First difference estimation of spatial dynamic panel data models with fixed effects

Fei Jin
Fudan University
jin.fei@live.com

Lung-fei Lee
Ohio State University
lee.1777@osu.edu

Jihai Yu*
Peking University
jihai.yu@gmail.com

October 18, 2019

Abstract

This paper investigates the first difference (FD) estimation of spatial dynamic panel data (SDPD) models with fixed effects using quasi-maximum likelihood (QML) approach, where both n and T are large. We show that the QML estimation for the SDPD with FD can be reduced to the direct estimation of individual effects, except for the estimation of variance parameter. After bias correction, these two approaches would yield asymptotically equivalent estimates for all parameters including the variance parameter. Our results extend the equivalence of LSDV estimate and GLS estimate of FD equation in the panel regression model to the spatial dynamic panel model, which includes the conventional dynamic panel as a special case. Our analysis highlights the importance of initial values rather than the many fixed effects in spatial panel models.

JEL classification: C3; C13

Keywords: Spatial autoregression, Dynamic panels, Fixed effects, First difference, Quasi-maximum likelihood estimation, Bias correction

*Corresponding author. Tel: 8610-62760702; fax: 8610-62753624. Address: Room 349, Guanghua School of Management Building #2, Peking University, Beijing, China 100871.

1 Introduction

For panel data, spatial dynamic panel data (SDPD) models take into account both dynamic and spatial interaction. A spatial panel model considers not only individual effects that allows for time invariant heterogeneity, but also spatial and time dependences. For spatial panel models, we specify spatial correlations by either spatially correlated disturbances or spatial lagged terms in a regression equation. For spatially correlated random components with individual heterogeneity, see Kapoor et al. (2007), Baltagi et al. (2007), Su and Yang (2015) etc. For the spatial lag model, see Korniotis (2010), Yu et al. (2008, 2012) and Yu and Lee (2010) among others.

The current paper aims to investigate asymptotic properties of the quasi-maximum likelihood estimate (QMLE) for SDPD models with first difference (FD) when both n and T are large.[1] With first differencing, individual effects are eliminated before estimation. Thus, one might expect the incidental parameter problem would disappear in estimation, for large T (and n) cases. However, we show that the QMLE for the SDPD model under time differencing by conditioning on initial dependent variable observations as if they were exogenously given, have the same spatial lag, time lag, spatial-time lag, and regressor coefficient estimates as those from direct estimation including individual effects; however, the variance parameter estimate is relatively more accurate. Thus, due to the presence of initial observations, which are conditional upon in estimation as given, asymptotic biases of the QMLEs of parameters of interest would appear in the same way as those from a direct estimation. We show that, it is possible to do bias correction on those estimates; and after bias correction, these two approaches can yield asymptotically equivalent estimates.

In the following, Section 2 introduces the model. Section 3 investigates asymptotic properties of the QMLE for the SDPD with FD, where initial period observations of the dependent variable are assumed exogenously given. It is shown that all the coefficient estimates are the same as those from direct estimation of individual effects. Hence, QMLEs of parameters of interest would exhibit asymptotic biases as those of the direct estimation. In that sense, the incidental parameter problem remains for the time differencing approach. However, their variance parameter estimates are different. We suggest bias correction procedures. After bias correction, both approaches can yield the same or essentially the same bias adjusted estimates including the variance parameter. Section 4 concludes this note and summarizes the contributions.

[1] For SDPD models with FD under the short T case, Kripganz (2017) extends Hsiao et al. (2002) by specifying initial observations, and Yang (2018) develops a unified M-estimation based on bias-corrected score vectors which does not need to specify the initial observations.

2 The Model

The SDPD model under consideration has the specification:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \quad t = 1, 2, ..., T, \tag{1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, ..., v_{nt})'$ are $n \times 1$ (column) vectors, v_{it}'s are *i.i.d.* $(0, \sigma^2_{v0})$ across i and t, and W_n is an $n \times n$ spatial weights matrix, which is predetermined and generates spatial dependence among cross-sectional units. Here, X_{nt} is an $n \times k_x$ matrix of time varying regressors, $\mathbf{c}_{n0}$ is $n \times 1$ column vector of fixed effects and α_{t0}'s are time effects with l_n being an n-dimensional column vector consisting of ones. We assume that the initial observation Y_{n0} is observable.

