Inference on Dynamic Spatial Autoregressive Models with Change Point Detection

Zetai Cen[1], Yudong Chen[2], Clifford Lam[3]

[1]School of Mathematics, University of Bristol
[2]Department of Statistics, University of Warwick
[3]Department of Statistics, London School of Economics

September 23, 2025

Abstract

We analyze a varying-coefficient spatial autoregressive model with spatial fixed effects. One salient feature of the model is the incorporation of multiple spatial weight matrices through their linear combinations with varying coefficients, which help solve the problem of choosing the most "correct" one for applied econometricians who often face the availability of multiple expert spatial weight matrices. We estimate and make inferences on the model coefficients and coefficients in basis expansions of the varying coefficients through penalized estimations, establishing the oracle properties of the estimators and the consistency of the overall estimated spatial weight matrix, which can be time-dependent. We further consider two applications of our model in change point detections in spatial autoregressive models, providing theoretical justifications in consistent change point locations estimation and practical implementations. Simulation experiments demonstrate the performance of our proposed methodology, and real data analyses are also carried out.

Keywords: Adaptive LASSO, Instrument, Spillover effect, Structural change, Time-varying spatial weight matrix

1. INTRODUCTION

The study of spatial dependence in regional science gives rise to the techniques in spatial econometrics that we commonly use nowadays. Restricting to cross-sectional data only, a very general form of a model describing spatial dependence can be $\mathbf{y} = f(\mathbf{y}) + \boldsymbol{\epsilon}$ (Anselin, 1988), where $\mathbf{y}$ denotes a vector of d observed units, and $\boldsymbol{\epsilon}$ denotes an error term. A prominent and widely used candidate model is the spatial autoregressive model (see for example LeSage and Pace (2009)), which assumes a known *spatial weight matrix* $\mathbf{W}$ with zero diagonal and $f(\mathbf{y})$ of the form $f(\mathbf{y}) = \rho \mathbf{W} \mathbf{y}$ (or $f(\mathbf{y}) = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta}$ for a model with matrix of covariates $\mathbf{X}$), where ρ is called the spatial correlation coefficient.

Users of these models need to specify the $d \times d$ spatial weight matrix $\mathbf{W}$, which can be a contiguity matrix of 0 and 1, a matrix of inverse distances between two cities/regions, relative amount of import export etc. An obvious shortcoming for practitioners is to specify an "accurate" spatial weight matrix for use, often in the face of too many potential choices. This leads to a series of attempts to estimate the spatial weight matrix itself from data. For instance, see Pinkse et al. (2002) and Sun (2016) for models dealing with cross-sectional data only, both allowing for nonlinear spatial weight matrix estimation. Beenstock and Felsenstein (2012), Bhattacharjee and Jensen-Butler (2013), Lam and Souza (2020) and Higgins and Martellosio (2023) use spatial panel data for spatial weight matrix estimation, with Lam and Souza (2020) and Higgins and Martellosio (2023) allowing for multiple specified spatial weight matrices by a linear combination of them with constant coefficients.

Recent advances in spatial econometrics allow researchers to specify more complex models with an observed panel $\{\mathbf{y}_t\}$. Zhang and Shen (2015) consider partially linear covariate effects and constant spatial interactions using a sieve method to estimate a nonlinear function, while Sun and Malikov (2018) consider varying coefficients in both the spatial correlation coefficient (with underlying variables differing over observed units) and the covariate effects, assuming smooth nonlinear functions for kernel estimations. Catania and Billé (2017) also consider a dynamic spatial model with time-varying spatial correlation coefficients. Liang et al. (2022) use kernel estimation on a model with constant spatial interactions but deterministic time-varying coefficient functions for the covariates, while Chang et al. (2025) generalize the model to include an unknown random time trend and deterministic time-varying spatial correlation coefficient, still using kernel estimation. Hong et al. (2024) investigate a model similar to Sun and Malikov (2018), but added dynamic terms involving $\mathbf{y}_{t-1}$.

However, all the above allow for one specified spatial weight matrix only. As mentioned before, practitioners often face with too many potential choices for a spatial weight matrix. Combining the flexibility of allowing for multiple specified spatial weight matrices as input in Lam and Souza (2020) and varying effects in spatial interactions over observed variables or time directly, we propose a model akin to that in Lam and Souza (2020), but with varying coefficients in the linear combination of spatial weight matrices. The varying coefficients can be varying over some observed variables (stochastic) or time directly (non-stochastic).

Meanwhile, change point detection is an important topic that is still actively researched in time series analysis, especially in high-dimensional time series. Spatio-temporal models certainly receive their fair share of attention in this regard. For example, recent works have studied change point detection in spatio-temporal point processes (Altieri et al., 2015), in generalized additive models for spatio-temporal processes (Hollaway and Killick, 2024) and in high-dimensional spatio-temporal processes (Moradi et al., 2023). However, in the context of spatio-temporal panel models commonly used in spatial econometrics, such as spatial autoregressive panel models akin to (2.1) (but with $p=1$ and $l_1=0$ there), there appears to be very few literature on change point detection or related problems, as far as we are aware.

In view of the above, our contributions in this paper are three-folds. Firstly, using basis representations, we allow for the varying coefficients to be either stochastic or directly time-varying, without the need for any smoothness conditions. Hence, the final estimated spatial weight matrix can be either stochastic or deterministic, e.g., directly time dependent. Secondly, our adaptive LASSO estimators are proved to have the oracle properties, so that ill-specified spatial weight matrices which are irrelevant in the end will be dropped with probability going to 1 as the dimension d and the sample size T go to infinity. Meanwhile, the effects of relevant spatial weight matrices can be seen to be truly varying or not, again with probability going to 1 as $d,T \to \infty$. This greatly facilitates the interpretability of the spillover effects over time. Last but not least, our framework includes special cases such as spatial autoregressive panel data models with structural changes (Li, 2018) or threshold variables (Deng, 2018; Li and Lin, 2024). Section 5 explores the applications to multiple change points detection in both spatial autoregressive models with structural changes or threshold variables, suggesting an applicable algorithm for consistent change points detection in both cases.

The rest of the paper is organized as follows. Section 2 introduces our spatial autoregressive model and presents a procedure using adaptive LASSO to estimate the spatial fixed effect, spatial autoregressive parameters in a basis expansion, and the regression coefficients. Section 3 includes the required assumptions and the theoretical guarantees on the parameter estimators. Section 4 covers the algorithm for practical implementations including model selection and covariance matrix estimation for our estimators. Section 5 focuses on applying our framework for change point detection for a spatial autoregressive model with threshold variables or structural changes. Finally, numerical results are presented in Section 6, with a case study of stock returns on the NYSE. Section 7 concludes. Technical proofs, additional lemmas, extra simulations and real data analysis are provided in the online supplementary material.

Throughout this paper and unless otherwise specified, we use a lower-case letter a, a bold lower-case letter $\mathbf{a}$, and a bold capital letter $\mathbf{A}$ to denote a scalar, a vector, and a matrix, respectively. We also use $a_i, A_{ij}, \mathbf{A}_{i\cdot}, \mathbf{A}_{\cdot j}$ to denote, respectively, the i-th element of $\mathbf{a}$, the (i,j)-th element of $\mathbf{A}$, the i-th row vector (as a column vector) of $\mathbf{A}$, and the j-th column vector of $\mathbf{A}$. Hereafter, given a positive integer m, define $[m] := \{1,2,...,m\}$, and we use $\otimes$ to represent the Kronecker product.

For a given set, we denote by $|\cdot|$ its cardinality, and $\mathbb{1}\{\cdot\}$ the indicator function. For a vector $\mathbf{v}$, we denote its L_2-norm by $\|\mathbf{v}\|$. For a given matrix $\mathbf{A}$, we use $\|\mathbf{A}\|$ to denote its spectral norm, $\|\mathbf{A}\|_F$ its Frobenius norm, $\|\mathbf{A}\|_{\max} := \max_{i,j}|A_{ij}|$ its max norm, $\|\mathbf{A}\|_1 := \max_j \sum_i |A_{ij}|$ its L_1-norm, and $\|\mathbf{A}\|_\infty := \max_i \sum_j |A_{ij}|$ its L_∞-norm. For $q > 0$, we define the L_q-norm of a given real-valued random variable x as $\|x\|_q := (\mathbb{E}|x|^q)^{1/q}$. Without loss of generality, we always assume that the eigenvalues of a matrix are arranged in descending order, with their eigenvectors ordered correspondingly.

2. MODEL AND ESTIMATION

2.1. Spatial autoregressive model with dynamic spatial weights

We propose a framework of spatial autoregressive models with fixed effects such that for each $t \in [T]$,

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t} \Big) \mathbf{W}_j \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{2.1}$$

where $\mathbf{y}_t \in \mathbb{R}^d$ is the observed vector at time t, and $\boldsymbol{\mu}^*$ is a constant vector of spatial fixed effects. Each $\mathbf{W}_j \in \mathbb{R}^{d \times d}$ is a pre-specified spatial weight matrix provided by researchers to feature the spillover effects of cross-sectional units from their neighbors. Each $\mathbf{W}_j$ has zero entries on its main diagonal with no restrictions on the signs of off-diagonal entries, and can be asymmetric. Each term $(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t})$ is essentially a spatial correlation coefficient for the spatial weight matrix $\mathbf{W}_j$ (see also Lam and Souza (2020)), which can be time-varying by being presented as either a basis expansion using some non-random pre-specified set of basis $\{z_{j,k,t}\}$, or an affine combination of random variables $\{z_{j,k,t}\}$. In either case, we call the $\{z_{j,k,t}\}$'s the dynamic variables hereafter. For $j \in [p], k \in [l_j]$, the parameters $\phi_{j,0}^*, \phi_{j,k}^*$ are unknown and need to be estimated. The covariate matrix $\mathbf{X}_t$ has size $d \times r$, with $\boldsymbol{\beta}^*$ the corresponding unknown regression coefficients of length r. Finally, $\boldsymbol{\epsilon}_t$ is the idiosyncratic noise with zero mean.

