Article

Second-Order Least Squares Method for Dynamic Panel Data Models with Application

Mustafa Salamh [1] and Liqun Wang [2,*]

1   Department of Statistics, Cairo University, Giza 12613, Egypt; must.kamal79@gmail.com
2   Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
*   Correspondence: Liqun.Wang@umanitoba.ca

Abstract: Management of financial risks and sound decision making rely on the accurate information and predictive models. Drawing useful information efficiently from big data with complex structures and building accurate models are therefore crucial tasks. Most commonly used methods for statistical inference in dynamic panel data models are based on the differencing transformation of data. However, differencing data may cause substantial loss of information, and therefore the subsequent analysis may fail to capture important features in the original level data. This point is demonstrated by a real data example where we use a semiparametrically efficient estimation method on the level data to reach a more favorable model. In particular, we study a second-order least squares approach which is based on the first two conditional moments of the response variable given the explanatory variables. This estimator is root-N consistent and its asymptotic variance reaches a lower bound semiparametric efficiency. Monte Carlo simulations show that this estimator performs favorably in finite sample situations compared to the first-differenced GMM and the random effects pseudo ML estimators. We also propose a new diagnostic test to check the working moments assumption based on the proposed estimator. A real data application is presented to further demonstrate the usage of this method.

Keywords: panel data; dynamic model; autoregressive processes; random effects; second order least squares; semiparametric efficiency; diagnostic test; econometric modeling; business economics

1. Introduction

Modern technology and data collection techniques have produced huge amount of data in business, economics, and many other fields. On the one hand, these data provide rich information to support decision making; however, on the other hand, the big sizes and complex structures of the data make the statistical analysis very challenging. How to draw useful information from data efficiently and how to identify accurate predictive models are important issues for the management of financial risks and decision making in general. A common data type is the repeated measurements data that are collected on a large number of units over certain period of time. In business and economics, the data are typically collected at regular time points (calendar time) and are usually called panel data. Statistical methodologies for the analysis of longitudinal data have been extensively studied in statistics as well as econometrics; however, the research follows two different directions.

In statistical literature, the main stream research focuses on the likelihood or generalized estimating equations (GEE) approaches in linear and nonlinear mixed effects models (e.g., Fitzmaurice et al. 2009). In contrast, in econometric literature, the emphasis is on the likelihood and generalized method of moments (GMM) approaches in dynamic panel data models (e.g., Arellano 2003; Baltagi 2008; Hsiao 2003 2011). In particular, The GMM approach is usually based on some suitable linear transformation such as the first differencing to eliminate the unobserved subject effects (Arellano and Bover 1995; Arellano and Bond 1991; Blundell and Bond 1998). However, differencing operation may cause substantial

loss of information in the data and therefore lead to the loss of estimation efficiency and inaccurate model identification. We will show through a real data example that more favorable model can be reached by applying a semiparametrically efficient estimation method to the original level data. Moreover, the model specification can be checked by our proposed diagnostic test.

In this paper, we propose a moments-based approach that is theoretically less restrictive than the likelihood-based methods and is fairly efficient and computationally tractable. This approach requires only the specification of the first two conditional moments of the unobserved subject effect given the process initial value and covariates, and does not require any other initial conditions or distributional assumptions. The data generating process can be either stationary or nonstationary, and does not need to be transformed. The so-called second-order least squares (SLS) estimator is consistent and asymptotically normally distributed when the cross section size N is large and the time series length T is fixed. In addition, this estimator reaches a semiparametric efficiency lower bound. We also propose a diagnostic test based on the SLS estimator to check the conditional moments assumption. Our extensive simulation studies show that the proposed estimator and its variants perform very well in finite sample situations and better than the GMM and likelihood based estimators in most cases.

The SLS method was first proposed by Wang (2003, 2004) to estimate the nonlinear measurement error models. It was extended to the nonlinear longitudinal data models with homoscedastic errors by Wang (2007) and to the censored linear models by Abarin and Wang (2009). Wang and Leblanc (2008) showed that under a nonlinear (homoscedastic) cross-sectional data model, the SLSE is asymptotically more efficient than the OLSE (pseudo MLE) when the error term has nonzero third moment, and both estimators are equally efficient otherwise. Kim and Ma (2012) proved that the SLSE attains the optimal semiparametric efficiency bound in general. More recently, the SLS method has also been used in optimal design problems by several researchers, e.g., Gao and Zhou (2014), Bose and Mukerjee (2015), Gao and Zhou (2017), Yin and Zhou (2017).

This paper is organized as follows. In Sections 2 and 3 we introduce the model and the semiparametric method for estimation and testing. In Section 4 we derive the asymptotically most efficient version of the proposed estimators. In Section 5 we carry out Monte Carlo simulations to study finite sample properties of the proposed estimators and compare them with other commonly used methods. In Section 6 we apply our method to a real data set to demonstrate its practical merits. Finally, conclusions and discussion are in Section 7 and regularity conditions and mathematical proofs are in the Appendices A–F.

2. Model and SLS Estimation

Let $(\mathbf{y}_i', \mathbf{x}_i', \eta_i)$, $i = 1, 2, \ldots, N$ be independent and identically distributed random vectors where $\mathbf{y}_i' = (y_{i0}, y_{i1}, \ldots, y_{iT})$ and $\mathbf{x}_i' = (\mathbf{x}_{i1}', \mathbf{x}_{i2}', \ldots, \mathbf{x}_{iT}')$ are respectively the measurements of the response variable and p covariates taken for the ith subject over T time periods. Suppose

$$y_{it} = \alpha_0 y_{i(t-1)} + \boldsymbol{\beta}_0' \mathbf{x}_{it} + \eta_i + \varepsilon_{it}, \quad t = 1, 2, \ldots, T, \tag{1}$$

where η_i is the unobserved subject effect and error term ε_{it} satisfies $E(\varepsilon_{it}|\eta_i, y_{i0}, \mathbf{x}_i) = 0$, $E(\varepsilon_{it}\varepsilon_{is}|\eta_i, y_{i0}, \mathbf{x}_i) = \sigma_0^2$ if $s = t$, and zero otherwise. In addition, we assume that the conditional moments $E(\eta_i^j|y_{i0}, \mathbf{x}_i) = f_j(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0)$, $j = 1, 2$ are known up to an ℓ-dimensional unknown parameter vector $\boldsymbol{\theta}_0$. This assumption is more general than the unrestricted initial conditions used by Blundell and Bond (1998) and Alvarez and Arellano (2003) to derive the conditional GLS (CGLS) and the random effects ML (RML) estimators, respectively. Note that our semiparametric assumption on η_i is not as restrictive as it appears because the functional forms of $f_j(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0)$, $j = 1, 2$ can be specified naturally based on some diagnostic tools as illustrated in Section 6. Moreover, any suggested specification can be tested using the test developed in Section 3. Also note that although we explicitly deal

with an $AR(1)$ model for the sake of simplicity of notation, our approach can be extended to more general $AR(p)$ models as discussed later.

We propose to estimate unknown parameters in model (1) based on the first two conditional moments of the response y_{it} given its initial value y_{i0} and covariates $\mathbf{x}_i$. Specifically, let $\boldsymbol{\gamma}_0 = (\alpha_0, \sigma_0^2, \boldsymbol{\beta}_0', \boldsymbol{\theta}_0')'$ and $\Gamma \subset \mathbf{R}^{p+\ell+2}$ be the corresponding parameter space. By backward substitution in Equation (1), we obtain the reduced form equation

$$y_{it} = \alpha_0^t y_{i0} + a_t(\alpha_0)\eta_i + \boldsymbol{\beta}_0' \tilde{\mathbf{x}}_{it}(\alpha_0) + \tilde{\varepsilon}_{it}(\alpha_0), \quad t = 1, 2, ..., T, \tag{2}$$

where $a_t(\alpha) = \sum_{r=0}^{t-1} \alpha^r$, $\tilde{\mathbf{x}}_{it}(\alpha) = \sum_{r=0}^{t-1} \alpha^r \mathbf{x}_{i(t-r)}$ and $\tilde{\varepsilon}_{it}(\alpha_0) = \sum_{r=0}^{t-1} \alpha_0^r \varepsilon_{i(t-r)}$. Under this model, the first two conditional moments of y_{it} are given by

$$
\begin{aligned}
\mu_{it}(\boldsymbol{\gamma}_0) &= E(y_{it}|y_{i0}, \mathbf{x}_i) = \alpha_0^t y_{i0} + \boldsymbol{\beta}_0' \tilde{\mathbf{x}}_{it}(\alpha_0) + a_t(\alpha_0) f_1(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0), \tag{3}\\
\nu_{its}(\boldsymbol{\gamma}_0) &= E(y_{it}y_{is}|y_{i0}, \mathbf{x}_i) = \alpha_0^{t+s} y_{i0}^2 + a_t(\alpha_0)a_s(\alpha_0) f_2(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0)\\
&\quad + \boldsymbol{\beta}_0' \tilde{\mathbf{x}}_{it}(\alpha_0) \tilde{\mathbf{x}}_{is}'(\alpha_0) \boldsymbol{\beta}_0 + \sigma_0^2 c_{ts}(\alpha_0) + d_{ts}(\alpha_0) y_{i0} f_1(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0)\\
&\quad + y_{i0} \boldsymbol{\beta}_0' \mathbf{w}_{its}(\alpha_0) + f_1(y_{i0}, \mathbf{x}_i, \boldsymbol{\theta}_0) \boldsymbol{\beta}_0' \mathbf{k}_{its}(\alpha_0), \tag{4}
\end{aligned}
$$

where $c_{ts}(\alpha) = \alpha^{t-s} \sum_{r=0}^{s-1} \alpha^{2r}$, $d_{ts}(\alpha) = \alpha^t a_s(\alpha) + \alpha^s a_t(\alpha)$, $\mathbf{w}_{its}(\alpha) = \alpha^t \tilde{\mathbf{x}}_{is}(\alpha) + \alpha^s \tilde{\mathbf{x}}_{it}(\alpha)$ and $\mathbf{k}_{its}(\alpha) = a_t(\alpha)\tilde{\mathbf{x}}_{is}(\alpha) + a_s(\alpha)\tilde{\mathbf{x}}_{it}(\alpha)$, $t \geq s$. Further, let

$$h_i(\boldsymbol{\gamma}) = (y_{it} - \mu_{it}(\boldsymbol{\gamma}), 1 \leq t \leq T, \ y_{it}y_{is} - \nu_{its}(\boldsymbol{\gamma}), 1 \leq s \leq t \leq T)', \boldsymbol{\gamma} \in \Gamma.$$