Depending on the eigenvalues structure of W_n and values of $(\lambda, \gamma_0, \rho_0)'$, we might have different stable or unstable cases (see Lee and Yu, 2015), which would have impact on the asymptotic distribution of QMLE. However, the likelihood functions to be maximized are the same for all models. Thus, we will work on one likelihood function for different SDPD models. We will compare properties of two estimation approaches: one is to make first difference (FD) to eliminate the individual effects, and the other is to directly estimate the individual effects as in Yu et al. (2008).

It is helpful to investigate the reduced form of (1) to understand dynamics of this general model. Define $S_n(\lambda) = I_n - \lambda W_n$ and $S_n \equiv S_n(\lambda_0) = I_n - \lambda_0 W_n$. Then, presuming that S_n is invertible and denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, (1) can be rewritten as

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_{t0} S_n^{-1} l_n + S_n^{-1} V_{nt}. \tag{2}$$

Under the stable case, by continuous substitution,

$$Y_{nt} = \sum\nolimits_{h=0}^{\infty} A_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}) = \mu_n + \mathcal{X}_{nt}\beta_0 + U_{nt}, \tag{3}$$

where $\mu_n \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1}\mathbf{c}_{n0}$, $\mathcal{X}_{nt} \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1} X_{n,t-h}$, and $U_{nt} \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1} V_{n,t-h}$.

To eliminate the time effects in the SDPD model in (1), we can make a data transformation as in Lee and Yu (2010). The transformed equation leads to a partial likelihood. This data transformation to eliminate the time effects requires a row-normalized W_n to have an SAR representation after the data transformation. This is not a strong additional requirement because the row-normalization of W_n is the preferred specification in empirical applications and it is tractable for a convenient analysis of stable or unstable conditions.

We denote $J_n = I_n - \frac{1}{n} l_n l_n'$ and let $F_{n,n-1}$ be the eigenvector matrix of J_n corresponding to the $(n-1)$ eigenvalues of one. By denoting $Y^*_{nt} = F'_{n,n-1} Y_{nt}$, we have

$$Y^*_{nt} = \lambda_0 W^*_n Y^*_{nt} + \gamma_0 Y^*_{n,t-1} + \rho_0 W^*_n Y^*_{n,t-1} + X^*_{nt}\beta_0 + \mathbf{c}^*_{n0} + V^*_{nt}, \tag{4}$$

where $W_n^* = F'_{n,n-1} W_n F_{n,n-1}$ holds because W_n is row-normalized. The resulting disturbance vector V_{nt}^* is an $(n-1)$ dimensional vector with zero mean and variance matrix $\sigma_0^2 I_{n-1}$. Thus, (4) is in the format of an SDPD model, where the number of observations is $T(n-1)$, reduced from the original sample observations by one for each period. The likelihood function can be constructed based on Y_{nt}^*, which can be treated as a partial likelihood. If V_{nt} is normally distributed $N(0, \sigma_0^2 I_n)$, the transformed V_{nt}^* will be normally distributed as $N(0, \sigma_0^2 I_{n-1})$. Therefore, all the variables hereafter are transformed by $F'_{n,n-1}$. For notational simplicity, we omit the superscript $*$ in Y_{nt}^*, X_{nt}^*, $\mathbf{c}_{n0}^*$, V_{nt}^* and W_n^*, and the effective cross sectional sample size n becomes $n^* = n - 1$.[2]

3 Conditional QMLE of SDPD with FD

3.1 Likelihood Function of SDPD under Time Differencing

By first differencing of (4) to eliminate the individual effects, we have

$$\Delta Y_{nt} = \lambda_0 W_n \Delta Y_{nt} + \gamma_0 \Delta Y_{n,t-1} + \rho_0 W_n \Delta Y_{n,t-1} + \Delta X_{nt}\beta_0 + \Delta V_{nt}, \quad t = 2, ..., T.$$