Without loss of generality, we assume $\mathbf{X}_t$ to have zero mean. Otherwise, we read $\boldsymbol{\mu}^* + \mathbf{X}_t \boldsymbol{\beta}^* = (\boldsymbol{\mu}^* + \mathbb{E}[\mathbf{X}_t]\boldsymbol{\beta}^*) + (\mathbf{X}_t - \mathbb{E}[\mathbf{X}_t])\boldsymbol{\beta}^*$, which leads to estimating $(\boldsymbol{\mu}^* + \mathbb{E}[\mathbf{X}_t]\boldsymbol{\beta}^*)$ as the spatial fixed effects instead. We can rewrite (2.1) as a traditional spatial autoregressive model

$\mathbf{y}_t = \boldsymbol{\mu}^* + \mathbf{W}_t^* \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t$ by defining the true spatial weight matrix at time t as

$$\mathbf{W}_t^* := \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t} \Big) \mathbf{W}_j, \text{ with } -1 < \rho_t^* := \sum_{j=1}^{p} \Big(\phi_{j,0}^* + \sum_{k=1}^{l_j} \phi_{j,k}^* z_{j,k,t} \Big) < 1. \tag{2.2}$$

The restrictions on ρ_t^* for all $t \in [T]$ ensure the model is stationary. See Assumptions (M2) and (M2') for technical details. We define $L := p + \sum_{j=1}^{p} l_j$ and reformulate (2.1) as

$$\begin{aligned}
&\mathbf{y}_t = \boldsymbol{\mu}^* + (\boldsymbol{\Lambda}_t \boldsymbol{\Phi}^*) \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \text{ where} \qquad (2.3)\\
&\boldsymbol{\Lambda}_t := (\boldsymbol{\Lambda}_{1,t}, \boldsymbol{\Lambda}_{2,t}, \ldots, \boldsymbol{\Lambda}_{p,t}) \in \mathbb{R}^{d \times dL}, \text{ with } \boldsymbol{\Lambda}_{j,t} := (\mathbf{W}_j, z_{j,1,t}\mathbf{W}_j, \ldots, z_{j,l_j,t}\mathbf{W}_j) \in \mathbb{R}^{d \times (d + dl_j)},\\
&\boldsymbol{\Phi}^* := \big(\boldsymbol{\Phi}_1^{*\top}, \boldsymbol{\Phi}_2^{*\top}, \ldots, \boldsymbol{\Phi}_p^{*\top} \big)^\top \in \mathbb{R}^{dL \times d}, \text{ with } \boldsymbol{\Phi}_j^* := \big(\phi_{j,0}^* \mathbf{I}_d, \phi_{j,1}^* \mathbf{I}_d, \ldots, \phi_{j,l_j}^* \mathbf{I}_d \big)^\top.
\end{aligned}$$

Due to the endogeneity in $\mathbf{y}_t$ and potentially $\mathbf{X}_t$, we assume that a set of valid instrumental variables $\mathbf{U}_t$ are available for $t \in [T]$. More specifically, each $\mathbf{U}_t$ is independent of $\boldsymbol{\epsilon}_t$ but is correlated with $\mathbf{y}_t$ and the endogenous $\mathbf{X}_t$. Note that if $\mathbf{X}_t$ is exogenous, we may simply have $\mathbf{U}_t = \mathbf{X}_t$. Following Kelejian and Prucha (1998), we can construct instruments $\mathbf{B}_t$ as a $d \times v$ matrix with $v \geq r$ by interacting each given spatial weight matrix with $\mathbf{U}_t$ such that $\mathbf{B}_t$ is composed of at least a subset of linearly independent columns in[1]

$$\Big\{ \mathbf{U}_t, \mathbf{W}_1 \mathbf{U}_t, \mathbf{W}_1^2 \mathbf{U}_t, \ldots, \mathbf{W}_p \mathbf{U}_t, \mathbf{W}_p^2 \mathbf{U}_t, \ldots \Big\}. \tag{2.4}$$

To enhance the interpretability of the true spatial weight matrix $\mathbf{W}_t^*$, we assume that the dynamic feature of model (2.1) is driven only by a few $\{z_{j,k,t}\}$. In other words, the vector of coefficients $\boldsymbol{\phi}^* := (\boldsymbol{\phi}_1^{*\top}, \boldsymbol{\phi}_2^{*\top}, \ldots, \boldsymbol{\phi}_p^{*\top})^\top$ (with $\boldsymbol{\phi}_j^* := (\phi_{j,0}^*, \phi_{j,1}^*, \ldots, \phi_{j,l_j}^*)^\top$) is assumed to be sparse. Using LASSO (Tibshirani, 1996), an L_1 penalty $\lambda \|\boldsymbol{\phi}\|_1$ can be included in the regression problem to shrink the estimators towards zero and some of them to exactly zero, where $\lambda > 0$ is a tuning

[1] Ideally, we should have $\mathbf{B}_t$ of the form $[\sum_{j=1}^{p} (\phi_{j,0} + \sum_{k=1}^{l_j} \phi_{j,k} z_{j,k,t}) \mathbf{W}_j]^m \mathbf{U}_t$ for $m = 0,1,2,\ldots$. However, similar to Lam and Souza (2020), each $\phi_{j,0}$ and $\phi_{j,k}$ is unknown and hence we exclude any cross-terms with more than one $\mathbf{W}_j$.

parameter. However, this form of regularization penalizes uniformly on each entry, which may lead to over- or under-penalization. The former induces bias while the latter fails sign-consistency, i.e., zeros are estimated exactly as zeros and non-zeros are estimated with the correct signs.

To ensure zero-consistency in variable selection, a necessary "irrepresentable condition" is often imposed (Zhao and Yu, 2006). Subsequently, Zou (2006) reweighs the regularization to be $\lambda\mathbf{u}^\top|\boldsymbol{\phi}|$ where $|\cdot|$ is applied entrywise and $\mathbf{u}$ contains the inverse of some initial estimators of $\boldsymbol{\phi}^*$. Now, the sign-consistency can be ensured even without the irrepresentable condition if the estimators in $\mathbf{u}$ are $\sqrt{T}$-consistent. Such a framework adaptively penalizes the magnitude of the estimators and is hence called "adaptive LASSO". To this end, we start by profiling out $\boldsymbol{\beta}$. To make use of the instruments, define $\bar{\mathbf{B}}:=T^{-1}\sum_{t=1}^T\mathbf{B}_t$. If $\boldsymbol{\phi}$ (and hence $\boldsymbol{\Phi}$) is given, by multiplying $(\mathbf{B}_t-\bar{\mathbf{B}})^\top$ and summing over all $t\in[T]$ on both sides of (2.3), we then have

$$\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top(\mathbf{I}_d-\boldsymbol{\Lambda}_t\boldsymbol{\Phi})\mathbf{y}_t=\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top\mathbf{X}_t\boldsymbol{\beta}+\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top\boldsymbol{\epsilon}_t,$$

where the true values $\boldsymbol{\Phi}^*$ and $\boldsymbol{\beta}^*$ are replaced by $\boldsymbol{\Phi}$ and $\boldsymbol{\beta}$ respectively. Note the spatial fixed effect $\boldsymbol{\mu}^*$ vanishes as $\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top\boldsymbol{\mu}^*=\mathbf{0}$. Thus, the least squares estimator of $\boldsymbol{\beta}^*$ given $\boldsymbol{\phi}$ can be denoted as

$$\boldsymbol{\beta}(\boldsymbol{\phi})=\Big\{\sum_{s=1}^T\mathbf{X}_s^\top(\mathbf{B}_s-\bar{\mathbf{B}})\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top\mathbf{X}_t\Big\}^{-1}\sum_{s=1}^T\mathbf{X}_s^\top(\mathbf{B}_s-\bar{\mathbf{B}})\sum_{t=1}^T(\mathbf{B}_t-\bar{\mathbf{B}})^\top(\mathbf{I}_d-\boldsymbol{\Lambda}_t\boldsymbol{\Phi})\mathbf{y}_t. \tag{2.5}$$

To facilitate formulating the adaptive LASSO problem by accommodating the instrumental variables, we write $\boldsymbol{\gamma}:=v^{-1}\mathbf{1}_v$ and the i-th row of $\mathbf{B}_t$ and $\bar{\mathbf{B}}$ by $\mathbf{B}_{t,i\cdot}$ and $\bar{\mathbf{B}}_{i\cdot}$, respectively. For $i\in[d]$, $t\in[T]$, define the outcome and covariates filtered through instrumental variables as $\mathbf{y}_{B,i,t}:=(\mathbf{B}_{t,i\cdot}-\bar{\mathbf{B}}_{i\cdot})^\top\boldsymbol{\gamma}\mathbf{y}_t$ and $\mathbf{X}_{B,i}:=\sum_{t=1}^T(\mathbf{B}_{t,i\cdot}-\bar{\mathbf{B}}_{i\cdot})^\top\boldsymbol{\gamma}\mathbf{X}_t$ respectively. The least squares problem is then

$$\widetilde{\boldsymbol{\phi}}=\operatorname*{argmin}_{\boldsymbol{\phi}}\frac{1}{2T}\sum_{i=1}^d\Big\|\sum_{t=1}^T(\mathbf{I}_d-\boldsymbol{\Lambda}_t\boldsymbol{\Phi})\mathbf{y}_{B,i,t}-\mathbf{X}_{B,i}\boldsymbol{\beta}(\boldsymbol{\phi})\Big\|^2. \tag{2.6}$$

