Regional Science and Urban Economics 61 (2016) 52–72

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regsciurbeco

Bias correction and refined inferences for fixed effects spatial panel data models

Zhenlin Yang [a], Jihai Yu [b], Shew Fan Liu [a,*]

[a] School of Economics, Singapore Management University, 90 Stamford Road, 178903 Singapore, Singapore
[b] Guanghua School of Management, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Received 3 March 2016
Received in revised form
12 August 2016
Accepted 24 August 2016
Available online 27 August 2016

JEL classification:
C10
C13
C21
C23
C15

Keywords:
Bias correction
Variance correction
Refined *t*-ratios
Bootstrap
Wild bootstrap
Spatial panels
Fixed effects

ABSTRACT

This paper first presents simple methods for conducting up to third-order bias and variance corrections for the quasi maximum likelihood (QML) estimators of the spatial parameter(s) in the fixed effects spatial panel data (FE-SPD) models. Then, it shows how the bias and variance corrections lead to refined *t*-ratios for spatial effects and for covariate effects. The implementation of these corrections depends on the proposed bootstrap methods of which validity is established. Monte Carlo results reveal that (i) the QML estimators of the spatial parameters can be quite biased, (ii) a second-order bias correction effectively removes the bias, and (iii) the proposed *t*-ratios are much more reliable than the usual *t*-ratios.

1. Introduction

Panel data models with spatial and social interactions have received a belated but recently increasing attention by econometricians, since Anselin (1988).[1] Spatial panel data (SPD) models are differentiated by whether they are static or dynamic and whether they contain random effects or fixed effects. The quasi maximum likelihood (QML) and the generalized method of moments (GMM) are the popular methods for estimation and inference of these models. See Lee and Yu (2010a), Lee and Yu (2015) and Anselin et al. (2008) for general accounts on issues related to SPD model specifications, parameter estimation, etc.

It has been recognized through the studies of spatial regression models that QML estimators of the spatial parameter(s), although efficient, can be quite biased (Lee, 2004; Bao and Ullah, 2007; Bao, 2013; Yang, 2015), and more so with a denser spatial weight matrix (Yang, 2015; Liu and Yang, 2015a). As a result the subsequent model inferences (based on *t*-ratios) can be seriously affected. Methods of bias-correcting the QML estimators of the spatial parameter(s) have been given for the spatial lag (SL) model (Bao and Ullah, 2007; Bao, 2013; Yang, 2015), the spatial error (SE) model (Liu and Yang, 2015a), and the spatial lag and error (SLE) model (Liu and Yang, 2015b). The improved *t*-ratios for the SL effect is given in Yang (2015), and improved *t*-ratios for the covariate effects are given in Liu and Yang (2015b) for the SL, SE and SLE models, respectively.

Evidently, the QML estimators of the SPD models are subjected to the same issues on the finite sample bias and finite sample performance of subsequent inferences, but these important issues have not been addressed.[2] Given the popularity of the SPD models among the applied researchers, it is highly desirable to have a set

* Corresponding author.
E-mail addresses: zlyang@smu.edu.sg (Z. Yang), sfliu@smu.edu.sg (S.F. Liu).

[1] See, among others, Baltagi et al. (2003, 2013), Kapoor et al. (2007), Yu et al. (2008, 2012), Yu and Lee (2010), Lee and Yu (2010a,b), Baltagi and Yang (2013a,b), and Su and Yang (2015).

[2] The importance of bias correction for models with nonlinear parameters is seen from the large literature on the regular dynamic panels (see, e.g., Nickell, 1981; Kiviet, 1995; Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Bun and Carree, 2005; Hahn and Moon, 2006; Arellano and Hahn, 2005).

http://dx.doi.org/10.1016/j.regsciurbeco.2016.08.003
0166-0462/

of simple and reliable methods for parameter estimation and model inference. In this paper, we focus on the SPD models with fixed effects to provide methods for bias and variance corrections (up to third-order) by extending the methods of Yang (2015),[3] and then to show how the bias and variance corrections lead to improved t-ratios for spatial and covariate effects. Lee and Yu (2010b) investigate the asymptotic properties for the QML estimation of this model based on *direct* and *transformation* approaches. The latter approach is more attractive as it provides consistent estimators for all the common parameters, which is crucial in the development of the methods for finite sample bias-corrections and refined inferences.[4]

We note that while the general stochastic expansions of Yang (2015) for nonlinear estimators are applicable to different models including the SPD models considered in this paper, the detailed developments of bias corrections, variance corrections and corrections on t-ratio vary from one model to another. Furthermore, the transformation approach induces errors that may no longer be independent and identically distributed (iid) even if the original errors are. Thus, the bootstrap method proposed by Yang (2015) under iid errors, may not be directly applicable. We demonstrate in this paper that when the original error distribution is not far from normality, the standard iid bootstrap method can still provide an excellent approximation, due to the fact that the transformed errors are homoskedastic and uncorrelated. When the original errors are extremely non-normal, we show that the wild bootstrap method can improve the approximation. Monte Carlo results reveal that the QMLEs of the spatial parameters can be quite biased, in particular for the models with spatial error dependence, and that a second-order bias correction effectively removes the bias. Furthermore, Monte Carlo results show that inferences for spatial and covariate effects based on the regular t-ratios can be misleading, but those based on the proposed t-ratios are very reliable. We emphasize that while corrections on the bias and variance of a point estimator are important, it is more important to correct the t-ratios so that practical applications of the models are more reliable. The methods presented in this paper show a plausible way to do so. They are simple and yet quite general as the spatial regression models are embedded as special cases.