As T is large, we treat (condition on) the first period ΔY_{n1} as if it were exogenously given. By denoting $\theta = (\lambda, \gamma, \rho, \beta', \sigma^2)'$, $\delta = (\gamma, \rho, \beta')'$ and $Z_{nt} = [Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt}]$, the log likelihood function is

$$\ln L_{nT}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln|S_n(\lambda)| - \frac{n}{2}\ln|H_{T-1}| - \frac{1}{2\sigma^2}\Delta\mathbf{V}'_{nT}(\theta)(H_{T-1}^{-1}\otimes I_n)\Delta\mathbf{V}_{nT}(\theta), \quad (5)$$

where $\Delta\mathbf{V}_{nT}(\theta) = (\Delta V'_{n2}(\theta), ..., \Delta V'_{nT}(\theta))'$ with $\Delta V_{nt}(\theta) = S_n(\lambda)\Delta Y_{nt} - \Delta Z_{nt}\delta$ and

$$H_{T-1} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}.$$

By denoting $\mathbf{V}_{nT} = (V'_{n1}, ..., V'_{nT})$ and the $(T-1)\times T$ difference operator

$$L_{T-1,T} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix},$$

[2] W_n^* might not have similar structures as the original spatial weights matrix W_n with zero diagonal and row-normalization.

we have $\Delta \mathbf{V}_{nT} = (L_{T-1,T} \otimes I_n)\mathbf{V}_{nT}$ so that $\Delta \mathbf{V}'_{nT}(H_{T-1}^{-1} \otimes I_n)\Delta \mathbf{V}_{nT} = \mathbf{V}'_{nT}(L'_{T-1,T}H_{T-1}^{-1}L_{T-1,T} \otimes I_n)\mathbf{V}_{nT}$. As the inverse of H_{T-1} is

$$H_{T-1}^{-1} = \frac{1}{T}\begin{pmatrix} T-1 & T-2 & T-3 & \cdots & 2 & 1 \\ T-2 & (T-2)2 & (T-3)2 & \cdots & 2\cdot 2 & 2 \\ T-3 & (T-3)2 & (T-3)3 & \cdots & 3\cdot 2 & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & 2\cdot 2 & 3\cdot 2 & \cdots & 2((T-3)2-(T-4)) & (T-3)2-(T-4) \\ 1 & 2 & 3 & \cdots & (T-3)2-(T-4) & (T-2)2-(T-3) \end{pmatrix},$$

by using the the first difference nature of $L_{T-1,T}$, we can verify that

$$L'_{T-1,T}H_{T-1}^{-1}L_{T-1,T} \text{ is equal to } J_T = I_T - \frac{1}{T}l_T l'_T.$$

Thus, the log likelihood function can then be written as

$$\ln L_{nT}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln|S_n(\lambda)| - \frac{n}{2}\ln|H_{T-1}| - \frac{1}{2\sigma^2}\mathbf{V}'_{nT}(\theta)(J_T \otimes I_n)\mathbf{V}_{nT}(\theta). \quad (6)$$

By denoting $\mathbf{Y}_{nT} = (Y'_{n1}, ..., Y'_{nT})'$, $\mathbf{Z}_{nT} = (Z'_{n1}, ..., Z'_{nT})'$, and $\mathbf{J}_{nT} = (J_T \otimes I_n)$, the QMLE of δ in terms of λ is

$$\hat{\delta}_{nT}(\lambda) = (\mathbf{Z}'_{nT}\mathbf{J}_{nT}\mathbf{Z}_{nT})^{-1}\mathbf{Z}'_{nT}\mathbf{J}_{nT}\mathbf{S}_{nT}(\lambda)\mathbf{Y}_{nT}.$$

After further concentrating out σ^2, the concentrated log likelihood function from (6) is

$$\ln L_{nT}(\lambda) = -\frac{n(T-1)}{2}\ln(2\pi\hat{\sigma}^2_{nT}(\lambda)) + (T-1)\ln|S_n(\lambda)| - \frac{n}{2}\ln|H_{T-1}| - \frac{n(T-1)}{2} \quad (7)$$

with $\hat{\sigma}^2_{nT}(\lambda) = \frac{1}{n(T-1)}\mathbf{V}'_{nT}(\lambda)\mathbf{J}_{nT}\mathbf{V}_{nT}(\lambda)$ and $\mathbf{V}_{nT}(\lambda) = \mathbf{S}_{nT}(\lambda)\mathbf{Y}_{nT} - \mathbf{Z}_{nT}\hat{\delta}(\lambda)$, where $\mathbf{S}_{nT}(\lambda) = I_T \otimes S_n(\lambda)$.