With the solution as an initial estimator, the adaptive LASSO problem now solves for

$$\widehat{\boldsymbol{\phi}}=\underset{\boldsymbol{\phi}}{\operatorname{argmin}}\ \frac{1}{2T}\sum_{i=1}^{d}\Big\|\sum_{t=1}^{T}(\mathbf{I}_d-\boldsymbol{\Lambda}_t\boldsymbol{\Phi})\mathbf{y}_{B,i,t}-\mathbf{X}_{B,i}\boldsymbol{\beta}(\boldsymbol{\phi})\Big\|^2+\lambda\mathbf{u}^\top|\boldsymbol{\phi}|, \tag{2.7}$$
$$\text{subj. to } \|\boldsymbol{\Lambda}_t\boldsymbol{\Phi}\|_\infty<1, \text{ with } |\mathbf{z}_t^\top\boldsymbol{\phi}|<1 \text{ for any } t\in[T],$$

where $\mathbf{z}_t := (\mathbf{z}_{1,t}^\top, \mathbf{z}_{2,t}^\top, \dots, \mathbf{z}_{p,t}^\top)^\top$, $\mathbf{z}_{j,t} := (1, z_{j,1,t}, \dots, z_{j,l_j,t})^\top$, $\mathbf{u} := (|\widetilde{\phi}_{1,0}|^{-1}, \dots, |\widetilde{\phi}_{p,l_p}|^{-1})^\top$, $|\boldsymbol{\phi}| := (|\phi_{1,0}|,\dots,|\phi_{p,l_p}|)^\top$ and λ is a tuning parameter. With $\widehat{\boldsymbol{\phi}}$ (and hence $\widehat{\boldsymbol{\Phi}}$), the adaptive LASSO estimators for $\boldsymbol{\beta}^*$ can be obtained by $\widehat{\boldsymbol{\beta}}:=\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$ and the fixed effect estimator by

$$\widehat{\boldsymbol{\mu}}:=\frac{1}{T}\sum_{t=1}^{T}\Big\{(\mathbf{I}_d-\boldsymbol{\Lambda}_t\widehat{\boldsymbol{\Phi}})\mathbf{y}_t-\mathbf{X}_t\widehat{\boldsymbol{\beta}}\Big\}. \tag{2.8}$$

2.2. Full matrix notations

To facilitate both the theoretical results and practical implementation, we present the least squares and the adaptive LASSO problems in matrix notations here. First, we introduce

$$\mathbf{B}:=T^{-1/2}d^{-a/2}(\mathbf{B}_\gamma-\bar{\mathbf{B}}_\gamma):=T^{-1/2}d^{-a/2}\mathbf{I}_d\otimes\big\{(\mathbf{I}_T\otimes\boldsymbol{\gamma}^\top)(\mathbf{B}_1-\bar{\mathbf{B}},\dots,\mathbf{B}_T-\bar{\mathbf{B}})^\top\big\}, \tag{2.9}$$

where a is a constant that gauges the correlation between $\mathbf{B}_t$ and $\mathbf{X}_t$ so that a larger a generally means that $\mathbf{B}_t$ is correlated with more covariates in $\mathbf{X}_t$. See Assumption (R4) for more technical details. As in Lam and Souza (2020), in practice we can set $a=1$ to compute $\mathbf{B}$, which would not change the optimal values of any tuning parameters or estimators in the adaptive LASSO problem below.

For ease of notation, denote $z_{j,0,t}=1$ for all $j\in[p], t\in[T]$. Now rewrite (2.1) as

$$\mathbf{y}=\boldsymbol{\mu}^*\otimes\mathbf{1}_T+\mathbf{V}\boldsymbol{\phi}^*+\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}\big(\mathbf{I}_d\big)+\boldsymbol{\epsilon}, \text{ where} \tag{2.10}$$
$$\mathbf{y}:=\mathrm{vec}\big((\mathbf{y}_1,\dots,\mathbf{y}_T)^\top\big),\ \boldsymbol{\epsilon}:=\mathrm{vec}\big((\boldsymbol{\epsilon}_1,\dots,\boldsymbol{\epsilon}_T)^\top\big),\ \mathbf{V}:=(\mathbf{V}_1,\dots,\mathbf{V}_p),$$
$$\mathbf{V}_j:=\Big[\boldsymbol{\Gamma}_{j,0}\mathrm{vec}\big(\mathbf{W}_j^\top\big),\boldsymbol{\Gamma}_{j,1}\mathrm{vec}\big(\mathbf{W}_j^\top\big),\dots,\boldsymbol{\Gamma}_{j,l_j}\mathrm{vec}\big(\mathbf{W}_j^\top\big)\Big],$$

$$\mathbf{\Gamma}_{j,k}:=\mathbf{I}_d\otimes(z_{j,k,1}\mathbf{y}_1,\ldots,z_{j,k,T}\mathbf{y}_T)^\top,\ \mathbf{X}_{\boldsymbol{\beta}^*}:=\mathbf{I}_d\otimes\big\{(\mathbf{I}_T\otimes(\boldsymbol{\beta}^*)^\top)(\mathbf{X}_1,\ldots,\mathbf{X}_T)^\top\big\}.$$

In this form, the model now has design matrix $\mathbf{V}$ in a classical linear regression setting, except that the endogenous variables $\mathbf{y}_t$ are present in $\mathbf{V}$. We thus obtain the augmented model by left-multiplying both sides of (2.10) by $\mathbf{B}^\top$:

$$\mathbf{B}^\top\mathbf{y}=\mathbf{B}^\top\mathbf{V}\boldsymbol{\phi}^*+\mathbf{B}^\top\mathbf{X}_{\boldsymbol{\beta}^*}\mathrm{vec}\big(\mathbf{I}_d\big)+\mathbf{B}^\top\boldsymbol{\epsilon}, \tag{2.11}$$

where the augmented spatial fixed effect vanishes since $\mathbf{B}^\top(\boldsymbol{\mu}^*\otimes\mathbf{1}_T)=T^{-1/2}d^{-a/2}\boldsymbol{\mu}^*\otimes\big\{(\mathbf{B}_1-\bar{\mathbf{B}},\ldots,\mathbf{B}_T-\bar{\mathbf{B}})(\mathbf{I}_T\otimes\boldsymbol{\gamma})\mathbf{1}_T\big\}=\mathbf{0}$. For any matrix $\mathbf{C}$, denote $\mathbf{C}^\otimes:=\mathbf{I}_T\otimes\mathbf{C}$. We can also rewrite (2.1) as

$$\mathbf{y}^\nu=\mathbf{1}_T\otimes\boldsymbol{\mu}^*+\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi^*_{j,k}\mathbf{W}_j^\otimes\mathbf{y}^\nu_{j,k}+\mathbf{X}\boldsymbol{\beta}^*+\boldsymbol{\epsilon}^\nu,\ \text{where} \tag{2.12}$$
$$\mathbf{y}^\nu:=(\mathbf{y}_1^\top,\ldots,\mathbf{y}_T^\top)^\top,\ \mathbf{y}^\nu_{j,k}:=(z_{j,k,1}\mathbf{y}_1^\top,\ldots,z_{j,k,T}\mathbf{y}_T^\top)^\top,\ \boldsymbol{\epsilon}^\nu:=(\boldsymbol{\epsilon}_1^\top,\ldots,\boldsymbol{\epsilon}_T^\top)^\top,\ \mathbf{X}:=(\mathbf{X}_1^\top,\ldots,\mathbf{X}_T^\top)^\top.$$

Thus with $\mathbf{B}^\nu:=(\mathbf{B}_1^\top-\bar{\mathbf{B}}^\top,\ldots,\mathbf{B}_T^\top-\bar{\mathbf{B}}^\top)^\top$, we may write (2.5) in matrix form as

$$\boldsymbol{\beta}(\boldsymbol{\phi})=\Big(\mathbf{X}^\top\mathbf{B}^\nu(\mathbf{B}^\nu)^\top\mathbf{X}\Big)^{-1}\mathbf{X}^\top\mathbf{B}^\nu(\mathbf{B}^\nu)^\top\Big(\mathbf{y}^\nu-\sum_{j=1}^{p}\sum_{k=0}^{l_j}\phi_{j,k}\mathbf{W}_j^\otimes\mathbf{y}^\nu_{j,k}\Big). \tag{2.13}$$

To rewrite the least squares and adaptive LASSO problem for $\widetilde{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\phi}}$, define

$$\mathbf{Y}_W:=(\mathbf{W}_1^\otimes\mathbf{y}^\nu_{1,0},\ldots,\mathbf{W}_1^\otimes\mathbf{y}^\nu_{1,l_1},\ldots,\mathbf{W}_p^\otimes\mathbf{y}^\nu_{p,0},\ldots,\mathbf{W}_p^\otimes\mathbf{y}^\nu_{p,l_p}),$$
$$\mathbf{\Xi}:=T^{-1/2}d^{-a/2}\Big(\sum_{t=1}^{T}\mathbf{X}_t\otimes(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}\Big)\big[\mathbf{X}^\top\mathbf{B}^\nu(\mathbf{B}^\nu)^\top\mathbf{X}\big]^{-1}\mathbf{X}^\top\mathbf{B}^\nu(\mathbf{B}^\nu)^\top.$$

