# Spatial panel data models with structural change

Li, Kunpeng

Capital University of Economics and Business

21 March 2018

# Spatial panel data models with structural changes*

Kunpeng Li

International School of Economics and Management
Capital University of Economics and Business

March 22, 2018

**Abstract**

Spatial panel data models are widely used in empirical studies. The existing theories of spatial models so far have largely confine the analysis under the assumption of parameters stabilities. This is unduely restrictive, since a large number of studies have well documented the presence of structural changes in the relationship of economic variables. This paper proposes and studies spatial panel data models with structural change. We consider using the quasi maximum likelihood method to estimate the model. Static and dynamic models are both considered. Large-$T$ and fixed-$T$ setups are both considered. We provide a relatively complete asymptotic theory for the maximum likelihood estimators, including consistency, convergence rates and limiting distributions of the regression coefficients, the timing of structural change and variance of errors. We study the hypothesis testing for the presence of structural change. The three super-type statistics are proposed. The Monte Carlo simulation results are consistent with our theoretical results and show that the maximum likelihood estimators have good finite sample performance.

**Key Words:** Spatial panel data models, structural changes, hypothesis testing, asymptotic theory.

**JEL:** C31; C33.

*This paper was presented at Advanced Econometric Forum of Wuhan University, March 19, 2018. I would like to thank Songnian Chen, Xiaohong Chen, Yixiao Sun, Edward J. Vytlacil and other participants of the forum for their helpful comments. The matlab code of this paper is available upon request. The author's email address: kunpenglithu@126.com.

# 1 Introduction

Since the seminal works of Cliff and Ord (1973) and Ord (1975), spatial econometric models have received much attention in the economic literature. A typical spatial autoregressive model specifies that one's outcome is directly affected by the outcomes of its spatial peers with some prespecified weights. With this particular specification, one has the chance to study the spatial interactions among a number of spatial units.

Spatial econometric models have been widely used in empirical studies. In microeconomics, evoked by the influential work of Manski (1993), spatial models are one of primary tools to study the endogenous effects or peer effects, see e.g., Bramoullé, Djebbari and Fortin (2009), Calvó-Armengol, Patacchini and Zenou (2009), Lin (2010). In public economics, spatial models are widely used to study the competitions for tax and fiscal expenditure among local governments, see e.g., Lyytikäinen (2012), Chirinko and Wilson (2007), etc. In international economics, spatial models are popular in the studies of the spillover effects of foreign direct investment and the gravity model of trade, see e.g., Bode, Nunnenkamp and Waldkirch (2012). In urban economics, spatial models are used to study the diffusion process of housing prices, see Holly, Pesaran and Yamagata (2011). In finance, Kou, Peng and Zhong (2017) show that the traditional capital asset pricing model or the arbitrary pricing theory augmented with spatial interactions can improve the model fitting to the real data. Spatial econometrics models are also used to study some social issues. As one of motivating example in Anselin (1988), spatial term is introduced to capture the social patten among different districts arising from the sphere of criminals.

The literature has also witnesses rapid developments on the theory of spatial econometric models. Early development has been summarized by a number of books, including Anselin (1988) and Cressie (1993). Due to the presence of endogenous spatial term, the ordinary least squares (OLS) method cannot deliver a consistent estimation. Generalized method of moments (GMM) and quasi maximum likelihood (ML) method are two popular estimation methods to address this issue. The GMM are studied by Kelijian and Prucha (1998, 1999, 2010), and Kapoor et al. (2007), among others. The ML method is investigated by Anselin (1988), Lee (2004), Yu et al. (2008) and Lee and Yu (2010), and so on.

The theoretical developments of spatial models so far have been much confined within the assumption of parameters stabilities over the sample horizon. This assumption is unduely restrictive, given that a large number of studies have well documented the presence of parameters instabilities or structural changes in econometric models of other types, see, for example, Bai (1997), Bai, Lumsdaine and Stock (1998) and Qu and Perron (2007) in linear regression models, and Perron (1989), Zivot and Andrews (1992), Bai and Carrion-I-Silvestre (2009) in unit root tests. This is particularly true in the sample with a long or moderately long horizon, since the longer time span of the sample, the more unobserved or observed factors affecting the parameters stabilities. Even for the sample with a small number of periods, it is still necessary to take into account structural change as a conservative modeling step since one may be unlucky to have a sample at hand, which has experienced one or more typical structural change events, such as presidential adminstration switching, policy-regime shift, financial crisis, etc. It is well known that failure to account for the structural change, if it does exist, would cause inconsistent estimations and incorrect statistical inferences, resulting in misleading economic implications. Introduction of structural change in model is not just for correct statistical analysis concern. In some applications such as policy evaluations, specifying a structural change allows one to quantify the effect of policy intervention, which has its own independent interests.

This paper proposes and studies spatial panel data models in the presence of structural change. We consider using the quasi ML method to estimate model. Our asymptotic analysis first focuses on a static spatial panel model under large-$T$ setup. The theory is next extended to a dynamic model, as well as to a model under fixed-$T$ setup. We also consider the hypothesis testing issue on the presence of structural change. The supW, supLM, supLR statistics, which are based on the classical Wald, Lagrange Multipliers (LM) and Likelihood ratio (LR) tests, are proposed to perform this work. The asymptotic properties of these three statistics are investigated. To my best knowledge, this paper is the first to develop a relatively complete asymptotic theory on spatial panel data models with structural change.

A main difficulty of the current theoretical analysis is to establish the global properties of the ML estimators. As far as I can see, one cannot directly apply the arguments well developed by the previous studies, such as Bai (1997), Bai and Perron (1998), Bai, Lumsdaine and Stock (1998), Qu and Perron (2007) and so forth, to the current model. The reasons are twofold. First, because of nonlinearity arising from the spatial lag term, we cannot concentrate out the regression coefficients from the objective function, which makes the arguments in Bai (1997) and Bai and Perron (1998) not suitable to our model. Second, the ML estimators would have biases due to the presence of incidental parameters. As a consequence, the ML estimators have two convergence rates, and which one is dominant depending on the values of $N$ and $T$, where $N$ is the number of spatial units and $T$ the number of periods. This feature makes the analysis of Bai, Lumsdaine and Stock (1998) much complicated since their analysis implicitly specifies the convergence rate. In addition, the particular identification issue in the spatial models requires some exclusive analysis to address it. In this paper, we develop a set of different and new arguments to establish the global properties, such as consistency and convergence rates.

The proposed model is related with spatial econometrics and structural change models, both of which have long histories and are very popular in empirical studies. As the mixture, we believe that the proposed model would inherit their popularity.

There are few studies related with this paper. Sengupta (2017) considers the hypothesis testing for a structural change in a spatial autoregressive models. An undesirable feature of this study is that the individual effects, which are the primary attractiveness of panel data models, are abstracted from the model. In addition, only supLR statistic is proposed in this study. In contrast, we propose three statistics. Among these three statistics, the supLM statistic is of more practical interests, since it only involves the estimator of the restricted model that is easier to compute. This can save much computation cost, especially when the sample size is large.

The rest of the paper is organized as follows. In Section 2, we propose a basic static model and write out the related likelihood function. The computation aspect of the model is also considered in this section. Section 3 lists the assumptions needed for the subsequent theoretical analysis. We also make a detailed discussion on the identification issue. Section 4 presents the asymptotic results. Section 5 makes an extension to the dynamic model. Section 6 gives the asymptotics under fixed-$T$ setup. Section 7 consider the hypothesis testing for the presence of structural change. Section 8 runs Monte Carlo simulations to investigate the finite performance of the ML estimators. Section 9 concludes the paper.

## 2 Basic model and likelihood function

The model studied in this paper is a spatial autoregressive panel data model with a structural change, which is written as

$$Y_t = \alpha^* + \rho^* W_N Y_t + \varrho^* W_N Y_t \mathbb{1}(t \leq [T\gamma^*]) + Z_t \beta^* + X_t \delta^* \mathbb{1}(t \leq [T\gamma^*]) + V_t, \tag{2.1}$$

or equivalently

$$\begin{aligned} Y_t &= \alpha^* + (\rho^* + \varrho^*) W_N Y_t + Z_t \beta^* + X_t \delta^* + V_t, && \text{for} \quad t \leq [T\gamma^*], \\ Y_t &= \alpha^* + \rho^* W_N Y_t + Z_t \beta^* + V_t, && \text{for} \quad t > [T\gamma^*], \end{aligned}$$

where $Y_t$ is an $N$-dimensional vector of observations at time $t$ for the dependent variable. $\rho^* W_N Y_t$ and $\varrho^* W_N Y_t \mathbb{1}(t \leq [T\gamma^*])$ are two spatial terms. $\mathbb{1}(\cdot)$ is an indicator function, which is 1 if the expression in the brackets is true, and 0 otherwise. $Z_t$ and $X_t$ are $N \times p$ and $N \times q$ observable data matrices with $p \geq q$ at time $t$ for explanatory variables. Here the columns of $X_t$ is a subset of those of $Z_t$. The model is a pure change model if $p = q$; and a partial change model if $p > q$. Our model specification implicitly assumes that there is one break point and the structural change always appears in the spatial term. The model with multiple structural changes is also of theoretical interests and practical relevance. But such a topic is beyond the scope of this paper and is left as a future work. In addition, the model with the structural change only present in the coefficients of $X_t$ is simpler than model (2.1). The analysis of this simpler model can be easily obtained by slightly modifying the analysis of the current paper.

For notational simplicity, we introduce the following symbols:

$$X_t(\gamma) = X_t \mathbb{1}(t \leq [T\gamma]), \qquad X_t(\gamma, \gamma^*) = X_t \mathbb{1}(t \leq [T\gamma]) - X_t \mathbb{1}(t \leq [T\gamma^*]),$$

$$Y_t(\gamma) = Y_t \mathbb{1}(t \le [T\gamma]), \qquad Y_t(\gamma, \gamma^*) = Y_t \mathbb{1}(t \le [T\gamma]) - Y_t \mathbb{1}(t \le [T\gamma^*]),$$
$$D_t(\rho, \varrho, \gamma) = I_N - \rho W_N - \varrho W_N \mathbb{1}(t \le [T\gamma]), \qquad D_t^* = D_t(\rho^*, \varrho^*, \gamma^*).$$

With these notations, the model now can be written as

$$D_t^* Y_t = \alpha^* + Z_t \beta^* + X_t(\gamma^*)\delta^* + V_t.$$

Let $\theta = (\rho, \varrho, \beta', \delta', \gamma, \sigma^2)'$. If $V_t$ is normally and independently distributed with with mean zero and variance $\sigma^2 I_N$, the gaussian log-likelihood function is

$$\mathcal{L}(\theta, \alpha) = -\frac{1}{2} \ln \sigma^2 + \frac{1}{NT} \sum_{t=1}^{T} \ln \left| D_t(\rho, \varrho, \gamma) \right| - \frac{1}{2NT\sigma^2} \sum_{t=1}^{T} \mathcal{X}_t(\theta, \alpha)' \mathcal{X}_t(\theta, \alpha) \tag{2.2}$$

where

$$\mathcal{X}_t(\theta, \alpha) = D_t(\rho, \varrho, \gamma) Y_t - \alpha - Z_t \beta - X_t(\gamma)\delta.$$

The first order condition for $\alpha$ gives

$$\alpha(\theta) = \frac{1}{T} \sum_{t=1}^{T} D_t(\rho, \varrho, \gamma) Y_t - \frac{1}{T} \sum_{t=1}^{T} Z_t \beta - \frac{1}{T} \sum_{t=1}^{T} X_t(\gamma)\delta.$$

Substituting the preceding formula into the likelihood function to concentrate out $\alpha$, we have

$$\mathcal{X}_t(\theta) = \widetilde{D_t Y_t} - \widetilde{Z}_t \beta - \widetilde{X_t(\gamma)}\delta = \widetilde{Y}_t - \rho W_N \widetilde{Y}_t - \varrho W_N \widetilde{Y_t(\gamma)} - \widetilde{Z}_t \beta - \widetilde{X_t(\gamma)}\delta. \tag{2.3}$$

where we use $\widetilde{A}_t$ to denote $A_t - T^{-1} \sum_{t=1}^{T} A_t$, for example

$$\widetilde{D_t Y_t} = D_t Y_t - \frac{1}{T} \sum_{t=1}^{T} D_t Y_t.$$

In addition, we suppress $\rho, \varrho$ and $\gamma$ from $D_t$ in the places where no confusion arises for notational simplicity. Now the likelihood function after concentrating out $\alpha$ is

$$\mathcal{L}(\theta) = -\frac{1}{2} \ln \sigma^2 + \frac{1}{NT} \sum_{t=1}^{T} \ln |D_t| - \frac{1}{2NT\sigma^2} \sum_{t=1}^{T} \mathcal{X}_t(\theta)' \mathcal{X}_t(\theta) \tag{2.4}$$

where $\mathcal{X}_t(\theta)$ is defined in (2.3). The MLE $\hat{\theta}$ is therefore defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}(\theta). \tag{2.5}$$

where $\Theta$ is the parameters space specified below.