We can compare the QMLE obtained from (5) to the QMLE from the direct estimation. Regarding that the initial Y_{n1} were exogenously given, the likelihood function of the SDPD model (4) (after time effects are eliminated) under the direct estimation approach is

$$\ln L^d_{nT}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^T \mathbf{V}'_{nT}(\theta, \mathbf{c}_n)\mathbf{V}_{nT}(\theta, \mathbf{c}_n), \quad (8)$$

where $\mathbf{V}_{nT}(\theta, \mathbf{c}_n) = (V'_{n1}(\theta, \mathbf{c}_n), ..., V'_{nT}(\theta, \mathbf{c}_n))'$ with $V_{nt}(\theta, \mathbf{c}_n) = S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n$. Using the first order condition for $\mathbf{c}_n$ in terms of λ and δ, the individual effects can be estimated by $\hat{\mathbf{c}}_n = \frac{1}{T}\sum_{t=1}^T(S_n(\lambda)Y_{nt} - Z_{nt}\delta)$. After concentrating out those individual effects, the corresponding concentrated log likelihood function is

$$\ln L^d_{nT}(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\mathbf{V}'_{nT}(\theta)\mathbf{J}_{nT}\mathbf{V}_{nT}(\theta), \quad (9)$$

where $\mathbf{V}_{nT}(\theta) = (V'_{n1}(\theta), ..., V'_{nT}(\theta))'$ with $V_{nt}(\theta) = S_n(\lambda)Y_{nt} - Z_{nt}\delta$.

Using the first order condition of (9) w.r.t. δ, the QMLE $\hat{\delta}^d_{nT}(\lambda)$ given λ from the maximization of (9) is

$$\hat{\delta}^d_{nT}(\lambda) = (\mathbf{Z}'_{nT}\mathbf{J}_{nT}\mathbf{Z}_{nT})^{-1}\mathbf{Z}'_{nT}\mathbf{J}_{nT}\mathbf{S}_{nT}(\lambda)\mathbf{Y}_{nT},$$

which is the same as $\hat{\delta}_{nT}(\lambda)$. After further concentration out σ^2, we have

$$\ln L^d_{nT}(\lambda) = -\frac{nT}{2}\ln(2\pi\hat{\sigma}^{2d}_{nT}(\lambda)) + T\ln|S_n(\lambda)| - \frac{nT}{2} \tag{10}$$

with $\hat{\sigma}^{2d}_{nT}(\lambda) = \frac{1}{nT}\mathbf{V}'_{nT}(\lambda)\mathbf{J}_{nT}\mathbf{V}_{nT}(\lambda)$. Thus, up to an additive constant and a scalar factor, $\frac{1}{n(T-1)}\ln L_{nT}(\lambda)$ in (7) and $\frac{1}{nT}\ln L^d_{nT}(\lambda)$ in (10) would be the same, so that the QMLEs of λ from (7) and (10) are numerically identical. As $\hat{\delta}_{nT}(\lambda)$ and $\hat{\delta}^d_{nT}(\lambda)$ are the same, estimates of (λ, δ) are also numerically the same as those from (6) and (9). The only difference is on the estimation of σ^2 in that $\hat{\sigma}^{2d}_{nT} = \frac{T-1}{T}\hat{\sigma}^2_{nT}$.