Then with all simplification steps relegated to the online supplement, (2.6) can be rewritten as

$$\begin{aligned}\widetilde{\boldsymbol{\phi}} &= \operatorname*{argmin}_{\boldsymbol{\phi}} \frac{1}{2T}\left\|\mathbf{B}^\top\mathbf{y} - \Xi\mathbf{y}^\nu - (\mathbf{B}^\top\mathbf{V} - \Xi\mathbf{Y}_W)\boldsymbol{\phi}\right\|^2 \\ &= \left[(\mathbf{B}^\top\mathbf{V} - \Xi\mathbf{Y}_W)^\top(\mathbf{B}^\top\mathbf{V} - \Xi\mathbf{Y}_W)\right]^{-1}(\mathbf{B}^\top\mathbf{V} - \Xi\mathbf{Y}_W)^\top(\mathbf{B}^\top\mathbf{y} - \Xi\mathbf{y}^\nu).\end{aligned} \tag{2.14}$$

The adaptive LASSO problem in (2.7) can be written as

$$\widehat{\boldsymbol{\phi}} = \operatorname*{argmin}_{\boldsymbol{\phi}} \frac{1}{2T}\left\|\mathbf{B}^\top\mathbf{y} - \Xi\mathbf{y}^\nu - (\mathbf{B}^\top\mathbf{V} - \Xi\mathbf{Y}_W)\boldsymbol{\phi}\right\|^2 + \lambda\mathbf{u}^\top|\boldsymbol{\phi}|, \tag{2.15}$$

$$\text{subj. to } \|\boldsymbol{\Lambda}_t\boldsymbol{\Phi}\|_\infty < 1, \text{ with } |\mathbf{z}_t^\top\boldsymbol{\phi}| < 1 \text{ for any } t \in [T].$$

3. ASSUMPTIONS AND THEORETICAL RESULTS

We first present some notations involving the measure of serial dependence of all time series variables, which is gauged by the functional dependence measure introduced by Wu (2005). We state all the assumptions in this paper in Section 3.1. Denote $\{\mathbf{x}_t\} := \{\mathrm{vec}(\mathbf{X}_t)\}$ and $\{\mathbf{b}_t\} := \{\mathrm{vec}(\mathbf{B}_t)\}$ to be the vectorized processes for $\{\mathbf{X}_t\}$ and $\{\mathbf{B}_t\}$ with length dr and dv, respectively. For $t \in [T]$, assume that

$$\mathbf{x}_t = \left[f_i(\mathcal{F}_t)\right]_{i\in[dr]}, \ \mathbf{b}_t = \left[g_i(\mathcal{G}_t)\right]_{i\in[dv]}, \ \boldsymbol{\epsilon}_t = \left[h_i(\mathcal{H}_t)\right]_{i\in[d]}, \tag{3.1}$$

where $f_i(\cdot)$'s, $g_i(\cdot)$'s, $h_i(\cdot)$'s are measurable functions defined on the real line, and $\mathcal{F}_t = (\dots, \mathbf{e}_{x,t-1}, \mathbf{e}_{x,t})$, $\mathcal{G}_t = (\dots, \mathbf{e}_{b,t-1}, \mathbf{e}_{b,t})$, $\mathcal{H}_t = (\dots, \mathbf{e}_{\epsilon,t-1}, \mathbf{e}_{\epsilon,t})$ are defined by i.i.d. processes $\{\mathbf{e}_{x,t}\}$, $\{\mathbf{e}_{b,t}\}$ and $\{\mathbf{e}_{\epsilon,t}\}$ respectively, with $\{\mathbf{e}_{b,t}\}$ independent of $\{\mathbf{e}_{\epsilon,t}\}$ but correlated with $\{\mathbf{e}_{x,t}\}$. For $q > 0$, we define

$$\begin{aligned}\theta^x_{t,q,i} &:= \left\|x_{t,i} - x'_{t,i}\right\|_q = (\mathbb{E}|x_{t,i} - x'_{t,i}|^q)^{1/q}, \ i \in [dr], \\ \theta^b_{t,q,i} &:= \left\|b_{t,i} - b'_{t,i}\right\|_q = (\mathbb{E}|b_{t,i} - b'_{t,i}|^q)^{1/q}, \ i \in [dv], \\ \theta^\epsilon_{t,q,i} &:= \left\|\epsilon_{t,i} - \epsilon'_{t,i}\right\|_q = (\mathbb{E}|\epsilon_{t,i} - \epsilon'_{t,i}|^q)^{1/q}, \ i \in [d],\end{aligned} \tag{3.2}$$

where $x'_{t,i} = f_i(\mathcal{F}'_t)$, $\mathcal{F}'_t = (\ldots,\mathbf{e}_{x,-1},\mathbf{e}'_{x,0},\mathbf{e}_{x,1},\ldots,\mathbf{e}_{x,t})$, with $\mathbf{e}'_{x,0}$ independent of all other $\mathbf{e}_{x,j}$'s. Hence $x'_{t,i}$ is a coupled version of $x_{t,i}$ with $\mathbf{e}_{x,0}$ replaced by an i.i.d. copy $\mathbf{e}'_{x,0}$. We define $b'_{t,i}$ and $\epsilon'_{t,i}$ similarly.

3.1. Assumptions

We present here the set of assumptions for our model. In summary, (I1) helps to identify the model; assumptions prefixed "M" renders the model framework; those prefixed "R" are more technical.

(I1) (Identification) *The matrix $\mathbf{Q}^\top\mathbf{Q}$ has all its eigenvalues uniformly bounded away from 0, where*

$$\mathbf{Q} = [\mathbb{E}(\mathbf{B}^\top\mathbf{V}),\mathbb{E}(\mathbf{B}^\top\widetilde{\mathbf{X}})],\ \widetilde{\mathbf{X}} = (\mathbf{x}_{1,1\cdot},\ldots,\mathbf{x}_{T,1\cdot},\ldots,\mathbf{x}_{1,d\cdot},\ldots,\mathbf{x}_{T,d\cdot})^\top.$$

(M1) (Time series in $\mathbf{X}_t$, $\mathbf{B}_t$ and $\boldsymbol{\epsilon}_t$) *The processes $\{\mathbf{X}_t\}$, $\{\mathbf{B}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ are second-order stationary and satisfy (3.1), with $\{\mathbf{X}_t\}$ and $\{\boldsymbol{\epsilon}_t\}$ having mean zero. The tail condition $\mathbb{P}(|Z| > z) \leq D_1\exp(-D_2 z^\ell)$ is satisfied by the variables $B_{t,ij}$, $X_{t,ij}$, $\epsilon_{t,i}$ for the same constants D_1, D_2 and ℓ.*

With (3.2), define the tail sums

$$\Theta^x_{m,q} = \sum_{t=m}^{\infty}\max_{i\in[dr]}\theta^x_{t,q,i},\ \Theta^b_{m,q} = \sum_{t=m}^{\infty}\max_{i\in[dv]}\theta^b_{t,q,i},\ \Theta^\epsilon_{m,q} = \sum_{t=m}^{\infty}\max_{i\in[d]}\theta^\epsilon_{t,q,i}.$$

For some $w>2$, assume $\Theta^x_{m,2w},\Theta^b_{m,2w},\Theta^\epsilon_{m,2w} \leq Cm^{-\alpha}$ with constants $\alpha,C>0$ that can depend on w.

(M2) (True spatial weight matrix $\mathbf{W}^*_t$ with non-random $z_{j,k,t}$) *$\mathbf{W}^*_t$ defined in (2.2) uses a uniformly bounded non-stochastic basis $\{z_{j,k,t}\}$ for $j\in[p],k\in[l_j]$. There exists a constant $\eta>0$ such that for all $t\in[T]$, $\|\mathbf{W}^*_t\|_\infty < \eta < 1$ uniformly as $d\to\infty$. The elements in $\mathbf{W}^*_t$ can be negative, and $\mathbf{W}^*_t$ can be asymmetric. Furthermore, ρ^*_t defined in (2.2) satisfies $|\rho^*_t| < 1$.*

(M2') (True spatial weight matrix $\mathbf{W}^*_t$ with random $z_{j,k,t}$) *Same as Assumption (M2), except that $\{z_{j,k,t}\}$ is a zero mean stochastic process with support $[-1,1]$, such that $z_{j,k,t} = u_{j,k}(\mathcal{U}_t)$, similar to (3.1), satisfies $\mathbb{E}(z_{j,k,t}\mathbf{X}_t) = \mathbf{0}$, $\mathbb{E}(z_{j,k,t}\boldsymbol{\epsilon}_t) = \mathbf{0}$, and $\Theta^z_{m,2w} \leq Cm^{-\alpha}$ as in Assumption (M1). Furthermore:*

1. *there exists* $\eta>0$ *such that* $\sum_{j=1}^{p}\sum_{k=0}^{l_j}\|\phi^*_{j,k}\mathbf{W}_j\|_\infty<\eta<1$ *uniformly as* $d\to\infty$*;*
2. $\sum_{j=1}^{p}\sum_{k=0}^{l_j}|\phi^*_{j,k}|<1$.

(R1) *Denote the* $d^2L\times L$ *block diagonal matrix* $\mathbf{D}_W:=\mathrm{diag}\{\mathbf{I}_{1+l_1}\otimes\mathrm{vec}(\mathbf{W}_1^\top),\ldots,\mathbf{I}_{1+l_p}\otimes\mathrm{vec}(\mathbf{W}_p^\top)\}$*. Then, there exists a constant* $u>0$ *such that the* L*-th largest singular value of* $\mathbf{D}_W$ *satisfies* $\sigma_L^2(\mathbf{D}_W)\geq du>0$ *uniformly as* $d\to\infty$.

