dynamic panel data models with highly persistent series and large variations in the fixed effects relative to the idiosyncratic errors (see Blundell and Bond [3]) since the instruments in those cases become less informative.

To improve the performance of the DIF GMM estimator, Blundell and Bond [3] propose taking into consideration extra moment conditions from the level equation that rely on certain restrictions on the initial observations, as suggested by Arellano and Bover [4]. The resulting system (SYS) GMM estimator has been shown to perform much better than the DIF GMM estimator in terms of finite sample bias and mean squared error, as well as with regard to coefficient estimator standard errors

since the instruments used for the level equation are still informative as the autoregressive coefficient approaches unity (see Blundell and Bond [3] and Blundell, Bond, and Windmeijer [5]). As a result, the SYS GMM estimator has been widely used for estimation of production functions, demand for addictive goods, empirical growth models, etc. However, it was pointed out later on (see Hayakawa [6] and Bun and Windmeijer [7]) that the weak instruments problem still remains in the SYS GMM estimator. Since the increase in the length of the panel leads to a quadratic increase in the number of instruments, the two-step DIF and SYS GMM estimators are both biased due to many weak moment conditions; see Newey and Windmeijer [8].

The work by Hansen, Heaton, and Yaron [9] suggests that the continuous-updating GMM estimator has smaller bias than the standard two-step GMM estimator. However, it involves high-dimensional optimizations when the number of regressors is large. Given the fact that the absolute value of the autoregressive parameter must be less than unity as a necessary requirement for the data-generating process to be stationary, we propose a computationally feasible variation on the two-step DIF and SYS GMM estimators, in which the idea of continuous-updating is applied solely to the autoregressive parameter; these two new estimators are denoted "SCUDIF" and "SCUSYS" below. Following the jackknife interpretation of the continuous-updating estimator in the work of Donald and Newey [10], we show that the subset-continuous-updating method that we propose in this paper does not alter the asymptotic distribution of the two-step GMM estimators, and it hence retains consistency. It is computationally advantageous relative to the continuous-updating estimator in that it replaces a relatively high-dimensional optimization over unbounded intervals by a one-dimensional optimization limited to the stationary domain $(-1, 1)$ of the autoregressive parameter. We conduct Monte Carlo experiments and show that the proposed subset-continuous-updating versions of the DIF and SYS GMM estimators outperform their standard two-step counterparts in small samples in terms of the estimation accuracy on the autoregressive parameter and the rejection frequency of the Sargan-Hansen test.

The layout of the paper is as follows: Section 2 describes the model specification and our proposed subset-continuous-updating method; Section 3 describes the Monte Carlo experiments and presents the results; and Section 4 concludes the paper.

2. Subset-Continuous-Updating GMM Estimator

Consider a linear panel data model with one dynamic dependent variable y_{it}, additional explanatory variables $\mathbf{X}_{it} = (x_{it}^1, ..., x_{it}^K)$, unobserved individual-specific fixed effects μ_i, and idiosyncratic errors v_{it}:

$$y_{it} = \theta y_{i,t-1} + \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}, \text{ where } u_{it} = \mu_i + v_{it}, \tag{1}$$

for $i = 1, ..., N$ and $t = 2, ..., T$, where N is large and T is small. Here, θ is the autoregressive parameter and we make a familiar assumption as in the literature that it satisfies $|\theta| < 1$ to ensure the stationarity of the model; $\boldsymbol{\beta}$ is a K-dimensional column vector of remaining coefficients. As Blundell, Bond, and Windmeijer [5] argue, this model specification is sufficient to cover most cases that researchers would encounter in linear dynamic panel applications.

While our discussion applies to the general setup of dynamic panel data models in Equation (1), for expositional clarity, we consider a special case with a unique, i.e., $K = 1$, additional explanatory variable x_{it}:

$$y_{it} = \theta y_{i,t-1} + \beta x_{it} + u_{it}, \text{ where } u_{it} = \mu_i + v_{it}, \tag{2}$$

for $i = 1, ..., N$ and $t = 2, ..., T$. We also follow Blundell, Bond, and Windmeijer [5] to allow for persistence and endogeneity in the explanatory variable x_{it}:

$$x_{it} = \rho x_{i,t-1} + \tau\mu_i + \lambda v_{it} + e_{it}, \tag{3}$$

where ρ captures the persistence of x_{it}, and τ and λ determine the correlation of x_{it} with the individual effects μ_i and the idiosyncratic errors ν_{it}, respectively.

We assume, at the outset, that μ_i, ν_{it}, and e_{it} have the following properties:

$$\mathbb{E}(\mu_i) = 0,\ \mathbb{E}(\nu_{it}) = 0,\ \mathbb{E}(e_{it}) = 0 \ \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T, \tag{4}$$

$$\mathbb{E}(\nu_{it}\mu_i) = 0,\ \mathbb{E}(e_{it}\mu_i) = 0 \ \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T, \tag{5}$$

$$\mathbb{E}(\nu_{it}\nu_{is}) = 0,\ \mathbb{E}(e_{it}e_{is}) = 0 \ \text{ for } i = 1, ..., N \text{ and } \forall t \neq s, \tag{6}$$

$$\mathbb{E}(\nu_{it}e_{is}) = 0 \ \text{ for } i = 1, ..., N \text{ and } \forall t, s. \tag{7}$$

Furthermore, we impose mean-stationarity restrictions on the initial conditions:

$$x_{i1} = \frac{\tau}{1-\rho}\mu_i + \epsilon_{i1} \ \text{ for } i = 1, ..., N, \tag{8}$$

$$y_{i1} = \frac{1}{1-\theta}\left(1 + \frac{\beta\tau}{1-\rho}\right)\mu_i + \xi_{i1} \ \text{ for } i = 1, ..., N, \tag{9}$$

and

$$\mathbb{E}(\epsilon_{i1}) = \mathbb{E}(\mu_i\epsilon_{i1}) = \mathbb{E}(\xi_{i1}) = \mathbb{E}(\mu_i\xi_{i1}) = 0 \ \text{ for } i = 1, ..., N, \tag{10}$$

$$\mathbb{E}(\epsilon_{i1}\nu_{it}) = \mathbb{E}(\xi_{i1}\nu_{it}) = 0 \ \text{ for } i = 1, ..., N \text{ and } t = 2, ..., T. \tag{11}$$

Under these conditions, we consider both the DIF GMM estimator of Arellano and Bond [1] and the SYS GMM estimator of Blundell and Bond [3], which are derived from the following moment conditions:

$$\mathbb{E}[\mathbf{g}(\mathbf{w}, \theta_0, \beta_0)] = \mathbf{0}, \tag{12}$$

where $\mathbf{w}$ denotes the data, (θ_0, β_0) are true parameters, and

$$\begin{array}{ll} \mathbf{g} = \mathbf{Z}_d'\Delta\mathbf{u} & \text{for the DIF GMM estimator,} \\ \mathbf{g} = \mathbf{Z}_s'\mathbf{p} & \text{for the SYS GMM estimator.} \end{array} \tag{13}$$