The above analysis implies that the asymptotic properties of QMLEs for SDPD models in Yu et al. (2008, 2012) and Yu and Lee (2010) will carry over for the approach under time differencing, except for the estimation for σ^2. To sum up,

$$\hat{\theta}_{nT} = \hat{\theta}^d_{nT} + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{T-1}\hat{\sigma}^{2d}_{nT} \end{pmatrix}, \tag{11}$$

where $\hat{\theta}^d_{nT}$ is obtained from the direct QML estimation of model (4) including individual effects, while $\hat{\theta}_{nT}$ is obtained from time differencing with the log likelihood (5).

3.2 Asymptotic Properties of QMLEs under FD and Direct Estimation

For the direct estimation for the SDPD model (4), when both n and T tend to infinity, rates of convergence of QMLEs are $\sqrt{nT}$ for the stable case, as shown in Yu et al. (2008). But the QMLE has asymptotic bias due to initial condition in the SDPD model with individual effects. Such biases can be eliminated by a bias correction procedure for the dynamic panel without spatial interactions in Hahn and Kuersteiner (2002). Here we shall consider bias correction procedures for the QMLE of the SDPD by restricting our attention to the stable case.[3] For the SDPD model in (4) under the direct approach, we can use

$$\hat{\theta}^d_{1,nT} = \hat{\theta}^d_{nT} - \frac{\hat{b}^d_{nT}}{T} \text{ with } \hat{b}^d_{nT} = \left.\left[\left(E\left(\frac{1}{nT}\frac{\partial^2 \ln L^d_{nT}(\theta)}{\partial\theta\partial\theta'}\right)\right)^{-1} a^d(\theta)\right]\right|_{\theta=\hat{\theta}^d_{nT}} \tag{12}$$

where $a^d(\theta)$ is the dominant component of the expected value of the score vector, and $\hat{b}^d_{nT}$ is the corresponding estimated bias. When T grows faster than $n^{1/3}$, the correction will eliminate the bias of order $O(T^{-1})$ and yield a properly centered asymptotic confidence interval.

[3] We may consider other unstable cases. The difference occurs at the information matrix, which would be singular for an unstable case. We shall leave those due to space limitation.

In the following, we investigate QML under time differencing and its corresponding bias corrected estimates. We compare them with those of the direct estimation as there are some differences on the estimates of σ_v^2. For notational purposes, we define $\tilde{Y}_{n,t-1}^{(-1)} = Y_{n,t-1} - \bar{Y}_{nT,-1}$ and $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$ for $t = 1, 2, \cdots, T$ where $\bar{Y}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T} Y_{n,t-1}$ and $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$. Furthermore, $\tilde{V}_{nt}(\theta) = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$ where $\tilde{Z}_{nt} = [\tilde{Y}_{n,t-1}^{(-1)}, W_n\tilde{Y}_{n,t-1}^{(-1)}, \tilde{X}_{nt}]$. From (3),

$$\tilde{Z}_{nt} = \tilde{Z}_{nt}^* - (\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x}), \tag{13}$$

where $\tilde{Z}_{nt}^* = ((\tilde{\bar{\mathcal{X}}}_{n,t-1} + U_{n,t-1}), (W_n\tilde{\bar{\mathcal{X}}}_{n,t-1} + W_n U_{n,t-1}), \tilde{X}_{nt})$ with $\tilde{\bar{\mathcal{X}}}_{n,t-1} = \mathcal{X}_{n,t-1} - \bar{\mathcal{X}}_{nT,-1}$. Hence, $\tilde{Z}_{nt}$ has two components: one is $\tilde{Z}_{nt}^*$, which is uncorrelated with V_{nt}; the other is $-(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})$, which is correlated with V_{nt} when $t \le T-1$.