Moreover, $\max_j\{\|\mathbf{W}_j\|_1,\|\mathbf{W}_j\|_\infty\}\leq c<\infty$ *uniformly as* $d\to\infty$ *for some constant* $c>0$.

(R2) *Write* $\boldsymbol{\epsilon}_t=\boldsymbol{\Sigma}_\epsilon^{1/2}\boldsymbol{\epsilon}_t^*$ *with* $\boldsymbol{\Sigma}_\epsilon$ *being the covariance matrix of* $\boldsymbol{\epsilon}_t$*. Assume* $\left\|\boldsymbol{\Sigma}_\epsilon\right\|_{\max}\leq\sigma^2_{\max}<\infty$ *uniformly as* $d\to\infty$*. The same applies to the variance of the elements in* $\mathbf{B}_t$.

Assume also $\left\|\boldsymbol{\Sigma}_\epsilon^{1/2}\right\|_\infty\leq S_\epsilon<\infty$ *uniformly as* $d\to\infty$*, with* $\{\epsilon^*_{t,i}\}_{i\in[d]}$ *being a martingale difference with respect to the filtration generated by* $\sigma(\epsilon_{t,1},\ldots,\epsilon_{t,i})$*. Furthermore,* $\{\boldsymbol{\epsilon}_t^*\}_{t\in[T]}$ *satisfies the tail condition and the functional dependence in Assumption (M1).*

(R3) *All singular values of* $\mathbb{E}(\mathbf{X}_t^\top\mathbf{B}_t)$ *are uniformly larger than du for some constant* $u>0$*, while the maximum singular value is of order d. Individual entries in the matrix* $\mathbb{E}(\mathbf{b}_t\mathbf{x}_t^\top)$ *are uniformly bounded away from infinity, with* $\mathbf{x}_t$ *and* $\mathbf{b}_t$ *defined in (3.1).*

(R4) *With the same* $a\in[0,1]$ *introduced in (2.9), we define*

$$\mathbf{G}:=d^{-a}\mathbf{I}_d\otimes\{\mathbb{E}(\check{\mathbf{G}})\mathbb{E}(\check{\mathbf{G}})^\top\},\ \check{\mathbf{G}}:=(\check{\mathbf{G}}_{1,0},\ldots,\check{\mathbf{G}}_{1,l_1},\ldots,\check{\mathbf{G}}_{p,0},\ldots,\check{\mathbf{G}}_{p,l_p}),$$
$$\check{\mathbf{G}}_{j,k}:=\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\gamma}\boldsymbol{\beta}^{*\top}\mathbf{X}_t^\top\boldsymbol{\Pi}_t^{*\top},\ \boldsymbol{\Pi}_t^*:=(\mathbf{I}_d-\mathbf{W}_t^*)^{-1}.$$

We assume that $\mathbf{G}$ *has full rank and that there exists a constant* $u>0$ *such that* $\lambda_{\min}(\mathbf{G})\geq u>0$ *and* $\lambda_{\max}(\mathbf{G})<\infty$ *uniformly as* $d\to\infty$.

(R5) *For the same constant a as in Assumption (R4), we have for each d,*

$$\max_{i\in[d]}\sum_{j=1}^{d}\left\|\mathbb{E}(\mathbf{B}_{t,i\cdot}\mathbf{X}_{t,j\cdot}^\top)\right\|_{\max},\ \max_{j\in[d]}\sum_{i=1}^{d}\left\|\mathbb{E}(\mathbf{B}_{t,i\cdot}\mathbf{X}_{t,j\cdot}^\top)\right\|_{\max}=O(d^a).$$

At the same time, assume $\mathbb{E}(\mathbf{X}_t\otimes\mathbf{B}_t\boldsymbol{\gamma})$ *has all singular values of order* d^{1+a}.

(R6) *With* $\boldsymbol{\Pi}_t^*$ *in Assumption (R4), define*

$$\ddot{\mathbf{G}}:=(\ddot{\mathbf{G}}_{1,0},\ldots,\ddot{\mathbf{G}}_{1,l_1},\ldots,\ddot{\mathbf{G}}_{p,0},\ldots,\ddot{\mathbf{G}}_{p,l_p}),\ \ddot{\mathbf{G}}_{j,k}:=\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}(\mathbf{B}_t-\bar{\mathbf{B}})^\top\big(\mathbf{I}_d\otimes\boldsymbol{\beta}^{*\top}\mathbf{X}_t^\top\boldsymbol{\Pi}_t^{*\top}\big).$$

Assume that $\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})^\top$ *has full rank and that there exists a constant* $u>0$ *such that* $\lambda_v\big(\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})^\top\big)\geq u>0$ *and* $\lambda_{\max}\big(\mathbb{E}(\ddot{\mathbf{G}})\mathbb{E}(\ddot{\mathbf{G}})^\top\big)<\infty$ *uniformly as* $d\to\infty$.

(R7) *Define the predictive dependence measures*

$$P_0^b(B_{t,jk}):=\mathbb{E}(B_{t,jk}\,|\,\mathcal{G}_0)-\mathbb{E}(B_{t,jk}\,|\,\mathcal{G}_{-1}),\ P_0^\epsilon(\epsilon_{t,j}):=\mathbb{E}(\epsilon_{t,j}\,|\,\mathcal{H}_0)-\mathbb{E}(\epsilon_{t,j}\,|\,\mathcal{H}_{-1}),$$

with $\mathcal{G}_t$ *and* $\mathcal{H}_t$ *specified after (3.1). Assume*

$$\sum_{t\geq 0}\max_{j\in[d]}\max_{k\in[v]}\big\|P_0^b(B_{t,jk})\big\|_2<\infty,\ \sum_{t\geq 0}\max_{j\in[d]}\big\|P_0^\epsilon(\epsilon_{t,j})\big\|_2<\infty.$$

(R8) *For* $b\in[0,1]$, *the eigenvalues of* $\mathrm{Var}(d^{-b/2}\mathbf{B}_{t,\cdot k})$ *and* $\mathrm{Var}(\boldsymbol{\epsilon}_t)$ *are uniformly bounded away from both zero and infinity, and respectively dominate the singular values of* $d^{-b}\mathbb{E}\big\{[\mathbf{B}_{t,\cdot k}-\mathbb{E}(\mathbf{B}_{t,\cdot k})][\mathbf{B}_{t+\tau,\cdot k}-\mathbb{E}(\mathbf{B}_{t,\cdot k})]^\top\big\}$ *and* $\mathbb{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t+\tau}^\top)$ *for any* $\tau\neq 0$. *The sum of the i-th largest singular values over all lags* $\tau\in\mathbb{Z}$ *for each* $i\in[d]$ *is assumed to be finite for both autocovariance matrices of* $\{d^{-b/2}\mathbf{B}_t\}$ *and* $\{\boldsymbol{\epsilon}_t\}$.

(R9) *Define* $c_T:=gT^{-1/2}\log^{1/2}(T\vee d)$ *for some constant* $g>0$. *The tuning parameter for the adaptive LASSO problem (2.7) is* $\lambda=Cc_T$ *for some constant* $C>0$.

(R10) (Rate assumptions) *We assume that as* $L,d,T\to\infty$,

$$c_TL^{3/2}d^{1-a},\ Ld^{-1},\ L^2d^3T^{2-w},\ d^{b+2a+1/w}T^{-1}=o(1),$$
$$d^{-1/w}\log(T\vee d),\ d^{b-a-1/w}\log^{-1}(T\vee d)=O(1).$$

Assumption (I1) facilitates identifying the model (2.11); see the online supplement for more details. In the sequel, we discuss in detail the structure of the true spatial weight matrix $\mathbf{W}_t^*$, and some technical assumptions made above.

The time series components in the model are described in Assumption (M1), allowing for weak serial dependence in addition to strict stationarity, as commonly seen in the spatial panel literature (e.g. Giordano et al. (2024) and Li and Lin (2024)). While mixing conditions are imposed in previous works, we adopt the notion of functional or physical dependence, following earlier studies such as Shao (2010) and Wu et al. (2024). We further assume exponential tails in (M1), enabling the application of a Nagaev-type inequality for functionally dependent data (Liu et al., 2013).

Assumptions (M2) and (M2') describe the structure of $\mathbf{W}_t^*$ under two different settings for $\{z_{j,k,t}\}$. The row sum condition for $\mathbf{W}_t^*$ ensures that the model (2.1) is uniformly stationary. Assumption (M2) treats $\{z_{j,k,t}\}$ as a non-random series, while (M2') allows $\{z_{j,k,t}\}$ to be stochastic. Note that in (M2'), the stationarity is guaranteed if $\|\mathbf{W}_t^*\|_\infty < \eta < 1$ and $|\rho_t^*| < 1$ hold with probability 1. The assumption of zero mean and support [-1,1] is equivalent to the general assumption of bounded support, as we can always rewrite each $\phi_{j,k}^* z_{j,k,t}$ as

$$\phi_{j,k}^* z_{j,k,t} = \phi_{j,k}^* \mathbb{E}(z_{j,k,t}) + \phi_{j,k}^* \big[z_{j,k,t} - \mathbb{E}(z_{j,k,t})\big] = \phi_{j,k}^* \mathbb{E}(z_{j,k,t}) + \big(\phi_{j,k}^* z_{j,k}^*\big)\big\{\big[z_{j,k,t} - \mathbb{E}(z_{j,k,t})\big]/z_{j,k}^*\big\},$$

where $[-z_{j,k}^*, z_{j,k}^*]$ is the support of $z_{j,k,t}$, and $\phi_{j,k}^* \mathbb{E}(z_{j,k,t})$, a constant term, can be absorbed into $z_{j,0,t}$. We note that, in fact, we may allow some $\{z_{j,k,t}\}$ for $j \in \mathcal{P} \subseteq [p], k \in \mathcal{L}_j \subseteq [l_j]$ satisfying (M2), while all other $\{z_{j,k,t}\}$ satisfying (M2'), as all theoretical results hold under either assumption.