In the above equations, $\mathbf{Z}_d'$ is the $m_d \times N(T-2)$ matrix $(\mathbf{Z}_{d1}', \mathbf{Z}_{d2}', ..., \mathbf{Z}_{dN}')$ and $\mathbf{Z}_s'$ is the $m_s \times 2N(T-2)$ matrix $(\mathbf{Z}_{s1}', \mathbf{Z}_{s2}', ..., \mathbf{Z}_{sN}')$, where the number of instruments for the DIF GMM estimator is $m_d = (T-2)(T-1)$ and the instrument count for the SYS GMM estimator is $m_s = m_d + 2(T-2)$ in the case of $K = 1$; $\Delta\mathbf{u}'$ is the $N(T-2)$ vector $(\Delta\mathbf{u}_1', \Delta\mathbf{u}_2', ..., \Delta\mathbf{u}_N')$ and $\mathbf{p}'$ is the $2N(T-2)$ vector $(\mathbf{p}_1', \mathbf{p}_2', ..., \mathbf{p}_N')$ with

$$\Delta\mathbf{u}_i = \begin{pmatrix} \Delta u_{i3} \\ \Delta u_{i4} \\ \vdots \\ \Delta u_{iT} \end{pmatrix}, \quad \mathbf{u}_i = \begin{pmatrix} u_{i3} \\ u_{i4} \\ \vdots \\ u_{iT} \end{pmatrix}, \quad \mathbf{p}_i = \begin{pmatrix} \Delta\mathbf{u}_i \\ \mathbf{u}_i \end{pmatrix}.$$

The instrument matrix for the differenced equation is $\mathbf{Z}_{di} = (\mathbf{Z}_{di}^y, \mathbf{Z}_{di}^x)$, where

$$\mathbf{Z}_{di}^y = \begin{pmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} \end{pmatrix},$$

and $\mathbf{Z}_{di}^x$ is similarly defined. The instrument for the system equation $\mathbf{Z}_{si}$ is a block matrix with $\mathbf{Z}_{di}$ and $\mathbf{Z}_{li}$ on the main diagonals and zeros otherwise, i.e.,

$$\mathbf{Z}_{si} = \begin{pmatrix} \mathbf{Z}_{di} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_{li} \end{pmatrix},$$

where $\mathbf{Z}_{li}$ is the instrument matrix for the level equation and $\mathbf{Z}_{li} = (\mathbf{Z}_{li}^{y}, \mathbf{Z}_{li}^{x})$, where

$$\mathbf{Z}_{li}^{y} = \begin{pmatrix} \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix},$$

and $\mathbf{Z}_{li}^{x}$ is defined similarly.

In words, the DIF GMM estimator is obtained from the moment conditions where endogenous differences of the variables are instrumented by their lagged levels and the SYS GMM estimator utilizes further moment conditions where endogenous level variables are instrumented by their lagged differences. The validity of these moment conditions is tested by the Sargan-Hansen test of overidentifying restrictions (see Sargan [11] and Hansen [12]).

Let $\mathbf{w}_i$ $(i = 1, ..., N)$ denote the i-th observation and $\mathbf{g}_i(\theta, \beta) = \mathbf{g}(\mathbf{w}_i, \theta, \beta)$. The sample first and second moments of the $\mathbf{g}$ are given by:

$$\widehat{\mathbf{g}}(\theta, \beta) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}_i(\theta, \beta), \tag{14}$$

$$\widehat{\mathbf{\Omega}}(\theta, \beta) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}_i(\theta, \beta) \mathbf{g}_i(\theta, \beta)'. \tag{15}$$

Two-Step GMM Estimator: The standard two-step GMM estimator is the solution to the following minimization problem:

$$\left(\widehat{\theta}, \widehat{\beta}\right) = \underset{\theta, \beta}{\operatorname{argmin}}\, \widehat{\mathbf{g}}(\theta, \beta)' \left[\widehat{\mathbf{\Omega}}\left(\widetilde{\theta}, \widetilde{\beta}\right)\right]^{-1} \widehat{\mathbf{g}}(\theta, \beta), \tag{16}$$

where $\left(\widetilde{\theta}, \widetilde{\beta}\right)$ is a preliminary estimator, e.g., the first-step estimator.[1] The first-order conditions are:

$$\widehat{\mathbf{g}}_{\theta}' \left[\widehat{\mathbf{\Omega}}\left(\widetilde{\theta}, \widetilde{\beta}\right)\right]^{-1} \widehat{\mathbf{g}} = 0, \tag{17}$$

---

[1] To obtain a consistent first-step estimator, we use

$$\mathbf{W}_d = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_{di}' \mathbf{H}_d \mathbf{Z}_{di}\right)^{-1} \text{ and } \mathbf{W}_s = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}_{si}' \mathbf{H}_s \mathbf{Z}_{si}\right)^{-1},$$

in place of $\left[\widehat{\mathbf{\Omega}}\left(\widetilde{\theta}, \widetilde{\beta}\right)\right]^{-1}$ in Equation (16), suggested by Arellano and Bond [1] and Blundell, Bond, and Windmeijer [5] for the DIF and SYS GMM estimators, respectively, where $\mathbf{H}_d$ is a (T-2) square matrix that has twos in the main diagonal, minus ones in the first subdiagonals, and zeros otherwise, i.e.,

$$\mathbf{H}_d = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{pmatrix},$$

and $\mathbf{H}_s$ is the matrix

$$\mathbf{H}_s = \begin{pmatrix} \mathbf{H}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T-2} \end{pmatrix}.$$

$$\widehat{\mathbf{g}}_{\beta}' \left[\widehat{\mathbf{\Omega}} \left(\widetilde{\theta}, \widetilde{\beta} \right) \right]^{-1} \widehat{\mathbf{g}} = \mathbf{0}, \tag{18}$$

where $\widehat{\mathbf{g}} = \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right)$, $\widehat{\mathbf{g}}_{\theta} = \partial \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right) / \partial \theta$ and $\widehat{\mathbf{g}}_{\beta} = \partial \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right) / \partial \beta$.

Subset-Continuous-Updating GMM Estimator: Motivated by the fact that the autoregressive parameter θ in a stationary dynamic panel data model lies in the bounded interval $(-1, 1)$, we propose to apply the idea of continuous-updating of Hansen, Heaton, and Yaron [9] solely to this bounded parameter θ. The subset-continuous-updating GMM estimator is obtained as the solution to the following minimization problem:

$$\widehat{\beta}(\theta) = \underset{\beta}{\operatorname{argmin}}\, \widehat{\mathbf{g}}(\theta, \beta)' \left[\widehat{\Omega} \left(\theta, \widetilde{\beta} \right) \right]^{-1} \widehat{\mathbf{g}}(\theta, \beta) \text{ conditional on } \theta, \tag{19}$$

and

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \widehat{\mathbf{g}} \left(\theta, \widehat{\beta}(\theta) \right)' \left[\widehat{\Omega} \left(\theta, \widehat{\beta}(\theta) \right) \right]^{-1} \widehat{\mathbf{g}} \left(\theta, \widehat{\beta}(\theta) \right), \tag{20}$$

where $\widetilde{\beta}$ is a preliminary estimator.[2] The first-order condition with respect to θ is:

$$\left(\widehat{\mathbf{g}}_{\theta} + \widehat{\mathbf{g}}_{\beta} \frac{\partial \widehat{\beta}}{\partial \theta} \right)' \left[\widehat{\mathbf{\Omega}} \left(\widehat{\theta}, \widehat{\beta}(\widehat{\theta}) \right) \right]^{-1} \widehat{\mathbf{g}} - \widehat{\mathbf{g}}' \left[\widehat{\mathbf{\Omega}} \left(\widehat{\theta}, \widehat{\beta}(\widehat{\theta}) \right) \right]^{-1} \widehat{\mathbf{\Lambda}} \left[\widehat{\mathbf{\Omega}} \left(\widehat{\theta}, \widehat{\beta}(\widehat{\theta}) \right) \right]^{-1} \widehat{\mathbf{g}} = 0, \tag{21}$$