The above maximization issue can be alternatively written as

$$\max_{\theta \in \Theta} \mathcal{L}(\theta) = \max_{\gamma \in [\gamma^L, \gamma^U]} \max_{\vartheta \in \mathrm{Par}(\vartheta)} \mathcal{L}(\vartheta, \gamma)$$

where $\vartheta = (\rho, \varrho, \beta', \delta', \sigma^2)'$ and $\mathrm{Par}(\vartheta)$ is the parameters space for $\vartheta$. We use the above formula to obtain the numerical values of the MLE, i.e., for each given $\gamma$, we get $\hat{\vartheta}(\gamma)$

by maximizing $\mathcal{L}(\vartheta, \gamma)$, then we get $\hat{\gamma}$ by maximizing $\mathcal{L}(\hat{\vartheta}(\gamma), \gamma)$. Once $\hat{\gamma}$ is obtained, the MLE $\hat{\theta}$ is $\hat{\theta} = (\hat{\vartheta}(\hat{\gamma}), \hat{\gamma})$. In practice, the maximization in the first step may be time-consuming, especially when $N$ is large, since it involves the calculation of determinant of large dimensional matrix. To economize the computation costs, we suggest the following estimation procedures. Let $W_N X_t$ and $W_N X_t(\gamma)$ be the instruments of $W_N Y_t$ and $W_N Y_t(\gamma)$. The estimation procedures consist of three steps. In the first step, for each given $\gamma$, we apply two-stage least square method to model (2.1) to obtain the sum of squared residual, which we denote by $\mathrm{SSR}(\gamma)$. In the second step, we sort $\mathrm{SSR}(\gamma)$ in ascending order and detect five $\gamma$ values which corresponds to the five smallest $\mathrm{SSR}(\gamma)$s. In the last step, we conduct the above maximization issue by restricting $\gamma$ to be these five values.

## 3 Assumptions and identification issue

We make the following assumptions for the subsequent theoretical analysis. Hereafter, we use $C$ to denote a generic constant, which need not to be the same at each appearance.

**Assumption A:** The errors $v_{it} (i = 1, 2, \ldots, N, t = 1, 2, \ldots, T)$ are identically and independently distributed with mean zero and variance $\sigma^{*2} > 0$. In addition, we assume that $\sup_{i,t} \mathbb{E}(|v_{it}|^{4+c}) < \infty$ for some $c > 0$.

**Assumption B:** $W_N$ is an exogenous spatial weights matrix whose diagonal elements are all zeros. In addition, $W_N$ is bounded by some constant $C$ for all $N$ under $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms.

**Assumption C:** Let $D_N(x) = I_N - xW_N$. We assume that $D_N(x)$ is invertible over $\mathbb{R}_\rho$ and $\mathbb{R}_\rho \oplus \mathbb{R}_\varrho$, where $\mathbb{R}_\rho$ and $\mathbb{R}_\varrho$ are the respective parameters space for $\rho$ and $\varrho$, which contain $\rho^*$ and $\varrho^*$ as interior points, and $\mathbb{R}_\rho \oplus \mathbb{R}_\varrho$ is the parameter space for $\rho + \varrho$ with $\rho \in \mathbb{R}_\rho$ and $\varrho \in \mathbb{R}_\varrho$. In addition, $D_N(\rho)^{-1}$ and $D_N(\rho + \varrho)^{-1}$ are bounded by some constant $C$ for all $N$ under $\|\cdot\|_1$ and $\|\cdot\|_\infty$ uniformly on $\mathbb{R}_\rho$ and $\mathbb{R}_\rho \oplus \mathbb{R}_\varrho$.

**Assumption D:** The underlying true value $\theta^* = (\rho^*, \varrho^*, \beta^{*\prime}, \delta^{*\prime}, \gamma^*, \sigma^{*2})'$ is an interior point of the parameter space $\Theta$ with $\Theta = \mathbb{R}_\rho \times \mathbb{R}_\varrho \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}_\gamma \times \mathbb{R}_{\sigma^2}$, where $\mathbb{R}_\rho$ and $\mathbb{R}_\varrho$ are both compact sets in $\mathbb{R}^1$, and $\mathbb{R}_\gamma = [\gamma^L, \gamma^U] \subset (0, 1)$, and $\mathbb{R}_{\sigma^2}$ is a compact set which is bounded away from zero, and $\mathbb{R}^d$ is $d$-dimensional Euclidean space. In addition, the parameters $\theta$ are estimated in the set $\Theta$.

**Assumption E:** Let $\psi^* = (\varrho^*, \delta^{*\prime})'$ and $C^*$ be a $(q+1)$-dimensional constant vector. We assume **Assumption E.1:** $\psi^* = C^*$, or **Assumption E.2:** $\psi^* = (NT)^{-\nu} C^*$ with $0 < \nu < \frac{1}{4}$.

**Assumption F:** Let $\mathcal{S}_t = \left[Z_t, X_t \mathbb{1}(t \leq [T\gamma^*])\right]$, we assume that $\mathcal{S}_t$ are nonrandom and $\sup_{i,t} \|Z_{it}\|^2 < \infty$ for all $i$ and $t$, and the sample covariance matrix $\frac{1}{NT} \sum_{t=1}^{T} \widetilde{\mathcal{S}}_t{}' \widetilde{\mathcal{S}}_t$ is positive definite.

Let $\mathcal{I}_1(\rho,\sigma^2)$ and $\mathcal{I}_2(\rho,\sigma^2)$ be both $2\times 2$ matrices, which are defined as

$$\mathcal{I}_1(\rho,\sigma^2)=\frac{1}{N}\begin{bmatrix}\frac{\sigma^{\circ 2}}{\sigma^2}\mathrm{tr}(S_1^{*\prime}S_1^*)+\mathrm{tr}\Big[\zeta_1^{-1}(\rho)S_1^*\zeta_1^{-1}(\rho)S_1^*\Big] & \frac{\sigma^{\circ 2}}{\sigma^4}\mathrm{tr}\Big[\zeta_1(\rho)'S_1^*\Big]\\ \frac{\sigma^{\circ 2}}{\sigma^4}\mathrm{tr}\Big[\zeta_1(\rho)'S_1^*\Big] & \frac{\sigma^{\circ 2}}{\sigma^6}\mathrm{tr}\Big[\zeta_1(\rho)'\zeta_1(\rho)\Big]-\frac{N}{2\sigma^4}\end{bmatrix}$$

and

$$\mathcal{I}_2(\rho,\sigma^2)=\frac{1}{N}\begin{bmatrix}\frac{\sigma^{\circ 2}}{\sigma^2}\mathrm{tr}(S_2^{*\prime}S_2^*)+\mathrm{tr}\Big[\zeta_2^{-1}(\rho)S_2^*\zeta_2^{-1}(\rho)S_2^*\Big] & \frac{\sigma^{\circ 2}}{\sigma^4}\mathrm{tr}\Big[\zeta_2(\rho)'S_2^*\Big]\\ \frac{\sigma^{\circ 2}}{\sigma^4}\mathrm{tr}\Big[\zeta_2(\rho)'S_2^*\Big] & \frac{\sigma^{\circ 2}}{\sigma^6}\mathrm{tr}\Big[\zeta_2(\rho)'\zeta_2(\rho)\Big]-\frac{N}{2\sigma^4}\end{bmatrix}$$

where $\zeta_1(\rho)=I_N-(\rho-\rho^*-\varrho^*)S_1^*$, $\zeta_2(\rho)=I_N-(\rho-\rho^*)S_2^*$ and $\sigma^{\circ 2}=\frac{T-1}{T}\sigma^{*2}$. We further define

$$J_t(\gamma)=\Big[S_t(\gamma,\gamma^*)\mu_t^*,\ \ X_t(\gamma,\gamma^*)\Big],\quad K_t(\gamma)=\Big[S_t^*\mu_t^*,\ \ S_t^*(\gamma)\mu_t^*,\ \ Z_t,\ \ X_t(\gamma)\Big],\tag{3.1}$$

where $\mu_t^*=\alpha^*+Z_t\beta^*+X_t(\gamma^*)\delta^*$, $S_t^*=W_ND_t^{*-1}$ and $S_t^*(\gamma,\gamma^*)=S_t^*\big[\mathbb{1}(t\leq[T\gamma])-\mathbb{1}(t\leq[T\gamma^*])\big]$. Let

$$\Pi_{JJ}(\gamma)=\frac{1}{NT}\sum_{t=1}^T\widetilde{J_t(\gamma)}'\widetilde{J_t(\gamma)},\qquad \Pi_{JK}(\gamma)=\frac{1}{NT}\sum_{t=1}^T\widetilde{J_t(\gamma)}'\widetilde{K_t(\gamma)}$$
$$\Pi_{KJ}(\gamma)=\frac{1}{NT}\sum_{t=1}^T\widetilde{K_t(\gamma)}'\widetilde{J_t(\gamma)},\qquad \Pi_{KK}(\gamma)=\frac{1}{NT}\sum_{t=1}^T\widetilde{K_t(\gamma)}'\widetilde{K_t(\gamma)}$$

We have the following assumption on the parameters identification.

**Assumption G:** One of the following assumption holds

**Assumption G.1 (a):** Matrices $\mathcal{I}_1(\rho,\sigma^2)$ and $\mathcal{I}_2(\rho,\sigma^2)$ are positive definite over the parameter space $(\mathbb{R}_\rho\oplus\mathbb{R}_\varrho)\times\mathbb{R}_{\sigma^2}$ and $\mathbb{R}_\rho\times\mathbb{R}_{\sigma^2}$.

**Assumption G.1 (b):** The following condition

$$\min\left(\frac{1}{N}\big\|S_1^{*\prime}+S_1^*-\varrho^*S_1^{*\prime}S_1^*\big\|^2,\frac{1}{N}\big\|S_2^{*\prime}+S_2^*-\varrho^*S_2^{*\prime}S_2^*\big\|^2\right)>0,$$

holds for all $N$.

**Assumption G.2 (a):** There exists a constant $c$ such that for all $N$ and $T$

$$\lambda_{\min}\Big(\Pi_{KK}(\gamma)\Big)=\lambda_{\min}\left(\frac{1}{NT}\sum_{t=1}^T\widetilde{K_t(\gamma)}'\widetilde{K_t(\gamma)}\right)\geq c.$$

**Assumption G.2 (b):** There exist a constant $c$ such that for all $N$ and $T$

$$\lambda_{\min}\Big(\Pi_{JJ}(\gamma)-\Pi_{JK}(\gamma)\Pi_{KK}(\gamma)^{-1}\Pi_{KJ}(\gamma)\Big)\geq c|\gamma-\gamma^*|,$$

or alternatively

$$\lambda_{\min}\left(\frac{1}{NT}\sum_{t=1}^T\widetilde{H_t(\gamma)}'\widetilde{H_t(\gamma)}\right)\geq c|\gamma-\gamma^*|,$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of $A$ and $H_t(\gamma) = [J_t(\gamma), K_t(\gamma)]$.

Assumption A requires that disturbances are drawn from a random sample. Similar assumption appears in a number of studies on QML estimations of spatial models, see Lee (2004), Yu et al. (2008) and Lee and Yu (2010). In spatial models, this assumption is not just for theoretical simplicity, but also serves as the base for the parameters identification in some special spatial models, see the discussions on Assumption G below. Assumption B is about spatial weights matrix, which is standard in spatial econometric literature. Our specification on spatial weights matrix implicitly assumes that it is time invariant, so the case of time-varying weights matrix is precluded. However, we note that the arguments in this paper can be easily extended to the time-varying case. Assumption C imposes the invertibilities of $D_N(\rho)$ and $D_N(\rho + \varrho)$. Invertibilities of $D_N(\rho^*)$ and $D_N(\rho^* + \varrho^*)$ are indispensable since they guarantee that the models before and after structural change are both well defined. Since $D_N(x)$ is a continuous function in $x$, the invertibeilities of $D_N(\rho)$ and $D_N(\rho + \varrho)$ can be maintained in some neighborhoods of $\rho^*$ and $\rho^* + \varrho^*$. However, this invertibility is a local property. Assumption C goes further to assume that this local invertibility can be extended over the parameters space. Assumption D assumes that the underlying true values are in a compact set, which is standard in econometric analysis. Assumption D also assumes that the parameters are estimated in a compact set. Such an assumption is often made when dealing with nonlinear objective functions, see, for example, Jennrich (1969). Our objective function is obviously nonlinear, due to the presence of spatial terms and structural change. This is the difficulty source of theoretical analysis. Assumption E is standard in the structural change literature. It gives the conditions under which the structural change is asymptotically identifiable. Assuming shrinking coefficients is important for developing the limiting theory of the break date estimate that does not depend on the distributions of the regressors and the errors.

Assumption F is the identification condition for $\beta$ and $\delta$. It assumes that the exogenous explanatory variable are non-random. Similar assumption also appears in various spatial studies, such as Lee (2004), Yu et al. (2008), Lee and Yu (2010), etc. If exogenous regressors are assumed to be random instead, the analysis of this paper can be conducted similarly with covaraince stationarity of $Z_t$ with appropriate mixing conditions and moment conditions. Assumption F also assumes that $\mathcal{S}_t$ is of full column rank in the sense that $\frac{1}{NT}\sum_{t=1}^{T} \widetilde{\mathcal{S}}_t{}' \widetilde{\mathcal{S}}_t$ is positive definite, which is standard in linear regression models.