The conditional likelihood for the equation in (4) with FD is that in (6), which can be rewritten as

$$\ln L_{nT}(\theta) = -\frac{n(T-1)}{2}\ln 2\pi - \frac{n(T-1)}{2}\ln\sigma^2 - \frac{n}{2}\ln|H_{T-1}| + (T-1)\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta)\tilde{V}_{nt}(\theta). \tag{14}$$

From the first order condition and the decomposition of $\tilde{Z}_{nt}$, by denoting $G_n = W_n S_n^{-1}$ so that $W_n\tilde{Y}_{nt} = G_n\tilde{Z}_{nt}\delta_0 + G_n\tilde{V}_{nt}$, the score vector can be decomposed as $\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{nT}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{nT}^*(\theta_0)}{\partial\theta} - \Delta_{nT}$ where

$$\frac{\partial \ln L_{nT}^*(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sum_{t=1}^{T}\tilde{Z}_{nt}^{*\prime}V_{nt} \\ \frac{1}{\sigma_0^2}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}^*\delta_0)'V_{nt} + \frac{1}{\sigma_0^2}\sum_{t=1}^{T}(\tilde{V}_{nt}'G_n'\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 trG_n) \\ \frac{1}{2\sigma_0^4}\sum_{t=1}^{T}(\tilde{V}_{nt}'\tilde{V}_{nt} - n\frac{T-1}{T}\sigma_0^2) \end{pmatrix}, \tag{15}$$

and

$$\Delta_{nT} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(G_n(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})\delta_0)'\bar{V}_{nT} \\ 0 \end{pmatrix}. \tag{16}$$

We have $\Delta_{nT} = \sqrt{\frac{n}{T}}a_n + O(\sqrt{\frac{n}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ where $a_n = a_n(\theta_0) = O(1)$ is the dominant component of the expected value of the score vector, which can be considered as the source of bias of MLE. Here,

$$a_n(\theta) = \begin{pmatrix} \frac{1}{n}tr\left(\left(\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)\right) \\ \frac{1}{n}tr\left(W_n\left(\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)\right) \\ \mathbf{0}_{k_x\times 1} \\ \frac{1}{n}\gamma tr(G_n(\lambda)\left(\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)) + \frac{1}{n}\rho tr(G_n(\lambda)W_n\left(\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)) \\ 0 \end{pmatrix}.$$

The asymptotic distribution of QMLE $\hat{\theta}_{nT}$ is

$$\sqrt{n(T-1)}\left(\hat{\theta}_{nT} - \theta_0\right) + \sqrt{\frac{n}{T-1}}\left(\Sigma_{\theta_0,nT}\right)^{-1}\varphi_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right)$$
$$\xrightarrow{d} N(0, \lim_{T\to\infty}\Sigma_{\theta_0,nT}^{-1}(\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT})\Sigma_{\theta_0,nT}^{-1}),$$

where $\Omega_{\theta_0,nT}$ captures the non-normality of the disturbances and

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT} & * \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n}\left[tr(G_n'G_n)+tr(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{\sigma_0^2 n}tr(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

with $\mathcal{H}_{nT} = \frac{1}{n(T-1)}\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$. The corresponding bias corrected estimate is

$$\hat{\theta}_{1,nT} = \hat{\theta}_{nT} - \frac{\hat{b}_{nT}}{T-1}, \text{ where } \hat{b}_{nT} = \left.\left[\left(E\left(\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}\right)\right)^{-1} a_n(\theta)\right]\right|_{\theta=\hat{\theta}_{nT}}.$$

For comparison, in the direct approach (see (20) in Appendix A), its score component corresponding to λ does not have the term $\frac{1}{n}trG_n$. The information matrix $-E\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}$ is the same as that of the direct approach. The bias adjusted estimate of the direct estimate is

$$\hat{\theta}_{1,nT}^d = \hat{\theta}_{nT}^d - \frac{1}{T}\left[\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}^d(\theta_0)}{\partial\theta\partial\theta'}\right]^{-1} a_n^d(\theta).$$

Because $\hat{\theta}_{nT} = \hat{\theta}_{nT}^d + (0, 0, \frac{1}{T-1}\hat{\sigma}_{nT}^{2d})'$ and $a_n^d(\theta) = a_n(\theta) + (0, \frac{1}{n}tr(G_n), \frac{1}{2\hat{\sigma}^{2d}})'$, it follows that the two bias adjusted estimates have the relation:

$$\hat{\theta}_{1,nT}^d - \hat{\theta}_{1,nT} = -\begin{pmatrix} 0 \\ 0 \\ \frac{1}{T-1}\hat{\sigma}_{nT}^{2d} \end{pmatrix} + \frac{1}{T-1}(\Sigma_{\theta,nT}^d)^{-1}\begin{pmatrix} 0 \\ \frac{1}{n}tr(G_n) \\ \frac{1}{2\hat{\sigma}^{2d}} \end{pmatrix} + O_p\left(\frac{1}{T^2}\right),$$

where the remainder term of order $O_p\left(\frac{1}{T^2}\right)$ is due to the difference of $(T-1)$ in the degrees of freedom under the FD approach and T in the degrees of freedom under the direct estimation approach on formulating the two bias-adjusted estimates. Thus, we may adopt a slightly altered bias corrected estimate for the approach with time differencing as

$$\tilde{\theta}_{1,nT} = \hat{\theta}_{1,nT}^d + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{T-1}\hat{\sigma}_{nT}^{2d} \end{pmatrix} - \frac{1}{T-1}(\Sigma_{\theta,nT}^d)^{-1}\begin{pmatrix} 0 \\ \frac{1}{n}tr(G_n) \\ \frac{1}{2\hat{\sigma}_{nT}^{2d}} \end{pmatrix}.$$

Here, the second term on the right hand side comes from the degrees of freedom adjustment before bias correction, and the third term is its impact from the bias correction procedure. However, the later impact will cancel out the degrees of freedom adjustment, i.e., $(0, 0, \hat{\sigma}_{nT}^{2d})' = (\Sigma_{\theta,nT}^d)^{-1}(0, \frac{1}{n}tr(G_n), \frac{1}{2\hat{\sigma}_{nT}^{2d}})'$.[4] As a result, $\hat{\theta}_{1,nT}^d = \tilde{\theta}_{1,nT}$, i.e., those bias corrected estimates for the approach under time differencing and the

[4]This equivalence can be obtained by the score expansion such that $\hat{\theta}_{nT}^d \doteq (\Sigma_{\theta,nT}^d)^{-1}\frac{\partial \ln L_{nT}^d(\theta_0)}{\partial\theta}$ and $\hat{\theta}_{nT} \doteq (\Sigma_{\theta,nT})^{-1}\frac{\partial \ln L_{nT}(\theta_0)}{\partial\theta}$ by ignoring higher order remainder terms of Taylor expansions. As $\hat{\theta}_{nT} - \hat{\theta}_{nT}^d = (0, 0, \frac{1}{T-1}\hat{\sigma}_{nT}^{2d})'$ from (11) and $\frac{\partial \ln L_{nT}^d(\theta_0)}{\partial\theta} - \frac{\partial \ln L_{nT}(\theta_0)}{\partial\theta} \doteq (0, tr(G_n), \frac{n}{2\hat{\sigma}_{nT}^{2d}})'$ from comparing score vectors of two approaches in (15) and (18), we have the asymptotic equivalence of $(0, 0, \hat{\sigma}_{nT}^{2d})'$ and $(\Sigma_{\theta,nT}^d)^{-1}(0, \frac{1}{n}tr(G_n), \frac{1}{2\hat{\sigma}_{nT}^{2d}})'$.

direct approach are numerically the same. On the other hand, for the previous bias corrected estimate $\hat{\theta}_{1,nT}$, it differs from $\tilde{\theta}_{1,nT}$ only by the possible smaller term $O_p\left(\frac{1}{T^2}\right)$. So, while $\hat{\theta}^d_{1,nT}$ might not be exactly equal to $\hat{\theta}_{1,nT}$, they are essentially the same.

4 Conclusion

This paper investigates estimation of SDPD models by using time differencing to eliminate the individual effects. We show that when T is large so that the sample observations of the dependent variable in the initial period can be regarded as exogenously given, except for the estimation of the variance parameter, the QMLEs for the SDPD under time differencing can be numerically the same as those of direct estimation which includes the estimation of individual effects. After bias correction, we illustrate that one may derive bias adjusted procedures which can yield the same bias corrected estimates for the two approaches. Our results extend the equivalence of LSDV estimate and GLS estimate of FD equation for the regression parameters in the panel linear regression models to the spatial dynamic panel setting.

