Dynamic spatial panel data models with common shocks

Jushan Bai* and and Kunpeng Li†

October 12, 2015

Abstract

Real data often have complicated correlation over cross section and time. Modeling, estimating and interpreting the correlations in data are particularly important in economic analysis. This paper integrates several correlation-modeling techniques and propose dynamic spatial panel data models with common shocks to accommodate possibly complicated correlation structure over cross section and time. A large number of incidental parameters exist within the model. The quasi maximum likelihood method (ML) is proposed to estimate the model. Heteroskedasticity is explicitly estimated. The asymptotic properties of the quasi maximum likelihood estimator (MLE) are investigated. Our analysis indicates that the MLE has a non-negligible bias. We propose a bias correction method for the MLE. The simulations further reveal the excellent finite sample properties of the quasi-MLE after bias correction.

Key Words: Panel data models, spatial interactions, common shocks, cross-sectional dependence, incidental parameters, maximum likelihood estimation

JEL Classification: C3; C13

*Department of Economics, Columbia University, NewYork

†School of International Economics and Management, Capital University of Economics and Business, Beijing, China,

1 Introduction

Real data often have complicated correlation over cross section and time. These correlations contain important information on the relationship among economic variables. Modeling, estimating and interpreting the correlations in data are particularly important in economic analysis. In econometric literature, the correlations over time are typically dealt with by the autoregressive models (e.g., Brockwell and Davis (1991), Fuller (1996), etc), among other models. The correlations over cross section are typically captured by spatial models or factor models (e.g., Anselin (1988), Bai and Li (2012), Fan et al. (2011), etc), among other models. In this paper, we integrate these correlation-modeling techniques and propose dynamic spatial panel data models with common shocks to accommodate possibly complicated correlation structure over cross section and time.

Spatial models are one of primary tools to study cross-sectional interactions among units. In these models, cross sectional dependence is captured by spatial weights matrices based either on physical distance, and relative position in a social network or on other types of economic distance[1]. Early development of spatial models has been summarized by a number of books, including Cliff and Ord (1973), Anselin (1988), and Cressie (1993). Generalized method of moments (GMM) estimation of spatial models are studied by Kelijian and Prucha (1998, 1999, 2010), and Kapoor et al. (2007), among others. The maximum likelihood method (ML) is considered by Ord (1975), Anselin (1988), Lee (2004a), Yu et al. (2008) and Lee and Yu (2010), and so on.

Cross-sectional dependence may also arises from the response of individuals to common shocks. This motivates common shocks models, which are widely used in applied studies, see, e.g., Ross (1976), Chamberlain and Rothschild (1983), Stock and Watson (1998), to name a few. For panel data models with multiple common shocks, Ahn et al. (2013) consider the fixed-T GMM estimation. Pesaran (2006) proposes the correlated random effects method by including additional regressors obtained from cross-sectionally averaging on dependent and the explanatory variables. The principal components method is studied by Bai (2009) and reinvestigated with perturbation theory by Moon and Weidner (2009). Bai and Li (2014b) consider the maximum likelihood method in the presence of heteroskedasticity.

A popular approach to dealing with temporal dependence is dynamic panel data models. In these models, the presence of individual time-invariant intercepts (fixed-effect) causes the so-called "incidental parameters problem" (Neyman and Scott (1948)), which is the primary concern in the related studies. A consequence of the incidental parameters problem is the inconsistency of the within group estimator under fixed-T (Nickell (1981)). Anderson and Hsiao (1981) suggests taking time difference to eliminate the fixed effects and use two-periods lagged dependent variable as instrument to estimate the model. Arellano and Bond (1991) extend the Anderson and Hsiao's idea with the GMM method. Under large-N and large-T setup, Hahn and Kuersteiner (2002) shows that the within-group estimator is still consistent but has a $O(\frac{1}{T})$ bias. After bias correction, the corrected estimator achieves the efficiency bound under normality assumption of errors. Alvarez and Arellano (2003)

[1] For spatial interaction and economic distance, see, e.g., Case (1991), Case et al. (1993), Conley (1999), Conley and Dupor (2003), and Topa (2001).

investigate the asymptotic properties of the within group, GMM and limited information ML estimators under large-N and large-T.

In this paper, we consider jointly modeling spatial interactions, dynamic interactions and common shocks within the following model:

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \delta y_{it-1} + x_{it}'\beta + \lambda_i' f_t + e_{it}. \tag{1.1}$$

where y_{it} is the dependent variable; $x_{it} = (x_{it1}, x_{it2}, \ldots, x_{itk})'$ is a k-dimensional vector of explanatory variables; f_t is an r-dimensional vector of unobservable common shocks; λ_i is the corresponding heterogenous response to the common shocks; $W_N = (w_{ij,N})_{N\times N}$ is a specified spatial weights matrix whose diagonal elements $w_{ii,N}$ are 0; and e_{it} are the idiosyncratic errors. In model (1.1), term $\lambda_i' f_t$ captures the common-shocks effects, $\rho\sum_{j=1}^{N} w_{ij,N} y_{jt}$ captures the spatial effects, and δy_{it-1} captures the dynamic effects. The joint modeling allows one to test which type of effects is present within data. We may test $\rho = 0$ while allowing common-shocks effects and dynamic effects; or similarly, we may determine if the number of factors is zero in a model with spatial effects and dynamic effects. It may be possible that all the three effects are present. The features of model (1.1) make it flexible enough to cover a wide range of applications. The applicability of the model is discussed in Section 2.

An additional feature of the model is the allowance of cross sectional heteroskedasticity. The importance of permitting heteroskedasticity is noted by Kelejian and Prucha (2010) and Lin and Lee (2010). The heteroskedastic variances can be empirically important, e.g., Glaeser et al. (1996) and Anselin (1988). In addition, if heteroskedasticity exists but homoskedasticity is imposed, then MLE can be inconsistent. Under large-N, the consistency analysis for MLE under heteroskedasticity is challenging even for spatial panel models without common shocks, owing to the simultaneous estimation of a large number of variance parameters along with (ρ, δ, β). The existing quasi maximum likelihood studies, such as Yu et al. (2008) and Lee and Yu (2010), typically assume homoskedasticity. These authors show that the limiting variance of MLE has a sandwich formula unless normality is assumed. Interestingly, we show that the limiting variance of the MLE is not of a sandwich form if heteroskedasticity is allowed.

Spatial correlations and common shocks are also considered by Pesaran and Tosetti (2011). Except that the dynamics is allowed in our model but not in theirs, another key difference is that they specify the spatial autocorrelation on the unobservable errors e_{it} while we specify the spatial autocorrelation on the observable dependent variable y_{it}. Both specifications are of practical relevance. Spatial specification on observable data makes explicit the empirical implication of the coefficient ρ. From a theoretical perspective, the spatial interaction on the dependent variable gives rise to the endogeneity problem, while the spatial interaction on the errors, in general, does not. As a result, under the Pesaran and Tosetti setup, existing estimation methods on the common shocks models such as Pesaran (2006) and Bai (2009) can be applied to estimate the model. As a comparison, these methods cannot be directly applied to model (1.1) due to the endogeneity from the spatial interactions.

In this study, we consider the pseudo-Gaussian maximum likelihood method (MLE), which simultaneously estimates all parameters of the model, including heteroskedasticity. We give a rigorous analysis of the MLE including the consistency, the rate of convergence and limiting distributions. Since the proposed model has several sources of incidental parameters (individual-dependent intercepts, interactive effects, heteroskedasticity), the incidental parameters problem exists and the MLE is shown to have a non-negligible bias. Following Hahn and Kuersteiner (2002), we conduct bias correction on the MLE to make it center around zero. The simulations show that the bias-corrected MLE has good finite sample performance.

The rest of the paper is organized as follows. Section 2 gives some potential showcase examples of the model. Section 3 lists the assumptions needed for the asymptotic analysis. Section 4 presents the objective function and the associated first order conditions. The asymptotic properties including the consistency, the convergence rates and the limiting distributions are derived in Section 5. Section 6 discusses the ML estimation on spatial models with heteroskedasticity. Section 7 reports simulation results. Section 8 discusses extensions of the model. The last section concludes. Technical proofs are given in a supplementary document. In subsequent exposition, the matrix norms are defined in the following way. For any $m \times n$ matrix A, $\|A\|$ denotes the Frobenius norm of A, i.e., $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$. In addition, $\|A\|_\infty$ is defined as $\|A\|_\infty = \max_{1\le i\le m}\sum_{j=1}^n |a_{ij}|$ and $\|A\|_1$ is defined as $\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^m |a_{ij}|$, where a_{ij} is the (i,j)th element of A. We use $\dot{a}_t$ to denote $\dot{a}_t = a_t - \frac{1}{T}\sum_{t=1}^T a_t$ for any column vector a_t. Throughout the paper, we assume the data of Y at time 0 are observed.

2 Some application examples

The proposed model can be applied in a variety of economic and social setups. In this section, we list two typical examples.