Then the second-order least squares (SLS) estimator for $\boldsymbol{\gamma}$ is defined by

$$\hat{\boldsymbol{\gamma}}_N = \operatorname*{argmin}_{\boldsymbol{\gamma} \in \Gamma} \frac{1}{N} \sum_{i=1}^{N} q_i(\boldsymbol{\gamma}), \tag{5}$$

where $q_i(\boldsymbol{\gamma}) = h_i'(\boldsymbol{\gamma}) \mathbf{W}_i h_i(\boldsymbol{\gamma})$ and $\mathbf{W}_i$ is a nonnegative definite matrix whose elements are real measurable functions of $(y_{i0}, \mathbf{x}_i)$. It follows from the standard M-estimation theory that the SLS estimator $\hat{\boldsymbol{\gamma}}_N$ has the following asymptotic properties under the regularity conditions given in Appendices A.

Theorem 1. *(1) Under Assumptions A1–A3, $\hat{\boldsymbol{\gamma}}_N \xrightarrow{a.s.} \boldsymbol{\gamma}_0$ as $N \to \infty$ and T is fixed.*

(2) Under Assumptions A1–A6, $\sqrt{N}(\hat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$ as $N \to \infty$ and T is fixed, where

$$\mathbf{A} = E\left\{ \frac{\partial h_1'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} \mathbf{W}_1 \frac{\partial h_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'} \right\}, \tag{6}$$

$$\mathbf{B} = E\left\{ \frac{\partial h_1'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} \mathbf{W}_1 h_1(\boldsymbol{\gamma}_0) h_1'(\boldsymbol{\gamma}_0) \mathbf{W}_1 \frac{\partial h_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'} \right\}. \tag{7}$$

3. A Diagnostic Test

The estimation approach in the previous section is based on the correct specification of the first two conditional moments of the response variables. In this section we propose a test for these moments, called SW for convenience. It is designed to test hypotheses

$$H_0 : E\{h_i(\boldsymbol{\gamma}_0)|y_{i0}, \mathbf{x}_i\} = \mathbf{0} \quad vs. \quad H_a : E\{h_i(\boldsymbol{\gamma}_0)|y_{i0}, \mathbf{x}_i\} \neq \mathbf{0}.$$

Specifically, let $h(\boldsymbol{\gamma}) = \frac{1}{N} \sum_{i=1}^{N} h_i(\boldsymbol{\gamma})$. Then the test statistic is defined as

$$\text{SW} = N\, h'(\hat{\boldsymbol{\gamma}}_N)\, \hat{\mathbf{G}}_N^{-1}\, h(\hat{\boldsymbol{\gamma}}_N), \tag{8}$$

where $\hat{\gamma}_N$ is defined by Equation (5) and $\hat{G}_N$ is given by

$$\hat{G}_N = \frac{1}{N}\sum_{i=1}^{N}\bar{P}_i(\hat{\gamma}_N)h_i(\hat{\gamma}_N)h_i'(\hat{\gamma}_N)\bar{P}_i'(\hat{\gamma}_N)$$

with

$$\bar{P}_i(\hat{\gamma}_N) = I - \bar{D}_N(\hat{\gamma}_N)A_N^{-1}\frac{\partial h_i'(\hat{\gamma}_N)}{\partial \gamma}W_i,$$

$$A_N = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial h_i'(\hat{\gamma}_N)}{\partial \gamma}W_i\frac{\partial h_i(\hat{\gamma}_N)}{\partial \gamma'},$$

and

$$\bar{D}_N(\hat{\gamma}_N) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial h_i(\hat{\gamma}_N)}{\partial \gamma'}.$$

The asymptotic distribution of the test statistic SW is given below.

Theorem 2. *Assume that $E\big\{P_i(\gamma_0)h_i(\gamma_0)h_i'(\gamma_0)P_i'(\gamma_0)\big\}$ has full rank $K = T(T+3)/2$. Then under Assumptions A1–A6 and H_0, SW $\xrightarrow{d} \chi_K^2$ as $N \to \infty$ and T is fixed.*

4. Optimal SLS Estimator

From Equations (6) and (7) we see that the asymptotic covariance matrix of $\hat{\gamma}_N$ depends on the weights W_i. It is therefore desirable to chose the optimal weight so that the asymptotic variance is minimized in the following sense.

Theorem 3. *Suppose $U_1 = E\big\{h_1(\gamma_0)h_1'(\gamma_0)\,|\,y_{10}, x_1\big\}$ is nonsingular with probability one, and Assumptions A2–A6 are satisfied with $W_1 = U_1^{-1}$. Then the SLS estimator $\hat{\gamma}_N^0$ obtained by taking $W_i = U_i^{-1}, i = 1, 2, ..., N$ has asymptotic covariance matrix*

$$A_0^{-1} = E^{-1}\left\{\frac{\partial h_i'(\gamma_0)}{\partial \gamma}U_i^{-1}\frac{\partial h_i(\gamma_0)}{\partial \gamma'}\right\}. \tag{9}$$

Furthermore, for any SLS estimator $\hat{\gamma}_N$, its asymptotic covariance matrix is such that $A^{-1}BA^{-1} - A_0^{-1}$ is non-negative definite.

However, since the optimal weight U_i^{-1} depends on γ_0 and conditional moments $E(\varepsilon_{it}^j|y_{i0}, x_i)$ and $E(\eta_i^j|y_{i0}, x_i), j = 3, 4$, the optimal SLS estimator $\hat{\gamma}_N^0$ is not feasible. The corresponding feasible SLS estimator can be calculated by plugging in consistent estimators of these unknown quantities in U_i^{-1}. This feasible optimal SLS (FOSLS) estimator is consistent under condition (A1) in Appendix A. Moreover, the FOSLS has the same asymptotic covariance matrix as the (infeasible) optimal SLS estimator (Newey and McFadden 1994, Th. 6.1).

We also suggest another version of the FOSLS, called FOSLS1, which may be more robust to any possible stochastic dependence between ε_{it} and η_i, and does not require initial estimates for the third and fourth conditional moments of ε_{it} and η_i. It is obtained by using the weight

$$\widehat{W}_i = C'\left(y_{i0}, x_i, \hat{\theta}_N^0, \hat{\alpha}_N^0\right)\left(\frac{1}{N}\sum_{i=1}^{N}h_i^*(\hat{\gamma}_N^0)h_i^{*'}(\hat{\gamma}_N^0)\right)^{-}C\left(y_{i0}, x_i, \hat{\theta}_N^0, \hat{\alpha}_N^0\right), \tag{10}$$

where $\hat{\gamma}_N^0$ is a preliminary consistent estimator of γ_0, and $C(y_{i0}, x_i, \theta, \alpha)$ is a transformation matrix that maps $h_i(\gamma)$ into

$$h_i^*(\gamma) = \left(u_{it}^*, 1 \le t \le T,\ u_{it}^*u_{is}^* - v_{its}^*(\sigma^2, \theta), 1 \le s \le t \le T\right)', \tag{11}$$

where $u_{it}^* = y_{it} - \alpha y_{it-1} - \boldsymbol{\beta}' \boldsymbol{x}_{it} - f_1(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta})$ and $v_{its}^*(\sigma^2, \boldsymbol{\theta}) = f_2(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}) - f_1^2(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}) + \sigma^2 \mathbf{1}_{\{s=t\}}$. It can be shown that the FOSLS1 has the same asymptotic covariance matrix as the optimal SLS estimator, if ε_{it} and η_i have constant second to fourth conditional central moments given the process initial value and covariate. It is worthwhile to note that the matrix C and vectors $\boldsymbol{h}_i^*(\boldsymbol{\gamma})$ do not involve the moments in (3) and (4) which are calculated based on the reduced form (2). Therefore it is straightforward to generalize them in $AR(p)$ models.

Some researchers have studied the semiparametric efficient estimators for various dynamic panel data models. Chamberlain (1992) derived the optimal instrumental variables for the first-difference equation of model (1) under the sequential conditional moment restrictions and showed that the GMM estimator based on these optimal instrumental variables attains the semiparametric efficiency lower bound. Hahn (1997) showed that the GMM estimator based on an increasing set of instruments as the sample size grows would achieve the semiparametric efficiency bound of Chamberlain (1992). More recently, Park et al. (2007) used the geometric approach of Bickel et al. (1993) to construct a semiparametric efficient estimator under the stationary model with normal error distribution.

A natural question is whether the optimal SLS estimator efficiently uses the information in conditional moments (3) and (4). By Lemma 2 of Chamberlain (1987) the minimum bound of the asymptotic variance under $E\{\boldsymbol{h}_i(\boldsymbol{\gamma}_0)|y_{i0}, \boldsymbol{x}_i\} = \boldsymbol{0}$ is given by A_0^{-1} in (9). Therefore by Theorem 3 the optimal SLS estimator attains Chamberlain's bound of variance and is a semiparametric optimal estimator in this sense.