Appendices

A QMLE under Direct Estimation of $\mathbf{c}_{n0}$

The concentrated likelihood in (9) can be written as

$$\ln L^d_{nT}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}'_{nt}(\theta)\tilde{V}_{nt}(\theta), \tag{17}$$

where $\tilde{V}_{nt}(\theta) = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$ with $\tilde{Z}_{nt} = (\tilde{Y}^{(-1)}_{n,t-1},\ W_n\tilde{Y}^{(-1)}_{n,t-1},\ \tilde{X}_{nt})$. From the first order condition and the decomposition of $\tilde{Z}_{nt}$ in (13), we have $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{nT}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d*}_{nT}(\theta_0)}{\partial\theta} - \Delta^d_{nT}$ where

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d*}_{nT}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \tilde{Z}^{*\prime}_{nt}V_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (G_n\tilde{Z}^*_{nt}\delta_0)'V_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (V'_{nt}G'_nV_{nt} - \sigma_0^2 trG_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (V'_{nt}V_{nt} - n\sigma_0^2) \end{pmatrix}, \tag{18}$$

and

$$\Delta^d_{nT} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(G_n(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})\delta_0)'\bar{V}_{nT} + \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}\bar{V}'_{nT}G'_n\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\sqrt{\frac{T}{n}}\bar{V}'_{nT}\bar{V}_{nT} \end{pmatrix}, \tag{19}$$

where $\Delta_{nT}^d = \sqrt{\frac{n}{T}} a_n^d + O(\sqrt{\frac{n}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ with $a_n^d = a_n^d(\theta_0) = O(1)$ and

$$a_n^d(\theta) = \begin{pmatrix} \frac{1}{n} tr\left(\left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)\right) \\ \frac{1}{n} tr\left(W_n \left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)\right) \\ \mathbf{0}_{k_x \times 1} \\ \frac{1}{n}\gamma tr(G_n(\lambda)\left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)) + \frac{1}{n}\rho tr(G_n(\lambda) W_n \left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)) + \frac{1}{n} tr G_n(\lambda) \\ \frac{1}{2\sigma^2} \end{pmatrix}. \quad (20)$$

Also, $\Sigma_{\theta_0,nT}^d = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT}^d & * \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n}\left[tr(G_n' G_n) + tr(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{\sigma_0^2 n} tr(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$ is the information matrix $-E\frac{1}{nT}\frac{\partial^2 \ln L_{nT}^d(\theta_0)}{\partial\theta\partial\theta'}$ with $\mathcal{H}_{nT}^d = \frac{1}{nT}\sum_{t=1}^T (\tilde{Z}_{nt},\, G_n\tilde{Z}_{nt}\delta_0)'(\tilde{Z}_{nt},\, G_n\tilde{Z}_{nt}\delta_0)$.

References

Baltagi, B., S.H. Song, B.C. Jung and W. Koh, 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5-51.

Hahn, J. and G. Kuersteiner, 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are Large. *Econometrica* 70, 1639-1657.

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

Kapoor, M., Kelejian, H.H. and I.R. Prucha, 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97-130.

Kripfganz, S., 2017. Unconditional transformed likelihood estimation of time-space dynamic panel data models. Working Paper.

Korniotis, G.M., 2010. Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics* 28, 145-158.

Lee, L.F. and J. Yu, 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.

Lee, L.F. and J. Yu, 2011. A unified transformation approach for the estimation of spatial dynamic panel data models: stability, spatial cointegration and explosive roots. In: Ullah, A., Giles, D.E.A. (Eds.), *Handbook on Empirical Economics and Finance*, 397-434. Taylor & Francis Group, New York.

Lee, L.F., and J. Yu, 2015. Estimation of fixed effects panel regression models with separable and nonseparable space-time filters. *Journal of Econometrics* 184, 174-192.

Su, L. and Z. Yang, 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.

Yang, Z., 2018. Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics* 205, 423-447.

Yu, J., R. de Jong and L.F. Lee, 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.

Yu, J., R. de Jong, and L.F. Lee, 2012. Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics* 167, 16-37.

Yu, J. and L.F. Lee, 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* 26, 1332-1362.