Assumption G is the identification condition for $\rho, \varrho, \sigma^2$ and $\gamma$. Consider the following model that the exogenous regressors are absent and the timing of structural change, $\gamma$, is observed. Now the model can be written as the following two models,

$$Y_t = \alpha + (\rho + \varrho) W_N Y_t + V_t, \quad \text{for } t \leq [T\gamma], \tag{3.2}$$

$$Y_t = \alpha + \rho W_N Y_t + V_t, \quad \text{for } t > [T\gamma] + 1, \tag{3.3}$$

Matrix $\mathcal{I}_1(\rho^* + \varrho^*, \sigma^{\circ 2})$ is the information matrix of the MLE for model (3.2), and matrix $\mathcal{I}_2(\rho^*, \sigma^{\circ 2})$ the information matrix for model (3.3). To make the MLE well defined, we need the positive definiteness of $\mathcal{I}_1(\rho^* + \varrho^*, \sigma^{\circ 2})$ and $\mathcal{I}_2(\rho^*, \sigma^{\circ 2})$. Since the matrix $\mathcal{I}_i(\rho, \sigma^2)$ for $i = 1, 2$ is a continuous function of $\rho$ and $\sigma^2$, the positive definiteness is maintained in some neighborhood of $\rho^*$ and $\sigma^{\circ 2}$. Assumption G.1 (a) is therefore made in this direction to assume that $\mathcal{I}_i(\rho, \sigma^2)$ for $i = 1, 2$ is positive definite over the parameters space. Now we see that Assumption G.1 (a) is the identification condition for $\rho, \varrho$ and $\sigma^2$ in pure panel spatial autoregressive models. Lee (2004) considers the identification issue for pure spatial autoregressive models. We can make a set of identification conditions analogous to the Lee's type. The conditions include: (i) for all $\rho + \varrho \neq \rho^* + \varrho^*$,

$$\lim_{N\to\infty} \left( \frac{1}{N} \ln \left| \sigma^{\circ 2} D_N^{-1\prime}(\rho^* + \varrho^*) D_N^{-1}(\rho^* + \varrho^*) \right| - \frac{1}{N} \ln \left| \sigma_1^2(\rho + \varrho) D_N^{-1\prime}(\rho + \varrho) D_N^{-1}(\rho + \varrho) \right| \right) \neq 0,$$

where

$$\sigma_1^2(\rho) = \frac{1}{N} \mathrm{tr} \Big[ \sigma^{\circ 2} D_N^{-1\prime}(\rho^* + \varrho^*) D_N(\rho)' D_N(\rho) D_N^{-1}(\rho^* + \varrho^*) \Big].$$

and (ii) for all $\rho \neq \rho^*$,

$$\lim_{N\to\infty} \left( \frac{1}{N} \ln \left| \sigma^{\circ 2} D_N^{-1\prime}(\rho^*) D_N^{-1}(\rho^*) \right| - \frac{1}{N} \ln \left| \sigma_2^2(\rho) D_N^{-1\prime}(\rho) D_N^{-1}(\rho) \right| \right) \neq 0,$$

where

$$\sigma_2^2(\rho) = \frac{1}{N} \mathrm{tr} \Big[ \sigma^{\circ 2} D_N^{-1\prime}(\rho^*) D_N(\rho)' D_N(\rho) D_N^{-1}(\rho^*) \Big].$$

Assumption G.1 can be viewed as variance identification conditions. These conditions are designed exclusively for some special models such as model (1.1) or

$$Y_t = \rho W_N Y_t + \varrho W_N Y_t \mathbb{1}(t \leq [T\gamma]) + V_t.$$

In these so-call pure spatial autoregressive models, the mean of $Y_t$ provides little or no information on the identification of parameters. So one has to resort to the variance and covariance structure of $Y_t$ to gain the identification. To achieve this goal, we must assume no cross sectional correlations in $V_t$. With this assumption, the cross sectional correlations pattern of $Y_t$ is totally due to the spatial terms and therefore identification is possible.

Assumption G.2 proposes an alternative set of identification conditions, which can be viewed as the mean identification conditions. Assumption G.2(a) is for the identification of $\vartheta$. Intuitively, Assumption G.2(a) uses the relationship between $Z_t, X_t$ and $\mathbb{E}(Y_t)$ to identify $\rho^*$ and $\varrho^*$ as well as other parameters. However, if $Z_t$ and $X_t$ have no effects on $\mathbb{E}(Y_t)$, which means $\beta^* = 0$ and $\delta^* = 0$, Assumption G.2(a) would break down. In addition, if $\beta^* = 0$ and $\delta^*$ shrinks to zero with the rate specified in Assumption E.2, it can be shown that the minimum eigenvalue in Assumption G.2 (a) is of magnitude $O_p[(NT)^{-\nu}]$, which also violates the required identification condition. Whether Assumption G.2(a) can be used for identification or not depends on the other parameters. So it is a local identification condition. Assumption G.2(b) is for the identification of $\gamma$. We note that Assumption G.2 (b) may hold even Assumption G.2 (a) break down. Consider the case that $\beta^* = 0$ and $\delta^*$ shrinks to zero. If $|\gamma - \gamma^*| = O[(NT)^{-\nu}]$, Assumption G.2 (b) still hold. This means we can use the mean information to identify $\gamma$ even this information fails to identify $\rho$ and $\varrho$.

# 4 Asymptotic properties

This section presents the asymptotic results of the MLE. We have the following proposition on the consistency.

**Proposition 4.1** *Let $\hat{\theta} = (\hat{\rho}, \hat{\varrho}, \hat{\beta}', \hat{\delta}', \hat{\gamma}, \hat{\sigma}^2)'$ be the MLE defined in (2.5). Under Assumptions A-G, as $N, T \to \infty$, we have $\hat{\theta} \xrightarrow{p} \theta^*$. Let $\vartheta^\circ = (\rho^*, \varrho^*, \beta^{*\prime}, \delta^{*\prime}, \sigma^{\circ 2})'$ and $\hat{\vartheta} = (\hat{\rho}, \hat{\varrho}, \hat{\beta}', \hat{\delta}', \hat{\sigma}^2)'$, we also have $(NT)^\nu(\hat{\vartheta} - \vartheta^\circ) = o_p(1)$, where $\sigma^{\circ 2} = \frac{T-1}{T}\sigma^{*2}$.*

Proposition 4.1 not only shows the consistency of the MLE, but also give some rough convergence rate on $\hat{\vartheta}$. As regard $\hat{\gamma}$, the estimator of break point location, we need the following assumption to establish its convergence rate.

**Assumption H:** Let

$$\Psi^*_{1,NT} = \frac{1}{N\ell\sigma^{*2}} \sum_{t=T^*+1}^{T^*+\ell} \begin{bmatrix} Y_t'W_N'W_N Y_t & Y_t'W_N'X_t \\ X_t'W_N Y_t & X_t'X_t \end{bmatrix} + \frac{1}{N} \begin{bmatrix} \mathrm{tr}(S_2^* S_2^*) & \underset{1\times k}{0} \\ \underset{k\times 1}{0} & \underset{k\times k}{0} \end{bmatrix}, \tag{4.1}$$

$$\Psi^\bullet_{1,NT} = \frac{1}{N\ell\sigma^{*2}} \sum_{t=T^*-\ell}^{T^*} \begin{bmatrix} Y_t'W_N'W_N Y_t & Y_t'W_N'X_t \\ X_t'W_N Y_t & X_t'X_t \end{bmatrix} + \frac{1}{N} \begin{bmatrix} \mathrm{tr}(S_1^* S_1^*) & \underset{1\times k}{0} \\ \underset{k\times 1}{0} & \underset{k\times k}{0} \end{bmatrix},$$

There exist $\ell_0 > 0$ and $c > 0$ such that for all $\ell > \ell_0$,

$$\min\left[\lambda_{\min}(\Psi^*_{1,NT}), \lambda_{\min}(\Psi^\bullet_{1,NT})\right] > c,$$

where $\lambda_{\min}(\cdot)$ is defined in Assumption G.2 and $T^* = T\gamma^*$.

The convergence rate of $\hat{\gamma}$ is given in the following proposition.

**Proposition 4.2** *Under Assumptions A-H, as $N, T \to \infty$ and $(NT)^{2\nu}/N \to \infty$, we have $\hat{\gamma} = \gamma^* + O_p(\frac{1}{(NT)^{1-2\nu}})$.*

**Remark 4.1** Proposition 4.2 is crucial for the subsequent analysis. With this result, it can be shown that the estimation error of $\hat{\gamma}$ would have no effect on the asymptotic properties of the remaining ML estimators. More specifically, the asymptotic representation and limiting distribution of $\hat{\vartheta}$ under $\gamma = \hat{\gamma}$ are the same with those under $\gamma = \gamma^*$, a property which is a primary step in proving Theorem 4.1. $\square$

To present the limiting distribution of $\vartheta$, we introduce the following notations. Let $\bar{k} = p + q + 1$, $\mathbf{h}_{t,1} = \mathrm{diag}[(S_1^*(\gamma^*) + S_2^* - S_2^*(\gamma^*)]$ and $\mathbf{h}_{t,2} = \mathrm{diag}[S_1^*(\gamma^*)]$, where $\mathrm{diag}(M)$

denote the operation which puts the diagonal elements of $M$ into a vector. Furthermore, define

$$\Xi_{1,NT} = \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S_2^* S_2^*) & \underset{1\times\bar{k}}{0} & \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_2^*) \\ 0 & \underset{1\times\bar{k}}{0} & 0 \\ \underset{(p+q)\times 1}{0} & \underset{(p+q)\times\bar{k}}{0} & \underset{(p+q)\times 1}{0} \\ \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_2^*) & \underset{1\times\bar{k}}{0} & 0 \end{bmatrix}, \Xi_{2,NT} = \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S_2^* \circ S_2^*) & \underset{1\times\bar{k}}{0} & \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_2^*) \\ 0 & \underset{1\times\bar{k}}{0} & 0 \\ \underset{(p+q)\times 1}{0} & \underset{(p+q)\times\bar{k}}{0} & \underset{(p+q)\times 1}{0} \\ \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_2^*) & \underset{1\times\bar{k}}{0} & \frac{N}{4\sigma^{\circ 4}} \end{bmatrix}$$

$$\Xi_{3,NT} = \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S_1^* S_1^*) & \mathrm{tr}(S_1^* S_1^*) & \underset{1\times(p+q)}{0} & \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_1^*) \\ \mathrm{tr}(S_1^* S_1^*) & \mathrm{tr}(S_1^* S_1^*) & \underset{1\times(p+q)}{0} & \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_1^*) \\ \underset{(p+q)\times 1}{0} & \underset{(p+q)\times 1}{0} & \underset{(p+q)\times(p+q)}{0} & \underset{(p+q)\times 1}{0} \\ \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_1^*) & \frac{1}{\sigma^{\circ 2}}\mathrm{tr}(S_1^*) & \underset{1\times(p+q)}{0} & 0 \end{bmatrix},$$

$$\Xi_{4,NT} = \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S_1^* \circ S_1^*) & \mathrm{tr}(S_1^* \circ S_1^*) & \underset{1\times(p+q)}{0} & \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_1^*) \\ \mathrm{tr}(S_1^* \circ S_1^*) & \mathrm{tr}(S_1^* \circ S_1^*) & \underset{1\times(p+q)}{0} & \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_1^*) \\ \underset{(p+q)\times 1}{0} & \underset{(p+q)\times 1}{0} & \underset{(p+q)\times(p+q)}{0} & \underset{(p+q)\times 1}{0} \\ \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_1^*) & \frac{1}{2\sigma^{\circ 2}}\mathrm{tr}(S_1^*) & \underset{1\times(p+q)}{0} & \frac{N}{4\sigma^{\circ 4}} \end{bmatrix}.$$

With the above notations, we define

$$\Omega_{1,NT} = \frac{1}{NT\sigma^{\circ 2}}\begin{bmatrix} \underset{(p+q+2)\times(p+q+2)}{\sum_{t=1}^{T}\widetilde{\mathcal{Z}}_t{}'\widetilde{\mathcal{Z}}_t} & \underset{(p+q+2)\times 1}{0} \\ \underset{1\times(p+q+2)}{0} & \underset{1\times 1}{NT/(2\sigma^{\circ 2})} \end{bmatrix} + \gamma^*\Xi_{3,NT} + (1-\gamma^*)\Xi_{1,NT}, \tag{4.2}$$

$$\Omega_{2,NT} = \frac{\kappa_4 - 3\sigma^{*4}}{\sigma^{*4}}\Big[\gamma^*\Xi_{4,NT} + (1-\gamma^*)\Xi_{2,NT}\Big], \tag{4.3}$$

$$\Omega_{3,NT} = \frac{\kappa_3}{NT\sigma^{*2}\sigma^{\circ 2}}\sum_{t=1}^{T}\begin{bmatrix} 2\widetilde{Y}_t{}'W_N'\mathbf{h}_{t,1} & \widetilde{Y}_t{}'W_N'\mathbf{h}_{t,2} + \widetilde{Y_t(\gamma^*)}{}'W_N'\mathbf{h}_{t,1} & \mathbf{h}_{t,1}'\widetilde{Z}_t & \mathbf{h}_{t,1}'\widetilde{X_t(\gamma^*)} & 0 \\ * & 2\widetilde{Y_t(\gamma^*)}{}'W_N'\mathbf{h}_{t,2} & \mathbf{h}_{t,2}'\widetilde{Z}_t & \mathbf{h}_{t,2}'\widetilde{X_t(\gamma^*)} & 0 \\ * & * & 0 & 0 & 0 \\ & * & * & 0 & 0 \\ & * & * & * & 0 \end{bmatrix} \tag{4.4}$$

where $\mathcal{Z}_t = [W_N Y_t, W_N Y_t(\gamma^*), Z_t, X_t(\gamma^*)]$, $\kappa_3 = \mathbb{E}(v_{it}^3)$ and $\kappa_4 = \mathbb{E}(v_{it}^4)$ and "$\circ$" denotes the Hadamard product. We have the following theorem on the limiting distribution of $\hat{\vartheta}$.