We conclude this section by comparing the asymptotic variance of the optimal SLS estimator with that of the RML estimator which is identical to the MLE conditional on the initial observation when the errors are normally distributed. Theoretically, the optimal SLSE is at least as efficient as the RMLE by (A14)–(A16) in Appendix F. Unfortunately it is difficult to evaluate the efficiency gain of the optimal SLSE analytically, hence we compare the asymptotic variances of the two estimators numerically. We considered a large number of scenarios with various data generating processes (stationary or nonstationary) and distributions for the error and unobserved random effects.

The percentage gain of efficiency in estimating α_0 as a function of T and α_0 is shown in Figure 1, where z-axis represents the percentage reduction in the variance of RMLE(α_0) by using the optimal SLSE(α_0). Our simulation results show convincingly that the asymptotic variance of the optimal SLSE(α_0) is strictly less than that of the RMLE(α_0) except for the case $\mu_{3(\varepsilon)} = 0$ and $\mu_{4(\varepsilon)} = 3\sigma_0^4$ (which is true under normal distribution), in which case both estimators have the same asymptotic variances.

(a)   (b)
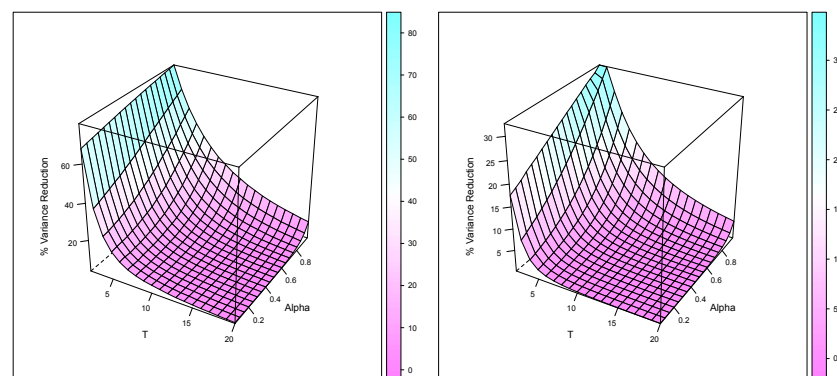
Figure 1. *Cont.*
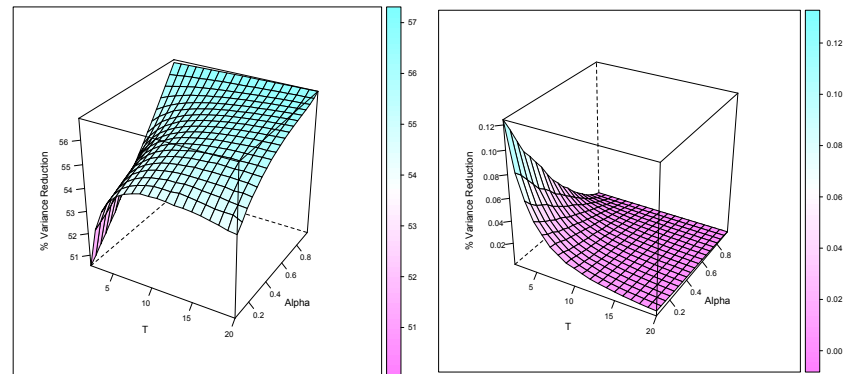
(c)   **(d)**

Figure 1. Reduction (%) in the variance of RML(α_0) gained by the optimal SLS(α_0). (**a**) Stationary, $\varepsilon_{it}|\eta_i, y_{i0} \sim \chi^2_{(1)}$. (**b**) Stationary, $\varepsilon_{it}|\eta_i, y_{i0} \sim t_{(5)}$. (**c**) Nonstationary, $\varepsilon_{it}|\eta_i, y_{i0} \sim \chi^2_{(1)}$. (**d**) Nonstationary, $\varepsilon_{it}|\eta_i, y_{i0} \sim t_{(5)}$.

5. Monte Carlo Simulation Studies

In this section we carry out extensive Monte Carlo simulations to examine the finite sample properties of FOSLS and FOSLS1 estimators, and compare them with some other popular estimators in the literature, including the linear first differenced GMM (Arellano and Bond 1991), the marginal pseudo maximum likelihood (MPML) (Arellano 2003) and the random effects pseudo maximum likelihood (RML) (Alvarez and Arellano 2003) estimators. For the sake of comparison we adopt the commonly used model setup in these literature, i.e., the model (1) with no covariates x_{it} but the following specification

$$E(\eta_i|y_{i0}) = \theta_{01} + \theta_{02}y_{i0}, \quad V(\eta_i|y_{i0}) = \exp(\theta_{03}), \tag{12}$$

$$E(\varepsilon^j_{it}|y_{i0}) = \mu_{j(\varepsilon)}, \quad E\left((\eta_i - \theta_{01} - \theta_{02}y_{i0})^j|y_{i0}\right) = \mu_{j(\eta)}, \quad j = 3, 4. \tag{13}$$

This model specification is also used by Blundell and Bond (1998) to derive conditional-type estimators such as the CGLS, see also Okui (2009).

5.1. Finite Sample Properties and Comparisons

Specifically, the data are generated according to $y_{i0} \sim$ i.i.d. $N(0, 2/(1 - \alpha_0^2)(1 - \alpha_0))$, $\eta_i|y_{i0} \sim \theta_{01} + \theta_{02}y_{i0} + \exp(\theta_{03})F_1$ and $\varepsilon_{it}|\eta_i, y_{i0} \sim \sigma_0^2 F_2$ with $\theta_{01} = 0$, $\theta_{02} = c(1 - \alpha_0^2)/2$ and $\theta_{03} = \log((1 - \alpha_0)/2)$, where c, F_1, F_2 are chosen as follows to be close to the setups in the literature.

Normal stationary process: $c = 1$ and $F_1, F_2 \sim N(0, 1)$. Under this setup the MPML estimator is the true MLE computed using all data including the initial observations.

Nonnormal stationary process: $c = 1$ and $F_1, F_2 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$.

Normal nonstationary process: $c = 20$ and $F_1, F_2 \sim N(0, 1)$. Under this setup the y process is nonstationary in first two moments and the MPML estimator is inconsistent.

Nonnormal nonstationary process: $c = 20$ and
(a) $F_1 \sim N(0, 1), F_2 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$;   (b) $F_1 \sim N(0, 1), F_2 \sim \sqrt{3/5}\, t_{(5)}$;
(c) $F_2 \sim N(0, 1), F_1 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$;   (d) $F_2 \sim N(0, 1), F_1 \sim \sqrt{3/5}\, t_{(5)}$.

In all scenarios the parameter values are $\alpha_0 = 0.2, 0.5, 0.8$ and the sample sizes are $N = 30, 300$ and $T = 5, 10, 15$. In each simulation 1000 Monte Carlo replications are done and the median estimates and median absolute deviation (MAD) are calculated to assess the performance of the estimators. To compute the FOSLS estimators, we use the RML to obtain preliminary consistent estimates of γ_0, $\mu_{j(\varepsilon)}$ and $\mu_{j(\eta)}, j = 3, 4$, and then plug

them in the optimal weight U_i^{-1}. As mentioned earlier this does not affect the asymptotic properties of the second-step estimators.

To save space we only present the numerical results for the nonnormal-nonstationary (a) scenario in Table 1, where we use the MAD of the FOSLS estimator as reference and report the relative MADs for all other estimators. For the other scenarios we provide a summary and discussion of the results below.

Table 1. The MAD of FOSLS and relative MAD (bold font) of all other estimators, and their Median (regular font) for nonnormal-nonstationary scenario.

	$\alpha_0 = 0.2$					$\alpha_0 = 0.5$					$\alpha_0 = 0.8$				
	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML
N								$T = 5$							
30	**0.006**	**1.07**	**1.29**	**1.29**	**120.57**	**0.003**	**1.07**	**1.44**	**1.44**	**124.24**	**0.002**	**1.18**	**1.56**	**1.54**	**89.56**
	0.200	0.20	0.20	0.20	0.90	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
300	**0.002**	**0.93**	**1.44**	**1.44**	**394.89**	**0.001**	**0.95**	**1.48**	**1.47**	**413.21**	**0.001**	**0.89**	**1.44**	**1.44**	**255.54**
	0.200	0.20	0.20	0.20	0.91	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
								$T = 10$							
30	**0.005**	**1.26**	**1.38**	**1.34**	**146.48**	**0.003**	**1.32**	**1.55**	**1.52**	**191.90**	**0.001**	**1.16**	**1.46**	**1.45**	**173.13**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
300	**0.001**	**0.95**	**1.44**	**1.47**	**468.00**	**0.001**	**0.95**	**1.47**	**1.46**	**554.61**	**0.000**	**1.01**	**1.53**	**1.53**	**590.03**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
								$T = 15$							
30	**0.005**	**2.00**	**1.37**	**1.40**	**156.16**	**0.003**	**2.38**	**1.43**	**1.42**	**193.59**	**0.001**	**2.18**	**1.53**	**1.54**	**239.64**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99
300	**0.001**	**1.06**	**1.50**	**1.52**	**497.46**	**0.001**	**1.10**	**1.56**	**1.57**	**641.36**	**0.000**	**1.18**	**1.59**	**1.60**	**748.73**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99

In the **normal-stationary** scenario, the MAD of the FOSLS is decreasing in T and nondecreasing in α_0. That the FOSLS and RML have equal asymptotic variance under the normal error components (see (A16)) appears clearly for $\alpha_0 = 0.2, 0.5$, while larger values of N are required to see this fact for larger α_0. As expected the MPML has the smallest MAD in almost all cases because it is the most efficient estimator under normality. However, the gap between the MPML and FOSLS is getting smaller as α_0 decreases or T increases. The first-difference GMM is generally inferior and the problem of weak instruments appears clearly for large α_0, which is consistent with Blundell and Bond (1998). The FOSLS1 is not reliable for small N because the numerical calculation of its weight matrix is unstable when $N < T(T+3)/2$. Moreover, the relative MADs of the FOSLS1 for $N = 300$ show that the estimator has a slower convergence rate than the FOSLS, which was expected. The downward bias in FOSLS vanishes quickly as T increases.