Assumption (R1) describes how sparse each spatial weight matrix candidate is. It is worthwhile pointing out that although each $\mathbf{W}_j$ is not necessarily linearly independent with each other by (R1), we actually implicitly impose such linear independence condition from (I1) through the combination of $\{z_{j,k,t}\}$ and $\mathbf{W}_j$ in $\mathbf{B}^\top \mathbf{V}$. (R2) is included as a technical addition to (M2).

Assumptions (R3) to (R6) all draw on the relation between $\mathbf{B}_t$ and $\mathbf{X}_t$. Their dependence structure is non-trivial due to the extra complication from spatial weight matrix and the time-varying components here. Naturally, $\mathbf{B}_t$ needs to be correlated with $\mathbf{X}_t$ to a certain extent, captured by an unknown constant a which facilitates the presentation of theoretical results. For instance, as an immediate consequence of (R5), we may derive $\left\|\mathbb{E}(\mathbf{X}_t\otimes\mathbf{B}_t\boldsymbol{\gamma})\right\|_1=O(d^{1+a})$, which is a key ingredient in obtaining the rates for $\left\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*\right\|_1$ in Theorem 1.

The predictive dependence measure defined in (R7) allows us to apply the central limit theorem for data with functional dependence and the Assumption (R7) can be satisfied by, for example, causal AR processes. Assumption (R8) further restricts the serial correlation in the noise. It also introduces another constant b that characterizes how elements in $\mathbf{B}_t$ are contemporarily and temporally correlated with each other. From the comparison of rates derived in the proofs, we conclude that b is actually bounded above by $1/w$, which is intuitive since a large w in (M1) generally implies light-tails and hence a small b. Lastly, (R9) sets the rate for λ and (R10) characterizes the relation between T, d and L. As an example, (R10) holds when $w=6$, $a=1/2$, $L=O(d^{1/3})$ and $T\asymp d^2$. More generally, (R10) can be satisfied as long as T grows at a polynomial rate in d.

3.2. Main results

In this subsection, we formally present the main results for the estimators of our model.

Theorem 1 *Let all assumptions in Section 3.1 hold ((M2) or (M2')). Given any $\boldsymbol{\phi}$ as an estimator of $\boldsymbol{\phi}^*$, with c_T defined in Assumption (R9), $\boldsymbol{\beta}(\boldsymbol{\phi})$ according to (2.13) satisfies*

$$\left\|\boldsymbol{\beta}(\boldsymbol{\phi})-\boldsymbol{\beta}^*\right\|_1=O_p\Big(\left\|\boldsymbol{\phi}-\boldsymbol{\phi}^*\right\|_1+c_T d^{-\frac{1}{2}+\frac{1}{2w}}\Big).$$

In particular, the least squares estimator $\widetilde{\boldsymbol{\phi}}$ in (2.14) and $\widetilde{\boldsymbol{\beta}}:=\boldsymbol{\beta}(\widetilde{\boldsymbol{\phi}})$ under $L=O(1)$ satisfy

$$\left\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*\right\|_1=O_P\big(c_T d^{-\frac{1}{2}+\frac{1}{2w}}\big)=\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\right\|_1.$$

Theorem 1 serves as a foundational step for the results hereafter. From Theorem 1, the error of our least squares estimator $\boldsymbol{\beta}(\boldsymbol{\phi})$ might be inflated by the plugged-in estimator for $\boldsymbol{\phi}^*$. With a dense estimator $\widetilde{\boldsymbol{\phi}}$, we arrive at the rate $c_T d^{-\frac{1}{2}+\frac{1}{2w}}$. The dependence of the rate on w confirms that weaker temporal dependence in the data results in better estimation, as expected. We now present the sign-consistency of our adaptive LASSO estimator.

Theorem 2 *(Oracle property for $\widehat{\boldsymbol{\phi}}$) Let all assumptions in Section 3.1 hold ((M2) or (M2')), except that (R4) and (R6) are satisfied with $\check{\mathbf{G}}$ and $\ddot{\mathbf{G}}$ respectively replaced by $\check{\mathbf{G}}_H$ and $\ddot{\mathbf{G}}_H$, where $H := \{i : (\boldsymbol{\phi}^*)_i \neq 0\}$ and for any matrix with L columns, $(\cdot)_H$ denotes the corresponding submatrix with columns restricted on H. Then, as $T, d \to \infty$, with probability approaching 1, $\mathrm{sign}(\widehat{\boldsymbol{\phi}}_H) = \mathrm{sign}(\boldsymbol{\phi}^*_H)$ and $\widehat{\boldsymbol{\phi}}_{H^c} = \mathbf{0}$.*

If we further assume the smallest eigenvalue of $\mathbf{R}_H \mathbf{S}_{\boldsymbol{\gamma}} \mathbf{S}_{\boldsymbol{\gamma}}^\top \mathbf{R}_H^\top$ is of constant order, where $\mathbf{R}_H$ and $\mathbf{S}_{\boldsymbol{\gamma}}$ are defined below, then $\widehat{\boldsymbol{\phi}}_H$ is asymptotically normal with rate $T^{-1/2} d^{-(1-b)/2}$ such that

$$T^{1/2}(\mathbf{R}_H \mathbf{S}_{\boldsymbol{\gamma}} \mathbf{R}_\beta \boldsymbol{\Sigma}_\beta \mathbf{R}_\beta^\top \mathbf{S}_{\boldsymbol{\gamma}}^\top \mathbf{R}_H^\top)^{-1/2}(\widehat{\boldsymbol{\phi}}_H - \boldsymbol{\phi}^*_H) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_{|H|}), \quad \textit{where}$$

$$\mathbf{R}_\beta = \big[\mathbb{E}(\mathbf{X}_t^\top \mathbf{B}_t)\mathbb{E}(\mathbf{B}_t^\top \mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t^\top \mathbf{B}_t),\ \boldsymbol{\Sigma}_\beta = \sum_\tau \mathbb{E}\big\{(\mathbf{B}_t - \mathbb{E}(\mathbf{B}_t))^\top \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_{t+\tau}^\top (\mathbf{B}_{t+\tau} - \mathbb{E}(\mathbf{B}_t))\big\},$$

$$\mathbf{S}_{\boldsymbol{\gamma}} = \Big(\mathbb{E}\big[\mathbf{X}_{t,1\cdot}(\mathbf{B}_{t,1\cdot} - \mathbb{E}(\mathbf{B}_{t,1\cdot}))^\top\big]\boldsymbol{\gamma}, \ldots, \mathbb{E}\big[\mathbf{X}_{t,1\cdot}(\mathbf{B}_{t,d\cdot} - \mathbb{E}(\mathbf{B}_{t,d\cdot}))^\top\big]\boldsymbol{\gamma},$$
$$\ldots, \mathbb{E}\big[\mathbf{X}_{t,d\cdot}(\mathbf{B}_{t,1\cdot} - \mathbb{E}(\mathbf{B}_{t,1\cdot}))^\top\big]\boldsymbol{\gamma}, \ldots, \mathbb{E}\big[\mathbf{X}_{t,d\cdot}(\mathbf{B}_{t,d\cdot} - \mathbb{E}(\mathbf{B}_{t,d\cdot}))^\top\big]\boldsymbol{\gamma}\Big)^\top,$$

$$\mathbf{R}_H = [(\mathbf{H}_{20} - \mathbf{H}_{10})_H^\top (\mathbf{H}_{20} - \mathbf{H}_{10})_H]^{-1}(\mathbf{H}_{20} - \mathbf{H}_{10})_H^\top,$$

$$\mathbf{H}_{10} = \Big\{\Big(\mathbf{I}_d \otimes (\boldsymbol{\gamma}^\top \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\mathbf{x},1,0}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\top}\big)\Big)\mathrm{vec}(\mathbf{W}_1^\top) + \Big(\mathbf{I}_d \otimes (\boldsymbol{\gamma}^\top \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\boldsymbol{\mu},1,0}\boldsymbol{\Pi}_t^{*\top}\big)\Big)\mathrm{vec}(\mathbf{W}_1^\top),$$
$$\ldots, \Big(\mathbf{I}_d \otimes (\boldsymbol{\gamma}^\top \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\mathbf{x},p,l_p}(\boldsymbol{\beta}^* \otimes \mathbf{I}_d)\boldsymbol{\Pi}_t^{*\top}\big)\Big)\mathrm{vec}(\mathbf{W}_p^\top) + \Big(\mathbf{I}_d \otimes (\boldsymbol{\gamma}^\top \otimes \mathbf{I}_d)\mathbb{E}\big(\mathbf{U}_{\boldsymbol{\mu},p,l_p}\boldsymbol{\Pi}_t^{*\top}\big)\Big)\mathrm{vec}(\mathbf{W}_p^\top)\Big\},$$