Finance. Recent studies pay much attention on financial network and financial contagion. Let y_{it} be the stock price (or profit) of firm i at period t. In financial market, one firm may hold shares of other firms and other firms may hold shares of this firm. This generates a financial network (Elliott et al. (2014)). Let $W_N = [w_{ij,N}]$ be some metric, which measures the cross-holding pattern among firms in market. Then $\rho\sum_{j=1}^N w_{ij,N}y_{jt}$ captures the cross-holding effects on firm i. In addition, as implied in asset pricing theory (see, e.g., Ross (1976), Conner and Korajczyk (1986, 1988), Geweke and Zhou (1996)), there are systematic shocks and risks affecting all the stocks, which we denote by f_t. The individual-dependent responses to these shocks are captured by λ_i. This leads to term $\lambda_i' f_t$. Furthermore, the adaptive expectation of firms gives rise to δy_{it-1}. Let x_{it} be a vector of explanatory variables, which are thought useful to explain the behaviors of stock prices. We allow that x_{it} has arbitrary correlations with systematic shocks f_t. Putting these ingredients together, we have the model specification like (1.1).

Macroeconomics. Standard economic theory asserts if other countries grow with high rates, the outside demand would drive up the growth rate of home country through trade. Recent studies shows that international trade exhibits some spatial pattern, not only due to the distance cost as illustrated by "gravity" theory, but also duo to regional trade agreement as well as ethnical, cultural and social network among the firms, see, e.g., Baltagi et al. (2008), Lawless (2009), Rauch and Trindade (2002), Defever et al. (2015), etc. Let y_{it} be growth rate of country i at period t, and $W_N = [w_{ij,N}]$ be some metric, which measures the closeness of countries based on the bilateral trade. Then term $\rho \sum_{j=1}^{N} w_{ij,N} y_{jt}$ captures the companion-driving effect in growth. Similarly as in the previous example, the growth rates of countries over the world are subject to global economic shocks, such as technological advances and financial crisis (Kose, Otrok and Whiteman 2003). We therefore introduce term $\lambda_i' f_t$ to adapt to this fact. Term δy_{it-1} is also necessary because of the inertia of growth. With these considerations, we have the specification of model (1.1).

Besides the above economic applications, the proposed model also has its applications in social science. In a pioneer study, Manski (1993) distinguishes three effects within social interactions, *endogenous effects, contextual effects* and *correlated effects*. In empirical studies, endogenous effects are estimated by the spatial term, controlling correlated effects through the usually additive fixed effects (Lin (2010)). In the proposed model, we can deal with correlated effects in a more general and plausible way by factor models. In addition, we allow the dynamics. In Appendix, we show that, with some slight modifications, our model specification can be motivated by the quadratic utility model of Calvó-Armengol et al. (2009).

We shall use $(\rho^*, \delta^*, \beta^*)$ to denote the true values for (ρ, δ, β), and we use (Λ^*, f_t^*) to denote the true values for (Λ, f_t). So the data generating process is

$$Y_t = \alpha^* + \rho^* W_N Y_t + \delta^* Y_{t-1} + X_t \beta^* + \Lambda^* f_t^* + e_t.$$

Let C be a generic constant large enough. We make following assumptions for the asymptotic analysis.

Assumption A: The x_{it} is either a fixed constant or a random variable. If x_{it} is fixed, we assume $\|x_{it}\| \le C$; if x_{it} is random, we assume $E(\|x_{it}\|^4) \le C$ for all i and t. If x_{it} is random, it is independent with the idiosyncratic error e_{js} for all i, j, t and s.

Assumption B: The λ_i^* and f_t^* can be either fixed constants and random variables. If λ_i^* is fixed, we assume that $\|\lambda_i^*\| \le C$ for all i and $\frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{*-1}\Lambda^* \to \Omega_\Lambda^*$ where $\Lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*)'$, otherwise we assume that $E(\|\lambda_i^*\|^4) \le C$ for all i and $\frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{*-1}\Lambda^* \xrightarrow{p} \Omega_\Lambda^*$, where Σ_{ee}^* is defined in Assumption C and Ω_Λ^* is some matrix positive definite. If f_t^* is fixed, we assume that $\|f_t^*\| \le C$ for all t and $\frac{1}{T}F^{*\prime}F^* \to \Omega_F^*$, otherwise we assume that $E\|f_t^*\|^4 \le C$ for all t and $\frac{1}{T}F^{*\prime}F^* \xrightarrow{p} \Omega_F^*$, where Ω_F^* is some matrix positive definite.

Assumption C: The e_{it} is independent and identically distributed over t and independent over i with $E(e_{it}) = 0$, $C^{-1} \le \sigma_i^{*2} \le C$ and $E(e_{it}^8) \le C$ for all i, where $\sigma_i^{*2} = E(e_{it}^2)$. Let $\Sigma_{ee}^* = \operatorname{diag}(\sigma_1^{*2}, \sigma_2^{*2}, \ldots, \sigma_N^{*2})$ be the variance of $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$. In addition, if $\{\lambda_i^*\}$ and $\{f_t^*\}$ are random, we assume that $\{e_{it}\}$ are independent with $\{\lambda_i^*\}$ and $\{f_t^*\}$.

Assumption D: The underlying value $\omega^* = (\rho^*, \delta^*, \beta^{*\prime})'$ is an interior point of parameters space $\Theta_\omega = (-1, 1) \times \mathrm{S}_\delta \times \mathrm{S}_\beta$, where S_δ and S_β are the two compact subsets of R and R^k.

Remark 3.1 Assumption A impose restrictions on the explanatory variables x_{it}. Although it requires that x_{it} be independent with e_{js}, it does allow x_{it} to have arbitrary correlations with λ_i or f_t or $\lambda_i' f_t$. This extends the traditional panel data analysis. Assumption B is about factors and factor loadings. This assumption is standard in pure factor analysis, see Bai (2003) and Bai and Li (2012). Assumption C assumes that the idiosyncratic error e_{it} is independent over the cross section and the time. In the present scenario, such an assumption is not restrictive as it looks to be since the weak correlations over the cross section and the time in data have been dealt with by the spatial term and the lag dependent term. However, if the cross sectional correlation of e_{it} is a major concern in empirical studies, our analysis can be extended to accommodate it, see the related discussion on SAR disturbances in Section 7. Assumptions D impose restrictions on the underlying coefficients. This assumption is standard.

Assumption E: The weights matrix W_N satisfies that $I_N - \rho^* W_N$ is invertible and

$$\limsup_{N\to\infty} \|W_M\|_\infty \le C; \qquad \limsup_{N\to\infty} \|W_N\|_1 \le C; \tag{3.2}$$

$$\limsup_{N\to\infty} \|(I_N - \rho^* W_N)^{-1}\|_\infty \le C; \qquad \limsup_{N\to\infty} \|(I_N - \rho^* W_N)^{-1}\|_1 \le C. \tag{3.3}$$

In addition, all the diagonal elements of W_N are zeros.

Assumption F: Let $G_N^* = (I_N - \rho^* W_N)^{-1}$. We assume

$$\limsup_{N\to\infty} \sum_{l=0}^{\infty} \left\|(\delta^* G_N^*)^l\right\|_\infty \leq C; \qquad \limsup_{N\to\infty} \sum_{l=0}^{\infty} \left\|(\delta^* G_N^*)^l\right\|_1 \leq C.$$

Remark 3.2 Assumptions E and F are imposed on the spatial weights matrix. Assumption E is standard in spatial econometrics, see Kelejian and Prucha (1998), Lee (2004a), Yu et al. (2008), Lee and Yu (2010), to name a few. Under this assumption, some key matrices, which play important roles in asymptotic analysis such as G_N^* in Assumption F and S_N^* in Assumption G, can be handled in a tractable way. Assumption F implicitly guarantees that y_{it} has a well-defined MA(∞) expression. Similar assumption also appears in Yu et al. (2008). A set of sufficient conditions for Assumptions E and F are $\limsup_{N\to\infty} \|W_N\|_\infty \leq 1$, $\limsup_{N\to\infty} \|W_N\|_1 \leq 1$ and $|\rho^*| + |\delta^*| < 1$ because

$$\limsup_{N\to\infty} \|G_N^*\|_\infty = \limsup_{N\to\infty} \|(I - \rho^* W_N)^{-1}\|_\infty \leq \limsup_{N\to\infty} \sum_{j=0}^{\infty} (\|\rho^* W_N\|_\infty)^j \leq \frac{1}{1 - |\rho^*|} < \infty,$$

and the argument for $\limsup_{N\to\infty} \|G_N^*\|_1 \leq \frac{1}{1-|\rho^*|} < \infty$ is the same. Similarly

$$\limsup_{N\to\infty} \sum_{l=0}^{\infty} \left\|(\delta^* G_N^*)^l\right\|_\infty \leq \limsup_{N\to\infty} \sum_{l=0}^{\infty} (|\delta^*| \cdot \|G_N^*\|_\infty)^l \leq \sum_{l=0}^{\infty} \left[\frac{|\delta^*|}{1 - |\rho^*|}\right]^l = \frac{1 - |\rho^*|}{1 - |\delta^*| - |\rho^*|} < \infty,$$

and the argument for $\limsup_{N\to\infty} \sum_{l=0}^{\infty} (|\delta^*| \cdot \|G_N^*\|_1)^l \leq \frac{1-|\rho^*|}{1-|\delta^*|-|\rho^*|} < \infty$ is the same.