**Theorem 4.1** *Under Assumptions A-G, as $N, T \to \infty$, we have*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta^\circ) \xrightarrow{d} N\Big(0, \Omega_1^{-1}(\Omega_1 + \Omega_2 + \Omega_3)\Omega_1^{-1}\Big),$$

*or equivalently*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta^* + \frac{\sigma^{*2}}{T}\mathbf{v}) \xrightarrow{d} N\Big(0, \Omega_1^{-1}(\Omega_1 + \Omega_2 + \Omega_3)\Omega_1^{-1}\Big),$$

*where* $\vartheta^* = (\rho^*, \varrho^*, \beta^{*\prime}, \delta^{*\prime}, \sigma^{*2})'$ *and* $\Omega_1 = \operatorname*{plim}_{N,T\to\infty} \Omega_{1,NT}$, $\Omega_2 = \lim_{N,T\to\infty} \Omega_{2,NT}$, $\Omega_3 = \operatorname*{plim}_{N,T\to\infty} \Omega_{3,NT}$, *and* $\mathbf{v}$ *is a* $(p+q+3)$*-dimensional vector, whose first* $(p+q+2)$ *elements are all 0 and the last one 1. If* $v_{it}$ *is normally distributed, then*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta^* + \frac{\sigma^{*2}}{T}\iota) \xrightarrow{d} N(0, \Omega_1^{-1}).$$

**Remark 4.2** Theorem 4.1 indicates that the limiting variance of the MLE involves both the skewness and the kurtosis of errors. This is in contrast with the results in standard spatial panel data models that the limiting variance only involves the kurtosis, see Yu et al. (2008), Lee and Yu (2010), Li (2017). When conducting the asymptotic analysis in spatial econometrics, one would encounter the following expression

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} A_t' V_t + \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} V_t' B_t V_t,$$

where $A_t$ is an $N$-dimensional vector which belongs to $\mathscr{F}_{t-1}$, where $\mathscr{F}_t$ is the $\sigma$-field generated by $V_1, V_2, \ldots, V_t$, and $B_t$ is an $N \times N$ nonrandom matrix. Generally, the limiting variance of this expression involves the skewness with the value $\mathbb{E}(v_{it}^3)\frac{1}{NT}\sum_{t=1}^{T} A_t' \text{diag}(B_t)$. However, with the conditions (i) $\sum_{t=1}^{T} A_t = 0$, and (ii) $B_t$ is a constant, we can easily check $\frac{1}{NT}\sum_{t=1}^{T} A_t'\text{diag}(B_t) = 0$, so the skewness term is gone. In standard spatial panel data models, the two conditions are both satisfied. But in the models with structural change, $B_t$ is a piecewise constant, so the skewness term is maintained. □

**Remark 4.3** We can use the plug-in method to estimate the bias and the limiting variance by replacing the unknown true parameters with the corresponding the ML estimators. The validity of this method can be easily verified by the following two basic facts (i) the bias and the limiting variance are both continuous functions of the unknown parameters, (ii) the MLE are consistent due to Proposition 4.1. So the plug-in estimators of the bias and the limiting variances are consistent due to the continuous mapping theorem. As regard the estimations of $\kappa_3$ and $\kappa_4$, they can be estimated by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^3$ and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^4$, where $\hat{v}_{it}$ is the estimated residual. □

To give the limiting distribution of $\hat{\gamma} - \gamma^*$, we introduce the following notations. Let $T^* = [T\gamma^*]$ and

$$\Psi_{2,N}^* = \frac{\kappa_4 - 3\sigma^{*4}}{N\sigma^{*4}} \begin{bmatrix} \text{tr}(S_2^* \circ S_2^*) & \underset{1\times q}{0} \\ \underset{q\times 1}{0} & \underset{q\times q}{0} \end{bmatrix}, \tag{4.5}$$

$$\Psi_{3,NT}^* = \frac{\kappa_3}{N\ell\sigma^{*4}} \sum_{t=T^*+1}^{T^*+\ell} \begin{bmatrix} 2S_2^{d*\prime}(\alpha^* + Z_t\beta^*) & \underset{1\times q}{S_2^{d*\prime}X_t} \\ \underset{q\times 1}{X_t' S_2^{d*}} & \underset{q\times q}{0} \end{bmatrix}, \tag{4.6}$$

with $S_2^{d*} = \text{diag}(S_2^*)$. For the limiting distribution of $\hat{\gamma}$, we make the following assumption.

**Assumption I:** The limits of $\Psi_{1,NT}^*, \Psi_{2,N}^*$ and $\Psi_{3,NT}^*$ exist, where $\Psi_{1,NT}^*$ is defined in (4.2) in Assumption H. Let $\Psi_1^* = \underset{N,T\to\infty}{\text{plim}}\, \Psi_{1,NT}^*, \Psi_2^* = \lim_{N\to\infty} \Psi_{2,N}^*$ and $\Psi_3^* = \lim_{N,T\to\infty} \Psi_{3,NT}^*$. We assume that $\Psi_1^*$ and $\Psi_3^*$ are also the limits of the expressions in (4.1) and (4.6) with $\sum_{t=T^*+1}^{T^*+\ell}$ replaced by $\sum_{t=T^*-\ell}^{T^*}$.

Under Assumption I, the asymptotic behavior of $\hat{\gamma} - \gamma^*$ adjusted with some appropriate scale factor would have a symmetric distribution, which makes the calculation of confidence interval somewhat easier. Similar assumptions also appear in a number of break point studies, e.g., Bai, Lumsdaine and Stock (1998), Bai and Perron (1998), etc. In some applications such as government intervention, Assumption I is plausible since the exogenous explanatory variables are unlikely subject to a structural change given the usually prudent behavior of the government. However, if the exogenous regressors are believed to experience a structural change, we can easily modify Assumption I to accommodate this general case. The analysis under this general case is almost the same with the one under Assumption I by analyzing the cases $\gamma > \gamma^*$ and $\gamma \leq \gamma^*$ separately. For a related treatment, see Bai (1997), Qu and Perron (2007).

We have the following theorem on the limiting distribution of $\hat{\gamma}$.

**Theorem 4.2** *Under Assumptions A-D, E.2, F-I, as $N, T \to \infty$ and $(NT)^{2\nu}/N \to \infty$, we have*

$$(NT)^{1-2\nu}(\hat{\gamma} - \gamma^*) \xrightarrow{d} \frac{C^{*\prime}(\Psi_1^* + \Psi_2^* + \Psi_3^*)C^*}{(C^{*\prime}\Psi_1^* C^*)^2} \underset{s}{\text{argmax}}\, M(s)$$

*with*

$$M(s) = -\frac{1}{2}|s| + B(s).$$

*where $B(s)$ is a two-sided Brownian motion on $(-\infty, \infty)$, which is defined as $B(s) = B_a(-s)$ for $s < 0$ and $B(s) = B_b(s)$ for $v \geq 0$, where $B_a(\cdot)$ and $B_b(\cdot)$ are two independent Brownian motion processes on $[0, \infty)$ with $B_a(0) = B_b(0) = 0$. If $v_{it}$ is normally distributed, then*

$$(NT)^{1-2\nu}(\hat{\gamma} - \gamma^*) \xrightarrow{d} \frac{1}{C^{*\prime}\Psi_1^* C^*} \underset{s}{\text{argmax}}\, M(s)$$

**Remark 4.4** The above limiting distribution can be written alternatively as

$$\left[\frac{(\psi^{*\prime}\Psi_1^*\psi^*)^2}{\psi^{*\prime}(\Psi_1^* + \Psi_2^* + \Psi_3^*)\psi^*}\right] N(\hat{T} - T^*) \xrightarrow{d} \underset{s}{\text{argmax}}\, M(s)$$

where $\psi^* = (\varrho^*, \delta^{*\prime})'$, a value depending on $N$ and $T$ according to Assumption E.2, and $\hat{T} = T\hat{\gamma}$ and $T^* = T\gamma^*$[①]. Again, we can use the plug-in method to estimate the unknown value of $\psi^*, \Psi_1^*, \Psi_2^*$ and $\Psi_3^*$.

[①]The difference between $T\gamma$ and $[T\gamma]$ are negligible due to the condition $(NT)^{2\nu}/N \to \infty$.

The $(1-\alpha)$ confidence interval, where $\alpha$ is the significance level, which should not be confounded with the intercept in the model, now can be constructed as

$$\left[\hat{T} - z_{\frac{\alpha}{2}} \frac{\hat{\psi}'(\hat{\Psi}_1 + \hat{\Psi}_2 + \hat{\Psi}_3)\hat{\psi}}{N(\hat{\psi}'\hat{\Psi}_1\hat{\psi})^2}, \qquad \hat{T} + z_{\frac{\alpha}{2}} \frac{\hat{\psi}'(\hat{\Psi}_1 + \hat{\Psi}_2 + \hat{\Psi}_3)\hat{\psi}}{N(\hat{\psi}'\hat{\Psi}_1\hat{\psi})^2}\right],$$

where $z_{\frac{\alpha}{2}}$ is the critical value such that $P(\text{argmax } M(s) > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, and $\hat{\psi}, \hat{\Psi}_1, \hat{\Psi}_2$ and $\hat{\Psi}_3$ are the respective plug-in estimators for $\psi^*, \Psi_1^*, \Psi_2^*$ and $\Psi_3^*$. The limiting distribution "argmax $M(s)$" have been well studied in the previous studies, see Picard (1985), Yao (1987) and Bai (1997). The 90th and 95th percentiles are 4.67 and 7.63, respectively. □

# 5 Dynamic model

This section considers the extension of the previous analysis to the dynamic spatial panel data model with structural change

$$\begin{aligned} Y_t = {} & \alpha^* + \rho^* W_N Y_t + \varrho^* W_N Y_t \mathbb{1}(t \le [T\gamma^*]) + \phi^* Y_{t-1} + \varphi^* Y_{t-1} \mathbb{1}(t \le [T\gamma^*]) \\ & + Z_t \beta^* + X_t \delta^* \mathbb{1}(t \le [T\gamma^*]) + V_t. \end{aligned}$$

Our model specification assumes that the structural change appears in the lag of dependent variable. If the lag of dependent variable is just introduced to capture the dynamics and no structural change appears on the lag, the analysis can be easily modified to accommodate this simpler case.

To conform with the analysis in Section 3 to the largest extent, we absorb $Y_{t-1}$ into $Z_t$ and $Y_{t-1}\mathbb{1}(t \le [T\gamma^*])$ into $X_t\mathbb{1}(t \le [T\gamma^*])$. So the columns of $Z_t$ and $X_t$ are augmented to $p+1$ and $q+1$, respectively. We arrange that $Y_{t-1}$ and $Y_{t-1}\mathbb{1}(t \le [T\gamma^*])$ are the first columns of $Z_t$ and $X_t\mathbb{1}(t \le [T\gamma^*])$. Note that $Z_t$ and $X_t$ now are not nonrandom matrices due to the presence of lag dependent variable. We additionally make the following assumption for theoretical analysis.

**Assumption J:** $|\phi^*| + |\varphi^*| < 1$ and

$$\sum_{l=1}^{\infty} (|\phi^*| + |\varphi^*|)^l \Big[ \max(\|D_1^{*-1}\|_1, \|D_1^{*-1}\|_\infty, \|D_2^{*-1}\|_1, \|D_2^{*-1}\|_\infty) \Big]^l < \infty,$$

where $D_1^* = D_N(\rho^* + \varrho^*)$ and $D_2^* = D_N(\rho^*)$.

Assumption J can be viewed as a variant of absolute summability condition in time series models, which guarantees that the stochastic part of $Y_t$ is stationary, a property which is needed for the large sample analysis. Similar conditions are also made in dynamic spatial panel data models, such as Assumption E in Li (2017) and Assumption 6 in Yu et al. (2008). A sufficient condition for Assumption J is

$$\Big[|\rho^*| + |\varrho^*| + |\phi^*| + |\varphi^*|\Big] \max(\|W_N\|_1, \|W_N\|_\infty, 1) < 1.$$

The consistency results of the dynamic model are summarized in the following proposition.