where

$$\widehat{\mathbf{\Lambda}} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{g}}_i \left(\widehat{\mathbf{g}}_{\theta,i} + \widehat{\mathbf{g}}_{\beta,i} \frac{\partial \widehat{\beta}}{\partial \theta} \right)', \tag{22}$$

and $\widehat{\mathbf{g}} = \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right)$, $\widehat{\mathbf{g}}_{\theta} = \partial \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right) / \partial \theta$, $\widehat{\mathbf{g}}_{\beta} = \partial \widehat{\mathbf{g}} \left(\widehat{\theta}, \widehat{\beta} \right) / \partial \beta$, $\widehat{\mathbf{g}}_i = \mathbf{g} \left(\mathbf{w}_i, \widehat{\theta}, \widehat{\beta} \right)$, $\widehat{\mathbf{g}}_{\theta,i} = \partial \mathbf{g} \left(\mathbf{w}_i, \widehat{\theta}, \widehat{\beta} \right) / \partial \theta$, and $\widehat{\mathbf{g}}_{\beta,i} = \partial \mathbf{g} \left(\mathbf{w}_i, \widehat{\theta}, \widehat{\beta} \right) / \partial \beta$.

We call the solution to the above minimization problem a subset-continuous-updating estimator to reflect the fact that we are applying the idea of continuous-updating of Hansen, Heaton, and Yaron [9] on a subset of model parameters. Given the linearity of the DIF and SYS GMM moment conditions, the minimization problem in Equation (19) yields a closed-form solution for $\widehat{\beta}$ as a function of θ. The $\widehat{\beta}$ estimator is of course consistent conditional on a consistent estimator $\widehat{\theta}$. The $\widehat{\theta}$ estimator is consistent according to the jackknife interpretation in the work of Donald and Newey [10]. More specifically, consider the regression of $\widehat{\mathbf{g}}_{\theta,i} + \widehat{\mathbf{g}}_{\beta,i} \partial \widehat{\beta} / \partial \theta$ on $\widehat{\mathbf{g}}_i$, let

$$\widehat{\mathbf{B}} = \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{g}}_i \widehat{\mathbf{g}}_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \widehat{\mathbf{g}}_i \left(\widehat{\mathbf{g}}_{\theta,i} + \widehat{\mathbf{g}}_{\beta,i} \frac{\partial \widehat{\beta}}{\partial \theta} \right)' \right) = \left[\widehat{\mathbf{\Omega}} \left(\widehat{\theta}, \widehat{\beta}(\widehat{\theta}) \right) \right]^{-1} \widehat{\mathbf{\Lambda}} \tag{23}$$

denote the matrix of coefficients. The vector of residuals follows:

$$\widehat{\boldsymbol{\eta}}_i = \widehat{\mathbf{g}}_{\theta,i} + \widehat{\mathbf{g}}_{\beta,i} \frac{\partial \widehat{\beta}}{\partial \theta} - \widehat{\mathbf{B}}' \widehat{\mathbf{g}}_i. \tag{24}$$

---

[2] In the Monte Carlo experiments, we conduct a bounded optimization over $\theta \in (-1, 1)$ using the two-step DIF or SYS GMM estimator of θ as the starting value until a convergence criterion is met. We set both the step tolerance and the function tolerance to a relatively small number, 10^{-8}.

By definition, the residual $\widehat{\boldsymbol{\eta}}_i$ is orthogonal to the regressor $\widehat{\mathbf{g}}_i$, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\eta}}_i\widehat{\mathbf{g}}_i' = 0. \tag{25}$$

Then, we can rewrite the first-order condition with respect to θ in Equation (21) as:

$$\left(\widehat{\mathbf{g}}_\theta + \widehat{\mathbf{g}}_\beta\frac{\partial\widehat{\beta}}{\partial\theta} - \widehat{\mathbf{B}}'\widehat{\mathbf{g}}\right)'\left[\widehat{\boldsymbol{\Omega}}\left(\widehat{\theta},\widehat{\beta}(\widehat{\theta})\right)\right]^{-1}\widehat{\mathbf{g}} = 0, \tag{26}$$

which simplifies to essentially the same equation given by Donald and Newey [10] (at the bottom of page 240):

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j\neq i}^{N}\widehat{\boldsymbol{\eta}}_j'\left[\widehat{\boldsymbol{\Omega}}\left(\widehat{\theta},\widehat{\beta}(\widehat{\theta})\right)\right]^{-1}\right)\widehat{\mathbf{g}}_i = 0. \tag{27}$$

Let $\widehat{\mathbf{A}}_i$ denote the term inside the parentheses in Equation (27):

$$\widehat{\mathbf{A}}_i = \frac{1}{N}\sum_{j\neq i}^{N}\widehat{\boldsymbol{\eta}}_j'\left[\widehat{\boldsymbol{\Omega}}\left(\widehat{\theta},\widehat{\beta}(\widehat{\theta})\right)\right]^{-1}. \tag{28}$$

This term converges to the same limit for all $i = 1,...,N$ as $N \to \infty$ since $\widehat{\mathbf{g}}$ converges to zero in probability, i.e., $\widehat{\mathbf{g}} = \mathbf{o}_p(1)$. As Donald and Newey [10] point out, Equation (27) is simply a modification of the usual interpretation of a GMM estimator that allows "a linear combination coefficient for each observation, which excludes its own observation from the Jacobian of the moments." Given that $\widehat{\mathbf{A}}_i$ has the same limit, this modification does not change the asymptotic distribution of the estimator, which implies that our subset-continuous-updating estimator retains consistency.

While we use a scalar parameter β for expositional clarity here, our subset-continuous-updating method applies to a K-dimensional vector of parameters $\boldsymbol{\beta}$ without any extra computational burden because the continuous-updating is imposed on the scalar parameter θ only; see Equation (20). Applying our subset-continuous-updating method to the dynamic panel data model, the resulting SCUDIF and SCUSYS GMM estimators have exactly the same asymptotic distributions as their two-step counterparts in terms of the first-order terms. They are computationally advantageous relative to the continuous-updating estimator in that they replace a $K+1$-dimensional numerical optimization over the unbounded domain of $(\theta,\boldsymbol{\beta})$ by a one-dimensional optimization over a bounded domain $\theta \in (-1,1)$. It is worth noting that our subset-continuous-updating estimator can be easily extended to the AR(2) case, where necessary. This extension leads to a two-dimensional, instead of one-dimensional, optimization, which is more computationally burdensome, but the optimization is at least still limited to the well-defined, bounded region of the parameter space corresponding to stationary dynamics. Per Box and Jenkins [13], this AR(2) stationary region is a triangle.

3. Monte Carlo Experiments

In this section, we conduct Monte Carlo experiments to compare the performance of our subset-continuous-updating estimators with the standard two-step estimators in finite samples. We consider the model and assumptions specified in Section 2. Without loss of generality, we consider one additional explanatory variable beyond the lagged dependent variable, i.e., we restrict $K = 1$. This restriction is made for expositional clarity here, as our proposed estimation method applies to models with multiple explanatory variables ($K > 1$) without any extra computational burden because the continuous-updating is imposed solely on the bounded scalar parameter θ. The model specification is described by Equations (2) and (3).