$$\mathbf{H}_{20} = \mathbb{E}(\mathbf{X}_t \otimes \mathbf{B}_t \boldsymbol{\gamma})\big[\mathbb{E}(\mathbf{X}_t^\top \mathbf{B}_t)\mathbb{E}(\mathbf{B}_t^\top \mathbf{X}_t)\big]^{-1}\mathbb{E}(\mathbf{X}_t^\top \mathbf{B}_t)$$
$$\cdot \Big\{\mathbf{V}_{\mathbf{W}_1^\top, v}^\top \mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\mathbf{x},1,0})\mathbf{V}_{\boldsymbol{\Pi}_t^*, r}\big]\boldsymbol{\beta}^* + \mathbf{V}_{\mathbf{W}_1^\top, v}^\top \mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\mu},1,0})\mathrm{vec}(\boldsymbol{\Pi}_t^{*\top})\big],$$
$$\ldots, \mathbf{V}_{\mathbf{W}_p^\top, v}^\top \mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\mathbf{x},p,l_p})\mathbf{V}_{\boldsymbol{\Pi}_t^*, r}\big]\boldsymbol{\beta}^* + \mathbf{V}_{\mathbf{W}_p^\top, v}^\top \mathbb{E}\big[(\mathbf{I}_d \otimes \mathbf{U}_{\boldsymbol{\mu},p,l_p})\mathrm{vec}(\boldsymbol{\Pi}_t^{*\top})\big]\Big\},$$

$$\mathbf{U}_{\mathbf{x},j,k}=\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}\mathrm{vec}(\mathbf{B}_t-\bar{\mathbf{B}})\mathbf{x}_t^\top,\ \mathbf{U}_{\boldsymbol{\mu},j,k}=\frac{1}{T}\sum_{t=1}^{T}z_{j,k,t}\mathrm{vec}(\mathbf{B}_t-\bar{\mathbf{B}})\boldsymbol{\mu}^{*\top},$$

with $\mathbf{V}_{\mathbf{H},K}=\big(\mathbf{I}_K\otimes\mathbf{h}_1^\top,\ ...,\ \mathbf{I}_K\otimes\mathbf{h}_n^\top\big)^\top$ *for any given* $n\times d$ *matrix* $\mathbf{H}=(\mathbf{h}_1,...,\mathbf{h}_n)^\top$.

Regarding the new assumptions in Theorem 2, the term $\check{\mathbf{G}}$ (resp. $\ddot{\mathbf{G}}$) in (R4) (resp. (R6)) is simply replaced by its H-restricted version. Moreover, we show in the proof that $\mathbf{R}_H\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{S}_{\boldsymbol{\gamma}}^\top\mathbf{R}_H^\top$ has its largest eigenvalue of constant order, and hence the requirement on its smallest eigenvalue is not particularly strong. Nonetheless, two key results are obtained: $\widehat{\boldsymbol{\phi}}$ consistently estimates the zeros in $\boldsymbol{\phi}^*$ as exact zeros, and is asymptotically normal on the nonzero entries in $\boldsymbol{\phi}^*$. The convergence rate is $T^{-1/2}d^{-(1-b)/2}$, which is slower if more variables in $\mathbf{B}_t$ are correlated. In the special case where $\mathbf{B}_t$ is cross-sectionally and serially uncorrelated with $b=0$, the convergence rate simplifies to $(Td)^{-1/2}$.

Theorem 2 enables us to perform inference on $\boldsymbol{\phi}_H$ in practice, with the covariance matrix replaced by the plug-in estimator; see Section 4 for more details. If $\{z_{j,k,t}\}$'s are non-stochastic, inference on ρ_t^* by $\widehat{\rho}_t:=(\mathbf{z}_t)_H^\top\widehat{\boldsymbol{\phi}}_H$ is also feasible, since

$$T^{1/2}((\mathbf{z}_t)_H^\top\mathbf{R}_H\mathbf{S}_{\boldsymbol{\gamma}}\mathbf{R}_\beta\boldsymbol{\Sigma}_\beta\mathbf{R}_\beta^\top\mathbf{S}_{\boldsymbol{\gamma}}^\top\mathbf{R}_H^\top(\mathbf{z}_t)_H)^{-1/2}(\widehat{\rho}_t-\rho_t^*)\xrightarrow{D}\mathcal{N}(0,1).$$

Lastly, we present the consistency of the spatial weight matrix estimator and the spatial fixed effect estimator.

Theorem 3 *Let assumptions in Theorems 1 and 2 hold. Then, for* $\widehat{\mathbf{W}}_t:=\sum_{j=1}^p\Big(\widehat{\phi}_{j,0}+\sum_{k=1}^{l_j}\widehat{\phi}_{j,k}z_{j,k,t}\Big)\mathbf{W}_j$ *and the spatial fixed effect estimator* $\widehat{\boldsymbol{\mu}}$ *defined in (2.8), we have*

$$\big\|\widehat{\mathbf{W}}_t-\mathbf{W}_t^*\big\|_\infty=O_P\big(T^{-1/2}d^{-(1-b)/2}\big)=\big\|\widehat{\mathbf{W}}_t-\mathbf{W}_t^*\big\|_1,\ \big\|\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}^*\big\|_{\max}=O_P(c_T).$$

Note that this result implies that the spectral norm error of the spatial weight matrix estimator $\widehat{\mathbf{W}}_t$ also satisfies $\big\|\widehat{\mathbf{W}}_t-\mathbf{W}_t^*\big\|=O_P\big(T^{-1/2}d^{-(1-b)/2}\big)$.

4. PRACTICAL IMPLEMENTATION

From Section 2.2, we estimate $(\boldsymbol{\mu}^*,\boldsymbol{\phi}^*,\boldsymbol{\beta}^*)$ by first obtaining a penalized estimator for $\boldsymbol{\phi}^*$, followed by the least squares estimator for $\boldsymbol{\beta}^*$ and $\boldsymbol{\mu}^*$. We now present the step-by-step algorithm.

<u>Algorithm for $(\boldsymbol{\mu}^*,\boldsymbol{\phi}^*,\boldsymbol{\beta}^*)$ Estimation</u>

1. Compute the least squares estimator stated in (2.14) and denote it as $\widetilde{\boldsymbol{\phi}}$.
2. Construct $\mathbf{u}$ using $\widetilde{\boldsymbol{\phi}}$. Using the Least Angle Regressions (LARS) (Efron et al., 2004), solve the adaptive LASSO problem stated in (2.15), and denote the solution by $\widehat{\boldsymbol{\phi}}$.
3. Using (2.13), obtain the least squares estimator for $\boldsymbol{\beta}^*$ as $\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$.
4. According to (2.3), construct $\widehat{\boldsymbol{\Phi}}$ using $\widehat{\boldsymbol{\phi}}$ and obtain the least squares estimator for $\boldsymbol{\mu}^*$ as $\widehat{\boldsymbol{\mu}}=T^{-1}\sum_{t=1}^{T}\big\{(\mathbf{I}_d-\boldsymbol{\Lambda}_t\widehat{\boldsymbol{\Phi}})\mathbf{y}_t-\mathbf{X}_t\widehat{\boldsymbol{\beta}}\big\}$.

The tuning parameter λ in step 2 can be determined via minimizing the following BIC, borrowed from equation (10) in Wang and Leng (2007):

$$\mathrm{BIC}(\lambda)=\log\Big(\frac{1}{T}\Big\|\mathbf{B}^\top\mathbf{y}-\mathbf{B}^\top\mathbf{V}\widehat{\boldsymbol{\phi}}-\mathbf{B}^\top\mathbf{X}_{\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})}\mathrm{vec}\big(\mathbf{I}_d\big)\Big\|^2\Big)+|\widehat{H}|\,\frac{\log(T)}{T}, \tag{4.1}$$

where $\widehat{\boldsymbol{\phi}}$ is the adaptive LASSO solution with parameter λ and $\widehat{H}$ is the set of indices on which $\widehat{\boldsymbol{\phi}}$ is nonzero. Note that although $\mathbf{B}$ contains the unknown constant a, the optimal λ is independent of it. A procedure for assessing the goodness of fit of a given set of dynamic variables $\{z_{j,k,t}\}$ can also be facilitated by (4.1). In detail, we can compare (2.1) with the null model $\mathcal{H}_0:\mathbf{y}_t=\boldsymbol{\mu}^*+\sum_{j=1}^{p}\phi_{j,0}^*\mathbf{W}_j\mathbf{y}_t+\mathbf{X}_t\boldsymbol{\beta}^*+\boldsymbol{\epsilon}_t$ also using (4.1), except that the variables are constructed under $\mathcal{H}_0$ and the above algorithm is implemented accordingly. More discussion is relegated to the online supplement, with extensive numerical results therein.

Finally, to utilize the asymptotic normality of $\widehat{\boldsymbol{\phi}}_H$ for feasible inference, we require estimators for $\mathbf{R}_H$, $\mathbf{R}_\beta$, $\boldsymbol{\Sigma}_\beta$ and $\mathbf{S}_\gamma$ in Theorem 2. By replacing the expected values by their sample estimates, using $\widehat{H}=\{i:(\widehat{\boldsymbol{\phi}})_i\neq 0\}$ to estimate the set H and leveraging all the consistency results for $\boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$,

$\widehat{\boldsymbol{\mu}}$ and $\widehat{H}$, we obtain estimators $\widehat{\mathbf{R}}_{\widehat{H}}$, $\widehat{\mathbf{R}}_\beta$ and $\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}$. For $\boldsymbol{\Sigma}_\beta$, we use a consistent estimator of $\boldsymbol{\epsilon}_t$ denoted by $\widehat{\boldsymbol{\epsilon}}_t := \mathbf{y}_t - \widehat{\boldsymbol{\mu}} - \widehat{\mathbf{W}}_t \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}(\widehat{\boldsymbol{\phi}})$. As $\boldsymbol{\Sigma}_\beta$ involves an infinite sum, we can sum up to a cut-off τ^* after which the sum changes little, and denote the constructed estimator $\widehat{\boldsymbol{\Sigma}}_\beta$. Putting everything together, the covariance matrix of $\widehat{\boldsymbol{\phi}}_H$ can be estimated by $T^{-1}\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}\widehat{\mathbf{R}}_\beta\widehat{\boldsymbol{\Sigma}}_\beta\widehat{\mathbf{R}}_\beta^\top\widehat{\mathbf{S}}_{\boldsymbol{\gamma}}^\top\widehat{\mathbf{R}}_{\widehat{H}}^\top$.