In the **nonnormal-stationary** scenario, the results shows wide outperformance of the FOSLS, specially for small T. The relative MADs of RML reveal that the true levels of variance reduction gained by FOSLS (see Figure 1a) require N to be larger than 300. The FOSLS competes well with the MPML for small N. Although the FOSLS1 is not reliable for small N/T ratio, it performs well for $N = 300$. The FOSLS has smaller bias for small N than other estimators.

In the **normal-nonstationary** scenario, the FOSLS, RML, and GMM compete very well for small and large N. The close performance of the FOSLS and RML is due to the normality of ε_{it}. The improvement in the GMM performance is due to the nonstationarity of y_{it} process. The results show clearly how the MPML breaks down everywhere under this scenario, demonstrating the consequence of misspecifying a nonstationary process. Again, the FOSLS1 requires large N/T ratio to get stable so it is not recommended in this scenario.

Finally, the results of **nonnormal-nonstationary** scenario presented in Table 1 show the effect of the skewness of the ε_{it} distribution on the performance of the FOSLS. The RML and GMM are less efficient than the FOSLS by at least 30% for $N = 30$, and by as high as 59% for $N = 300$. The relative MADs of the RML for large N are consistent with Figure 1c. Although the true levels of variance reduction gained by FOSLS require N to be larger than 300, the gain of efficiency for small N is much larger than the corresponding gain in the nonnormal-stationary scenario. The numerical results of the three remaining

cases (b,c,d) in this scenario clearly show the outperformance of the FOSLS(α_0) when ε_{it} has nonnormal distribution.

We conclude this subsection by examining to what extent the asymptotic variance formula for the OSLS in (9) can be used to approximate its finite sample counterpart. To this end we use the asymptotic formula to calculate the standard deviation of the OSLS estimator for α_0 and compare with the sample RMSE calculated through Monte Carlo simulations for various values of N, T, and α_0. Table 2 shows that both results are fairly close and therefore the asymptotic formula can be safely used in further inference about α_0 such as confidence intervals or testing hypothesis.

Table 2. The RMSE of FOSLS calculated by simulation and asymptotic formula (9) for nonnormal-nonstationary scenario.

	$\alpha_0 = 0.2$		$\alpha_0 = 0.5$		$\alpha_0 = 0.8$	
	Simulation	**Asy Formula**	**Simulation**	**Asy Formula**	**Simulation**	**Asy Formula**
N			$T = 5$			
30	0.00866	0.00815	0.00526	0.00501	0.00268	0.00288
300	0.00265	0.00258	0.00159	0.00158	0.00094	0.00091
			$T = 15$			
30	0.00729	0.00703	0.00369	0.00364	0.00119	0.00122
300	0.00229	0.00222	0.00112	0.00115	0.00038	0.00039

5.2. Robustness of FOSLS against Unit-Root

It is well known in the literature that linear first-difference GMM is generally inferior when the autoregressive parameter α_0 is close to one due to the problem of weak instruments, see e.g., Blundell and Bond (1998). On the other hand, our simulation results in the previous subsection show that the performance of FOSLS is not affected by large values of α_0. Furthermore, it becomes closer to the MPML and sometimes is even better due to the efficiency gain in the case of nonnormal errors. Hence it is interesting to investigate the performance of FOSLS when α_0 is very close to one (unit root case) and see if it breaks down like the GMM or it is stable like the MPML.

To answer these questions we carry out a simulation using the nonnormal-stationary model defined before. The results in Table 3 show that the FOSLS is less biased and more efficient than the RML due to the skewness in the within group errors. Although the FOSLS has larger downward bias than the MPML, the difference vanishes quickly as T and α_0 increase. These results demonstrate that the FOSLS is robust in the case of near unit root and is even more efficient than the MPML for $T = 6$ and $\alpha_0 = 0.99$. This may be due to nonnormality.

Table 3. The RMSE of FOSLS and relative RMSE (bold font) of all other estimators, and their Mean (regular font) for nonnormal near unit root scenario.

	$\alpha_0 = 0.9$				$\alpha_0 = 0.95$				$\alpha_0 = 0.99$			
	FOSLS	**RML**	**GMM**	**MPML**	**FOSLS**	**RML**	**GMM**	**MPML**	**FOSLS**	**RML**	**GMM**	**MPML**
N							$T = 3$					
100	**0.216**	**1.23**	**4.94**	**0.45**	**0.240**	**1.13**	**5.04**	**0.38**	**0.123**	**2.36**	**10.28**	**0.66**
	0.82	0.71	0.30	0.88	0.87	0.75	0.12	0.92	0.95	0.77	0.08	0.95
300	**0.151**	**1.21**	**5.85**	**0.43**	**0.130**	**1.52**	**8.32**	**0.44**	**0.060**	**3.59**	**21.85**	**0.81**
	0.83	0.78	0.48	0.90	0.89	0.81	0.27	0.94	0.97	0.83	0.09	0.97
							$T = 6$					
100	**0.086**	**1.27**	**5.27**	**0.67**	**0.094**	**1.32**	**6.21**	**0.49**	**0.037**	**3.75**	**17.10**	**1.07**
	0.87	0.83	0.52	0.90	0.91	0.86	0.43	0.94	0.98	0.88	0.43	0.97
300	**0.039**	**1.85**	**7.44**	**0.88**	**0.064**	**1.36**	**7.83**	**0.48**	**0.013**	**7.50**	**47.90**	**1.69**
	0.90	0.86	0.68	0.90	0.92	0.89	0.53	0.95	0.99	0.91	0.43	0.98

5.3. Performance in the Presence of Covariates

In the previous simulation studies we considered models without covariates in order to make our numerical results comparable with existing methods in the literature. Now we investigate if the finite sample performance of the FOSLS changes when the model includes covariates. A well-known work in this respect is Kiviet (1995) who compared some least squares and instrumental variable (IV) type estimators (among them the linear first-difference GMM of Arellano and Bond 1991).

Following Kiviet (1995, Section 5 and Appendix B) we consider model (1) with a covariate and y_t being generated from the reduced form equation (here we omit the subscript i) $y_t = \beta\phi_t + \psi_t + \eta/(1-\alpha)$, $t = 0, 1, ..., T$, where $\phi_t \sim AR(2)$ and $\psi_t \sim AR(1)$ are mutually independent stationary processes and both are independent of $\eta \sim N(0, \sigma_\eta^2)$. The orthogonality and normality assumptions of Kiviet (1995) imply $E(\eta|y_0, x_1, \ldots, x_T) = \theta_2 y_0 + \theta_4 x_1$ and $V(\eta|y_0, x_1, \ldots, x_T) = \exp(\theta_3)$ as in our notation in (12).

Kiviet (1995) reported simulation results for 14 different combinations (designs) of parameter values and sample sizes. We have simulated all the 14 designs but only present the results for four designs to save space. Table 4 contains the bias, standard deviation and root mean squared error of the FOSLS, RML, and linear first-difference GMM (Arellano and Bond 1991) for α and β. As expected, the RML (which is the conditional MLE) is the best in all cases because the data are generated using the normal distribution. However, the FOSLS (which is asymptotically as efficient as the RML) is very close to the RML. These results show that in the presence of strictly exogenous variable the FOSLS is a good alternative to the most efficient RML under normality. The calculated summary statistics of GMM are different than the corresponding values of GMM1 in Kiviet (1995), probably due to two errors in his equation B6 where ξ_0 and ξ_1 should be standardized. We have also done another simulation under nonnormal-nonstationary setup and obtained similar results as in Table 1. They also clearly show the outperformance of FOSLS compared to RML due to deviation from normality.

Table 4. Bias, standard deviation and RMSE of estimators for Kiviet (1995) model with a covariate, $N = 100$.

		Bias		Std		RMSE	
Design	**Estimator**	α	β	α	β	α	β
II	GMM	−0.029	0.002	0.388	0.130	0.396	0.131
	RML	−0.005	−0.006	0.106	0.093	0.107	0.094
	FOSLS	−0.005	−0.006	0.107	0.098	0.107	0.099
VI	GMM	−0.060	−0.026	0.304	0.065	0.313	0.066
	RML	−0.021	0.023	0.059	0.046	0.059	0.047
	FOSLS	−0.020	0.023	0.063	0.047	0.063	0.048
X	GMM	−0.074	−0.011	0.065	0.309	0.067	0.309
	RML	−0.004	0.009	0.048	0.182	0.048	0.183
	FSOLS	−0.005	0.009	0.048	0.183	0.048	0.184
XII	GMM	−0.038	−0.019	0.088	0.344	0.096	0.344
	RML	0.000	0.006	0.058	0.178	0.058	0.179
	FOSLS	−0.030	0.003	0.058	0.180	0.058	0.180

6. Application

In this section we use a real data example to demonstrate the practical usefulness of the SLS approach in comparison with the IV approach and to assess the practical gain of efficiency over the RML estimator. In particular, we use a data set published in Wooldridge (2010) and downloadable at http://mitpress.mit.edu/sites/default/files/titles/content/wooldridge/statafiles.zip (as of 28 April 2014). The dataset airfare.dta contains data on airfares, number of passengers, distance, and the market share of the largest carrier for each of the top 1149 city-pair markets within the contiguous 48 US states for the fourth quarters of 1997 through 2000. A detailed description of the data can be

found at http://academic.reed.edu/economics/parker/s10/312/Asgns/proj3.html (as of 28 April 2014).