For each Monte Carlo replication, μ_i, ν_{it}, and e_{it} are all drawn from the normal distribution with zero means and standard deviations σ_μ, σ_ν, and σ_e. The initial observations are drawn from the mean stationary distribution as in Equations (8) and (9).[3] Then, we generate the data x_{it} and y_{it} and discard the first 30 observations before selecting our sample. We keep the following parameters fixed in the various Monte Carlo simulations:

$$\beta = 1,\ \tau = 0.25,\ \sigma_\nu^2 = 1,\ \sigma_e^2 = 0.16.$$

These parameter values are taken from Blundell, Bond, and Windmeijer [5]. The parameters that are varied in the Monte Carlo experiments include:

$\theta = \{0.5, 0.8\}$ for a moderate vs. high persistence in y_{it},
$\rho = \{0.5, 0.8\}$ for a moderate vs. high persistence in x_{it},
$\lambda = \{-0.1, -0.4\}$ for a low vs. high level of idiosynchratic-error endogeneity in x_{it},
$\sigma_\mu^2 = \{1/4, 4\}$ for a small vs. large variance of fixed effects relative to idiosyncratic errors.

Following Blundell, Bond, and Windmeijer [5], we fix the sample size N at 500 and consider $T = \{4, 8\}$. We also further consider an even longer panel, i.e., $T = 12$. These results are presented in Tables 1–3, respectively, where we compare our subset-continuous-updating estimators (denoted SCUDIF and SCUSYS) to the standard two-step estimators of Arellano and Bond [1] and Blundell and Bond [3] (denoted DIF and SYS) from three perspectives: (1) the estimation accuracy, quantified by median absolute errors (MAE), (2) the sampling standard deviations across all simulation repetitions (SD), (3) the Windmeijer [14] corrected standard errors (SE),[4] (4) the size of the two-tailed t-test under the null hypothesis that the parameter equals the true value at the 5% significance level, and (5) the rejection frequency of the 5% Sargan-Hansen test (denoted RF_{SH}). All results are obtained from 10,000 Monte Carlo repetitions. The following findings are worth noting.

Firstly, we find evidence that the standard two-step DIF and SYS GMM estimates of θ are sensitive to the variance ratio $\sigma_\mu^2/\sigma_\nu^2$, which is consistent with the results of Hayakawa [6] and Bun and Windmeijer [7]. In particular, for any given combination of θ, ρ, and λ, the DIF and SYS GMM estimates of the autoregressive parameter become more biased as the variance ratio increases. In most cases, the bias in the β estimates also increases with the variance ratio. For example, with $(\theta, \rho, \lambda) = (0.8, 0.8, -0.4)$ and $T = 4$, the MAEs of $\widehat{\theta}_{DIF}$ and $\widehat{\theta}_{SYS}$ both become double when the variance ratio increases from $1/4$ to 4; the MAEs of $\widehat{\beta}_{DIF}$ and $\widehat{\beta}_{SYS}$ become twice and four thirds bigger, respectively. In contrast, while the variance ratio also affects the estimation accuracy of the SCUDIF and SCUSYS GMM estimators, the MAE values for these estimators deteriorate less quickly than do the corresponding values for the standard two-step DIF and SYS GMM estimators.

Secondly, compared to the standard two-step DIF and SYS GMM estimators, our proposed subset-continuous-updating counterparts are noticeably less biased in estimating the autoregressive parameter, especially in the case of large variance ratios and relatively long panels, where the problem of too many weak instruments becomes more prominent. Consequently, for example, with

---

[3] After we draw μ_i, ν_{it}, and e_{it} from the normal distribution, we simply set the initial values of x_{it} and y_{it} to

$$x_{i1} = \frac{\tau}{1-\rho}\mu_i + \epsilon_{i1} \text{ where } \epsilon_{i1} = \lambda\nu_{i1} + e_{i1},$$
$$y_{i1} = \frac{1}{1-\theta}\left(1 + \frac{\beta\tau}{1-\rho}\right) + \xi_{i1} \text{ where } \xi_{i1} = \beta(\lambda\nu_{i1} + e_{i1}) + \nu_{i1}$$

to ensure mean stationarity.

[4] The standard error estimates for the two-step DIF and SYS GMM estimators of θ and β are calculated as in Windmeijer [14]. The same is obtained for the SCUDIF and SCUSYS GMM estimators of β conditional on θ. The standard error estimates for the SCUDIF and SCUSYS GMM estimators of θ are instead obtained from the Hessian matrix of the single-dimensional optimization problem.

$(\theta, \rho, \lambda) = (0.8, 0.8, -0.4)$, $\sigma_\mu^2/\sigma_v^2 = 4$, and $T = 12$, the MAE of $\widehat{\theta}_{SCUDIF}$ is two-thirds that of $\widehat{\theta}_{DIF}$ and the MAE of $\widehat{\theta}_{SCUSYS}$ is only one-fourth of that of $\widehat{\theta}_{SYS}$. Summarizing the θ MAE results in Tables 1–3, the following matrix displays the frequencies in the cases corresponding to the large variance ratio of $\sigma_\mu^2/\sigma_v^2 = 4$ with which our proposed subset-continuous-updating estimators have smaller MAE than the corresponding two-step estimators across the different (θ, ρ, λ) combinations:

Estimation accuracy on θ	$T = 4$	$T = 8$	$T = 12$
SCUDIF outperforms DIF	75.0%	100.0%	75.0%
SCUSYS outperforms SYS	87.5%	100.0%	100.0%

With regard to the β estimation, we do not observe any clear pattern in terms of the relative performance of these two sets of estimators. For example, with $(\theta, \rho, \lambda) = (0.8, 0.8, -0.4)$, $\sigma_\mu^2/\sigma_v^2 = 4$, and $T = 12$, both SCUDIF and SCUSYS are respective improvements on DIF and SYS in estimating β. However, when x_{it} becomes less persistent, i.e., $\rho = 0.5$, while SCUDIF still significantly outperforms DIF, SCUSYS performs worse than SYS in terms of the estimation accuracy on β.

Turning to the size estimates for the t-tests, we find that none of the standard two-step or the subset-continuous-updating estimators consistently yields well-sized t-tests for either θ or for β. For the DIF and SYS tests, this is a bias problem: the Windmeijer [14] corrected standard errors are generally good estimates of the sampling standard deviations for these standard two-step estimators, but this is no guarantee of a well-sized t-test when the parameter estimates suffer substantial biases. For example, with $(\theta = 0.8, \rho = 0.5, \lambda = -0.4)$ and $\sigma_\mu^2/\sigma_v^2 = 4$, the empirical rejection frequencies of the t-tests of $H_0 : \theta = 0.8$ and $H_0 : \beta = 1$ are both greater than 50% at all of the dataset lengths. The SCUDIF and SCUSYS estimators are less biased, but these subset-continuous-updating estimators lead to standard error estimates which are somewhat downward biased; hence, these tests are also over-sized. These results suggest that bootstrap-simulation based methods are necessary for credible structural parameter inference with samples of the lengths considered here, regardless of which of these estimators one chooses. We note, however, that the more precise estimation provided by the SCUDIF and SCUSYS theta estimators is also advantageous using simulation-based inference.