To state Assumption G, we first introduce some notations for ease of exposition. Let $\ddot{Y} = (\ddot{y}_{it})_{N\times T}$ be the data matrix for $\ddot{y}_{it}$ with $\ddot{y}_{it} = \sum_{j=1}^N w_{ij,N} \dot{y}_{jt}$ and $\dot{y}_{jt} = y_{jt} - T^{-1}\sum_{s=1}^T y_{js}$, $\dot{Y}_{-1} = (\dot{y}_{it-1})_{N\times T}$ with $\dot{y}_{it-1} = y_{it-1} - T^{-1}\sum_{s=1}^T y_{is-1}$ and $\dot{X}_1, \dot{X}_2, \ldots, \dot{X}_k$ be defined similarly as $\dot{Y}_{-1}$. Furthermore, let $(k+1)\times(k+1)$ matrix D_b be defined as

$$\mathrm{D}_b = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}\dot{X}_kM_{F^*}) \\ \mathrm{tr}(\dot{X}_1'\ddot{M}\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{X}_1'\ddot{M}\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{X}_1'\ddot{M}\dot{X}_kM_{F^*}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\dot{X}_k'\ddot{M}\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{X}_k'\ddot{M}\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{X}_k'\ddot{M}\dot{X}_kM_{F^*}) \end{bmatrix}.$$

Assumption G: Let $S_N^* = W_N(I_N - \rho^* W_N)^{-1}$ and $S_{ij,N}^*$ be the (i,j)th element of S_N^*. Let be parameters space for Λ and Σ_{ee}, which satisfies the normalization conditions, i.e.,

$$= \left\{(\Lambda, \Sigma_{ee}) \;\middle|\; C^{-1} \leq \sigma_i^2 \leq C, \quad i; \text{ and } \frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = I_r\right\},$$

We assume one of the following conditions:

(i) $\delta^* = 0$ or $\beta^* = 0$. Let $\widetilde{Y} = \delta^*\dot{Y}_{-1} + \sum_{p=1}^k \beta_p^* \dot{X}_p$ and

$$\zeta = \left[\frac{1}{NT}\mathrm{tr}(\widetilde{Y}'\ddot{M}\dot{Y}_{-1}M_{F^*}), \frac{1}{NT}\mathrm{tr}(\widetilde{Y}'\ddot{M}\dot{X}_1M_{F^*}), \cdots, \frac{1}{NT}\mathrm{tr}(\widetilde{Y}'\ddot{M}\dot{X}_kM_{F^*})\right],$$

where ζ is a $(k+1)$-dimensional row vector. The matrix $\mathrm{D}_a = \begin{bmatrix} \frac{1}{NT}\mathrm{tr}(\widetilde{Y}'\ddot{M}\widetilde{Y}M_{F^*}) & \zeta \\ \zeta' & \mathrm{D}_b \end{bmatrix}$ is positive definite on , where $M_{F^*} = I_T - F^*(F^{*\prime}F^*)^{-1}F^{*\prime}$ and $\ddot{M} = \Sigma_{ee}^{-1} - N^{-1}\Sigma_{ee}^{-1}\Lambda\Lambda'\Sigma_{ee}^{-1}$.

(ii) For all ρ S_ρ and all N,

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\Big(S_{ij,N}^*\sigma_j^{*2} + S_{ji,N}^*\sigma_i^{*2} - (\rho-\rho^*)\sum_{p=1}^{N}S_{ip,N}^*S_{jp,N}^*\sigma_p^{*2}\Big)^2 > 0, \tag{3.4}$$

and D_b is positive definite on , where $\ddot{M}$ and M_{F^*} are defined the same as in (i).

Remark 3.3 Assumption G imposes the conditions for the identification of ρ and δ, β. The identification for the coefficient of spatial term is a non-trivial problem in spatial econometrics. This problem is investigated in a thorough way in Lee (2004a). Assumption G(i) can be viewed as a version of Assumption 8 of Lee (2004a) in the common shocks setting. Since the identification of ρ in Assumption G(i) depends on the underlying value of δ and β, it is a local identification condition. In contrast, Assumption G(ii) is a global identification condition. Condition (3.4) corresponds to Assumption 9 in Lee (2004a) and the condition in Theorem 2 of Yu et al. (2008), but it is different from theirs because we allow heteroskedasticity. To see this, we show in Appendix A that condition (3.4) is related to the unique solution of $T_{1N}(\rho, \sigma_1^2, \dots, \sigma_N^2) = 0$ with

$$T_{1N}(\rho, \sigma_1^2, \dots, \sigma_N^2) = -\frac{1}{2N}\mathrm{tr}[R\Sigma_{ee}^*R'\Sigma_{ee}^{-1}] + \frac{1}{2N}\ln|R\Sigma_{ee}^*R'\Sigma_{ee}^{-1}| + \frac{1}{2},$$

where $R = (I_N - \rho W_N)(I_N - \rho^* W_N)^{-1}$. When homoskedasticity is assumed, T_{1N} reduces to $T_{1,n}$ in Yu et al. (2008). After concentrating out the common variance σ^2, $T_{1,n}$ leads to Assumption 9 in Lee (2004a) and the assumption of Theorem 2 in Yu et al. (2008). Because of heteroskedasticity our identification condition takes a different form.

Assumption H: The parameters ω and σ_i^2 for $i = 1, 2, \dots, N$ are estimated in compact sets.

Remark 3.4 Assumption H assumes that partial parameters are estimated in compact sets. This assumption guarantees that the maximizer of the objective function is well defined. In pure factor analysis, it is known that the global maximizer of the quasi likelihood function with allowance of cross sectional heteroskedasticity do not exist, but the local maximizers are well defined and are consistent estimators for the underlying parameters under large N and large T, see, e.g., Andreson (2003). The objective function in the present paper is an extended version of the one in pure factor models and inherits the same problem. We therefore impose Assumption H to confine our analysis on local maximizers.

4 Objective function and first order conditions

Let $Z_t(\alpha, \omega, \Lambda, F) = Y_t - \alpha - \rho W_N Y_t - \delta Y_{t-1} - X_t\beta - \Lambda f_t$ with $\omega = (\rho, \delta, \beta')'$. Conditional on Y_0 which we assume are observed, the quasi likelihood function, by assuming the normality of e_{it}, is

$$L^*(\theta) = -\frac{1}{2NT}\sum_{t=1}^{T} Z_t(\alpha,\omega,\Lambda,F)'\Sigma_{ee}^{-1}Z_t(\alpha,\omega,\Lambda,F) - \frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|I_N - \rho W_N|.$$

where $\theta = (\omega, \Lambda, \mathrm{diag}(\Sigma_{ee})$.[2] Given Σ_{ee}, ω and Λ, it is seen that α and f_t maximize the above function at

$$\alpha = \overline{Y} - \rho W\overline{Y} - \delta\overline{Y}_{-1} - \overline{X}\beta - \Lambda\overline{f}$$

and

$$f_t = (\Lambda'\Sigma_{ee}^{-1}\Lambda)^{-1}\Lambda'\Sigma_{ee}^{-1}(\dot{Y}_t - \rho W\dot{Y}_t - \delta\dot{Y}_{t-1} - \dot{X}_t\beta).$$

Substituting the above two equation into the preceding likelihood function to concentrate out α and f_t, the objective function can therefore be simplified as

$$\begin{aligned} L(\theta) = & -\frac{1}{2NT}\sum_{t=1}^{T}(\dot{Y}_t - \rho\ddot{Y}_t - \delta\dot{Y}_{t-1} - \dot{X}_t\beta)'\ddot{M}(\dot{Y}_t - \rho\ddot{Y}_t - \delta\dot{Y}_{t-1} - \dot{X}_t\beta) \\ & - \frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|I_N - \rho W_N|. \end{aligned}$$

where $\ddot{M} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}\Lambda(\Lambda'\Sigma_{ee}^{-1}\Lambda)^{-1}\Lambda'\Sigma_{ee}^{-1} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Lambda\Lambda'\Sigma_{ee}^{-1}$ and $\ddot{Y}_t = W_N\dot{Y}_t$. The maximizer, defined by

$$\hat{\theta} = \underset{\theta\in\Theta}{\mathrm{argmax}}\ L(\theta),$$

is referred to as the quasi maximum likelihood estimator or MLE, where Θ is the parameters space specified by Assumptions G and H. More specifically, Θ is defined as

$$\Theta = \left\{\theta = (\omega, \Sigma_{ee}, \Lambda)\ \middle|\ \omega \le C;\ C^{-1} \le \sigma_i^2 \le C,\ \ i;\ \frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = I_r\right\}.$$

The first order condition for Λ gives

$$\left[\frac{1}{NT}\sum_{t=1}^{T}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta})(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta})'\right]\hat{\Sigma}_{ee}^{-1}\hat{\Lambda} = \hat{\Lambda}\hat{V}. \qquad (4.1)$$

where $\hat{V}$ is a diagonal matrix. The first order condition for σ_i^2 gives

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\left[\dot{y}_{it} - \hat{\rho}\ddot{y}_{it} - \hat{\delta}\dot{y}_{it-1} - \dot{x}_{it}'\hat{\beta} - \hat{\lambda}_i'\hat{f}_t\right]^2$$

where $\ddot{y}_{it} = \sum_{j=1}^{N} w_{ij,N}\dot{y}_{jt}$ and

$$\hat{f}_t = (\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta}) = \frac{1}{N}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta}).$$

The first order condition for ρ is

$$\frac{1}{NT}\sum_{t=1}^{T}\ddot{Y}_t'\widehat{\ddot{M}}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta}) - \frac{1}{N}\mathrm{tr}[W_N(I_N - \hat{\rho}W_N)^{-1}] = 0.$$

[2] Strictly speaking, θ should be written as θ_N since it also depends on N. But we drop this dependence from the symbol for notational simplicity. The symbols Θ and below are treated in a similar way.