**Proposition 5.1** *Let $\theta^* = (\rho^*, \varrho^*, \phi^*, \psi^*, \beta^{*\prime}, \delta^{*\prime}, \gamma^*, \sigma^{*2})'$ and $\hat{\theta}$ be the corresponding MLE. Under Assumptions A-G and J, as $N, T \to \infty$ and $(NT)^{2\nu}/T \to 0$, we have $(NT)^{\nu}(\hat{\vartheta} - \vartheta^*) = o_p(1)$ and $\hat{\gamma} - \gamma^* = o_p(1)$. Under Assumptions A-H, and J, as $N, T \to \infty$, $(NT)^{2\nu}/N \to \infty$ and $(NT)^{2\nu}/T \to 0$, we have the same conclusion with Proposition 4.2.*

**Remark 5.1** In dynamic models, we additionally impose $(NT)^{2\nu}/T \to 0$ to obtain the consistency. This condition makes sure that the effect of the within group transformation on errors can be ignored asymptotically when deriving the consistency. We note that such an effect also exists in the static model. As seen in Proposition 4.1, the preliminary convergence rate is given by $(NT)^{\nu}(\hat{\vartheta} - \vartheta^{\circ}) = o_p(1)$ instead of $(NT)^{\nu}(\hat{\vartheta} - \vartheta^*) = o_p(1)$. However, in static models, the within group transformation only affects the estimation of $\sigma^{*2}$, the remaining parameters are unaffected, see Theorem 4.1. But in dynamic models, all the parameters estimations are affected by the within group transformation, see Theorem 5.1 below. This is the reason why we present the consistency result in different ways in static and dynamic models, although they are essentially the same under the condition $(NT)^{2\nu}/T \to 0$. $\square$

For ease of exposition, we define the following notations. Let

$$Z_t^{\diamond} = \Big[Y_{t-1}, Z_t\Big], \quad X_t^{\diamond}(\gamma^*) = \Big[Y_{t-1}(\gamma^*), X_t(\gamma^*)\Big], \quad \mathcal{Z}_t^{\diamond} = \Big[W_N Y_t, W_N Y_t(\gamma^*), Z_t^{\diamond}, X_t^{\diamond}(\gamma^*)\Big]$$

$$\Delta_N^{\diamond} = \frac{1}{N}\Big[\mathrm{tr}(\mathcal{S}_{\gamma}^*), \gamma^* \mathrm{tr}(\mathcal{S}_1^*), \gamma^* \mathrm{tr}(\mathcal{D}_1^{*-1}) + (1-\gamma^*)\mathrm{tr}(\mathcal{D}_2^{*-1}), \underset{1\times p}{0}, \gamma^* \mathrm{tr}(\mathcal{D}_1^{*-1}), \underset{1\times q}{0}, \frac{1}{2\sigma^{*2}}\Big]',$$

where

$$\mathcal{D}_1^* = (1-\phi^*-\varphi^*)I_N - (\rho^*+\varrho^*)W_N, \qquad \mathcal{D}_2^* = (1-\phi^*)I_N - \rho^* W_N,$$
$$\mathcal{S}_1^* = W_N \mathcal{D}_1^{*-1}, \qquad \mathcal{S}_2^* = W_N \mathcal{D}_2^{*-1}, \qquad \mathcal{S}_{\gamma}^* = \gamma^* \mathcal{S}_1^* + (1-\gamma^*)\mathcal{S}_2^*.$$

Furthermore, let $\Omega_{1,NT}^{\diamond}, \Omega_{2,NT}^{\diamond}$ and $\Omega_{3,NT}^{\diamond}$ be defined similarly as $\Omega_{1,NT}, \Omega_{2,NT}$ and $\Omega_{3,NT}$ in (4.2), (4.3) and (4.4) except that $\mathcal{Z}_t$, $Z_t$, $X_t(\gamma^*)$, $S_1^*$ and $S_2^*$ are replaced with $\mathcal{Z}_t^{\diamond}, Z_t^{\diamond}$, $X_t^{\diamond}(\gamma^*)$, $\mathcal{S}_1^*$ and $\mathcal{S}_2^*$, respectively. Likewise, let $\Psi_{1,NT}^{\diamond}$ and $\Psi_{3,NT}^{\diamond}$ be defined similarly as $\Psi_{1,NT}^*$ and $\Psi_{3,NT}^*$ with $Z_t$ and $X_t$ replaced by $Z_t^{\diamond}$ and $X_t^{\diamond}$. Given these definitions, Assumption I should be modified by replacing $\Psi_{1,NT}^*$ and $\Psi_{3,NT}^*$ with $\Psi_{1,NT}^{\diamond}$ and $\Psi_{3,NT}^{\diamond}$.

We have the following theorem on the limiting distribution of the MLE.

**Theorem 5.1** *Under Assumptions A-H and J, as $N, T \to \infty$ and $(NT)^{2\nu}/T \to 0$ and $N/T^3 \to 0$, we have*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta^* + \frac{1}{T}\Omega_{1,NT}^{\diamond-1}\Delta_N^{\diamond}) \xrightarrow{d} N\Big(0, \Omega_1^{\diamond-1}(\Omega_1^{\diamond} + \Omega_2^{\diamond} + \Omega_3^{\diamond})\Omega_1^{\diamond-1}\Big).$$

*Under Assumptions A-D, E.2 and F-J, as $N, T \to \infty$, $(NT)^{2\nu}/N \to \infty$ and $(NT)^{2\nu}/T \to 0$, we have*

$$(NT)^{1-2\nu}(\hat{\gamma} - \gamma^*) \xrightarrow{d} \frac{C^{*\prime}(\Psi_1^{\diamond} + \Psi_2^* + \Psi_3^{\diamond})C^*}{(C^{*\prime}\Psi_1^{\diamond}C^*)^2} \operatorname*{argmax}_s M(s)$$

*with*

$$M(s) = -\frac{1}{2}|s| + B(s).$$

*where* $\Psi_1^\diamond = \underset{N,T\to\infty}{\text{plim}}\, \Psi_{1,NT}^\diamond$ *and* $\Psi_3^\diamond = \underset{N,T\to\infty}{\text{plim}}\, \Psi_{3,NT}^\diamond$. $\Psi_2^*$ *and* $B(s)$ *are defined the same as in Theorem 4.2.*

# 6 Fixed-$T$ setup

In microeconomics, panel data often exhibits a large-$N$, small-$T$ feature, which raises the necessity to investigate the asymptotics under fixed-$T$ setup. This section addresses this concern. It is well known that the within group estimators for dynamic panel data models are inconsistent due to the incidental parameters issue, see, e.g., Anderson and Hsiao (1981). So we only consider the static model. We note, however, that the within-group estimators with some carefully designed bias correction method would have remarkable finite sample performance in dynamic panel models even it is inconsistent, see Dhaene and Jochmans (2015). So the ML estimators, which reduces to the within group estimators, are still useful in practical application in this viewpoint.

A close investigation on the analysis of the theoretical results in Section 3 verifies that all the analysis continue to hold under fixed-$T$. The only difference is that when $T$ is fixed, the condition $(NT)^{2v}/N \to \infty$ breaks down. A concomitant consequence is that we are now capable of estimating the location of break point $T^*$ instead of the fraction value $\gamma^*$, a result that we are happy to see since it means that we can estimate the break point more accurately.

We therefore have the following theorem on the MLE under fixed-$T$ setup.

**Proposition 6.1** *Let $\hat{\theta}$ be the MLE defined the same with Proposition 4.1. Under Assumptions A-G, as $N \to \infty$, we have the same conclusions with Proposition 4.1. Under Assumptions A-H, $N \to \infty$, we have $\mathbb{P}(\hat{T} = T^*) \to 1$, where $\hat{T} = [T\hat{\gamma}]$ and $T^* = [T\gamma^*]$.*

Let $\Omega_{4,NT}$ be defined as

$$\Omega_{4,NT} = \frac{\sigma^{*4}}{NT\sigma^{\circ 2}} \begin{bmatrix} \gamma^*\text{tr}(S_1^*S_1^*) + (1-\gamma^*)\text{tr}(S_2^*S_2^*) - \text{tr}(S_\gamma^*S_\gamma^*) & \gamma^*(1-\gamma^*)\text{tr}(S_1^*S_1^* - S_1^*S_2^*) & \underset{1\times\bar{k}}{0} \\ \gamma^*(1-\gamma^*)\text{tr}(S_1^*S_1^* - S_1^*S_2^*) & \gamma^*(1-\gamma^*)\text{tr}(S_1^*S_1^*) & \underset{1\times\bar{k}}{0} \\ \underset{\bar{k}\times 1}{0} & \underset{\bar{k}\times 1}{0} & \underset{\bar{k}\times\bar{k}}{0} \end{bmatrix}$$

We have following limiting result under fixed $T$.

**Theorem 6.1** *Under Assumptions A-H, as $N \to \infty$, we have*

$$\sqrt{NT}(\hat{\vartheta} - \vartheta^\circ) \xrightarrow{d} N\left(0, \Omega_1^{\dagger-1}\left(\frac{T}{T-1}\Omega_1^\dagger + \Omega_2^\dagger + \Omega_3^\dagger - \Omega_4^\dagger\right)\Omega_1^{\dagger-1}\right).$$

*where* $\Omega_1^\dagger = \underset{N\to\infty}{\text{plim}}\Omega_{1,NT}$, $\Omega_2^\dagger = \underset{N\to\infty}{\text{plim}}\Omega_{2,NT}$, $\Omega_3^\dagger = \underset{N\to\infty}{\text{plim}}\Omega_{3,NT}$ *and* $\Omega_4^\dagger = \underset{N\to\infty}{\text{plim}}\Omega_{4,NT}$, *where* $\Omega_{1,NT}, \Omega_{2,NT}$ *and* $\Omega_{3,NT}$ *are defined the same as in (4.2), (4.3) and (4.4).*

**Remark 6.1** Parameters $\sigma^{*2}$ and $\sigma^{\circ 2}$ can be estimated by $\frac{T}{T-1}\hat{\sigma}^2$ and $\hat{\sigma}^2$, respectively. Note that the limiting variance of $\frac{T}{T-1}\hat{\sigma}^2$ would be enlarged with $\frac{T^2}{(T-1)^2}$. So when conducting the hypothesis testing, we should take into account this factor in the calculation of the standard deviation.

**Remark 6.2** The limiting variances can be estimated by the method given in Remark 4.3. As regard $\kappa_3$ and $\kappa_4$, they should be estimated in a different way. This is because that the estimated residual $\hat{V}_t$ is a consistent estimator for $\widetilde{V}_t$ instead of $V_t$. When $T$ is large, the difference between $\widetilde{V}_t$ and $V_t$ is negligible. But for a fixed $T$, the difference is considerably large and cannot be ignored. To adjust this difference, we estimate these two parameters by the following formulas,

$$\hat{\kappa}_3 = \frac{T^3}{(T-1)^3-(T-1)}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^3,$$
$$\hat{\kappa}_4 = \left[\frac{T^4}{(T-1)^4+T-1}\right]\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^4 - 3\frac{2(T-1)^3+(T-1)(T-2)}{T^4}\Big(\frac{T}{T-1}\hat{\sigma}^2\Big)^2\right].$$

Apparently, when $T$ tends to infinity, the above two expressions converge to $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^3$ and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^4$, respectively, the estimating formulas which are given in Remark 4.3 under large-$T$ setup.

# 7 Testing on the presence of structural change

Testing the presence of structural change is an important issue in structural change models, which has received much attention in econometric literature. A striking feature of this hypothesis testing is that the parameter governing the timing of structural change only appears under the alternative hypothesis, which makes the test essentially different from the classical one. Andrews (1993) gives a comprehensive treatment on this issue. The current analysis departures somewhat from the Andrews's setup in that the model suffers model misspecification issue if the disturbance $e_{it}$ is not gaussian. In addition, the presence of incidental parameters make the partial assumptions such as Assumption 1(c) in Andrews (1993) does not hold.

When the timing of structural change is given under the alternative, the issue reduces to the one covered by the well-known Chow test. The classical Wald (W), Lagrange multiplers (LM) and likelihood ratio (LR) tests are all applicable in this simpler case. To formulate these three classical tests, we first define the ML estimators for the unrestricted and restricted models. Note that the likelihood function for the restricted model is

$$\mathcal{L}^r(\vartheta) = -\frac{1}{2}\ln\sigma^2 + \frac{1}{N}\ln|D_N(\rho)| - \frac{1}{2NT\sigma^2}\sum_{t=1}^{T}\Big[\widetilde{Y}_t - \rho W_N\widetilde{Y}_t - \widetilde{Z}_t\beta\Big]'\Big[\widetilde{Y}_t - \rho W_N\widetilde{Y}_t - \widetilde{Z}_t\beta\Big], \tag{7.1}$$

where $D_N(\rho) = I_N - \rho W_N$, which is defined in Assumption C. The likelihood function for the unrestricted model is given in (2.4). Let $\hat{\vartheta}^r = (\hat{\rho}^r, \hat{\beta}^{r\prime}, \hat{\sigma}^{r2})'$ be the MLE which maximizes the likelihood function (7.1) and $\hat{\vartheta}^u_\gamma = (\hat{\rho}^u_\gamma, \hat{\varrho}^u_\gamma, \hat{\beta}^{u\prime}_\gamma, \hat{\delta}^{u\prime}_\gamma, \hat{\sigma}^{u2}_\gamma)'$ the MLE that maximizes the likelihood function (2.4) with the timing of structural change set to $\gamma$. We attach the superscript "$u$" to the MLE for the unrestricted model and the superscript "$r$" for the resstricted one. Let $\hat{\vartheta}^{ur} = (\hat{\rho}^r, 0, \hat{\beta}^{r\prime}, 0_{1\times q}, \hat{\sigma}^{r2})'$ be the restricted MLE adapted to the unrestricted model. Furthermore, we define the $(p+q+3)\times(q+1)$ matrix $\mathbf{e}$ as

$$\mathbf{e} = \begin{bmatrix} 0 & 1 & \underset{1\times p}{0} & \underset{1\times q}{0} & 0 \\ \underset{q\times 1}{0} & \underset{q\times 1}{0} & \underset{q\times p}{0} & I_q & 0 \end{bmatrix}'.$$