Lastly and most importantly, the standard two-step estimators tend to over-reject the Sargan-Hansen test and our subset-continuous-updating counterparts are generally well-sized. Given the fact that the endogenous variable x_{it} and the dependent variable y_{it} are both mean stationary, lagged levels and lagged differences of x_{it} and y_{it} are valid instruments for the first-difference equation and the level equation, respectively. Thus, presuming that the model is in all other respects well-specified, the null hypothesis of the Sargan-Hansen test is satisfied. With any panel length that we consider, $T = \{4, 8, 12\}$, the rejection frequency of the Sargan-Hansen test for our subset-continuous-updating estimators lies around 5% whereas it sometimes exceeds 15%, for instance with $\theta = 0.8$, $\sigma_\mu^2/\sigma_v^2 = 4$, and $T = \{8, 12\}$, when the standard two-step estimators are adopted. In the worst case scenario, $(\theta, \rho, \lambda) = (0.8, 0.5, -0.1)$, $\sigma_\mu^2/\sigma_v^2 = 4$, and $T = 8$, the rejection frequency is even higher than 25%. In other words, the Sargan–Hansen associated with the standard two-step SYS GMM estimator has a 25 percent chance of incorrectly rejecting a true null hypothesis. In contrast, the 5% Sargan-Hansen test associated with the subset-continuous-updating SYS GMM estimator has a reasonable rejection frequency (size) of 6.2%.

In conclusion, the subset-continuous-updating method that we propose in this paper is shown to improve the estimation accuracy on the autoregressive parameter and the size of the Sargan-Hansen test in a dynamic panel data model when the variance ratio becomes large. No extra computational burden is incurred when we apply this method to models with multiple explanatory variables.

Table 1. Monte Carlo results, $N = 500$ and $T = 4$.

$\lambda = -0.1$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0625	0.0914	0.0922	5.5	0.2341	0.3479	0.3463	5.4	5.2	0.1240	0.1707	0.1721	8.7	0.3543	0.5018	0.5057	7.2	5.7
	SCUDIF	0.0627	0.0945	0.0665	16.7	0.2328	0.3538	0.3180	7.4	4.8	0.1287	0.2004	0.1447	25.8	0.3679	0.5760	0.3384	23.0	4.1
	SYS	0.0366	0.0525	0.0525	5.4	0.1565	0.2321	0.2309	4.9	4.6	0.0652	0.0749	0.0753	17.4	0.2203	0.3025	0.3019	9.7	10.2
	SCUSYS	0.0377	0.0546	0.0375	18.4	0.1569	0.2327	0.2282	5.3	4.4	0.0515	0.0792	0.0499	26.8	0.2191	0.3196	0.2978	9.5	6.0
(0.5, 0.8)	DIF	0.0593	0.0872	0.0877	5.5	0.2811	0.4285	0.4309	4.5	4.9	0.1139	0.1598	0.1612	8.6	0.5146	0.7863	0.7799	6.9	6.2
	SCUDIF	0.0592	0.0917	0.0642	16.9	0.2859	0.4456	0.3519	11.1	4.3	0.1316	0.2141	0.1482	34.0	0.6021	1.0479	0.4074	40.6	4.1
	SYS	0.0335	0.0488	0.0485	6.0	0.1037	0.1592	0.1632	4.2	4.3	0.0568	0.0639	0.0612	20.0	0.1986	0.2217	0.2214	19.0	14.2
	SCUSYS	0.0340	0.0509	0.0346	19.0	0.1042	0.1605	0.1605	4.7	4.1	0.0462	0.0767	0.0400	33.7	0.2124	0.2565	0.2249	22.2	10.3
(0.8, 0.5)	DIF	0.1423	0.2010	0.2058	8.5	0.3243	0.4624	0.4726	7.2	6.2	0.3336	0.3647	0.3702	19.5	0.6584	0.7767	0.7656	17.5	7.2
	SCUDIF	0.1442	0.1778	0.2230	14.9	0.2950	0.4431	0.3114	16.5	4.5	0.2000	0.3446	0.4377	26.5	0.4862	0.7555	0.3129	39.5	4.3
	SYS	0.0388	0.0593	0.0578	7.2	0.1475	0.2223	0.2270	4.4	4.5	0.0836	0.0652	0.0605	48.0	0.1695	0.2529	0.2671	4.9	12.8
	SCUSYS	0.0422	0.0683	0.0419	23.6	0.1508	0.2268	0.2213	5.5	4.1	0.0738	0.1466	0.0980	48.9	0.2066	0.3466	0.2877	9.0	5.7
(0.8, 0.8)	DIF	0.1073	0.1592	0.1598	6.7	0.4205	0.6486	0.6479	6.0	5.4	0.2202	0.2791	0.2827	12.6	0.8422	1.1380	1.1334	11.2	4.9
	SCUDIF	0.1124	0.1658	0.1786	18.1	0.4179	0.7014	0.3375	28.8	4.0	0.2000	0.3246	0.3684	32.6	0.7872	1.3599	0.3669	56.6	2.6
	SYS	0.0301	0.0461	0.0459	6.4	0.0910	0.1433	0.1450	4.3	4.7	0.0586	0.0457	0.0429	44.3	0.1138	0.1752	0.1871	4.8	9.5
	SCUSYS	0.0330	0.0539	0.0338	23.5	0.0935	0.1528	0.1400	6.0	4.1	0.0542	0.0893	0.0389	53.6	0.1523	0.2717	0.2051	13.4	5.5
$\lambda = -0.4$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0974	0.1288	0.1319	10.0	0.2623	0.3349	0.3407	12.3	7.8	0.1909	0.1961	0.2020	22.4	0.4626	0.4750	0.4872	22.8	11.6
	SCUDIF	0.0936	0.1500	0.1037	18.2	0.2483	0.3835	0.2241	24.5	5.3	0.1776	0.2575	0.1893	34.2	0.4215	0.6052	0.2334	49.1	4.9
	SYS	0.0452	0.0668	0.0666	6.1	0.1237	0.1712	0.1727	6.9	5.2	0.0921	0.1027	0.0981	23.5	0.1753	0.2489	0.2474	9.9	10.5
	SCUSYS	0.0483	0.0718	0.0479	20.4	0.1235	0.1743	0.1574	9.9	4.9	0.0735	0.1242	0.0678	33.8	0.1864	0.2912	0.2082	16.0	4.2
(0.5, 0.8)	DIF	0.0631	0.0888	0.0899	7.6	0.2501	0.3409	0.3472	8.6	6.6	0.1144	0.1404	0.1425	14.4	0.4260	0.5365	0.5426	14.0	7.7
	SCUDIF	0.0626	0.0993	0.0682	17.9	0.2390	0.3787	0.2361	20.7	5.2	0.1116	0.1890	0.1283	33.9	0.4030	0.7097	0.2556	43.9	4.0
	SYS	0.0331	0.0484	0.0484	5.5	0.0705	0.1057	0.1059	4.7	4.8	0.0611	0.0681	0.0639	22.8	0.1265	0.1637	0.1656	15.5	10.8
	SCUSYS	0.0349	0.0513	0.0345	19.7	0.0713	0.1073	0.0997	6.6	4.6	0.0459	0.0837	0.0415	32.7	0.1371	0.2080	0.1618	15.8	5.5
(0.8, 0.5)	DIF	0.2941	0.2778	0.2659	30.4	0.6348	0.5928	0.5655	31.6	12.9	0.5076	0.3054	0.2996	53.7	1.0708	0.6656	0.6335	53.0	10.3
	SCUDIF	0.2000	0.2754	0.2961	21.6	0.3877	0.6099	0.2122	42.9	5.8	0.2216	0.3771	0.4255	33.1	0.6169	0.8182	0.1970	60.8	5.9
	SYS	0.0503	0.0809	0.0782	6.9	0.1254	0.1861	0.1861	5.4	4.9	0.0966	0.0698	0.0643	54.1	0.1238	0.1931	0.2059	3.9	8.4
	SCUSYS	0.0593	0.0950	0.0624	28.2	0.1312	0.1992	0.1567	13.3	4.0	0.1053	0.1853	0.1103	59.9	0.1845	0.3784	0.2153	15.8	3.3
(0.8, 0.8)	DIF	0.1067	0.1350	0.1360	13.0	0.4377	0.5473	0.5444	13.9	8.5	0.2115	0.1941	0.1923	29.4	0.8441	0.7932	0.7753	28.7	9.2
	SCUDIF	0.1034	0.1491	0.1360	21.7	0.3989	0.6164	0.2341	44.3	4.6	0.2000	0.2705	0.2288	38.1	0.6728	1.1154	0.2419	67.7	3.3
	SYS	0.0272	0.0429	0.0428	5.9	0.0601	0.0940	0.0945	3.5	5.2	0.0540	0.0459	0.0430	42.0	0.0820	0.1300	0.1378	6.4	8.4
	SCUSYS	0.0309	0.0499	0.0317	23.6	0.0619	0.1029	0.0837	8.1	4.6	0.0510	0.0943	0.0385	53.0	0.1167	0.2754	0.1502	16.4	4.1