The first order condition for δ is

$$\frac{1}{NT}\sum_{t=1}^{T}\dot{Y}_{t-1}'\widehat{\ddot{M}}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta}) = 0.$$

The first order condition for β is

$$\frac{1}{NT}\sum_{t=1}^{T}\dot{X}_t'\widehat{\ddot{M}}(\dot{Y}_t - \hat{\rho}\ddot{Y}_t - \hat{\delta}\dot{Y}_{t-1} - \dot{X}_t\hat{\beta}) = 0.$$

We emphasize that in computing the MLE, we do not need to solve the first order conditions. They are for theoretical analysis.

5 Asymptotic properties of the MLE

In this section, we first show that the MLE is consistent, we then derive the convergence rates, the asymptotic representation and the limiting distributions.

Proposition 5.1 *Under Assumptions A-H, when $N, T \to \infty$[3], we have*

$$\hat{\omega} \xrightarrow{p} \omega^*;$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^{*2})^2 \xrightarrow{p} 0;$$

$$\frac{1}{N}\Lambda^{*\prime}\widehat{\ddot{M}}\Lambda^* \xrightarrow{p} 0.$$

where $\omega^* = (\rho^*, \delta^*, \beta^{*\prime})'$ and $\widehat{\ddot{M}} = \hat{\Sigma}_{ee}^{-1} - N^{-1}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}$.

In the analysis of panel data models with common shocks but without spatial effects, a difficult problem is to establish consistency. The parameters of interest (δ, β) are simultaneously estimated with high dimensional nuisance parameters Λ and Σ_{ee}. The usual arguments need some modifications to accommodate this feature. The presence of spatial effects further compounds the difficult. Our proof of Proposition 5.1 consists of three steps. First we show there exists a function $L_1(\theta)$ such that

$$\sup_{\theta\in\Theta}|L(\theta) - L_1(\theta)| \xrightarrow{p} 0.$$

Then we show that the function $L_1(\theta)$ possesses the property that there exists an $\epsilon > 0$, which depends on the $N^c(\omega^*)$, such that

$$\sup_{(\Lambda,\Sigma_{ee})\in\Im}\ \sup_{\omega\in\mathcal{N}^c(\omega^*)} L_1(\theta) - L_1(\theta^*) < -\epsilon,$$

[3] In this paper, when we say the limit we mean the joint limit, which is the limit by letting N and T pass to infinity simultaneously, without naming the order that which index diverges first and which one diverges next. The latter case is called the sequential limit in the literature. Readers are referred to Phillips and Moon (1999) for a formal and precise definition of the two types of limit. See also the definition O_p and o_p in Appendix A.

where $N^c(\omega^*)$ is the complement of an open neighborhood of ω^*. Given the above two results, we have $\hat\omega \xrightarrow{p} \omega^*$. After obtaining the consistency of $\hat\omega$, in the third step we show the remaining two results in Proposition 5.1.

Notice that ω is low-dimensional but Σ_{ee} and Λ are high dimensional. So the usual consistency concept applies for ω. But for Σ_{ee} and Λ, their consistencies can only be defined under some chosen norm. The second result is equivalent to $\frac{1}{N}\ \hat\Sigma_{ee} - \Sigma^*_{ee}\ ^2 \xrightarrow{p} 0$. So the chosen norm is dimension-adjusted frobenious norm. The norm used in the last result can be viewed as a extension of generalized square coefficient between two high-dimensional vectors. We choose this norm to take account of rotational indeterminacy on factor and factor loadings, see Bai and Li (2012) for discussions on rotational indeterminacy in factor analysis.

The consistency result allows us to further derive the rates of convergence.

Theorem 5.1 *Let* $H = \frac{1}{NT}\hat V^{-1}(\hat\Lambda'\hat\Sigma_{ee}^{-1}\Lambda^*)(F^{*\prime}F^*)$. *Under Assumptions A-H, when* $N, T \to \infty$, *we have*

$$\frac{1}{N}\sum_{i=1}^{N}\ \hat\lambda_i - H\lambda_i^*\ ^2 = O_p(N^{-2}) + O_p(T^{-1});$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat\sigma_i^2 - \sigma_i^{*2})^2 = O_p(N^{-2}) + O_p(T^{-1});$$

$$\hat\omega - \omega^* = O_p(N^{-1}) + O_p(T^{-1}).$$

where $\hat V$ *is defined in (4.1).*

It is well documented in econometric literature that the MLE for dynamic panel data models has a $O(\frac{1}{T})$ bias term, see, for example, Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003). The case with inclusion of spatial term and lag spatial term has been investigated by Yu et al. (2008), which shows that the bias term is still $O(\frac{1}{T})$ but the expression is related with spatial weights matrix. This bias term is inherited by our MLE, as we can see that model (1.1) is an extension of classical spatial dynamic models. Apart from this $O(\frac{1}{T})$ bias term, our analysis indicates that there is another $O(\frac{1}{N})$ bias arising from common shocks part Λf_t. The presence of biases in the MLE is due to incidental parameters problem, see Neyman and Scott (1948) for a general discussion.

To state the asymptotic properties of the MLE, we define the following notations:

$$B_t = \sum_{l=0}^{\infty}(\delta^* G_N^*)^l G_N^* \dot X_{t-l}\beta^* + \sum_{l=0}^{\infty}(\delta^* G_N^*)^l G_N^* \Lambda^* \dot f_{t-l}^*, \qquad \dot B_t = B_t - \frac{1}{T}\sum_{s=1}^{T} B_s$$

$$\ddot B_t = W_N \dot B_t, \qquad Q_t = \sum_{l=0}^{\infty}(\delta^* G_N^*)^l G_N^* e_{t-l}, \qquad J_t = S_N^* \sum_{l=1}^{\infty}(\delta^* G_N^*)^l e_{t-l}.$$

Now we state the main theorem in this paper, which gives the asymptotic representation of $\hat\omega - \omega$.

Theorem 5.2 *Under Assumptions A-H, when* $N, T \to \infty$ *and* $\sqrt{N}/T \to 0$, $\sqrt{T}/N \to 0$, *we have*

$$\sqrt{NT}(\hat\omega - \omega^* + b) = \mathrm{D}^{-1}\xi + o_p(1),$$

where

$$\xi = \frac{1}{\sqrt{NT}} \begin{bmatrix} \sum_{t=1}^{T} \ddot{B}_t' \ddot{M}^* e_t - \sum_{t=1}^{T}\sum_{s=1}^{T} \ddot{B}_t' \ddot{M}^* e_s \pi_{st}^* + \sum_{t=1}^{T} J_t' \Sigma_{ee}^{*-1} e_t + \eta \\ \sum_{t=1}^{T} \dot{B}_{t-1}' \ddot{M}^* e_t - \sum_{t=1}^{T}\sum_{s=1}^{T} \dot{B}_{t-1}' \ddot{M}^* e_s \pi_{st}^* + \sum_{t=1}^{T} Q_{t-1}' \Sigma_{ee}^{*-1} e_t \\ \sum_{t=1}^{T} \dot{X}_t' \ddot{M}^* e_t - \sum_{t=1}^{T}\sum_{s=1}^{T} \dot{X}_t' \ddot{M}^* e_s \pi_{st}^* \end{bmatrix}$$

with $\pi_{st}^* = f_s^{*\prime}(F^{*\prime}F^*)^{-1} f_t^*$ *and* $\ddot{M}^* = \Sigma_{ee}^{*-1} - \frac{1}{N}\Sigma_{ee}^{*-1}\Lambda^*\Lambda^{*\prime}\Sigma_{ee}^{*-1}$. *The* $(k+2)\times(k+2)$ *matrix* D *is defined as*

$$\mathrm{D} = \frac{1}{NT} \times \begin{bmatrix} \mathrm{tr}(\ddot{Y}'\ddot{M}^*\ddot{Y}M_{F^*})+\Phi & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \mathrm{tr}(\dot{X}_1'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\dot{X}_1'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{X}_1'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{X}_1'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\dot{X}_k'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\dot{X}_k'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{X}_k'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{X}_k'\ddot{M}^*\dot{X}_kM_{F^*}) \end{bmatrix}$$

with $\Phi = T[\mathrm{tr}(S_N^{*2}) - 2\sum_{i=1}^{N} S_{ii,N}^{*2}]$. *The* $(k+2)$*-dimensional vector* b *is defined as*

$$b = \mathrm{D}^{-1} \begin{bmatrix} \frac{1}{NT}\mathrm{tr}[\delta^* S_N^* G_N^*(I_N - \delta^* G_N^*)^{-1}] + \frac{1}{N}\mathrm{tr}[\Lambda^{*\prime} S_N^{\circ\prime} \Sigma_{ee}^{*-1}\Lambda^*(\Lambda^{*\prime}\Sigma_{ee}^{*-1}\Lambda^*)^{-1}] \\ \frac{1}{NT}\mathrm{tr}[G_N^*(I_N - \delta^* G_N^*)^{-1}] \\ 0_{k\times 1} \end{bmatrix}$$

and

$$\eta = \sum_{t=1}^{T} e_t' S_N^{*\circ\prime}\Sigma_{ee}^{*-1} e_t = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^{*2}} 1(i=j) e_{it} e_{jt} S_{ij,N}^*.$$

Here $S_N^{*\circ}$ *is an* $N \times N$ *matrix which is obtained by setting all the diagonal elements of* S_N^* *to zeros.*

Although ξ has a relatively complicated expression, it can be shown that $\mathrm{D}^{-1/2}\xi \xrightarrow{d} N(0,1)$ by resorting to the martingale difference central limit theorem (see Corollary 3.1 in Hall and Heyde (1980)). Given this result, we have the following corollary.