The main factors that influence airfares are the flight distance (ldist), average number of passengers (lpassne), and market concentration (concen). All variables except concen are measured in logarithmic scale. However, while the first factor is time invariant exogenous variable, the other two factors are clearly endogenous because they are also influenced by the airfare. Since the airlines usually set the current airfare by adjusting the previous year's fare, a linear dynamic regression model with unobserved route heterogeneity and time dummies may be appropriate to measure the effect of the determinants. Indeed, from bivariate scatterplots one can clearly see that the linear dependence between the current and previous airfares, lfare and lfare1 respectively, and moderate positive linear association between lfare and ldist. However, although theoretically concen and lfare should be positively correlated, this is not clearly seen. A possible explanation is that the relation is masked by the negative correlation between concen and ldist, given the positive correlation between lfare and ldist.

In light of these considerations, it is reasonable to start with the following model which includes time dummies $D99$, $D00$ for years 1999, 2000, respectively:

$$
\begin{aligned}
lfare_{it} = {} & \alpha_0 + \alpha_1\, lfare_{i(t-1)} + \beta_1\, ldist_i + \beta_2\, concen_{it} + \beta_3\, concen_{i(t-1)} \\
& + \beta_4\, lpassen_{it} + \beta_5\, lpassen_{i(t-1)} + \beta_6 D99 + \beta_7\, D00 + \eta_i + \varepsilon_{it}.
\end{aligned}
\tag{14}
$$

We first calculate the naive ordinary least squares (OLS) estimates of the proposed Equation (14) which are shown in the OLS column of Table 5. Although the OLSE are not consistent due to the correlation between η_i and $lfare_{i(t-1)}$, the high value of $R^2 = 0.95$ reflects strong explanatory power of the regressors. So we refit the model using the two stage linear first differenced GMM (GMM2) of Arellano and Bond (1991) and the lags of order at least two of $(lfare, concen, lpassen)$ as instrumental variables to deal with the possible endogeneity of these variables. The estimates are calculated using STATA 13.0 and the results are shown in Table 5 under Model-I.

Wooldridge (2010, p. 373) used the GMM2 to fit a first order linear dynamic model with only concen (treated as strictly exogenous) and dummy variables and obtained the estimated autoregressive parameter 0.333 which is not far from ours. According to the reported value of Sargan test (see Table 5), the GMM sequential moments based on the reduced form of model (14) are not rejected. However, the estimated elasticity of lfare with respect to concen is negative (p-value $= 0.059$). A possible explanation for this unexpected sign is the problem of weak instruments which is likely to occur in identifying β_2 given the strong positive linear correlation between concen and concen1. This is similar to the situation where the first differenced GMM estimator is used to estimate the autoregressive parameter when it is close to one (Blundell and Bond 1998). This problem causes inflation in the coefficient variance and leads to unreliable estimates. This is an example where the RE approach with the level data is preferred over the FE approach with differenced data, because the former does not depend on the instrumental variables and hence is able to give more reliable estimates.

However, since the sequential moments are correctly specified in Model-I, the GMM estimates can still be used to recover the within group errors ε_{it} and route effect η_i. We follow Arellano (2003, p. 118–19) to estimate the time effect for year 1998 and subsequently η_i. Then realizations of ε_{it} are obtained directly from Equation (14). Since bivariate scatterplots of $\hat{\eta}_i$ and the initial values (1997) of the variables (lfare0, concen0, lpassen0, and ldsit0) show possible linear relationships between η_i and these initial values, we fit the following auxiliary equation where all coefficients are significant at 0.01 level of significance:

$$
\hat{\eta}_i = -5.507 + 0.641\, lfare_{i0} + 0.464\, concen_{i0} - 0.041\, ldist_{i0} + 0.286\, lpassen_{i0}.
\tag{15}
$$

Further, simple diagnostics show that $\hat{\varepsilon}_{it}$ have constant variance across time and no significant autocorrelation, hence the assumptions on η_i and ε_{it} seem reasonable for model (14). These assumptions will be implicitly tested by using our SW test latter.

Based on Equation (15) we use the 'lme' command in R to calculate the RML estimates of the following equation and the results are given in Table 5 under Model-I.

$$
\begin{aligned}
lfare_{it} =\ & \theta_0 + \alpha_1\, lfare_{i(t-1)} + \beta_1\, ldist_i + \beta_2\, concen_{it} + \beta_3\, concen_{i(t-1)} \\
& + \beta_4\, lpassen_{it} + \beta_5\, lpassen_{i(t-1)} + \beta_6 D99 + \beta_7\, D00 + \theta_1\, lfare_{i0} \\
& + \theta_2\, concen_{i0} + \theta_3\, lpassen_{i0} + \eta_i^* + \varepsilon_{it}.
\end{aligned}
\tag{16}
$$

A normal QQ plot of the corresponding residuals $\hat{\varepsilon}_{it}$ shows a flat tail symmetric distribution that can be approximated by a Student t distribution with df $= 5$, indicating the possibility of efficiency gain by applying the FOSLS estimator. Hence we use the sample skewness and kurtosis of $\hat{\varepsilon}_{it}$ and $\hat{\eta}_i^*$ to calculate the optimal weight matrix for the FOSLS. To save space we report only the estimated coefficients of the main factors in Table 5, where the standard errors of GMM2 are computed using the robust formula, while the standard errors of FOSLS and RML are computed using Formula (9) and (A14), respectively.

Table 5. The models fitted to airfare data with coefficient estimates in bold and standard errors in regular font. The standard errors of FOSLS and RML are computed using (9) and (A14), respectively.

Coef.	Model-I				Model-II			Model-III			Model-IV		
	OLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS
Const: θ_0	**0.153**	—	**0.256**	**−0.122**	—	**0.248**	**0.723**	—	**0.243**	**0.786**	—	**0.279**	**0.708**
	—	—	0.032	0.032	—	0.026	0.025	—	0.010	0.010	—	0.051	0.050
lfare1: α_1	**0.929**	**0.216**	**0.374**	**0.536**	**0.216**	**0.377**	**0.461**	**0.157**	**0.373**	**0.817**	**0.169**	**0.147**	**0.823**
	—	0.100	0.039	0.033	0.100	0.014	0.013	0.080	0.009	0.009	0.072	0.017	0.017
ldist: β_1	**0.031**	—	**0.049**	**0.066**	—	**0.051**	**0.091**	—	**0.052**	**−0.007**	—	**0.062**	**−0.019**
	—	—	0.004	0.004	—	0.003	0.003	—	0.001	0.001	—	0.006	0.006
concen: β_2	**0.119**	**−0.849**	**0.108**	**0.048**	**−0.849**	**0.091**	**0.162**	**−0.217**	**0.080**	**0.070**	**−0.842**	**0.095**	**0.077**
	—	0.449	0.018	0.018	0.449	0.011	0.011	0.167	0.003	0.003	0.389	0.020	0.019
concen1: β_3	**−0.076**	**0.325**	**0.013**	**0.040**	**0.325**	**−0.015**	**−0.032**	—	—	—	**0.320**	**0.000**	**−0.106**
	—	0.215	0.022	0.022	0.215	0.011	0.011	—	—	—	0.187	0.020	0.019
lpassen: β_4	**−0.374**	**−0.404**	**−0.367**	**−0.556**	**−0.404**	**−0.367**	**−0.108**	**−0.450**	**−0.367**	**−0.094**	**−0.382**	**−0.333**	**−0.084**
	—	0.081	0.006	0.006	0.081	0.004	0.004	0.057	0.002	0.002	0.060	0.007	0.007
lpassen1: β_5	**0.374**	**0.103**	**0.175**	**0.427**	**0.103**	**0.176**	**−0.014**	**0.106**	**0.175**	**0.005**	—	—	—
	—	0.141	0.021	0.019	0.141	0.006	0.006	0.111	0.003	0.003	—	—	—
D99: β_6	**0.000**	**0.003**	**0.016**	**0.009**	**0.003**	**0.016**	**0.037**	**0.020**	**0.016**	**−0.017**	**0.005**	**0.023**	**−0.014**
	—	0.013	0.003	0.003	0.013	0.002	0.002	0.006	0.001	0.001	0.012	0.003	0.003
D00: β_7	**0.042**	**0.070**	**0.074**	**0.060**	**0.070**	**0.073**	**0.092**	**0.087**	**0.073**	**−0.003**	**0.078**	**0.090**	**0.073**
	—	0.015	0.003	0.003	0.015	0.003	0.002	0.011	0.001	0.001	0.012	0.003	0.003
Diagnostic Statistics													
SW/Sargan		**6.41**		**20.98**	**6.41**		**4.01**	**13.90**		**335.80**	**8.16**		**8545.16**
p-value		0.268		0.013	0.268		0.911	0.031		0.000	0.226		0.000
RMSE	0.09	5.12	4.02	3.23	5.12	4.04	2.66	5.12	4.07	0.69	5.12	6.01	0.74

In Model-I, while the Sargan test does not reject the GMM sequential moments (p-value $= 0.268$), our SW test shows fairly strong evidence against the first and second moment specification given by Equations (3) and (4) (p-value $= 0.013$). This motivates us to check the reliability of SW test by examining its empirical sampling distribution under (3) and (4) as follows. First, we use the RML estimates of Model-I to generate data from (16), and draw η_i^* and ε_{it} from Student t distribution with df $= 6$ and 5 respectively to be as close as possible to the estimated residuals and random effect of Model-I. Second, we use the generated data to fit (16) using the RML followed by FOSLS and then calculate the SW statistic. We repeat these two steps 1000 times to obtain an approximation to the sampling distribution of SW statistics under H_0. Figure 2 confirms that a sample size of 1149 routes is sufficient for the empirical sampling distribution to be close to the asymptotic one χ_9^2

under H_0. Consequently Model-I needs to be modified by adding or eliminating some variables to pass the SW test.