Note: Statistics are based on 10,000 Monte Carlo replications. MAE stands for the median absolute error; SD is the sampling standard deviation across 10,000 simulation replications and SE stands for the average standard error estimate. Size (in percent) is the rejection frequency of the two-tailed t-test, and RF_{SH} (in percent) is the rejection frequency of the 5% Sargan-Hansen test. DIF and SYS stand for the standard two-step Arellano-Bond and Blundell-Bond GMM estimators. Their subset-continuous-updating counterparts are denoted SCUDIF and SCUSYS. The total number of instruments is $m_d = 6$ for DIF and SCUDIF and $m_s = 10$ for SYS and SCUSYS.

Table 2. Monte Carlo results, $N = 500$ and $T = 8$.

$\lambda = -0.1$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0251	0.0345	0.0347	7.1	0.0969	0.1304	0.1289	7.6	5.5	0.0353	0.0443	0.0443	8.9	0.1080	0.1381	0.1381	9.1	5.6
	SCUDIF	0.0242	0.0361	0.0245	19.3	0.0943	0.1305	0.1263	7.7	5.2	0.0319	0.0479	0.0327	32.4	0.1011	0.1397	0.1288	9.2	5.0
	SYS	0.0167	0.0244	0.0241	5.8	0.0778	0.1116	0.1101	6.1	5.4	0.0383	0.0349	0.0320	22.6	0.1174	0.1435	0.1340	12.7	11.0
	SCUSYS	0.0171	0.0252	0.0171	19.7	0.0773	0.1117	0.1093	6.2	5.3	0.0197	0.0294	0.0205	29.4	0.1017	0.1421	0.1360	8.1	5.6
(0.5, 0.8)	DIF	0.0235	0.0325	0.0328	7.0	0.0880	0.1203	0.1197	6.9	5.2	0.0319	0.0414	0.0415	8.5	0.1098	0.1444	0.1431	8.1	5.3
	SCUDIF	0.0229	0.0343	0.0238	21.8	0.0837	0.1211	0.1111	8.1	4.7	0.0305	0.0460	0.0291	39.6	0.1009	0.1506	0.1167	13.0	4.7
	SYS	0.0158	0.0228	0.0225	6.2	0.0512	0.0760	0.0751	5.7	4.8	0.0462	0.0300	0.0275	40.2	0.1493	0.0962	0.0909	41.9	17.1
	SCUSYS	0.0161	0.0237	0.0162	21.7	0.0513	0.0763	0.0744	6.2	4.8	0.0200	0.0278	0.0178	39.7	0.1441	0.1135	0.1056	31.2	9.6
(0.8, 0.5)	DIF	0.0568	0.0584	0.0595	13.8	0.1322	0.1465	0.1454	12.6	6.4	0.1023	0.0800	0.0820	22.8	0.1952	0.1738	0.1715	19.4	7.0
	SCUDIF	0.0431	0.0650	0.0439	32.4	0.1069	0.1520	0.1246	12.3	5.0	0.0658	0.0994	0.0542	52.4	0.1306	0.1921	0.1258	20.3	4.6
	SYS	0.0194	0.0264	0.0263	8.7	0.0757	0.1083	0.1073	6.3	5.0	0.0813	0.0275	0.0240	80.2	0.0870	0.1137	0.1124	8.0	26.0
	SCUSYS	0.0194	0.0281	0.0188	29.6	0.0762	0.1090	0.1065	6.9	4.9	0.0290	0.0477	0.0227	52.0	0.0976	0.1365	0.1286	8.3	6.2
(0.8, 0.8)	DIF	0.0410	0.0456	0.0455	11.5	0.1347	0.1591	0.1553	10.8	5.6	0.0666	0.0624	0.0617	17.6	0.2074	0.2137	0.2049	16.0	5.8
	SCUDIF	0.0340	0.0531	0.0345	34.8	0.1140	0.1752	0.1065	21.8	4.5	0.0543	0.0864	0.0417	55.2	0.1656	0.2754	0.1102	39.3	4.0
	SYS	0.0154	0.0201	0.0201	9.3	0.0447	0.0675	0.0672	5.1	5.6	0.0573	0.0196	0.0168	81.5	0.0642	0.0699	0.0693	14.4	22.4
	SCUSYS	0.0149	0.0222	0.0145	31.2	0.0462	0.0695	0.0673	6.1	5.0	0.0237	0.0403	0.0130	59.7	0.1003	0.0963	0.0901	23.0	8.4
$\lambda = -0.4$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0447	0.0449	0.0454	15.8	0.1514	0.1007	0.1007	33.8	9.3	0.0637	0.0531	0.0534	22.5	0.1826	0.1102	0.1096	39.9	9.6
	SCUDIF	0.0341	0.0499	0.0342	22.8	0.1078	0.1059	0.0848	28.3	7.3	0.0413	0.0615	0.0395	40.6	0.1150	0.1190	0.0850	31.8	6.9
	SYS	0.0216	0.0313	0.0309	5.9	0.0937	0.0809	0.0804	21.6	6.7	0.0624	0.0496	0.0437	31.5	0.0810	0.1121	0.1025	10.7	10.9
	SCUSYS	0.0226	0.0334	0.0222	23.1	0.0871	0.0817	0.0752	22.5	6.4	0.0266	0.0402	0.0273	35.6	0.0837	0.1072	0.0917	14.7	4.1
(0.5, 0.8)	DIF	0.0244	0.0318	0.0318	8.3	0.0868	0.0879	0.0853	16.9	6.1	0.0313	0.0377	0.0374	10.7	0.1026	0.0978	0.0957	19.2	6.8
	SCUDIF	0.0224	0.0338	0.0234	22.9	0.0741	0.0897	0.0728	18.8	5.4	0.0272	0.0409	0.0253	40.4	0.0790	0.1016	0.0739	22.7	6.0
	SYS	0.0154	0.0222	0.0220	6.1	0.0405	0.0523	0.0507	9.1	5.0	0.0497	0.0319	0.0285	42.2	0.0782	0.0740	0.0687	25.2	14.2
	SCUSYS	0.0161	0.0236	0.0161	23.8	0.0405	0.0523	0.0498	9.9	4.9	0.0184	0.0281	0.0185	37.6	0.0609	0.0840	0.0755	11.6	5.2
(0.8, 0.5)	DIF	0.1627	0.0820	0.0797	54.4	0.3560	0.1615	0.1526	65.9	13.5	0.2329	0.0974	0.0941	70.2	0.4788	0.1926	0.1782	77.0	14.0
	SCUDIF	0.0665	0.0994	0.0607	42.9	0.1477	0.1833	0.0848	44.2	6.9	0.0895	0.1302	0.0653	57.6	0.1725	0.2341	0.0851	50.8	6.0
	SYS	0.0257	0.0370	0.0358	7.9	0.0936	0.0869	0.0857	18.9	6.3	0.0940	0.0287	0.0248	84.1	0.0591	0.0805	0.0792	7.2	17.5
	SCUSYS	0.0281	0.0405	0.0249	38.8	0.0903	0.0891	0.0737	26.9	5.7	0.0452	0.0755	0.0305	61.7	0.0988	0.1274	0.0900	21.1	4.1
(0.8, 0.8)	DIF	0.0432	0.0388	0.0383	19.4	0.1645	0.1288	0.1248	26.6	7.5	0.0642	0.0483	0.0468	27.5	0.2301	0.1623	0.1524	34.1	7.8
	SCUDIF	0.0284	0.0437	0.0277	37.8	0.1012	0.1395	0.0720	34.7	5.4	0.0376	0.0589	0.0294	53.8	0.1259	0.1864	0.0730	44.6	5.3
	SYS	0.0136	0.0188	0.0183	8.9	0.0348	0.0479	0.0471	7.2	5.2	0.0522	0.0190	0.0165	78.0	0.0552	0.0502	0.0499	21.7	16.9
	SCUSYS	0.0140	0.0206	0.0133	33.9	0.0362	0.0492	0.0445	10.2	4.9	0.0195	0.0362	0.0135	54.7	0.0628	0.0881	0.0695	16.0	5.4