Corollary 5.1 *Under the assumptions in Theorem 5.2, when* $N, T \to \infty$ *and* $N/T \to \kappa^2$, *we have*

$$\sqrt{NT}(\hat{\omega} - \omega^*) \xrightarrow{d} N\Big(-b^{\diamond}, \Big[\operatorname*{plim}_{N,T\to\infty} \mathrm{D}\Big]^{-1}\Big),$$

where

$$b^{\diamond} = \operatorname*{plim}_{N,T\to\infty} \left\{ \mathrm{D}^{-1} \begin{bmatrix} \kappa\frac{1}{N}\mathrm{tr}[\delta^* S_N^* G_N^*(I_N - \delta^* G_N^*)^{-1}] + \frac{1}{\kappa}\mathrm{tr}[\Lambda^{*\prime} S_N^{\circ\prime} \Sigma_{ee}^{*-1}\Lambda^*(\Lambda^{*\prime}\Sigma_{ee}^{*-1}\Lambda^*)^{-1}] \\ \kappa\frac{1}{N}\mathrm{tr}[G_N^*(I_N - \delta^* G_N^*)^{-1}] \\ 0_{k\times 1} \end{bmatrix} \right\}.$$

Theorem 5.2 include some important models as special cases. If there are no lag dependent term and spatial term in model (1.1), i.e.,

$$y_{it} = \alpha_i + x_{it}'\beta + \lambda_i' f_t + e_{it},$$

the present analysis indicates that under $\sqrt{N}/T \to 0$, $\sqrt{T}/N \to 0$ as well as other regularity conditions, the asymptotic representation of $\hat{\beta} - \beta$ is

$$\sqrt{NT}(\hat{\beta} - \beta^*) = \mathsf{D}_{\beta}^{-1} \frac{1}{\sqrt{NT}} \left(\sum_{t=1}^{T} \dot{X}_t' \ddot{M}^* e_t - \sum_{t=1}^{T} \sum_{s=1}^{T} \dot{X}_t' \ddot{M}^* e_s \pi_{st}^* \right) + o_p(1),$$

where

$$\mathsf{D}_{\beta} = \frac{1}{NT} \Bigg[$$

where we use $[M]_{ij}$ to denote the (i, j)th element of M. So the marginal e□ects of $x_{j(t-s)p}$ and $e_{j(t-s)}$ on y_{it} can be estimated according to the above formulas by plug-in method. The limiting distributions of the marginal e□ects can be easily calculated by the delta method via Theorem 5.2.

Remark 5.3 The limiting variance and the bias term can be estimated by plug-in method. More specifically, matrix D can be consistently estimated by

$$\widehat{\mathrm{D}} = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}(\ddot{Y}'\widehat{\ddot{M}}\ddot{Y}M_{\hat{F}}) + \hat{\Phi} & \mathrm{tr}(\ddot{Y}'\widehat{\ddot{M}}\dot{Y}_{-1}M_{\hat{F}}) & \mathrm{tr}(\ddot{Y}'\widehat{\ddot{M}}\dot{X}_1M_{\hat{F}}) & \cdots & \mathrm{tr}(\ddot{Y}'\widehat{\ddot{M}}\dot{X}_kM_{\hat{F}}) \\ \mathrm{tr}(\dot{Y}_{-1}'\widehat{\ddot{M}}\ddot{Y}M_{\hat{F}}) & \mathrm{tr}(\dot{Y}_{-1}'\widehat{\ddot{M}}\dot{Y}_{-1}M_{\hat{F}}) & \mathrm{tr}(\dot{Y}_{-1}'\widehat{\ddot{M}}\dot{X}_1M_{\hat{F}}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'\widehat{\ddot{M}}\dot{X}_kM_{\hat{F}}) \\ \mathrm{tr}(\dot{X}_1'\widehat{\ddot{M}}\ddot{Y}M_{\hat{F}}) & \mathrm{tr}(\dot{X}_1'\widehat{\ddot{M}}\dot{Y}_{-1}M_{\hat{F}}) & \mathrm{tr}(\dot{X}_1'\widehat{\ddot{M}}\dot{X}_1M_{\hat{F}}) & \cdots & \mathrm{tr}(\dot{X}_1'\widehat{\ddot{M}}\dot{X}_kM_{\hat{F}}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\dot{X}_k'\widehat{\ddot{M}}\ddot{Y}M_{\hat{F}}) & \mathrm{tr}(\dot{X}_k'\widehat{\ddot{M}}\dot{Y}_{-1}M_{\hat{F}}) & \mathrm{tr}(\dot{X}_k'\widehat{\ddot{M}}\dot{X}_1M_{\hat{F}}) & \cdots & \mathrm{tr}(\dot{X}_k'\widehat{\ddot{M}}\dot{X}_kM_{\hat{F}}) \end{bmatrix}$$

where

$$\hat{F} = \frac{1}{N}\Big(\dot{Y} - \hat{\delta}\dot{Y}_{-1} - \hat{\rho}\ddot{Y} - \sum_{p=1}^{k}\dot{X}_p\hat{\beta}_p\Big)'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}$$

and $\hat{\Phi} = T \cdot \mathrm{tr}[(\hat{S}_N^2) - 2\sum_{i=1}^{N}\hat{S}_{ii,N}^2]$ with $\hat{S}_N = W_N\hat{G}_N$, $\hat{G}_N = (I_N - \hat{\rho}W_N)^{-1}$ and $\hat{S}_{ii,N}$ being the ith diagonal element of $\hat{S}_N$. In addition, the bias term ξ can be consistently estimated by

$$\hat{b} = \hat{\mathrm{D}}^{-1}\begin{bmatrix} \frac{1}{NT}\mathrm{tr}[\hat{\delta}\hat{S}_N\hat{G}_N(I_N - \hat{\delta}\hat{G}_N)^{-1}] + \frac{1}{N}\mathrm{tr}[\hat{\Lambda}'\hat{S}_N^{\circ\prime}\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1}] \\ \frac{1}{NT}\mathrm{tr}[\hat{G}_N(I_N - \hat{\delta}\hat{G}_N)^{-1}] \\ 0_{k\times 1} \end{bmatrix}.$$

Bias correction in the simulation uses this formula.

6 Discussions on spatial models with heteroskedasticity

Allowance of heteroskedasticy in pure spatial models is of theoretical and practical relevance. As pointed out by Kelejian and Prucha (2010) and Lin and Lee (2010) among others, if heteroskedasticity exists but homoskedasticity is imposed, the MLE generally is inconsistent. In viewpoint of applied studies, assuming homoskedasticity seems too restrictive to be true. However, to the best of our knowledge, the MLE under heteroskedasticity has not been investigated so far in the literature. In this section, we give some discussions on this issue, which is of independent interest.

6.1 Dynamic spatial models

Consider the following dynamic spatial model,

$$y_{it} = \alpha_i + \rho\sum_{j=1}^{N} w_{ij,N}y_{jt} + \delta y$$

The above model is special case of model (1.1). Under some regularity conditions stated in Section 2, the analysis of Theorem 5.2 indicates that the MLE for (6.1) has the following asymptotic representation:

$$\sqrt{NT}(\hat{\omega} - \omega + \upsilon_1) = D_1^{-1}\frac{1}{\sqrt{NT}}\begin{bmatrix} \sum_{t=1}^{T}(\ddot{B}_t + J_t + S_N^{*\circ}e_t)'\Sigma_{ee}^{*-1}e_t \\ \sum_{t=1}^{T}(\dot{B}_{t-1} + J_{t-1})'\Sigma_{ee}^{*-1}e_t \\ \sum_{t=1}^{T}\dot{X}_t'\Sigma_{ee}^{*-1}e_t \end{bmatrix} + o_p(1), \qquad (6.2)$$

where

$$D_1 = \frac{1}{NT}\begin{bmatrix} \sum_{t=1}^{T}\ddot{Y}_t'\Sigma_{ee}^{*-1}\ddot{Y}_t + \Phi & \sum_{t=1}^{T}\ddot{Y}_t'\Sigma_{ee}^{*-1}\dot{Y}_{t-1} & \sum_{t=1}^{T}\ddot{Y}_t'\Sigma_{ee}^{*-1}\dot{X}_t \\ \sum_{t=1}^{T}\dot{Y}_{t-1}'\Sigma_{ee}^{*-1}\ddot{Y}_t & \sum_{t=1}^{T}\dot{Y}_{t-1}'\Sigma_{ee}^{*-1}\dot{Y}_{t-1} & \sum_{t=1}^{T}\dot{Y}_{t-1}'\Sigma_{ee}^{*-1}\dot{X}_t \\ \sum_{t=1}^{T}\dot{X}_t'\Sigma_{ee}^{*-1}\ddot{Y}_t & \sum_{t=1}^{T}\dot{X}_t'\Sigma_{ee}^{*-1}\dot{Y}_{t-1} & \sum_{t=1}^{T}\dot{X}_t'\Sigma_{ee}^{*-1}\dot{X}_t \end{bmatrix},$$

with Φ defined the same as in Theorem 5.2 and

$$\upsilon_1 = D_1^{-1}\begin{bmatrix} \frac{1}{NT}\mathrm{tr}[\delta^* S_N^* G_N^*(I_N - \delta^* G_N^*)^{-1}] \\ \frac{1}{NT}\mathrm{tr}[G_N^*(I_N - \delta^* G_N^*)^{-1}] \\ 0_{k\times 1} \end{bmatrix}.$$