To proceed, we fit (16) again without concen0 because it is not of main interest and has weak correlation with $\hat{\eta}_i$. This leads to Model-II which contains the same explanatory variables as Model-I and of course the same value of Sargan test (6.41). On the other hand, the value of SW statistic drops to 4.01 which is insignificant at any reasonable level of significance. This suggests that Model-II should be used for testing purpose provided that the signs of FOSLS estimates are consistent with the economic theory. The robust standard errors of the FOSLS are computed by Theorem 1 and are reported in Table 6, according to which two variables concen1 and lpassen1 are the candidates to be dropped from the model.
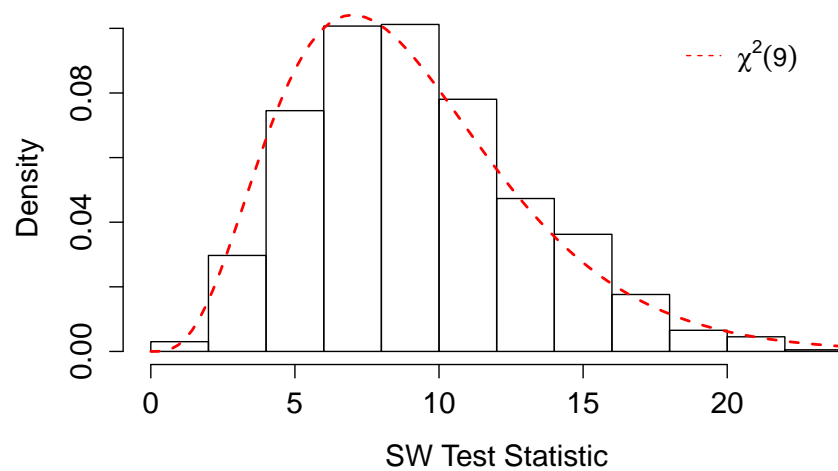
Figure 2. The empirical vs. asymptotic sampling distribution of SW by 1000 data cloning.

Table 6. Model-II robust standard errors of FOSLS calculated by Theorem 1.

θ_0 Const.	α_1 lfare1	β_1 ldist	β_2 concen	β_3 concen1	β_4 lpassen	β_5 lpassen1	β_6 D99	β_7 D00
0.061	0.041	0.007	0.034	0.034	0.026	0.023	0.004	0.005

Model-III is obtained by dropping concen1 only from Model-II, but this new specification is rejected by both Sargan and SW test. Consequently Model-IV is obtained by dropping lpassen1 only from Model-II. Interestingly, while SW test rejects Model-IV specification, Sargan test doesn't reject the GMM sequential moments of this model. However, the GMM estimate of β_2 is negative, which contradicts the economic theory. It follows that Model-IV is not desirable and instead Model-II is preferred. It does include both concen1 and lpassen1 although they may be insignificant according to the robust standard errors. On the other hand, if the standard errors in Table 5 are used, then both concen1 and lpassen1 are significant according to FOSLS and only lpassen1 is according to RML. Since Model-III is rejected by SW test, it follows that using the information inherent in the fourth moment through FOSLS is effective in keeping the variable concen1 in Model-II.

7. Conclusions and Discussion

We studied the SLS approach as an alternative to the commonly used random effects ML (RML) or differenced-GMM estimation for linear dynamic panel data models. Our approach is based on the first two conditional moments of the outcome process and does not postulate any distributional assumptions on the error components in the model. The asymptotic and finite sample properties and practical merits of the proposed estimators are thoroughly investigated.

This research reveals the following interesting new findings. First, differencing data and using instrumental variables may cause substantial loss of information and produce misleading relations, specially when the time varying variables are generated from autoregressive process with high autocorrelation (which is common in economic data). In such a case the linear first-differenced GMM or other similar methods not only are unable to estimate the effect of the time invariant variables, but also, more harmfully, weakly identify the effect of the time variant explanatory variables. In contrast the SLS approach makes use of the information inherent in the level data and therefore can improve the estimation precision and the goodness of fit considerably.

Second, the information in the sample skewness and kurtosis of the within group residuals can be utilized by our OSLS to gain more efficiency over the RML and consequently save important explanatory variables from being wrongly eliminated. In other words, by using the extra efficiency of the OSLS one can avoid falling into misspecification traps. Third, our newly proposed diagnostic test proved to be very useful not only in validating the working conditional moments but also for model selection purpose, while usual goodness of fit criteria such as RMSE, R^2 or AIC may be misleading when the model is incorrectly specified.

We have explicitly dealt with an $AR(1)$ model for the simplicity of notation. However, our approach can be extended to more general $AR(p)$ models. For example, It is straightforward to calculate the FOSLS1 in Section 4 using a generalized form of the transformation matrix C and the deviation form of the moments as in (10) and (11).

This work also raises some interesting points for future research. Although our diagnostic test provides a systematic built-in tool to validate the model conditional moments assumption, in practice it would be interesting to run some sensitivity analysis to investigate the consequences of possible non-rejected deviations from this assumption. It would also be interesting to study the properties of the SW test and extend it to check the working assumption on the third and fourth moments. Last but not least, other general methods such as GMM or equivalently estimating equations applied on level data could have similar asymptotic properties as our approach. However, it is important and worthwhile to compare their finite sample properties and practical implications in real data analysis. The current work provides a complete set of tools for the inference in linear dynamic models with level data which has not been studied so far in the literature.

Appendix A. Regularity Conditions

To establish the asymptotic properties of the SLS estimator $\hat{\gamma}_N$, the following regularity conditions are assumed, where $\|\cdot\|$ denotes the Euclidean norm.

Assumption A1. *The parameter space* $\mathbf{\Gamma}$ *is a compact subset of* $\mathbb{R}^{p+\ell+2}$.

Assumption A2. $f_j(y_{i0}, x_i, \theta), j = 1, 2$ *are Borel measurable functions of* (y_{i0}, x_i) *for each* θ *in the corresponding parameter space* Θ, *and are continuous functions of* θ *with probability one. Furthermore, for all* t,

$$E\|\mathbf{W}_1\| \left(y_{10}^4 + \sup_{\Theta} f_2^2(y_{10}, x_1, \theta) + \|x_{1t}\|^4 + \eta_1^4 + \varepsilon_{1t}^4 + 1 \right) < \infty.$$

Assumption A3. $E[h_1(\gamma) - h_1(\gamma_0)]' \mathbf{W}_1[h_1(\gamma) - h_1(\gamma_0)] = 0$ *if and only if* $\gamma = \gamma_0$.

Assumption A4. *With probability one,* $f_j(y_{i0}, x_i, \theta), j = 1, 2$ *are twice continuously differentiable in* $int(\Theta)$. *Furthermore, for* $j = 1, 2$,

$$E\|\mathbf{W}_1\|(y_{10}^2 + \|x_1\|^2 + 1)^{2-j} \sup_{\mathcal{N}(\theta_0)} \left(\left\| \frac{\partial f_j(y_{10}, x_1, \theta)}{\partial \theta} \right\|^2 + \left\| \frac{\partial^2 f_j(y_{10}, x_1, \theta)}{\partial \theta \partial \theta'} \right\|^2 \right) < \infty,$$

where $\mathcal{N}(\theta_0) \subset int(\Theta)$ *is a closed neighborhood of* θ_0.

Assumption A5. *The matrix*

$$A = E\left\{ \frac{\partial h_1'(\gamma_0)}{\partial \gamma} \mathbf{W}_1 \frac{\partial h_1(\gamma_0)}{\partial \gamma'} \right\}$$

in (6) *is nonsingular.*

Assumption A6. *It holds*

$$E\|\mathbf{W}_1\|^2 \left(y_{10}^8 + \|x_1\|^8 + f_2^4(y_{10}, \theta_0) + \eta_1^8 + \varepsilon_{1t}^8 \right.$$

$$\left. + \left(y_{10}^4 + \|x_1\|^4 + 1 \right) \left\| \frac{\partial f_1(y_{10}, x_1, \theta_0)}{\partial \theta} \right\|^4 + \left\| \frac{\partial f_2(y_{10}, x_1, \theta_0)}{\partial \theta} \right\|^4 + 1 \right) < \infty.$$

Note that Assumptions A1–A6 are standard regularity conditions in the M-estimation literature. Assumption A3 is necessary and sufficient for parameter identification, while Assumptions A2, A4 and A6 are sufficient but not necessary.

Furthermore, for the FOSLSE to be consistent, it is sufficient that

$$E \sup_{\Gamma^*} \left\| \mathbf{U}_1^{-1}(\gamma^*) \right\| \left(y_{10}^4 + \|x_1\|^4 + \sup_{\Theta} f_2^2(y_{10}, x_1, \theta) + \eta_1^4 + \varepsilon_{1t}^4 + 1 \right) < \infty, \quad \text{(A1)}$$

where $\mathbf{U}_1^{-1}$ is as in Theorem 3, γ^* is the vector containing all generic parameters in $\mathbf{U}_1^{-1}$ including γ, and $\mathbf{\Gamma}^*$ is the corresponding compact parameter space.