Note: Statistics are based on 10,000 Monte Carlo replications. MAE stands for the median absolute error; SD is the sampling standard deviation across 10,000 simulation replications and SE stands for the average standard error estimate. Size (in percent) is the rejection frequency of the two-tailed t-test, and RF_{SH} (in percent) is the rejection frequency of the 5% Sargan-Hansen test. DIF and SYS stand for the standard two-step Arellano-Bond and Blundell-Bond GMM estimators. Their subset-continuous-updating counterparts are denoted SCUDIF and SCUSYS. The total number of instruments is $m_d = 42$ for DIF and SCUDIF and $m_s = 54$ for SYS and SCUSYS.

Table 3. Monte Carlo Results, $N = 500$ and $T = 12$.

$\lambda = -0.1$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0166	0.0231	0.0230	7.2	0.0752	0.0896	0.0879	12.0	4.5	0.0205	0.0265	0.0266	8.5	0.0813	0.0918	0.0900	12.8	4.8
	SCUDIF	0.0171	0.0257	0.0175	21.6	0.0738	0.0894	0.0872	11.9	4.4	0.0197	0.0299	0.0195	37.0	0.0770	0.0918	0.0884	12.2	4.5
	SYS	0.0125	0.0177	0.0175	5.9	0.0620	0.0822	0.0808	7.9	4.3	0.0327	0.0234	0.0210	35.8	0.0807	0.0988	0.0924	12.1	8.1
	SCUSYS	0.0133	0.0193	0.0133	22.0	0.0616	0.0822	0.0801	8.2	4.3	0.0145	0.0209	0.0145	35.7	0.0733	0.0980	0.0936	8.9	3.9
(0.5, 0.8)	DIF	0.0156	0.0219	0.0217	6.9	0.0530	0.0688	0.0684	8.7	5.2	0.0192	0.0251	0.0250	8.1	0.0590	0.0749	0.0739	10.2	5.3
	SCUDIF	0.0167	0.0246	0.0173	25.2	0.0517	0.0688	0.0669	8.7	5.0	0.0197	0.0290	0.0174	44.1	0.0553	0.0760	0.0694	10.5	5.1
	SYS	0.0119	0.0167	0.0164	6.5	0.0370	0.0545	0.0536	5.8	4.9	0.0430	0.0218	0.0195	57.7	0.1160	0.0683	0.0632	46.3	13.0
	SCUSYS	0.0126	0.0184	0.0126	26.8	0.0368	0.0545	0.0527	6.1	5.0	0.0155	0.0214	0.0130	45.7	0.1134	0.0767	0.0705	38.0	7.5
(0.8, 0.5)	DIF	0.0361	0.0335	0.0336	17.6	0.0953	0.0914	0.0921	16.5	6.1	0.0541	0.0410	0.0414	24.4	0.1149	0.0977	0.0978	21.0	6.3
	SCUDIF	0.0267	0.0399	0.0260	37.9	0.0753	0.0933	0.0861	13.0	5.2	0.0347	0.0527	0.0270	53.8	0.0791	0.1022	0.0866	15.1	5.1
	SYS	0.0136	0.0187	0.0187	7.7	0.0612	0.0795	0.0791	8.3	4.7	0.0760	0.0203	0.0180	91.7	0.0663	0.0825	0.0816	9.9	22.1
	SCUSYS	0.0139	0.0204	0.0136	34.3	0.0610	0.0797	0.0786	8.4	4.6	0.0175	0.0271	0.0143	52.0	0.0701	0.0948	0.0908	8.9	4.6
(0.8, 0.8)	DIF	0.0242	0.0257	0.0255	12.3	0.0718	0.0817	0.0789	12.8	5.2	0.0343	0.0317	0.0313	18.1	0.0958	0.0964	0.0911	16.7	5.2
	SCUDIF	0.0202	0.0307	0.0192	39.2	0.0585	0.0865	0.0647	16.0	4.6	0.0270	0.0419	0.0197	56.6	0.0721	0.1109	0.0658	23.7	4.4
	SYS	0.0111	0.0143	0.0142	9.8	0.0329	0.0489	0.0490	5.1	4.6	0.0545	0.0139	0.0123	94.8	0.0630	0.0476	0.0485	25.3	22.6
	SCUSYS	0.0108	0.0162	0.0105	36.7	0.0332	0.0494	0.0490	5.5	4.5	0.0168	0.0258	0.0102	59.9	0.0874	0.0648	0.0629	29.5	7.0
$\lambda = -0.4$		$\sigma_\mu^2/\sigma_v^2 = 1/4$									$\sigma_\mu^2/\sigma_v^2 = 4$								
(θ, ρ)		MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}	MAE(θ)	SD(θ)	SE(θ)	Size(θ)	MAE(β)	SD(β)	SE(β)	Size(β)	RF_{SH}
(0.5, 0.5)	DIF	0.0319	0.0294	0.0291	18.2	0.1493	0.0638	0.0628	65.6	9.1	0.0393	0.0321	0.0319	23.2	0.1591	0.0660	0.0646	68.0	9.4
	SCUDIF	0.0227	0.0346	0.0230	26.6	0.1247	0.0666	0.0577	56.9	8.0	0.0251	0.0386	0.0219	44.6	0.1253	0.0700	0.0579	56.7	7.7
	SYS	0.0158	0.0227	0.0225	6.2	0.1107	0.0582	0.0571	49.0	6.2	0.0534	0.0340	0.0291	45.1	0.0498	0.0738	0.0674	7.9	9.0
	SCUSYS	0.0176	0.0257	0.0173	27.5	0.1027	0.0592	0.0544	47.7	6.0	0.0190	0.0285	0.0180	43.2	0.0714	0.0696	0.0627	22.5	3.3
(0.5, 0.8)	DIF	0.0148	0.0210	0.0206	6.7	0.0624	0.0472	0.0455	29.1	6.2	0.0168	0.0230	0.0226	8.0	0.0667	0.0495	0.0474	30.1	6.4
	SCUDIF	0.0160	0.0238	0.0159	28.4	0.0579	0.0480	0.0434	29.1	6.1	0.0174	0.0264	0.0157	45.0	0.0595	0.0509	0.0439	30.3	5.8
	SYS	0.0117	0.0165	0.0161	7.4	0.0372	0.0371	0.0360	17.7	5.2	0.0471	0.0226	0.0199	62.7	0.0476	0.0499	0.0454	18.4	10.8
	SCUSYS	0.0130	0.0186	0.0125	31.2	0.0363	0.0370	0.0357	17.4	5.2	0.0141	0.0212	0.0131	43.9	0.0368	0.0529	0.0488	8.5	4.3
(0.8, 0.5)	DIF	0.1016	0.0456	0.0444	63.9	0.2555	0.0851	0.0820	88.0	13.4	0.1297	0.0522	0.0502	74.1	0.2967	0.0954	0.0895	91.5	14.1
	SCUDIF	0.0386	0.0580	0.0327	46.5	0.1287	0.0962	0.0576	57.2	8.3	0.0448	0.0696	0.0315	59.2	0.1331	0.1104	0.0577	57.6	8.3
	SYS	0.0177	0.0256	0.0252	7.4	0.1147	0.0614	0.0610	46.7	7.0	0.0897	0.0209	0.0184	94.8	0.0384	0.0572	0.0564	5.4	18.4
	SCUSYS	0.0210	0.0302	0.0170	45.2	0.1062	0.0638	0.0536	50.7	6.6	0.0254	0.0422	0.0174	58.9	0.0790	0.0754	0.0628	26.1	4.1
(0.8, 0.8)	DIF	0.0231	0.0215	0.0210	18.1	0.0947	0.0631	0.0587	37.3	6.7	0.0294	0.0247	0.0238	23.2	0.1103	0.0710	0.0647	41.6	6.6
	SCUDIF	0.0170	0.0255	0.0153	40.9	0.0619	0.0690	0.0431	35.4	5.7	0.0201	0.0303	0.0146	55.2	0.0673	0.0800	0.0434	39.0	5.5
	SYS	0.0095	0.0131	0.0127	8.8	0.0321	0.0362	0.0347	13.5	4.8	0.0495	0.0135	0.0119	92.3	0.0470	0.0358	0.0353	28.9	17.1
	SCUSYS	0.0105	0.0154	0.0094	40.2	0.0329	0.0369	0.0331	17.3	4.9	0.0122	0.0208	0.0092	54.6	0.0383	0.0550	0.0482	10.9	4.3