Given the above asymptotic representation, invoking the central limiting theorem for quadratic form (Kelejian and Prucha (2001), Giraitis and Taqqu (1998)), we have

$$\sqrt{NT}(\hat{\omega} - \omega^* + \upsilon_1) \xrightarrow{d} N\Big(0, \Big[\operatorname*{plim}_{N,T\to\infty} D_1\Big]^{-1}\Big).$$

6.2 Spatial panel data models with SAR disturbances

Another interesting spatial model, which receive much attention in practice, is spatial panel data model with SAR disturbances, i.e.,

$$\begin{aligned} Y_t &= \alpha + \rho W_N Y_t + X_t\beta + u_t; \\ u_t &= \ M_N u_t + e_t. \end{aligned} \qquad (6.3)$$

where M_N is another spatial weights matrix. Lee and Yu (2010) make a rigorous analysis for the ML estimation of (6.3) under the assumption that e_{it} is cross-sectionally homoskedastic. Using the method in this paper to deal with high dimensional variance parameters, [4] we can extend Lee and Yu's analysis to heteroskedasticity. For ease of exposition, we further introduce the following notations. Let

$$F = M_N S_N^*(I_N - \ {}^*M_N)^{-1}, \ \ G = (I_N - \ {}^*M_N)S_N^*(I_N - \ {}^*M_N)^{-1}, \ \ H = M_N(I_N - \ {}^*M_N)^{-1};$$

$$P_t = (I_N - \ {}^*M_N)W_N\dot{Y}_t, \quad Q_t = M_N[(I_N - \rho^* W_N)\dot{Y}_t - \dot{X}_t\beta^*], \quad R_t = (I_N - \ {}^*M_N)\dot{X}_t.$$

Define the $(k + 2) \times (k + 2)$ matrix D_2 as

$$D_2 = \frac{1}{NT}\begin{bmatrix} \sum_{t=1}^{T}P_t'\Sigma_{ee}^{-1}P_t + \varsigma_1 & \sum_{t=1}^{T}P_t'\Sigma_{ee}^{-1}Q_t + \varsigma_2 & \sum_{t=1}^{T}P_t'\Sigma_{ee}^{-1}R_t \\ \sum_{t=1}^{T}Q_t'\Sigma_{ee}^{-1}P_t + \varsigma_2 & \sum_{t=1}^{T}Q_t'\Sigma_{ee}^{-1}Q_t + \varsigma_3 & \sum_{t=1}^{T}Q_t'\Sigma_{ee}^{-1}R_t \\ \sum_{t=1}^{T}R_t'\Sigma_{ee}^{-1}P_t & \sum_{t=1}^{T}R_t'\Sigma_{ee}^{-1}Q_t & \sum_{t=1}^{T}R_t'\Sigma_{ee}^{-1}R_t \end{bmatrix}$$

[4] The method to deal with high dimensional variance parameters σ_i^2 is as follows: First show $\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 = o_p(1)$, see Proposition 5.1; then derive its convergence rate, see Propositions B.4 and B.6; then use this result to show that the magnitude of the di□erence between the term involving $\hat{\Sigma}_{ee}$ and the term involving Σ_{ee} is asymptotically negligible.

with $\varsigma_1 = T[\mathrm{tr}(G^2) - 2\mathrm{tr}(G \circ G)]$, $\varsigma_2 = T[\mathrm{tr}(F) - 2\mathrm{tr}(G \circ H)]$ and $\varsigma_3 = T[\mathrm{tr}(H^2) - 2\mathrm{tr}(H \circ H)]$, where "$\circ$" denotes the Hadamard product.

Under some regularity conditions, we can show that the MLE for $\omega = (\rho, \ , \beta')'$ in (6.3) under cross sectional heteroskedasticity has the following asymptotic representation,

$$\sqrt{NT}(\hat{\omega} - \omega^*) = D_2^{-1} \frac{1}{\sqrt{NT}} \begin{bmatrix} \sum_{t=1}^T [\beta^{*\prime} \dot{X}_t' S_N^{*\prime} (I_N - \ ^* M_N)' + e_t' G^{\circ\prime}] \Sigma_{ee}^{*-1} e_t \\ \sum_{t=1}^T e_t' H^{\circ\prime} \Sigma_{ee}^{*-1} e_t \\ \sum_{t=1}^T \dot{X}_t' (I_N - \ ^* M_N)' \Sigma_{ee}^{*-1} e_t \end{bmatrix} + o_p(1), \quad (6.4)$$

where G° and H° are defined similarly as $S_N^{*\circ}$. Given the above result, invoking the central limit theorem for quadratic form, we have

$$\sqrt{NT}(\hat{\omega} - \omega) \xrightarrow{d} N\Big(0, \Big[\operatorname*{plim}_{N,T\to\infty} D_2\Big]^{-1}\Big).$$

6.3 Homoskedasticity versus heteroskedasticity

It is seen from the above that the limiting variance of the MLE is not a sandwich form. This result contrasts with the existing results in the literature such as Yu et al. (2008) and Lee and Yu (2010), in which the limiting variance of the MLE has a sandwich formula. The reason for the difference is the heteroskedasticity estimation. In the present paper we allow cross-sectional heteroskedasticity, while Yu et al. (2008) assume homoskedasticity. Under heteroskedasticity, the asymptotic expression does not involve e_{it}^2, as seen in (6.2) and (6.4). But under homoskedasticity, the situation is different. Still consider model (6.1). If homoskedasticity is assumed and is imposed in estimation (let $\sigma^{*2} = E(e_{it}^2)$), the asymptotic expression for the MLE is

$$\sqrt{NT}(\tilde{\omega} - \omega^* + \upsilon_2) = D_3^{-1} \frac{1}{\sqrt{NT}\sigma^{*2}} \begin{bmatrix} \sum_{t=1}^T e_t' S_N^{*\circ\prime} e_t + \sum_{t=1}^T J_t' e_t + \sum_{t=1}^T \ddot{B}_t^{\star\prime} e_t + \vartheta \\ \sum_{t=1}^T \dot{B}_{t-1}^{\star\prime} e_t + \sum_{t=1}^T Q_{t-1}' e_t \\ \sum_{t=1}^T \dot{X}_t' e_t \end{bmatrix} + o_p(1),$$

where

$$\vartheta = \sum_{i=1}^N \sum_{t=1}^T \Big[S_{ii,N}^* - \frac{1}{N}\mathrm{tr}(S_N^*)\Big](e_{it}^2 - \sigma^{*2}), \quad B_t^\star = \sum_{l=0}^\infty (\delta^* G_N^*)^l X_{t-l} \beta^*, \quad \ddot{B}_t^\star = W_N \dot{B}_t^\star,$$

$$\upsilon_2 = D_3^{-1} \Big[\frac{1}{N}\mathrm{tr}\Big(\delta^* S_N^* G_N^* (I_N - \delta^* G_N^*)^{-1}\Big), \ \frac{1}{N}\mathrm{tr}\Big(G_N^* (I_N - \delta^* G_N^*)^{-1}\Big), 0_{1\times k}\Big]',$$

and

$$D_3 = \frac{1}{NT\sigma^{*2}} \begin{bmatrix} \sum_{t=1}^T \ddot{Y}_t' \ddot{Y}_t + T\sigma^{*2}\{\mathrm{tr}(S_N^{*2}) - \frac{2}{N}[\mathrm{tr}(S_N^*)]^2\} & \sum_{t=1}^T \ddot{Y}_t' \dot{Y}_{t-1} & \sum_{t=1}^T \ddot{Y}_t' \dot{X}_t \\ \sum_{t=1}^T \dot{Y}_{t-1}' \ddot{Y}_t & \sum_{t=1}^T \dot{Y}_{t-1}' \dot{Y}_{t-1} & \sum_{t=1}^T \dot{Y}_{t-1}' \dot{X}_t \\ \sum_{t=1}^T \dot{X}_t' \ddot{Y}_t & \sum_{t=1}^T \dot{X}_t' \dot{Y}_{t-1} & \sum_{t=1}^T \dot{X}_t' \dot{X}_t \end{bmatrix}$$

From the above, we can see that the asymptotic expression under the homoskedasticity involves e_{it}^2. So the limiting variance of $\tilde{\omega} - \omega^*$ will depend on the kurtosis of e_{it}. Because D_3 does not depend on the kurtosis, the limiting variance of $\tilde{\omega} - \omega^*$ has a sandwich formula. In contrast, the MLE under heteroskedasticity has a limiting variance not of a sandwich form, regardless of normality. The same phenomenon also occurs for the spatial panel data models with SAR disturbances, see Lee and Yu (2010) for the asymptotic result of the MLE under homoskedasticity. This results is interesting. Thus estimating heteroskedasticity is desirable from two considerations: the limiting distribution is robust to the underlying distributions; it avoids potential inconsistency when homoskedasticity is incorrectly imposed.