The Wald, LM and LR statistics on testing the null hypothesis $\psi^* = 0$ versus the alternative that $\psi^* \neq 0$ and the timing of break point is $\gamma$, are

$$W_{NT}(\gamma) = NT\hat{\psi}^{u\prime}_\gamma \left\{ \mathbf{e}'[\hat{\Omega}^u_1(\gamma)]^{-1}\Big[\hat{\Omega}^u_1(\gamma) + \hat{\Omega}^u_2(\gamma) + \hat{\Omega}^u_3(\gamma)\Big][\hat{\Omega}^u_1(\gamma)]^{-1}\mathbf{e}\right\}^{-1} \hat{\psi}^u_\gamma, \tag{7.2}$$

$$\begin{aligned} LM_{NT}(\gamma) = {} & NT\widehat{\mathbb{J}(\gamma)}'\Big[\mathbf{e}'[\hat{\Omega}^r_1(\gamma)]^{-1}\mathbf{e}\Big]\Big\{\mathbf{e}'[\hat{\Omega}^r_1(\gamma)]^{-1}[\hat{\Omega}^r_1(\gamma) + \hat{\Omega}^r_2(\gamma) + \hat{\Omega}^r_3(\gamma)] \\ & \times [\hat{\Omega}^r_1(\gamma)]^{-1}\mathbf{e}\Big\}^{-1}\Big[\mathbf{e}'[\hat{\Omega}^r_1(\gamma)]^{-1}\mathbf{e}\Big]\widehat{\mathbb{J}(\gamma)}, \end{aligned} \tag{7.3}$$

$$\begin{aligned} LR_{NT}(\gamma) = {} & 2NT\Big[\mathcal{L}(\hat{\vartheta}^u_\gamma, \gamma) - \mathcal{L}^r(\hat{\theta}^r)\Big] - NT\Big[(\hat{\vartheta}^{ur} - \hat{\vartheta}^u_\gamma)'\mathbf{e}\Big]\Big[\mathbf{e}'[\hat{\Omega}^u_1(\gamma)]^{-1}\mathbf{e} \\ & + \mathbf{e}'[\hat{\Omega}^u_1(\gamma)]^{-1}\mathbf{e}\Big(\mathbf{e}'[\hat{\Omega}^u_1(\gamma)]^{-1}[\hat{\Omega}^u_2(\gamma) + \hat{\Omega}^u_3(\gamma)][\hat{\Omega}^u_1(\gamma)]^{-1}\mathbf{e}\Big)^{-1} \\ & \times \mathbf{e}'[\hat{\Omega}^u_1(\gamma)]^{-1}\mathbf{e}\Big]^{-1}\Big[\mathbf{e}'(\hat{\vartheta}^{ur} - \hat{\vartheta}^u_\gamma)\Big], \end{aligned} \tag{7.4}$$

where $\widehat{\mathbb{J}(\gamma)} = [\widehat{\mathbb{J}_1(\gamma)}, \quad \widehat{\mathbb{J}_2(\gamma)}']'$ with

$$\begin{aligned} \widehat{\mathbb{J}_1(\gamma)} &= \frac{\partial\mathcal{L}(\hat{\vartheta}^{ur}, \gamma)}{\partial\varrho} = \frac{1}{NT\hat{\sigma}^{r2}}\sum_{t=1}^T Y_t(\gamma)'W_N'(\hat{D}^r\widetilde{Y}_t - \widetilde{Z}_t\hat{\beta}^r) - \gamma\frac{1}{N}\mathrm{tr}\Big[W_N(\hat{D}^r)^{-1}\Big], \\ \widehat{\mathbb{J}_2(\gamma)} &= \frac{\partial\mathcal{L}(\hat{\vartheta}^{ur}, \gamma)}{\partial\delta} = \frac{1}{NT\hat{\sigma}^{r2}}\sum_{t=1}^T X_t(\gamma)'(\hat{D}^r\widetilde{Y}_t - \widetilde{Z}_t\hat{\beta}^r). \end{aligned}$$

where $\hat{\Omega}^u_1(\gamma), \hat{\Omega}^u_2(\gamma)$ and $\hat{\Omega}^u_3(\gamma)$ [$\hat{\Omega}^r_1(\gamma), \hat{\Omega}^r_2(\gamma)$ and $\hat{\Omega}^r_3(\gamma)$] are the respective estimators of $\Omega_1(\gamma), \Omega_2(\gamma)$ and $\Omega_3(\gamma)$, by replacing the unknown parameters by the MLE for the unrestricted (restricted) model.

Note that the above three statistics would change with $\gamma$. If $\gamma$ is not specified under the alternative, to make our statistics possess the most conservative size, we should choose a $\gamma$ which is the most favorable to the alternative. Obviously we would accept the alternative with a large $W_{NT}(\gamma)$ (or $LM_{NT}(\gamma)$ and $LR_{NT}(\gamma)$). As a result, we should maximize the above $W_{NT}(\gamma)$ (or $LM_{NT}(\gamma)$ and $LR_{NT}(\gamma)$), which leads to the so-called supW, supLM and supLR statistics, which are defined as

$$\text{supWald} = \sup_{\gamma\in[\gamma^L,\gamma^U]} W_{NT}(\gamma), \ \text{supLM} = \sup_{\gamma\in[\gamma^L,\gamma^U]} LM_{NT}(\gamma), \ \text{supLR} = \sup_{\gamma\in[\gamma^L,\gamma^U]} LR_{NT}(\gamma).$$

To establish the asymptotics of the above three super tests under the null, we need to establish the convergences of $W_{NT}(\cdot), LM_{NT}(\cdot)$ and $LR_{NT}(\cdot)$ under the Skorohod metric, otherwise we cannot invoke the continuous mapping theorem to derive the final limiting distributions. To present the results, we first introduce some notations.

$$\begin{aligned}
\Psi^{\star}_{1,NT} &= \frac{1}{NT\sigma^{*2}}\sum_{t=1}^{T}\begin{bmatrix} Y_t'W_N'W_NY_t & Y_t'W_N'X_t \\ X_t'W_NY_t & X_t'X_t \end{bmatrix} + \frac{1}{N}\begin{bmatrix} \mathrm{tr}(S_N^*S_N^*) & \underset{1\times q}{0} \\ \underset{q\times 1}{0} & \underset{q\times q}{0} \end{bmatrix} \\
\Psi^{\star}_{2,N} &= \frac{\kappa_4 - 3\sigma^{*4}}{N\sigma^{*4}}\begin{bmatrix} \mathrm{tr}(S_N^*\circ S_N^*) & \underset{1\times q}{0} \\ \underset{q\times 1}{0} & \underset{q\times q}{0} \end{bmatrix} \\
\Psi^{\star}_{3,NT} &= \frac{\kappa_3}{NT\sigma^{*4}}\sum_{t=1}^{T}\begin{bmatrix} 2S_N^{d*\prime}(\alpha^* + Z_t\beta^*) & \underset{1\times q}{S_N^{d*\prime}X_t} \\ \underset{q\times 1}{X_t'S_N^{d*}} & \underset{q\times q}{0} \end{bmatrix}
\end{aligned}$$

Now we have the following theorem on the basic results.

**Theorem 7.1** *Under Assumptions A-H and G.1, if the null hypothesis holds, as $N, T \to \infty$, we have*

$$\begin{aligned}
&\mathbf{e}'[\hat{\Omega}_1^u(\gamma)]^{-1}\Big[\hat{\Omega}_1^u(\gamma) + \hat{\Omega}_2^u(\gamma) + \hat{\Omega}_3^u(\gamma)\Big][\hat{\Omega}_1^u(\gamma)]^{-1}\mathbf{e} \\
&\qquad\qquad = \frac{1}{\gamma(1-\gamma)}\Psi_1^{\star-1}(\Psi_1^\star + \Psi_2^\star + \Psi_3^\star)\Psi_1^{\star-1} + o_{p\gamma}(1),
\end{aligned}$$

*and*

$$\sqrt{NT}\hat{\psi}_\gamma^u \Rightarrow \frac{1}{\gamma(1-\gamma)}\Psi_1^{\star-1}(\Psi_1^\star + \Psi_2^\star + \Psi_3^\star)^{1/2}\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big],$$

*where $o_{p\gamma}(1)$ denotes some term which is $o_p(1)$ uniformly on $\gamma \in [\gamma^L, \gamma^U]$, $B_q(\cdot)$ is the $q$-dimensional standard Brownian motion on $[0,1]$, and "$\Rightarrow$" denotes the weak convergence in the Skorohod topology. In addition, $\Psi_1^\star = \operatorname*{plim}_{N,T\to\infty}\Psi^\star_{1,NT}$, $\Psi_2^\star = \lim_{N\to\infty}\Psi_{2,N}$ and $\Psi_3^\star = \lim_{N,T\to\infty}\Psi^\star_{3,NT}$. Given the above results, we have*

$$W_{NT}(\gamma) \Rightarrow \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big].$$

*In addition, we also have*

$$\begin{aligned}
LM_{NT}(\gamma) &\Rightarrow \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big], \\
LR_{NT}(\gamma) &\Rightarrow \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big].
\end{aligned}$$

**Corollary 7.1** *Under the assumptions in Theorem 7.1, we have*

$$\sup_{\gamma\in[\gamma^L,\gamma^U]} W_{NT}(\gamma) \xrightarrow{d} \sup_{\gamma\in[\gamma^L,\gamma^U]} \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma) - \gamma B_{q+1}(1)\Big],$$

$$\sup_{\gamma\in[\gamma^L,\gamma^U]} LM_{NT}(\gamma) \xrightarrow{d} \sup_{\gamma\in[\gamma^L,\gamma^U]} \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma)-\gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma)-\gamma B_{q+1}(1)\Big],$$
$$\sup_{\gamma\in[\gamma^L,\gamma^U]} LR_{NT}(\gamma) \xrightarrow{d} \sup_{\gamma\in[\gamma^L,\gamma^U]} \frac{1}{\gamma(1-\gamma)}\Big[B_{q+1}(\gamma)-\gamma B_{q+1}(1)\Big]'\Big[B_{q+1}(\gamma)-\gamma B_{q+1}(1)\Big].$$

**Remark 7.1** The powers of the supW, supLM and supLR tests can be studied in the same way under the local alternative $\varpi^\star = \varpi^* + \mu(t/T)/\sqrt{NT}$, where $\varpi^* = (\rho^*, \beta_a^{*\prime})'$ with $\beta_a^*$ is the coefficients of the exogenous regressors that are suspected to experience a structural change, and $\mu(\cdot)$ is a bounded real-valued function on $[0,1]$, which satisfies some regularity conditions. For more details, see Assumption 1-LP and the related discussions in Andrews (1993). We will not pursuit this work for the sake of space. □

**Remark 7.2** The critical values for the supW, supLM and supLR tests are given in Andrews (2003) and Estrella (2003). The critical values depend on the column dimension of $X_t$ (i.e., the value $q$) and the chosen interval $[\gamma^L, \gamma^U]$. As $\gamma^L$ decreases to zero and $\gamma^U$ increases to one, the critical values would diverge to infinity, a result which is shown in Corollary 1 of Andrews (1993). Many studies recommend that the interval is chosen to be [0.15, 0.85], e.g., Andrews (1993). We will adopt this interval in our simulations investigation. Given this interval, according to Estrella (2003), the critical values for 10%, 5% and 1% significance levels are 10.14, 11,87 and 15.69 when $p = 1$, and 12.46, 14.31 and 18.36 when $p = 2$. □

**Remark 7.3** The supW, supLM and supLR statistics in the dynamic model can be constructed similarly as those in the static model. The only caveat is that these statistics should be computed through the bias-corrected MLE, instead of the original one, to remove the effect of bias. The analyses on the three statistics in the dynamic model are similar as in the static model, by treating $Y_{t-1}$ as a part of $Z_t$. Under the condition that $N/T^3 \to 0$, we can show that these three statistics have the same limiting distribution as in Corollary 7.1. □

# 8 Simulations

We run Monte Carlo simulations to investigate the finite sample performance of the ML estimators in this section.

## 8.1 Static spatial panel data models

The data are generated according to

$$Y_t = \alpha + \rho^* W_N Y_t + \varrho^* W_N Y_t \mathbb{1}(t \le [T\gamma^*]) + X_{t1}\beta_1^* + X_{t2}\beta_2^* + X_{t1}\mathbb{1}(t \le [T\gamma^*])\delta^* + V_t,$$

with $(\rho^*, \varrho^*, \beta_1^*, \beta_2^*, \delta^*, \gamma^*, \sigma^{*2}) = (0.4, -0.1, 2, 1, -1, 0.25, 0.36)$. Our data generating process specifies two exogenous regressors and one regressor experiences structural change, the other does not. The spatial weights matrices used in simulations are "$q$ ahead and $q$ behind" spatial weights matrix as in Kelejian and Prucha (1999), which is obtained as follows: 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 equal to 1. In our simulations, we consider "3 ahead and 3 behind".