The rest of the paper is organized as follows. Section 2 introduces the spatial panel data model allowing both spatial lag and spatial error, and both time-specific effects and individual-specific effects, and its QML estimation based on the transformed likelihood function. Section 3 presents a third-order stochastic expansion for the QML estimators of the spatial parameters, a third-order expansion for the bias, and a third-order expansion for the variance of the QML estimators of the spatial parameters. Section 3 also addresses issues on the bias of QMLEs of other model parameters, and on the inferences following bias and variance corrections. Section 4 introduces the bootstrap methods for estimating various quantities in the expansions, and presents theories for the validity of these methods. Section 5 presents Monte Carlo results. Section 6 discusses and concludes the paper.

2. The model and its QML estimation

For the spatial panel data (SPD) model with fixed effects (FE), we investigate the case with both spatial lag and spatial error, where n is large and T could be finite or large. We include both individual effects and time effects to have a robust specification. The FE-SPD model under consideration is,

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt},$$
$$U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}, \tag{2.1}$$

for $t = 1, 2, \ldots, T$, where, for a given t, $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the response variable, X_{nt} is an $n \times k$ matrix containing the values of k nonstochastic, individually and time varying regressors, $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of errors where $\{v_{it}\}$ are independent and identically distributed (iid) for all i and t with mean 0 and variance σ_0^2, $\mathbf{c}_{n0}$ is an $n \times 1$ vector of fixed individual effects, and α_{t0} is the fixed time effect with l_n being an $n \times 1$ vector of ones. W_{1n} and W_{2n} are given $n \times n$ spatial weights matrices where W_{1n} generates the 'direct' spatial effects among the spatial units in their response values Y_{nt}, and W_{2n} generates cross-sectional dependence among the disturbances U_{nt}. In practice, W_{1n} and W_{2n} may be the same.

In Lee and Yu (2010b), QML estimation of (2.1) is considered by using either a direct approach or a transformation approach. The direct approach is to estimate the regression parameters jointly with the individual and time effects, which yields a bias of order $O(T^{-1})$ due to the estimation of individual effects and a bias of order $O(n^{-1})$ due to the estimation of time effects. The transformation approach eliminates the individual and time effects and then implements the estimation, which yields consistent estimates of the common parameters when either n or T is large. In the current paper, we follow the transformation approach so that it is free from the incidental parameter problem.

To eliminate the individual effects, define $J_T = \left(I_T - \frac{1}{T} l_T l_T'\right)$ and let $\left[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T\right]$ be the orthonormal eigenvector matrix of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ submatrix corresponding to the eigenvalues of one, I_T is a $T \times T$ identity matrix and l_T is a $T \times 1$ vector of ones.[5] To eliminate the time effects, let J_n and $F_{n,n-1}$ be similarly defined, and W_{1n} and W_{2n} be row normalized.[6] For any $n \times T$ matrix $\left[Z_{n1}, \ldots, Z_{nT}\right]$, define the $(n-1) \times (T-1)$ transformed matrix as

$$\left[Z_{n1}^*, \ldots, Z_{n,T-1}^*\right] = F_{n,n-1}'[Z_{n1}, \ldots, Z_{nT}]F_{T,T-1}. \tag{2.2}$$

This leads to, for $t = 1, \ldots, T-1$, Y_{nt}^*, U_{nt}^*, V_{nt}^*, and $X_{nt,j}^*$ for the jth regressor. As in Lee and Yu (2010b), let $X_{nt}^* = [X_{nt,1}^*, X_{nt,2}^*, \ldots, X_{nt,k}^*]$, and $W_{hn}^* = F_{n,n-1}' W_{hn} F_{n,n-1}$, $h = 1, 2$. The transformed model we will work on thus takes the form:

$$Y_{nt}^* = \lambda_0 W_{1n}^* Y_{nt}^* + X_{nt}^*\beta_0 + U_{nt}^*, \quad U_{nt}^* = \rho_0 W_{2n}^* U_{nt}^* + V_{nt}^*,$$
$$t = 1, \ldots, T-1. \tag{2.3}$$

After the transformations, the effective sample size becomes $N = (n-1)(T-1)$. Stacking the vectors and matrices, i.e., letting $\mathbf{Y}_N = \left(Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime}\right)'$, $\mathbf{U}_N = (U_{n1}^{*\prime}, \ldots, U_{n,T-1}^{*\prime})'$, $\mathbf{V}_N = (V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})'$,

[3] The fixed effects model has the advantage of robustness because fixed effects are allowed to depend on included regressors. It also provides a unified model framework for different random effects models considered in, e.g., Anselin (1988), Kapoor et al. (2007) and Baltagi et al. (2013). However, fixed effects model encounters incidental parameter problem (Neyman and Scott, 1948; Lancaster, 2000).

[4] Lee and Yu (2010b) observe that when conducting a direct estimation using the likelihood function where all the common parameters and the fixed effects are estimated together, the estimate of the variance parameter is inconsistent when T is finite while n is large. With data transformations to eliminate the fixed effects, the incidental parameter problem is avoided, and the ratio of n and T does not affect the asymptotic properties of estimates as the data are pooled. The QMLEs so derived are shown to be consistent, and, except for the variance estimate, are identical to those from the direct approach.

[5] As discussed in Lee and Yu (2010b, Footnote 12), the first difference and Helmert transformation have often been used to eliminate the individual effects. A special selection of $F_{T,T-1}$ gives rise to the Helmert transformation where $\{V_{nt}\}$ are transformed to $\left(\frac{T-t}{T-t+1}\right)^{1/2}[V_{nt} - \frac{1}{T-t}(V_{n,t+1} + \cdots + V_{nT})]$, which is of particular interest for dynamic panel data models.

[6] When W_{jn} are not row normalized, the linear SARAR representation of (2.4) for the spatial panel model will no longer hold. In that case, a likelihood formulation would not be feasible.