7 Finite sample properties

In this section, we run Monte Carlo simulations to investigate the finite sample properties of the MLE. The data are generated according to

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \delta y_{it-1} + x_{it1}\beta_1 + x_{it2}\beta_2 + \lambda_i' f_t + e_{it}$$

with $(\rho, \delta, \beta_2, \beta_2) = (0.5, 0.4, 1, 2)$. The number of factors is fixed to 2. The explanatory variable x_{itp} is generated according to

$$x_{itp} = \left[(\lambda_i + \gamma_{ip})' f_t + u_{itp}\right] 1\left[(\lambda_i + \gamma_{ip})' f_t + u_{itp} \geq -3.5\right]$$

for $p = 1, 2$. All the elements of α_i, λ_i, f_t, γ_{ip} and u_{itp} are all generated independently from $N(0, 1)$. The way to generate the explanatory variables here is similar as in Moon and Weidner (2013). To generate the errors and heteroskedasticity, we consider the method similar as in Bai and Li (2014b). More specifically, we set $e_{it} = \sqrt{\psi_i}\varepsilon_{it}$ where ψ_i is defined as

$$\psi_i = 0.5 + \frac{1 - \nu_i}{\nu_i} \lambda_i' \lambda_i,$$

where ν_i is drawn independently from $U[0.2, 0.8]$. The error ε_{it} is equal to $(\chi_2^2 - 2)/2$, where χ_2^2 denotes the chi-squared distribution with two degrees of freedom, which is normalized to zero mean and unit variance.

The generated data exhibit heteroskedasticity. The generated x_{it} does not have a factor structure and is correlated with the factors and factor loadings, and the two regressors x_{it1} and x_{it2} are also correlated; the errors are non-normal and skewed.

The spatial weights matrices generated in the simulation are similar to Kelejian and Prucha (1999) and Kapoor et al. (2007). More specifically, all the units are arranged in a circle and each unit is aﬀected only by the q units immediately before it and immediately after it with equal weight. Following Kelejian and Prucha (1999), we normalize the spatial weights matrix by letting the sum of each row be equal to 1 (so the weight is $\frac{1}{2q}$) and call this specification of spatial weights matrix "q ahead and q behind."

Adapting a criterion in Bai and Li (2014b), the number of factors is determined by

$$\hat{r} = \underset{0 \leq m \leq r_{max}}{\text{argmin}} \; IC(m)$$

with

$$IC(m) = \frac{1}{2N} \sum_{i=1}^{N} \ln|(\hat{\sigma}_i^m)^2| - \frac{1}{N} \ln|I_N - \hat{\rho}^m W_N| + m \frac{N + T}{2NT} \ln[\min(N, T)].$$

and

$$(\hat{\sigma}_i^m)^2 = \frac{1}{T}\sum_{t=1}^{T}(\dot{y}_{it} - \hat{\rho}^m \ddot{y}_{it} - \hat{\delta}^m \dot{y}_{it-1} - \dot{x}_{it}'\hat{\beta} - \hat{\lambda}_i^{m\prime}\hat{f}_t^m)^2,$$

where the hat symbols with superscript "m" denotes the MLE when the number of factors is set to m. We set $r_{\max} = 4$.

The following four tables present the simulation results from 1000 repetitions under the combinations of $N = 100, 200, 300$ and $T = 50, 100, 150$. Biases and root mean square errors (RMSE) are both computed to measure the performance. In all the simulations, the number of factors can be correctly estimated with probability almost one. The first two tables report the performance of the MLE before and after the bias correction under "1 ahead and 1 behind" spatial weights matrix. From Table 1, we see that the MLE are consistent. As the sample size becomes larger, the RMSEs of the MLE decrease stably. However, we also find that the ratio of the bias relative to the RMSE for the MLE of δ is considerably large, the ratio for ρ, albeit not as large as δ, is still pronounced, especially when N/T is large. This causes problems in statistical inference. We then investigates the performance of the bias-corrected MLE. From Table 2, we see that the bias-correct estimator performs well. The biases of the original estimators have been effectively reduced. The next two tables report the performance of the MLE under '3 ahead and 3 behind" spatial weights matrix. The simulation results are similar as the case under '1 ahead and 1 behind" weights matrix. So we do not repeat the detailed analysis.

Table 1: The performance of the MLE before bias correction with "1 ahead and 1 behind" spatial weights matrix

N	T	ρ		δ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	50	0.0007	0.0043	-0.0018	0.0042	0.0012	0.0160	-0.0001	0.0162
100	100	0.0002	0.0030	-0.0007	0.0027	0.0005	0.0113	0.0002	0.0109
100	150	0.0002	0.0023	-0.0006	0.0021	0.0002	0.0089	0.0002	0.0086
200	50	0.0008	0.0030	-0.0019	0.0031	-0.0003	0.0106	-0.0001	0.0110
200	100	0.0005	0.0021	-0.0010	0.0020	-0.0003	0.0077	-0.0004	0.0080
200	150	0.0003	0.0017	-0.0007	0.0015	-0.0001	0.0063	-0.0002	0.0064
300	50	0.0010	0.0026	-0.0020	0.0029	-0.0001	0.0091	-0.0010	0.0092
300	100	0.0004	0.0016	-0.0009	0.0016	-0.0004	0.0063	-0.0004	0.0063
300	150	0.0003	0.0013	-0.0007	0.0013	-0.0001	0.0050	0.0002	0.0051

Table 2: The performance of the MLE after bias correction with "1 ahead and 1 behind" spatial weights matrix

N	T	ρ		δ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	50	-0.0001	0.0042	0.0001	0.0037	0.0018	0.0160	0.0008	0.0162
100	100	-0.0002	0.0030	0.0002	0.0026	0.0008	0.0113	0.0007	0.0109
100	150	-0.0001	0.0023	0.0000	0.0020	0.0004	0.0089	0.0005	0.0086
200	50	-0.0000	0.0028	0.0000	0.0024	0.0003	0.0106	0.0008	0.0110
200	100	0.0000	0.0020	-0.0001	0.0017	-0.0000	0.0077	0.0001	0.0080
200	150	0.0000	0.0016	-0.0000	0.0014	0.0001	0.0063	0.0001	0.0064
300	50	0.0002	0.0023	-0.0001	0.0021	0.0005	0.0091	-0.0000	0.0091
300	100	-0.0000	0.0016	0.0001	0.0013	-0.0001	0.0063	0.0000	0.0062
300	150	-0.0000	0.0013	0.0000	0.0011	0.0001	0.0050	0.0002	0.0051

Table 3: The performance of the MLE before bias correction
with "3 ahead and 3 behind" spatial weights matrix

N	T	ρ		δ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	50	0.0012	0.0051	-0.0022	0.0045	-0.0002	0.0158	0.0010	0.0159
100	100	0.0006	0.0033	-0.0011	0.0029	0.0002	0.0106	0.0001	0.0103
100	150	0.0003	0.0027	-0.0007	0.0022	0.0001	0.0091	0.0000	0.0093
200	50	0.0013	0.0038	-0.0023	0.0036	-0.0002	0.0108	0.0001	0.0111
200	100	0.0006	0.0024	-0.0012	0.0022	-0.0000	0.0075	0.0002	0.0079
200	150	0.0004	0.0019	-0.0007	0.0016	0.0001	0.0063		

Theorem 8.1 *Under Assumptions A-E, F′, G-H, when $N, T \to \infty$, $\sqrt{N}/T \to 0$ and $\sqrt{T}/N \to 0$, we have*

$$\sqrt{NT}(\hat{\varphi} - \varphi^* + b_\phi) \xrightarrow{d} N\Big(0, \Big[\operatorname*{plim}_{N,T\to\infty} \mathrm{D}_\phi\Big]^{-1}\Big),$$

where

$$\mathrm{D}_\phi = \frac{1}{NT}$$
$$\times \begin{bmatrix} \mathrm{tr}(\ddot{Y}'\ddot{M}^*\ddot{Y}M_{F^*}) + \Phi & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\ddot{Y}_{-1}M_{F^*}) & \cdots & \mathrm{tr}(\ddot{Y}'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\ddot{Y}_{-1}M_{F^*}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \mathrm{tr}(\ddot{Y}_{-1}'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\ddot{Y}_{-1}'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\ddot{Y}_{-1}'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\ddot{Y}_{-1}'\ddot{M}^*\dot{X}_kM_{F^*}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\dot{X}_k'\ddot{M}^*\ddot{Y}M_{F^*}) & \mathrm{tr}(\dot{X}_k'\ddot{M}^*\dot{Y}_{-1}M_{F^*}) & \mathrm{tr}(\ddot{Y}_{-1}'\ddot{M}^*\dot{X}_1M_{F^*}) & \cdots & \mathrm{tr}(\dot{X}_k'\ddot{M}^*\dot{X}_kM_{F^*}) \end{bmatrix}$$

with Φ defined the same as in Theorem 5.2 and

$$b_\phi = \mathrm{D}_\phi^{-1} \begin{bmatrix} \frac{1}{NT}\mathrm{tr}[W_N(\delta^*G_N^* + \ \ ^*G_N^*M_N)(I_N - \delta^*G_N^* + \ \ ^*G_N^*M_N)^{-1}G_N^*] + \varsigma \\ \frac{1}{NT}\mathrm{tr}[(I_N - \delta^*G_N^* + \ \ ^*G_N^*M_N)^{-1}G_N^*] \\ \frac{1}{NT}\mathrm{tr}[M_N(I_N - \delta^*G_N^* + \ \ ^*G_N^*M_N)^{-1}G_N^*] \\ 0_{k\times 1} \end{bmatrix}$$

with $\varsigma = \frac{1}{N}\mathrm{tr}[\Lambda^{*\prime}S_N^{\circ\prime}\Sigma_{ee}^{*-1}\Lambda^*(\Lambda^{*\prime}\Sigma_{ee}^{*-1}\Lambda^*)^{-1}]$.