All the elements of the exogenous regressors $X_{t1}, X_{t2}$ and the intercept $\alpha$ are drawn independently from $N(0,1)$. The disturbance $v_{it}$, the $i$th element of $V_t$, is 0.6 times of a normalized $\chi^2(2)$, i.e., $[\chi^2(2)-2]/2$. Once $X_{t1}, X_{t2}, \alpha$ and $V_t$ are generated, we calculate $Y_t$ by

$$Y_t = \Big[I_N - \rho^* W_N - \varrho^* W_N \mathbb{1}(t \leq [T\gamma^*])\Big]^{-1}\Big[\alpha + X_{t1}\beta_1^* + X_{t2}\beta_2^* + X_{t1}\mathbb{1}(t \leq [T\gamma^*])\delta^* + V_t\Big].$$

Throughout this section, we use bias and root mean square error (RMSE) as the measures of the performance of the ML estimators. To investigate the estimation accuracy of asymptotic variances, we calculate the empirical sizes of the $t$-statistic for 5% nominal level. Table 1 presents the simulation results under the combinations of $N = 50, 75, 100$ and $T = 50, 75, 100$, which are obtained by 1000 repetitions. In this section, we do not evaluate the performance of $\hat{\gamma}$, partly because the bias and RMSE are not appropriate measures with respect to $\hat{\gamma}$, and partly because the performance of $\hat{\gamma}$ is implicitly shown in the performance of the ML estimators for the other parameters. From Table 1, we have following findings. First, the ML estimators are consistent. As $N$ and $T$ tends to large, the RMSE decrease stably. Second, the ML estimators for $\rho, \varrho, \beta$ and $\delta$ are unbiased. In all the sample sizes, the biases of these ML estimators are very small in terms of both absolute values and relative values, where the relative value is defined by the ratio of bias and RMSE. Third, the ML estimator for $\sigma^2$ is biased, and the bias is loosely related with $N$ and closely related with $T$. Consider the case of $T = 50$, the biases are $-0.0073, -0.0077$ and $-0.0074$ for $N = 50, 75$ and 100. Obviously, the increase of $N$ has no effect on the bias, but when $T$ grows larger, the bias decrease dramatically. This is consistent with our theoretical result in Theorem 4.2. Table 2 presents the results under the combination of $N = 300, 500, 700$ and $T = 8, 12$. The results are similar as those in Table. So we do not repeat the analysis.

Tables 3 and 4 present the empirical sizes of the $t$-test for nominal 5% significance level. We see that all the empirical sizes are close to the nominal one except the estimator

Table 1: The performance of the MLE with moderate large $T$ in the static model

| $N$ | $T$ | $\rho$ | | $\varrho$ | | $\beta_1$ | | $\beta_2$ | | $\delta$ | | $\sigma^2$ (before) | | $\sigma^2$ (after) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 30 | 50 | -0.0008 | 0.0101 | -0.0007 | 0.0273 | 0.0004 | 0.0140 | 0.0005 | 0.0122 | 0.0003 | 0.0290 | -0.0073 | 0.0213 | -0.0001 | 0.0204 |
| 75 | 50 | 0.0000 | 0.0085 | -0.0018 | 0.0231 | 0.0002 | 0.0121 | -0.0001 | 0.0100 | 0.0003 | 0.0247 | -0.0077 | 0.0179 | -0.0005 | 0.0165 |
| 100 | 50 | 0.0000 | 0.0074 | -0.0005 | 0.0202 | -0.0003 | 0.0097 | 0.0003 | 0.0083 | 0.0015 | 0.0204 | -0.0074 | 0.0160 | -0.0002 | 0.0144 |
| 50 | 75 | -0.0006 | 0.0084 | -0.0007 | 0.0222 | 0.0003 | 0.0116 | -0.0001 | 0.0095 | 0.0001 | 0.0237 | -0.0047 | 0.0165 | 0.0001 | 0.0161 |
| 75 | 75 | -0.0001 | 0.0069 | -0.0009 | 0.0189 | -0.0000 | 0.0095 | 0.0002 | 0.0078 | -0.0005 | 0.0193 | -0.0056 | 0.0145 | -0.0008 | 0.0136 |
| 100 | 75 | 0.0001 | 0.0057 | -0.0003 | 0.0156 | -0.0004 | 0.0080 | -0.0000 | 0.0070 | -0.0000 | 0.0162 | -0.0051 | 0.0122 | -0.0003 | 0.0112 |
| 50 | 100 | -0.0002 | 0.0073 | 0.0002 | 0.0195 | 0.0002 | 0.0101 | 0.0002 | 0.0087 | -0.0009 | 0.0196 | -0.0032 | 0.0147 | 0.0004 | 0.0145 |
| 75 | 100 | -0.0001 | 0.0060 | -0.0007 | 0.0154 | -0.0002 | 0.0081 | -0.0004 | 0.0071 | -0.0001 | 0.0164 | -0.0040 | 0.0124 | -0.0004 | 0.0119 |
| 100 | 100 | -0.0002 | 0.0052 | 0.0000 | 0.0136 | -0.0001 | 0.0068 | -0.0002 | 0.0058 | -0.0001 | 0.0140 | -0.0037 | 0.0105 | -0.0001 | 0.0099 |

Table 2: The performance of the MLE with small $T$ in the static model

| $N$ | $T$ | $\rho$ | | $\varrho$ | | $\beta_1$ | | $\beta_2$ | | $\delta$ | | $\sigma^2$ (before) | | $\sigma^2$ (after) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 300 | 8 | -0.0006 | 0.0114 | -0.0024 | 0.0280 | 0.0003 | 0.0155 | -0.0002 | 0.0132 | -0.0004 | 0.0315 | -0.0456 | 0.0491 | -0.0006 | 0.0210 |
| 500 | 8 | -0.0000 | 0.0088 | -0.0009 | 0.0222 | -0.0001 | 0.0117 | -0.0001 | 0.0100 | -0.0009 | 0.0243 | -0.0452 | 0.0473 | -0.0003 | 0.0160 |
| 700 | 8 | 0.0002 | 0.0075 | -0.0014 | 0.0186 | -0.0002 | 0.0100 | 0.0000 | 0.0085 | -0.0000 | 0.0201 | -0.0448 | 0.0466 | 0.0002 | 0.0144 |
| 300 | 12 | -0.0005 | 0.0088 | -0.0017 | 0.0230 | 0.0006 | 0.0121 | 0.0003 | 0.0109 | -0.0010 | 0.0238 | -0.0306 | 0.0345 | -0.0007 | 0.0173 |
| 500 | 12 | 0.0002 | 0.0068 | -0.0001 | 0.0181 | 0.0002 | 0.0095 | -0.0000 | 0.0080 | 0.0004 | 0.0185 | -0.0298 | 0.0322 | 0.0002 | 0.0132 |
| 700 | 12 | -0.0003 | 0.0059 | -0.0004 | 0.0146 | -0.0003 | 0.0080 | 0.0004 | 0.0067 | -0.0004 | 0.0157 | -0.0308 | 0.0324 | -0.0008 | 0.0113 |

Note: $\sigma^2$ (before) and $\sigma^2$ (after) denote the estimators before and after the bias correction, respectively.

Table 3: The empirical sizes of $t$-test under nominal 5% significance level in moderate large-$T$ setup

| $N$ | $T$ | $\rho$ | $\varrho$ | $\beta_1$ | $\beta_2$ | $\delta$ | $\sigma^2$ before | $\sigma^2$ after |
|---|---|---|---|---|---|---|---|---|
| 50 | 50 | 4.6% | 5.3% | 5.1% | 5.0% | 5.1% | 10.1% | 6.8% |
| 75 | 50 | 5.4% | 5.6% | 6.1% | 6.2% | 6.0% | 9.7% | 6.1% |
| 100 | 50 | 5.8% | 5.9% | 5.3% | 4.2% | 5.1% | 10.2% | 4.8% |
| 50 | 75 | 5.2% | 4.4% | 4.6% | 4.4% | 5.7% | 5.8% | 4.9% |
| 75 | 75 | 5.3% | 5.5% | 5.7% | 4.3% | 5.5% | 9.6% | 6.4% |
| 100 | 75 | 3.8% | 5.0% | 4.3% | 5.8% | 5.1% | 7.9% | 4.0% |
| 50 | 100 | 5.4% | 5.2% | 5.4% | 5.7% | 5.1% | 6.0% | 4.7% |
| 75 | 100 | 4.9% | 4.1% | 5.0% | 5.5% | 5.2% | 6.8% | 5.5% |
| 100 | 100 | 5.3% | 5.4% | 4.1% | 4.4% | 5.1% | 7.0% | 4.7% |

Table 4: The empirical sizes of $t$-test under nominal 5% significance level in small-$T$ setup

| $N$ | $T$ | $\rho$ | $\varrho$ | $\beta_1$ | $\beta_2$ | $\delta$ | $\sigma^2$ before | $\sigma^2$ after |
|---|---|---|---|---|---|---|---|---|
| 300 | 8 | 4.7% | 4.8% | 5.5% | 6.5% | 5.1% | 67.8% | 9.6% |
| 500 | 8 | 5.7% | 5.0% | 4.5% | 4.7% | 6.5% | 85.4% | 8.8% |
| 700 | 8 | 5.6% | 4.4% | 4.8% | 4.3% | 5.2% | 92.6% | 8.5% |
| 300 | 12 | 4.6% | 5.6% | 4.7% | 5.0% | 5.6% | 52.5% | 8.3% |
| 500 | 12 | 4.9% | 5.3% | 5.5% | 4.3% | 5.3% | 68.2% | 7.6% |
| 700 | 12 | 5.2% | 4.8% | 5.2% | 4.4% | 4.7% | 82.9% | 7.7% |

$\hat{\sigma}^2$, which suffers a mild size distortion under moderate large $T$ and a severe size distortion under fixed $T$. But after conducting bias correction, the performance has been much improved. Overall, the performance of the MLE after bias correction are satisfactory.

### 8.2 Dynamic spatial panel data models

We next examine the performance of the ML estimators in the dynamic model. The data are generated according to

$$\begin{aligned} Y_t = &\alpha + \rho^* W_N Y_t + \varrho^* W_N Y_t \mathbb{1}(t \le [T\gamma^*]) + Y_{t-1}\phi^* + X_{t1}\beta_1^* + X_{t2}\beta_2^* \\ &+ Y_{t-1}\mathbb{1}(t \le [T\gamma^*])\varphi^* + X_{t1}\mathbb{1}(t \le [T\gamma^*])\delta^* + V_t, \end{aligned}$$

with $(\rho^*, \varrho^*, \phi^*, \beta_1^*, \beta_2^*, \varphi^*, \delta^*, \gamma^*, \sigma^{*2}) = (0.4, -0.1, 0.5, 2, 1, -0.2, -1, 0.25, 0.36)$. The intercept $\alpha^*$, the exogenous regressors $X_{t1}$ and $X_{t2}$ and the disturbance $V_t$ are generated by the same way in the static model. The data of dependent variable is calculated recursively by

$$Y_t = \Big[I_N - \rho^* W_N - \varrho^* W_N \mathbb{1}(t \le [T\gamma^*])\Big]^{-1} \Big[\alpha^* + Y_{t-1}\phi^* + X_{t1}\beta_1^* + X_{t2}\beta_2^*$$

$$+ Y_{t-1}\mathbb{1}(t \le [T\gamma^*])\varphi^* + X_{t1}\mathbb{1}(t \le [T\gamma^*])\delta^* + V_t\Big].$$

with $Y_1 = 0$. To eliminate the effect of initial values, we generate $T + 200$ periods for data and discard the first 200 periods.

Tables 5 and 6 present the simulation results measured by the bias and RMSE. Tables 7 and 8 present the simulation results measured by the empirical sizes. From table 5, we see that the ML estimators have the bias issue. This issue is relatively severe for the estimators of $\phi, \varphi$ and $\sigma^2$. The other estimators seem fine for this issue. If we conduct bias correction, the bias issue is effectively removed since, as seen, all the estimators are centered approximately around zero. The results for empirical sizes of the $t$-tests in Tables 7 and 8 echo those in Tables 5 and 6. The $t$-statistics for $\phi$ and $\sigma^2$ suffer mild size distortion. But after bias corrections, the empirical sizes are improved with some extent.

### 8.3 Three super statistics

In this subsection, we investigate the performance of the three super statistics. We only consider the size of tests. The data are generated according to

$$Y_t = \alpha^* + \rho^* W_N Y_t + X_t\beta^* + V_t.$$

with $(\rho^*, \beta^*, \sigma^{*2}) = (0.5, 1, 0.36)$. The intercept $\alpha^*$, the exogenous regressor $X_t$ and the disturbance $V_t$ are generated by the same way in the static model. The dependent variables are calculated according to the preceding equation. We calculate the three super statistics through (7.2), (7.3) and (7.4) by setting $[\gamma^L, \gamma^U] = [0.15, 0.85]$. The critical values, according to Estrella (2003), is 10.14, 11.87 and 15.69 for 10%, 5% and 1%. Note that the critical values of Estrella (2003) are calculated by the numerical method. Due to the nature of super statistics, Estrella's critical values are greater than those under finite sample sizes. As a result, the empirical sizes of the simulations would be downward relative to the nominal size.

Table 9 presents the empirical sizes of three super statistics. As expected, all the empirical sizes are smaller than the nominal ones and suffer mild size distortion, due to the reason of critical values. If one is unpleasant with such mild size distortions, he can use simulation method to get the critical values of finite sample, see Andrews (1993) for a related discussion. Beside this result, we also find that the three statistics almost have the same empirical sizes. However, the computation times of the three statistics are quite different. The supLM statistics takes the shortest time, the supW one next, and the supLR takes the longest time. Given the close performance of these three statistics, as well as the computation times, we recommend the supLM statistics in real data applications.