Although Assumptions A2–A6 and condition (A1) look complicated for the sake of generality, they can be simplified by specifying the functional forms of $E(\eta_i^j | y_{i0}, x_i), j = 1, 2, 3, 4$ and $E(\varepsilon_{it}^j | y_{i0}, x_i), j = 3, 4$. For example, under (12)–(13), $\mathbf{U}_i^{-1}$ has a special structure so that assumptions A2, A4, A6 and condition (A1) are implied by $E(y_{i0}^4) < \infty$.

Appendix B. Proof of Theorem 1 (1)

For the simplicity of notation, we use f_j, y_t, $\tilde{x}_t$, a_t, c_{ts}, d_{ts}, w_{ts}, k_{ts} for $f_j(y_{10}, x_1, \theta)$, y_{1t}, $\tilde{x}_{1t}(\alpha)$, $a_t(\alpha)$, $c_{ts}(\alpha)$, $d_{ts}(\alpha)$, $w_{1ts}(\alpha)$, and $k_{1ts}(\alpha)$ respectively.

First, by Cauchy-Schwarz inequality we have

$$\|h_1(\gamma)\|^2 \le 2\sum_{t=1}^{T} y_t^2 + 8y_0^2 \sum_{t=1}^{T} \alpha^{2t} + 8f_1^2 \sum_{t=1}^{T} a_t^2 + 4\sum_{t=1}^{T} \beta' \tilde{x}_t \tilde{x}_t' \beta + 2\sum_{1 \le s \le t}^{T} y_t^2 y_s^2$$

$$+ 16y_0^4 \sum_{1 \le s \le t}^{T} \alpha^{2(t+s)} + 16f_2^2 \sum_{1 \le s \le t}^{T} a_t^2 a_s^2 + 16\sigma^4 c_{ts}^2 + 16y_0^2 f_1^2 \sum_{1 \le s \le t}^{T} d_{ts}^2$$

$$+ 8\sum_{1 \le s \le t}^{T} (\beta' \tilde{x}_t \tilde{x}_s' \beta)^2 + 16y_0^2 \sum_{1 \le s \le t}^{T} (\beta' w_{ts})^2 + 16f_1^2 \sum_{1 \le s \le t}^{T} (\beta' k_{ts})^2,$$

$E\|W_1\|(y_t y_s)^2 \le E\|W_1\|y_t^4 E\|W_1\|y_s^4$ and $E\|W_1\|y_0^2 f_1^2 \le E\|W_1\|y_0^4 E\|W_1\|f_2^2$. Therefore by Assumptions A1 and A2 we have $E\sup_{\Gamma} |q_1(\gamma)| \le E\|W_1\| \sup_{\Gamma} \|h_1(\gamma)\|^2 < \infty$. It follows from the uniform law of large numbers (ULLN Amemiya 1985; Jennrich 1969) that

$$\sup_{\Gamma} \left| \frac{1}{N} \sum_{i=1}^{N} q_i(\gamma) - Eq_1(\gamma) \right| \xrightarrow{a.s.} 0 \quad \text{as } N \to \infty \text{ and } T \text{ is fixed.} \tag{A2}$$

Second, since $Eq_1(\gamma) = Eh_1'(\gamma_0) W_1 h_1(\gamma_0) + 2E[h_1(\gamma) - h_1(\gamma_0)]' W_1 h_1(\gamma_0)$, we have $Eq_1(\gamma) \ge Eq_1(\gamma_0)$ and the equality holds if and only if $\gamma = \gamma_0$. Finally the result follows from (A2) and Lemma 1 of Wang and Leblanc (2008).

Appendix C. Proof of Theorem 1 (2)

First, by the mean value theorem for random functions (Jennrich 1969), Assumptions A1–A4 guarantee that

$$1_{\hat{\gamma}_N \in \mathcal{N}(\gamma_0)} \left[\sum_{i=1}^{N} s_i(\gamma_0) + \left(\sum_{i=1}^{N} \bar{H}_i \right)(\hat{\gamma}_N - \gamma_0) \right] = 0, \tag{A3}$$

where

$$s_i(\gamma) = \frac{\partial q_i(\gamma)}{\partial \gamma} = 2\frac{\partial h_i'(\gamma)}{\partial \gamma} W_i h_i(\gamma),$$

matrix $\bar{H}_i$ has rows given by $\dfrac{\partial^2 q_i(\bar{\gamma}_N^r)}{\partial \gamma_{(r)} \partial \gamma'}$, $r = 1, 2, \ldots, (p + \ell + 2)$, and $\bar{\gamma}_N^r$ are measurable mappings into $\mathcal{N}(\gamma_0)$ and lying on the segment joining $\hat{\gamma}_N$ and γ_0. By the triangle inequality we have

$$\left\| \frac{\partial^2 q_1(\gamma)}{\partial \gamma \partial \gamma'} \right\| \le 2 \left\| \frac{\partial h_1'(\gamma)}{\partial \gamma} W_1 \frac{\partial h_1(\gamma)}{\partial \gamma'} \right\| + 2 \left\| \left(\frac{\partial^2 h_1'(\gamma)}{\partial \gamma_{(i)} \partial \gamma_{(j)}} W_1 h_1(\gamma) \right)_{i,j} \right\|, \tag{A4}$$

and further by Cauchy-Schwarz inequality we have

$$E \sup_{\mathcal{N}(\gamma_0)} \left\| \frac{\partial h_1'(\gamma)}{\partial \gamma} W_1 \frac{\partial h_1(\gamma)}{\partial \gamma'} \right\| \le E\|W_1\| \sup_{\mathcal{N}(\gamma_0)} \left\| \frac{\partial h_1'(\gamma)}{\partial \gamma} \right\|^2$$

and the following inequalities for $1 \leq s \leq t \leq T$,

$$\left(\frac{\partial \mu_{1t}(\gamma)}{\partial \alpha}\right)^2 \leq 4t^2 \alpha^{2(t-1)} y_0^2 + 4\left(\frac{\partial a_t}{\partial \alpha}\right)^2 f_2 + 2\left(\beta' \frac{\partial \tilde{x}_t}{\partial \alpha}\right)^2,$$

$$\left(\frac{\partial \nu_{1ts}(\gamma)}{\partial \alpha}\right)^2 \leq 8(t+s)^2 \alpha^{2(t+s-1)} y_0^4 + 8f_2^2\left(a_s \frac{\partial a_t}{\partial \alpha} + a_t \frac{\partial a_s}{\partial \alpha}\right)^2 + 8\sigma^4\left(\frac{\partial c_{ts}}{\partial \alpha}\right)^2$$

$$+ 8y_0^2 f_2\left(\frac{\partial d_{ts}}{\partial \alpha}\right)^2 + 8y_0^2\left(\beta' \frac{\partial w_{ts}}{\partial \alpha}\right)^2 + 8f_2\left(\beta' \frac{\partial k_{ts}}{\partial \alpha}\right)^2$$

$$+ 8\left(\beta' \frac{\partial \tilde{x}_t}{\partial \alpha}\right)^2 (\beta' \tilde{x}_s)^2 + 8\left(\beta' \frac{\partial \tilde{x}_s}{\partial \alpha}\right)^2 (\beta' \tilde{x}_t)^2,$$

$$\left\|\frac{\partial \nu_{1ts}(\gamma)}{\partial \theta}\right\|^2 \leq 4a_t^2 a_s^2 \left\|\frac{\partial f_2}{\partial \theta}\right\|^2 + 4d_{ts}^2 y_0^2 \left\|\frac{\partial f_1}{\partial \theta}\right\|^2 + 2(\beta' k_{ts})^2 \left\|\frac{\partial f_1}{\partial \theta}\right\|^2,$$

$$\left\|\frac{\partial \nu_{1ts}(\gamma)}{\partial \beta}\right\|^2 \leq 4\left\|\tilde{x}_t \tilde{x}_s'\right\|^2 \|\beta\|^2 + 4y_0^2 \|w_{ts}\|^2 + 4f_2 \|k_{ts}\|^2.$$

Hence it follows from Assumptions A2 and A4 that

$$E\|W_1\| \sup_{\mathcal{N}(\gamma_0)} \left\|\frac{\partial h_1'(\gamma)}{\partial \gamma}\right\|^2 < \infty. \tag{A5}$$

Similarly by the Cauchy-Schwarz inequality and Assumptions A2 and A4 we have

$$E \sup_{\mathcal{N}(\gamma_0)} \left\|\left(\frac{\partial^2 h_1'(\gamma)}{\partial \gamma_{(i)} \partial \gamma_{(j)}} W_1 h_1(\gamma)\right)_{i,j}\right\|$$

$$\leq E\|W_1\| \sup_{\mathcal{N}(\gamma_0)} \|h_1(\gamma)\|^2 E\|W_1\| \sup_{\mathcal{N}(\gamma_0)} \left(\sum_{i,j}^{p+\ell+2} \left\|\frac{\partial^2 h_1'(\gamma)}{\partial \gamma_{(i)} \partial \gamma_{(j)}}\right\|^2\right) < \infty. \tag{A6}$$

Combining inequalities (A4), (A5), (A6), Assumptions A2, A4 and the ULLN we have

$$\sup_{\mathcal{N}(\gamma_0)} \left\|\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 q_i(\gamma)}{\partial \gamma \partial \gamma'} - E \frac{\partial^2 q_1(\gamma)}{\partial \gamma \partial \gamma'}\right\| \xrightarrow{a.s.} 0 \quad \text{as} \quad N \to \infty \quad \text{for fixed } T. \tag{A7}$$