We use simulations to illustrate the performance of the MLE. The data are generated according to (8.1) with $(\rho, \delta, \ \ ) = (0.2, 0.4, 0.3)$. The factors, factor loadings, errors and heteroskedasticity are generated in the same way as in Section 7. Other prespecified parameters such as the number of factors, the number of regressors and the true values of β are also the same. W_N is a "3 ahead and 3 behind" weights matrix and M_N is a "1 ahead and 1 behind" one. For simplicity, the number of factors is assumed to be known. Tables 5 and 6 reports the simulation results based on 1000 repetitions.

Tables 5 and 6 show that the maximum likelihood method continue to perform well. The RMSE decreases as the sample size becomes larger, implying that the MLE is consistent. The bias has been e□ectively reduced after the bias correction.

Table 5: The performance of the MLE before bias correction

N	T	ρ		δ		ϱ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	50	0.0005	0.0068	-0.0025	0.0049	-0.0000	0.0043	0.0008	0.0155	0.0003	0.0157
100	100	0.0002	0.0044	-0.0012	0.0031	0.0001	0.0030	0.0005	0.0107	0.0001	0.0108
100	150	0.0004	0.0036	-0.0008	0.0024	-0.0000	0.0024	0.0004	0.0091	0.0002	0.0089
200	50	0.0009	0.0045	-0.0027	0.0040	0.0002	0.0029	0.0005	0.0110	-0.0007	0.0114
200	100	0.0004	0.0031	-0.0013	0.0024	0.0000	0.0020	-0.0001	0.0076	0.0002	0.0077
200	150	0.0004	0.0025	-0.0009	0.0018	0.0001	0.0017	0.0000	0.0060	-0.0004	0.0063
300	50	0.0010	0.0040	-0.0025	0.0034	-0.0001	0.0024	0.0001	0.0088	0.0001	0.0089
300	100	0.0004	0.0026	-0.0013	0.0021	0.0001	0.0017	-0.0000	0.0059	0.0003	0.0064
300	150	0.0004	0.0021	-0.0009	0.0016	-0.0000	0.0014	-0.0001	0.0049	0.0001	0.0051

Table 6: The performance of the MLE after bias correction

N	T	ρ		δ		ϱ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	50	-0.0001	0.0067	-0.0001	0.0042	-0.0001	0.0043	0.0011	0.0155	0.0007	0.0157
100	100	-0.0002	0.0044	0.0000	0.0028	0.0001	0.0030	0.0006	0.0107	0.0003	0.0108
100	150	0.0001	0.0035	-0.0000	0.0023	-0.0001	0.0024	0.0005	0.0091	0.0003	0.0089
200	50	0.0003	0.0044	-0.0003	0.0029	0.0001	0.0029	0.0008	0.0110	-0.0003	0.0113
200	100	0.0000	0.0031	-0.0001	0.0020	-0.0001	0.0020	0.0000	0.0076	0.0003	0.0077
200	150	0.0001	0.0025	-0.0001	0.0016	0.0000	0.0017	0.0001	0.0060	-0.0003	0.0063
300	50	0.0004	0.0039	-0.0001	0.0024	-0.0002	0.0024	0.0004	0.0088	0.0006	0.0089
300	100	0.0001	0.0026	-0.0000	0.0017	0.0000	0.0017	0.0001	0.0059	0.0005	0.0065
300	150	0.0001	0.0020	-0.0000	0.0013	-0.0001	0.0014	-0.0000	0.0049	0.0002	0.0051

The present analysis can be also extended to allow SAR disturbance. Suppose $e_t = M_N e_t + \varepsilon_t$, where ε satisfies Assumption C. Under this specification, e_t has weak cross sectional correlation. To derive a tractable expression, pre-multiplying $I_N - \quad M_N$ on both sides of (3.1), we have

$$Y_t = \alpha^\star + \rho W_N Y_t + \quad M_N Y_t - \rho \quad M_N W_N Y_t + \delta Y_{t-1} - \quad \delta M_N Y_{t-1} + X_t \beta - M_N X_t \beta \quad + \Lambda^\star f_t + \varepsilon_t$$

where $\alpha^\star = (I_N - \quad M_N)\alpha$ and $\Lambda^\star = (I_N - \quad M_N)\Lambda$. Now we see that the above model is similar as (3.1) except for high order spatial lags. The analysis of the MLE for the above model is similar as that of (3.1).

9 Conclusion

This paper considers spatial panel data models with common shocks, in which the spatial lag term is endogenous and the explanatory variables are correlated with the unobservable common factors and factor loadings. The proposed maximum likelihood estimator is capable of handling of both types of cross sectional dependence. The results make it possible to determine which type of cross-section dependence or both are present. Heteroskedasticity is explicitly allowed. It is found that when heteroskedasticity is estimated, the limiting variance of MLE is no longer of a sandwich form regardless of normality. We provide a rigorous analysis for the asymptotic theory of the MLE, demonstrating its desirable properties. The Monte Carlo simulations show that the MLE has good finite sample properties.

References

Ahn, S. G., Lee, Y. H. and Schmidt, P. (2013): Panel data models with multiple time-varying effects, *Journal of Econometrics*, 174, 1-14.

Alvarez, J., and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121-1159.

Anderson, T.W. (2003). *Introduction to Multivariate Statistical Analysis.* Hoboken, New Jersey: John Wiley & Sons.

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

Anselin, L. (1988). Spatial econometrics: methods and models. The Netherlands: Kluwer Academic Publishers.

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277-297.

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica*, 71(1), 135-171.

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Bai, J. and Li, K. (2012) Statistical analysis of factor models of high dimension, *The Annals of Statistics*, 40(1), 436–465.

Bai, J. and Li, K. (2014a) Spatial panel data models with common shocks, *Manuscript*, Columbia University.

Bai, J. and Li, K. (2014b) Theory and methods of panel data models with interactive effects, *The Annals of Statistics*, 42(1), 142–170.

Baltagi, B., Song, S.H., and Koh, W. (2003) Testing panel data regression models with spatial error correlation, *Journal of Econometrics*, 117, 123-150.

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

Baltagi, B., Egger, P., Pfaffermayr, M. (2008). Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics*, 145(1), 194-208.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1), 41-55.

Brockwell, P. J., and Davis, R. A. (1991). *Time series: Theory and methods*. Springer Series in Statistics, Berlin, New York: Springer, 2nd edition.

Calvó-Armengol, A., Patacchini, E., And Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 1239-1267.

Case, A. C. (1991) Spatial patterns in household demand. *Econometrica*, 59(4), 953–965.

Case, A. C., Rosen, H. S., and Hines, J. R. (1993) Budget spillovers and

Cli, A. D., and Ord, J. K. (1973) *Spatial autocorrelation,* London: Pion Ltd.

Conley, T. G. (1999). GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics*, 92, 1-45.

Conley, T.G., and Dupor, B. (2003) “A Spatial Analysis of Sectoral Complementarity," *Journal of Political Economy*, 111, 311-352.

Connor, G., and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3), 373-394.

Connor, G., and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT ... Journal of Financial Economics ... 3(3)

,A. (0503don:Sttim8(Eomi(,)-339(fo.)-333(Spatia,)-3Data3(K.)-3N8(ne3)4(Y)84)28(don:)-4Johtio3

Kelejian, H. H., and Prucha, I. R. (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99-121.

Kelejian, H. H., and Prucha, I. R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509-533.

Kelejian, H. H., and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), 53-67.

Kelejian, H. H., and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2), 219-257.

Lawless, M. (2009). Firm export dynamics and the geography of trade. *Journal of International Economics*, 77(2), 245-254.

Lee, L. (2004a) Asymptotic distribution of quasi-maximum likelihood estiators for spatial augoregressive models, *Econometrica*, 72(6), 1899–1925.

Lee, L. (2004b) A supplement to “Asymptotic distribution of quasi-maximum likelihood estiators for spatial augoregressive models”, *Manuscipt*.

Lee, L. F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2), 333-374.

Lee, L. and Yu, J. (2010) Estimation of spatial autoregressive panel data models with fixed effects, *Journal of Econometrics*, 154(2), 165-185.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4), 825-860.

Lin, X., and Lee, L. F. (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), 34-52.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3), 531-542.

Moon, H. and Weidner, M. (2009) Likelihood expansion for panel regression models with factors, *manuscipt, University of Southern California*.

Newey, W. and McFadden, D. (1994): Large Sample Estimation and Hypothesis Testing, in Engle, R.F. and D. McFadden (eds.) *Handbook of Econometrics,* North Holland.

Neyman, J., and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*. 1-32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6), 1417-1426.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349), 120-126.

Pesaran, M. H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica*, 74(4), 967–1012.

Pesaran, M. H., and Tosetti, E. (2011). Large panels with spatial correlations and common factors,