Note: Statistics are based on 10,000 Monte Carlo replications. MAE stands for the median absolute error; SD is the sampling standard deviation across 10,000 simulation replications and SE stands for the average standard error estimate. Size (in percent) is the rejection frequency of the two-tailed t-test, and RF_{SH} (in percent) is the rejection frequency of the 5% Sargan-Hansen test. DIF and SYS stand for the standard two-step Arellano-Bond and Blundell-Bond GMM estimators. Their subset-continuous-updating counterparts are denoted SCUDIF and SCUSYS. The total number of instruments is $m_d = 110$ for DIF and SCUDIF and $m_s = 130$ for SYS and SCUSYS.

4. Conclusions

The two-step GMM estimators of Arellano and Bond [1] and Blundell and Bond [3] for dynamic panel data models have been widely used in empirical work. However, neither of them performs well in small samples with weak instruments. The continuous-updating GMM estimator proposed by Hansen, Heaton, and Yaron [9] is in principle able to reduce the small-sample bias, but it involves high-dimensional optimizations when the number of regressors is large. Given the fact that the absolute value of the autoregressive parameter is less than unity for a dynamic panel data model to be stationary, we propose a computationally feasible variation on the standard two-step GMM estimators by applying the idea of continuous-updating to the autoregressive parameter only. We show that our subset-continuous-updating method does not alter the asymptotic distribution of the two-step GMM estimators and hence retains consistency. According to our Monte Carlo simulation results, the subset-continuous-updating GMM estimators for dynamic panel data models outperform their standard two-step counterparts in finite samples in terms of the estimation accuracy on the autoregressive parameter and the size of the Sargan–Hansen test.

Acknowledgments: We would like to thank editor Kerry Patterson and two anonymous referees for constructive comments.

Author Contributions: Overall, both authors contributed equally to this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arellano, M.; Bond, S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.* **1991**, *58*, 277–297.
2. Ahn, S.C.; Schmidt, P. Efficient estimation of models for dynamic panel data. *J. Econom.* **1995**, *68*, 5–27.
3. Blundell, R.; Bond, S. Initial conditions and moment restrictions in dynamic panel data models. *J. Econom.* **1998**, *87*, 115–143.
4. Arellano, M.; Bover, O. Another look at the instrumental variable estimation of error-component models. *J. Econom.* **1995**, *68*, 29–51.
5. Blundell, R.; Bond, S.; Windmeijer, F. Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator. In *Nonstationary Panels, Panel Cointegration and Dynamic Panels*; Baltagi, B.H., Fomby, T.B., Hill, R.C., Eds.; (Book Series: Advances in Econometrics); Emerald Group Publishing Limited: Bingley, UK, 2001; Volume 15, pp. 53–91.
6. Hayakawa, K. Small sample bias properties of the system GMM estimator in dynamic panel data models. *Econ. Lett.* **2007**, *95*, 32–38.
7. Bun, M.J.G.; Windmeijer, F. The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econom. J.* **2010**, *13*, 95–126.
8. Newey, W.K.; Windmeijer, F. Generalized method of moments with many weak moment conditions. *Econometrica* **2009**, *77*, 687–719.
9. Hansen, L.P.; Heaton, J.; Yaron, A. Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Stat.* **1996**, *14*, 262–280.
10. Donald, S.G.; Newey, W.K. A Jackknife Interpretation of the Continuous Updating Estimator. *Econ. Lett.* **2000**, *67*, 239–243.
11. Sargan, J.D. The estimation of economic relationships using instrumental variables. *Econometrica* **1958**, *26*, 393–415.
12. Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, *50*, 1029–1054.
13. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1976.

14. Windmeijer, F. A finite sample correction for the variance of linear efficient two-step GMM estimators. *J. Econom.* **2005**, *126*, 25–51.