It follows from Lemma 2 of Wang and Leblanc (2008) and the strong consistency of $\hat{\gamma}_N$ that

$$\frac{1}{N} \sum_{i=1}^N \bar{H}_i \xrightarrow{a.s.} E \frac{\partial^2 q_1(\gamma_0)}{\partial \gamma \partial \gamma'} = 2A \quad \text{as} \quad N \to \infty \quad \text{for fixed } T. \tag{A8}$$

Further, by Assumption A6 and the central limit theorem (CLT) we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N s_i(\gamma_0) \xrightarrow{d} N(0, 4B) \quad \text{as} \quad N \to \infty \quad \text{for fixed } T,$$

where B is given in (7). Hence by Slutzky theorem and (A3), (A8) we have, for fixed T,

$$\sqrt{N}(\hat{\gamma}_N - \gamma_0) = -(2A)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N s_i(\gamma_0) + o_p(1). \tag{A9}$$

Appendix D. Proof of Theorem 2

First, by Theorem 1, Assumption A2 and the mean value theorem, for sufficiently large N,

$$h(\hat{\gamma}_N) = \frac{1}{N} \sum_{i=1}^{N} h_i(\gamma_0) + \frac{\partial h(\tilde{\gamma}_N)}{\partial \gamma'} (\hat{\gamma}_N - \gamma_0), \tag{A10}$$

where $\|\tilde{\gamma}_N - \gamma_0\| \leq \|\hat{\gamma}_N - \gamma_0\|$. By (A3), for sufficiently large N such that $\hat{\gamma}_N \in \mathcal{N}(\gamma_0)$,

$$\left(\sum_{i=1}^{N} \bar{H}_i \right) (\hat{\gamma}_N - \gamma_0) = -2 \sum_{i=1}^{N} \frac{\partial h_i'(\gamma_0)}{\partial \gamma} W_i h_i(\gamma_0). \tag{A11}$$

Further, by (A8),

$$\frac{1}{N} \sum_{i=1}^{N} \bar{H}_i = 2A + o_p(1) \tag{A12}$$

and analog to the proof of (A8), by Assumptions A2, A4 and the ULLN we can verify that

$$\frac{\partial h(\tilde{\gamma}_N)}{\partial \gamma'} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial h_i(\tilde{\gamma}_N)}{\partial \gamma'} = E\left\{ \frac{\partial h_1(\gamma_0)}{\partial \gamma'} \right\} + o_p(1). \tag{A13}$$

Combining (A10)–(A13) we obtain

$$\begin{aligned}
\sqrt{N} \, h(\hat{\gamma}_N) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(\gamma_0) + E\left\{ \frac{\partial h_1(\gamma_0)}{\partial \gamma'} \right\} \sqrt{N}(\hat{\gamma}_N - \gamma_0) + o_p(1) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} P_i(\gamma_0) h_i(\gamma_0) + o_p(1), \quad \text{[see Equation (A9)]}
\end{aligned}$$

where

$$P_i(\gamma_0) = I - E\left\{ \frac{\partial h_1(\gamma_0)}{\partial \gamma'} \right\} A^{-1} \frac{\partial h_i'(\gamma_0)}{\partial \gamma} W_i.$$

Since $E\{P_i(\gamma_0) h_i(\gamma_0) h_i'(\gamma_0) P_i'(\gamma_0)\}$ is nonsingular, the result follows from the CLT.

Appendix E. Proof of Theorem 3

Let

$$R = U_1^{1/2} W_1 \frac{\partial h_1(\gamma_0)}{\partial \gamma'}, \qquad Q = U_1^{-1/2} \frac{\partial h_1(\gamma_0)}{\partial \gamma'}.$$

Then by matrix form of Cauchy-Schwartz inequality it is straightforward to see that $E(R'R) - E(R'Q)E^{-1}(Q'Q)E(Q'R)$ is nonnegative definite and is zero matrix if $W_1 = U_1^{-1}$.

Appendix F. Asymptotic Variance of the RMLE

Using the GMM setup the asymptotic variance of the RML estimator under model (1) is

$$F_0 = E^{-1}(K'VK) E(K'VMVK) E^{-1}(K'VK), \tag{A14}$$

where

$$K' = \left(\begin{array}{cc} \dfrac{\partial \mu'(\gamma_0)}{\partial \gamma} & \dfrac{\partial vech'(S(\gamma_0))}{\partial \gamma} \end{array} \right),$$

$$V = \left(\begin{array}{cc} S(\gamma_0)^{-1} & 0 \\ 0 & \frac{1}{2} L' [S(\gamma_0)^{-1} \otimes S(\gamma_0)^{-1}] L \end{array} \right),$$

$$M = \left(\begin{array}{cc} S(\gamma_0) & E\{u_1 vech'(u_1 u_1')|y_{10}, x_1\} \\ . & E\{vech(u_1 u_1') vech'(u_1 u_1')|y_{10}, x_1\} - vech(S(\gamma_0)) vech'(S(\gamma_0)) \end{array} \right),$$

with $\boldsymbol{\mu}'(\boldsymbol{\gamma}) = (\mu_{1t}(\boldsymbol{\gamma}), 1 \le t \le T)$, $vech'(\boldsymbol{S}(\boldsymbol{\gamma})) = (a_t a_s(f_2 - f_1^2) + \sigma^2 c_{ts}, 1 \le s \le t \le T)$ and $\boldsymbol{u}_1 = (y_{1t} - \mu_{1t}(\boldsymbol{\gamma}_0), 1 \le t \le T)'$. Here *vec* and *vech* are the usual operators that stack the columns and lower triangle columns respectively of a matrix to form a vector, and L is the so-called selection matrix such as $vec(\boldsymbol{S}(\boldsymbol{\gamma})) = \boldsymbol{L}\, vech(\boldsymbol{S}(\boldsymbol{\gamma}))$. Using the same notations, the asymptotic variance of the optimal SLS estimator is given by

$$\boldsymbol{F}_0^* = E^{-1}\left(\boldsymbol{K}'\boldsymbol{M}^{-1}\boldsymbol{K}\right). \tag{A15}$$

Again, by the matrix form of Cauchy-Schwartz inequality $\boldsymbol{F}_0 - \boldsymbol{F}_0^*$ is nonnegative definite and is zero matrix if and only if

$$\boldsymbol{M}\boldsymbol{V}\boldsymbol{K} = \boldsymbol{K}E^{-1}\left(\boldsymbol{K}'\boldsymbol{M}^{-1}\boldsymbol{K}\right)E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right). \tag{A16}$$

References

Abarin, Taraneh, and Liqun Wang. 2009. Second-order least squares estimation of censored regression models. *Journal of Statistical Planning and Inference* 139: 125–35. [CrossRef]

Alvarez, Javier, and Manuel Arellano. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71: 1121–59. [CrossRef]

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press.

Arellano, Manuel. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.

Arellano, Manuel, and Olympia Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29–51. [CrossRef]

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies* 58: 277–97. [CrossRef]

Baltagi, Badi. 2008. *Econometric Analysis of Panel Data*. New York: Wiley.

Bickel, Peter J., Chris A. Klaassen, Ya'acov Ritov, and Jon A. Wellner. 1993. *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

Blundell, Richard, and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–43. [CrossRef]

Bose, Mausumi, and Rahul Mukerjee. 2015. Optimal design measures under asymmetric errors, with application to binary design points. *Journal of Statistical Planning and Inference* 159: 28–36. [CrossRef]

Chamberlain, Gary. 1987. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34: 305–34. [CrossRef]

Chamberlain, Gary. 1992. Comment: Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics* 10: 20–26.

Fitzmaurice, Garrett, Marie Davidian, Geert Verbeke, and Geert Molenberghs. 2009. *Longitudinal Data Analysis*. Boca Raton: CRC Press.

Gao, Lucy L., and Julie Zhou. 2014. New optimal design criteria for regression models with asymmetric errors. *Journal of Statistical Planning and Inference* 149: 140–51. [CrossRef]

Gao, Lucy L., and Julie Zhou. 2017. D-optimal designs based on the second-order least squares estimator. *Statistical Papers* 58: 77–94. [CrossRef]

Hahn, Jinyong. 1997. Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics* 79: 1–21. [CrossRef]

Hsiao, Cheng. 2003. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Hsiao, Cheng. 2011. Dynamic panel data models. *Handbook of Empirical Economics and Finance* 13: 373–96.

Jennrich, Robert I. 1969. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics* 40: 633–43. [CrossRef]

Kim, Mijeong, and Yanyuan Ma. 2012. The efficiency of the second-order nonlinear least squares estimator and its extension. *Annals of the Institute of Statistical Mathematics* 64: 751–64. [CrossRef]

Kiviet, Jan F. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68: 53–78. [CrossRef]

Newey, Whitney K., and Daniel McFadden. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* 4: 2111–245.

Okui, Ryo. 2009. The optimal choice of moments in dynamic panel data models. *Journal of Econometrics* 151: 1–16. [CrossRef]

Park, Byeong U, Robin C Sickles, and Léopold Simar. 2007. Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics* 136: 281–301. [CrossRef]

Wang, Liqun. 2003. Estimation of nonlinear berkson-type measurement error models. *Statistica Sinica* 13: 1201–10.

Wang, Liqun. 2004. Estimation of nonlinear models with berkson measurement errors. *The Annals of Statistics* 32: 2559–79. [CrossRef]

Wang, Liqun. 2007. A unified approach to estimation of nonlinear mixed effects and berkson measurement error models. *Canadian Journal of Statistics* 35: 233–48. [CrossRef]

Wang, Liqun, and Alexandre Leblanc. 2008. Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics* 60: 883–900. [CrossRef]

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Yin, Yue, and Julie Zhou. 2017. Optimal designs for regression models using the second-order least squares estimator. *Statistica Sinica* 27: 1841–56. [CrossRef]