Table 5: The performance of the MLE in the dynamic model before bias correction

| $N$ | $T$ | $\rho$ | | $\varrho$ | | $\phi$ | | $\beta_1$ | | $\beta_2$ | | $\varphi$ | | $\delta$ | | $\sigma^2$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 50 | 50 | 0.0003 | 0.0077 | -0.0039 | 0.0242 | -0.0022 | 0.0053 | 0.0005 | 0.0138 | 0.0005 | 0.0125 | -0.0022 | 0.0130 | -0.0003 | 0.0292 | -0.0080 | 0.0224 |
| 75 | 50 | 0.0004 | 0.0063 | -0.0020 | 0.0199 | -0.0021 | 0.0047 | -0.0003 | 0.0115 | 0.0003 | 0.0099 | -0.0018 | 0.0107 | 0.0008 | 0.0245 | -0.0082 | 0.0186 |
| 100 | 50 | 0.0002 | 0.0054 | -0.0018 | 0.0167 | -0.0022 | 0.0041 | -0.0003 | 0.0097 | -0.0004 | 0.0089 | -0.0014 | 0.0089 | -0.0002 | 0.0213 | -0.0081 | 0.0159 |
| 50 | 75 | -0.0003 | 0.0058 | -0.0027 | 0.0193 | -0.0014 | 0.0043 | -0.0005 | 0.0110 | -0.0001 | 0.0098 | -0.0011 | 0.0102 | 0.0006 | 0.0235 | -0.0056 | 0.0171 |
| 75 | 75 | 0.0001 | 0.0050 | -0.0011 | 0.0152 | -0.0014 | 0.0035 | 0.0001 | 0.0095 | -0.0007 | 0.0082 | -0.0014 | 0.0085 | 0.0007 | 0.0200 | -0.0058 | 0.0148 |
| 100 | 75 | 0.0004 | 0.0043 | -0.0008 | 0.0136 | -0.0014 | 0.0032 | 0.0002 | 0.0078 | -0.0003 | 0.0072 | -0.0013 | 0.0073 | 0.0001 | 0.0174 | -0.0053 | 0.0127 |
| 50 | 100 | -0.0001 | 0.0053 | -0.0006 | 0.0161 | -0.0010 | 0.0036 | 0.0002 | 0.0100 | -0.0006 | 0.0086 | -0.0008 | 0.0087 | -0.0002 | 0.0190 | -0.0036 | 0.0148 |
| 75 | 100 | -0.0001 | 0.0044 | -0.0003 | 0.0131 | -0.0009 | 0.0030 | -0.0000 | 0.0081 | 0.0001 | 0.0067 | -0.0011 | 0.0072 | -0.0003 | 0.0161 | -0.0044 | 0.0125 |
| 100 | 100 | 0.0002 | 0.0036 | -0.0008 | 0.0116 | -0.0010 | 0.0026 | -0.0004 | 0.0070 | -0.0001 | 0.0060 | -0.0010 | 0.0060 | 0.0006 | 0.0141 | -0.0043 | 0.0114 |

Table 6: The performance of the MLE in the dynamic model after bias correction

| $N$ | $T$ | $\rho$ | | $\varrho$ | | $\phi$ | | $\beta_1$ | | $\beta_2$ | | $\varphi$ | | $\delta$ | | $\sigma^2$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| 50 | 50 | 0.0000 | 0.0077 | -0.0030 | 0.0240 | -0.0000 | 0.0048 | 0.0008 | 0.0138 | 0.0006 | 0.0126 | -0.0005 | 0.0128 | -0.0005 | 0.0292 | -0.0008 | 0.0214 |
| 75 | 50 | 0.0001 | 0.0063 | -0.0011 | 0.0198 | 0.0001 | 0.0042 | -0.0000 | 0.0115 | 0.0004 | 0.0099 | -0.0001 | 0.0106 | 0.0006 | 0.0245 | -0.0011 | 0.0170 |
| 100 | 50 | -0.0001 | 0.0054 | -0.0009 | 0.0165 | -0.0001 | 0.0034 | 0.0000 | 0.0097 | -0.0003 | 0.0089 | 0.0002 | 0.0088 | -0.0003 | 0.0213 | -0.0009 | 0.0140 |
| 50 | 75 | -0.0006 | 0.0059 | -0.0021 | 0.0192 | -0.0000 | 0.0041 | -0.0004 | 0.0110 | -0.0001 | 0.0098 | 0.0000 | 0.0102 | 0.0005 | 0.0235 | -0.0008 | 0.0164 |
| 75 | 75 | -0.0001 | 0.0050 | -0.0005 | 0.0151 | 0.0000 | 0.0032 | 0.0003 | 0.0095 | -0.0006 | 0.0082 | -0.0004 | 0.0084 | 0.0006 | 0.0200 | -0.0010 | 0.0139 |
| 100 | 75 | 0.0001 | 0.0043 | -0.0002 | 0.0135 | -0.0001 | 0.0028 | 0.0003 | 0.0078 | -0.0002 | 0.0072 | -0.0003 | 0.0072 | -0.0000 | 0.0173 | -0.0005 | 0.0118 |
| 50 | 100 | -0.0003 | 0.0053 | -0.0001 | 0.0161 | 0.0001 | 0.0035 | 0.0002 | 0.0100 | -0.0005 | 0.0086 | 0.0001 | 0.0087 | -0.0003 | 0.0191 | 0.0000 | 0.0145 |
| 75 | 100 | -0.0003 | 0.0044 | 0.0001 | 0.0130 | 0.0001 | 0.0029 | 0.0001 | 0.0081 | 0.0001 | 0.0067 | -0.0003 | 0.0071 | -0.0004 | 0.0161 | -0.0008 | 0.0118 |
| 100 | 100 | -0.0000 | 0.0036 | -0.0003 | 0.0116 | -0.0000 | 0.0024 | -0.0003 | 0.0070 | -0.0000 | 0.0060 | -0.0002 | 0.0060 | 0.0005 | 0.0141 | -0.0007 | 0.0106 |

Table 7: The empirical sizes of $t$-test under nominal 5% significance level before bias correction

| $N$ | $T$ | $\rho$ | $\varrho$ | $\phi$ | $\beta_1$ | $\beta_2$ | $\varphi$ | $\delta$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 50 | 50 | 5.8% | 5.1% | 6.6% | 5.1% | 5.8% | 4.7% | 5.4% | 10.6% |
| 75 | 50 | 5.9% | 5.1% | 9.6% | 4.5% | 5.2% | 6.0% | 6.2% | 11.5% |
| 100 | 50 | 4.7% | 4.9% | 9.4% | 5.3% | 5.8% | 4.8% | 6.0% | 10.9% |
| 50 | 75 | 3.4% | 5.0% | 8.0% | 4.7% | 5.4% | 5.3% | 4.7% | 8.5% |
| 75 | 75 | 5.0% | 4.8% | 7.1% | 6.9% | 4.2% | 6.5% | 6.5% | 10.3% |
| 100 | 75 | 5.1% | 5.2% | 8.5% | 5.0% | 5.8% | 5.2% | 6.0% | 8.8% |
| 50 | 100 | 4.4% | 4.7% | 6.0% | 6.1% | 5.3% | 5.8% | 4.4% | 7.9% |
| 75 | 100 | 5.7% | 4.5% | 6.7% | 5.0% | 4.1% | 5.3% | 5.0% | 8.5% |
| 100 | 100 | 5.2% | 5.5% | 6.6% | 6.0% | 5.1% | 4.8% | 5.6% | 9.9% |

Table 8: The empirical sizes of $t$-test under nominal 5% significance level after bias correction

| $N$ | $T$ | $\rho$ | $\varrho$ | $\phi$ | $\beta_1$ | $\beta_2$ | $\varphi$ | $\delta$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 50 | 50 | 5.8% | 4.6% | 4.2% | 5.0% | 5.9% | 5.1% | 5.3% | 7.9% |
| 75 | 50 | 6.1% | 5.1% | 5.8% | 4.3% | 5.1% | 6.2% | 6.1% | 7.1% |
| 100 | 50 | 4.3% | 5.0% | 5.0% | 5.2% | 5.8% | 4.3% | 6.0% | 5.6% |
| 50 | 75 | 4.6% | 4.7% | 6.4% | 4.5% | 5.2% | 4.6% | 4.7% | 6.1% |
| 75 | 75 | 5.3% | 4.9% | 5.5% | 6.6% | 4.4% | 5.9% | 6.5% | 6.3% |
| 100 | 75 | 5.1% | 5.1% | 5.3% | 5.0% | 5.8% | 5.0% | 6.0% | 5.4% |
| 50 | 100 | 4.3% | 4.5% | 5.1% | 6.1% | 5.4% | 5.5% | 4.4% | 6.9% |
| 75 | 100 | 5.6% | 4.5% | 5.5% | 5.1% | 4.2% | 5.0% | 5.0% | 5.9% |
| 100 | 100 | 5.1% | 5.2% | 4.0% | 5.8% | 5.0% | 4.9% | 5.6% | 6.0% |

Table 9: The empirical sizes of three super statistics

| $N$ | $T$ | Nominal 1% | | | Nominal 5% | | | Nominal 10% | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | supW | supLM | supLR | supW | supLM | supLR | supW | supLM | supLR |
| 100 | 100 | 0.7% | 0.7% | 0.7% | 3.4% | 3.3% | 3.4% | 7.8% | 7.7% | 7.9% |
| 125 | 100 | 0.7% | 0.7% | 0.7% | 3.4% | 3.4% | 3.4% | 7.8% | 7.8% | 7.8% |
| 150 | 100 | 0.9% | 0.9% | 0.9% | 4.1% | 4.1% | 4.1% | 7.9% | 7.9% | 7.9% |
| 100 | 150 | 0.8% | 0.8% | 0.8% | 3.7% | 3.7% | 3.7% | 7.7% | 7.6% | 7.7% |
| 125 | 150 | 0.9% | 0.9% | 0.9% | 3.9% | 3.9% | 3.9% | 8.2% | 8.2% | 8.2% |
| 150 | 150 | 0.8% | 0.8% | 0.8% | 3.9% | 3.9% | 4.0% | 8.1% | 8.1% | 8.1% |
| 100 | 200 | 0.7% | 0.7% | 0.7% | 3.6% | 3.6% | 3.6% | 7.0% | 7.0% | 7.0% |
| 125 | 200 | 0.7% | 0.7% | 0.7% | 4.0% | 4.0% | 4.0% | 8.5% | 8.3% | 8.3% |
| 150 | 200 | 0.8% | 0.8% | 0.8% | 4.2% | 4.2% | 4.2% | 8.5% | 8.5% | 8.4% |

# 9 Conclusion

Spatial models are widely used in empirical studies. So far spatial models are developed under the assumption of parameters stabilities, which greatly limits the potential applicabilities of models. This paper proposes and studies spatial panel data models with structural change. Quasi maximum likelihood method are considered to estimate the model. We build up a relatively complete asymptotic theory, including the consistencies, convergence rates and limiting distributions of the regression coefficient, the timing of structural change and the variance of regression errors. Static and Dynamic models are both considered. Large-$T$ and fixed-$T$ setups are both considered. We also investigate the hypothesis testing issue on the presence of structural change. The super statistics are proposed and the associated asymptotic properties are studied.

As the first step to study spatial models with structural change, we only consider one structural change in our model. This assumption is restrictive and implausible in some real data applications. Extension to the allowance of multiple structural changes is of both theoretical and practical interests. We will investigate this issue in the future work.

## References


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

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 821-856.

Andrews, D. W. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica*, 395-397.

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

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4), 551-563.

Bai, J., and Carrion-I-Silvestre, J. L. (2009). Structural changes, common stochastic trends, and unit roots in panel data. *The Review of Economic Studies*, 76(2), 471-501.

Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998). Testing for and dating common breaks in multivariate time series. *The Review of Economic Studies*, 65(3), 395-432.

Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47-78.

Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons.

Bode, E., Nunnenkamp, P., and Waldkirch, A. (2012). Spatial effects of foreign direct investment in US states. *Canadian Journal of Economics*, 45(1), 16-40.

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

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.

Chirinko, R. S., and Wilson, D. J. (2017). Tax competition among US states: Racing to the bottom or riding on a seesaw?. *Journal of Public Economics*, 155, 147-163.

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

Cressie, N. (1993): Statistics for Spatial Data. New York: John Wiley & Sons.

Dhaene, G., and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3), 991-1030.

Estrella, A. (2003). Critical values and p values of bessel process distributions: computation and application to structural break tests. *Econometric Theory*, 19(6), 1128-1143.

Hall, P., and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press.

Holly, S., Pesaran, M. H., and Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, 69(1), 2-23.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 633-643.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97-130.

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.

Kou, S., Peng, X., and Zhong, H. (2017). Asset pricing with spatial interaction. Forthcoming in *Management Science*.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 1899-1925.

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

LeSage, J. P., and Pace, R. K. (2009). *Introduction to Spatial Econometrics (Statistics, textbooks and monographs)*. CRC Press.

Li, K. (2017). Fixed-effects dynamic spatial panel data models and impulse response analysis. *Journal of Econometrics*, 198(1), 102-121.

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.

Lyytikäinen, T. (2012). Tax competition among local governments: Evidence from a property tax reform in Finland. *Journal of Public Economics*, 96(7-8), 584-595.

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

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

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6) 1361-1401.

Picard, D. (1985). Testing and estimating change-points in time series. *Advances in applied probability*, 17(4), 841-867.

Sengupta, A. (2017). Testing for a structural break in a spatial panel model. *Econometrics*, 5(1), 12.

Qu, Z., and Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2), 459-502.

Van Der Vaart A. W., and Wellner J. A. (1996). *Weak Convergence and Empirical Processes with Application to Statistics*. Springer.

Yao, Y. C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, 1321-1328.

Yu, J., de Jong, R., and Lee, L. F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118-134.

Zivot, E., and Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, 10(1), 251-270.