Establishing the theoretical validity of inference based on plug-in estimators would require additional assumptions to rigorously justify each step outlined above. As this is not the primary focus of our work, we do not pursue a formal proof of consistency for the plug-in estimator. Instead, we provide empirical evidence supporting its adequacy through simulation studies in Section 6.1.

5. CHANGE POINT DETECTION AND ESTIMATION IN SPATIAL AUTOREGRESSIVE MODELS

5.1. Threshold spatial autoregressive models

The early work by Tong (1978) proposes a regime switching mechanism via the threshold autoregressive model. Since then, it has been studied extensively for panel data in the past few decades (Hansen, 1999). More recently, such threshold structure is used by researchers in spatial econometrics; see, for example, threshold spatial autoregressive models for cross-sectional data by Deng (2018) and Li and Lin (2024), and spatial panel data models with threshold effects also on regression coefficient by Meng and Yang (2023). One benefit of the framework introduced in this work is that threshold variables can be directly adapted into (2.1) under a panel data setup. As a simple example, we consider

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \Big(\phi_1^* \mathbf{W}_1 \mathbb{1}\{q_t \le \gamma^*\} + \phi_2^* \mathbf{W}_2 \mathbb{1}\{q_t > \gamma^*\}\Big)\mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \tag{5.1}$$

This is a spatial autoregressive model with regime switching on the spatial weight matrix, where q_t is some observed threshold variable with an unknown threshold value γ^*. By rewriting (5.1) in the form of (2.1), we have

$$\mathbf{y}_t = \boldsymbol{\mu}^* + z_{1,1,t}\phi_1^*\mathbf{W}_1\mathbf{y}_t + z_{2,1,t}\phi_2^*\mathbf{W}_2\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \text{ where } z_{1,1,t} := \mathbb{1}\{q_t \le \gamma^*\}, z_{2,1,t} := \mathbb{1}\{q_t > \gamma^*\}. \tag{5.2}$$

We consider the estimation of the threshold value γ^*. Suppose there is a domain of possible threshold values $\Gamma = [\gamma_{\min}, \gamma_{\max}]$. A standard approach in threshold models is to search the minimum regression error over the intersection $\Gamma \cap \{q_1, ..., q_T\}$; see, for example, Deng (2018). Our framework provides an alternative approach. Denote the elements in the intersection by $\gamma_1 \le \gamma_2 \le \cdots \le \gamma_L$, with $L \equiv |\Gamma \cap \{q_1, ..., q_T\}|$, and let γ^* be identified as one of them. Let $z_{1,l,t} := \mathbb{1}\{q_t \le \gamma_l\}$ and $z_{2,l,t} := \mathbb{1}\{q_t > \gamma_l\}$ for all $l \in [L]$. Then, we can consider a spatial autoregressive model:

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\mathbf{W}_1\mathbf{y}_t + \sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^*\mathbf{W}_2\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{5.3}$$

where $\phi_{1,l}^*$ and $\phi_{2,l}^*$ would be nonzero[2] only for $\gamma_l = \gamma^*$. The threshold value can be selected consistently in one step by the oracle property of our adaptive LASSO estimator in Theorem 2. We present this result in Corollary 1. Given the sparse solution $(\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l})_{l \in [L]}$ of (5.3), we can re-estimate all model parameters. Note that such an approach remains applicable for L growing with T. In practice, the order of L may not fulfill Assumption (R10), but we can circumvent this issue by a two-step procedure, as we will discuss later in Section 5.2.

Our framework also allows us to consider a spatial autoregressive model with regime switching on the spatial correlation coefficients, similar to Li and Lin (2024) except that their regimes change over spatial units while ours over timestamps as follows,

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \Big(\phi_1^*\mathbb{1}\{q_t \le \gamma^*\} + \phi_2^*\mathbb{1}\{q_t > \gamma^*\}\Big)\mathbf{W}_1\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \tag{5.4}$$

[2] When there is no prior information on which spatial weight matrix corresponds to the regime $q_t \le \gamma^*$, in (5.3) we can write $\big(\sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\big)$ and $\big(\sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^*\big)$ as $\big(\phi_{1,0}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\big)$ and $\big(\phi_{2,0}^* + \sum_{l=1}^{L} z_{2,l,t}\phi_{2,l}^*\big)$, respectively.

To estimate the parameters, Li and Lin (2024) uses quasi maximum likelihood estimators (QMLE) and traverses over a finite parameter space for the threshold value γ^*. In contrast, a one-step estimation is again feasible by our framework. To this end, we may read (5.4) in the form,

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \Big(\phi_{1,0}^* + \sum_{l=1}^{L} z_{1,l,t}\phi_{1,l}^*\Big)\mathbf{W}_1\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{5.5}$$

where $z_{1,l,t}$, $l\in[L]$, are as previously defined. With our adaptive LASSO estimators, only $\widehat{\phi}_{1,0}$ and one $\widehat{\phi}_{1,l}$ such that $\gamma_l = \gamma^*$ should be nonzero. The consistency of such estimator for γ^* is left in Corollary 1.

Corollary 1 *(Consistency on threshold value estimation) Let all assumptions in Theorem 2 hold.*
(a) For model (5.1), the threshold value γ^* can be consistently estimated by the set of estimators $\{\gamma_l : \widehat{\phi}_{1,l} \neq 0, \widehat{\phi}_{2,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l}\}_{l\in[L]}$ is the adaptive LASSO solution for $\{\phi_{1,l}^*, \phi_{2,l}^*\}_{l\in[L]}$ in (5.3).
(b) For model (5.4), the threshold value γ^* can be consistently estimated by the set of estimators $\{\gamma_l : \widehat{\phi}_{1,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}\}_{l\in[L]}$ is the adaptive LASSO solution for $\{\phi_{1,l}^*\}_{l\in[L]}$ in (5.5).

In fact, our framework (2.1) can be applied to spatial autoregressive models with more complicated threshold structures in the spatial weight matrix $\mathbf{W}_t^*$, e.g., regimes from multiple threshold variables with multiple threshold values. We leave this discussion to the online supplement.

5.2. Spatial autoregressive models with structural change points

Structural changes in the relationship of variables in econometric models have been studied extensively in the literature; see, for example, Sengupta (2017) and Barigozzi and Trapani (2020). As a second example demonstrating the applicability of our framework, we consider the spatial autoregressive model with a structural change:

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \Big(\phi_1^*\mathbf{W}_1\mathbb{1}\{t\leq t^*\} + \phi_2^*\mathbf{W}_2\mathbb{1}\{t>t^*\}\Big)\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \tag{5.6}$$

where t^* is some unknown change location. Similar to the threshold model example, this can also be expressed in the form of (5.2), now with $z_{1,1,t} := \mathbb{1}\{t \le t^*\}$ and $z_{2,1,t} := \mathbb{1}\{t > t^*\}$. It is important to note that, despite the models taking the same form, the dynamic variables $z_{1,1,t}$ and $z_{2,1,t}$ are random in the threshold model but non-random in the change point model here. Recall that our main results hold for both types of $\{z_{j,k,t}\}$; see Section 3.1 for a more detailed discussion.

To estimate the change location t^*, Li (2018) calculates the quasi maximum likelihood for each possible change location and sets the maximizer as the estimator. When the set of possible change locations is large, this approach requires a significant number of model fittings. Using our framework, it is again possible to estimate t^* in one go. Let $\mathcal{T}$ denote the set of all candidate change point locations such that $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$. Then, we rewrite model (5.6) as

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{l=1}^{|\mathcal{T}|} z_{1,l,t} \phi^*_{1,l} \mathbf{W}_1 \mathbf{y}_t + \sum_{l=1}^{|\mathcal{T}|} z_{2,l,t} \phi^*_{2,l} \mathbf{W}_2 \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t, \tag{5.7}$$

where $z_{1,l,t} := \mathbb{1}\{t \le t_l\}$ and $z_{2,l,t} := \mathbb{1}\{t > t_l\}$. We then follow the same argument used in the threshold model below (5.3) and the consistency of our change location estimate is again guaranteed.

Corollary 2 *(Consistency on change location estimation) Let all assumptions in Theorem 2 hold, with L replaced by $|\mathcal{T}|$. Consider model (5.6) and assume that $t^* \in \mathcal{T}$. Then, the change location t^* can be consistently estimated by the set of estimators $\{l \in [|\mathcal{T}|] : \widehat{\phi}_{1,l} \neq 0, \widehat{\phi}_{2,l} \neq 0\}$, where $\{\widehat{\phi}_{1,l}, \widehat{\phi}_{2,l}\}_{l \in [|\mathcal{T}|]}$ is the adaptive LASSO solution for $\{\phi^*_{1,l}, \phi^*_{2,l}\}_{l \in [|\mathcal{T}|]}$ in (5.7).*

We can also consider a multiple change model:

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p} \Big(\phi^*_{j,1} \mathbb{1}\{t \le t^*_1\} + \phi^*_{j,2} \mathbb{1}\{t^*_1 < t \le t^*_2\} + \cdots + \phi^*_{j,k} \mathbb{1}\{t > t^*_k\} \Big) \mathbf{W}_j \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t. \tag{5.8}$$

This model allows for k change points in $\mathbf{W}^*_t$ consisting of p spatial weight candidates, with the number of change points k unknown. The result below confirms the consistency of the estimations in both change point numbers and locations.

Corollary 3 *(Consistency on the estimations for the number of changes and the change locations) Given a set of all candidate change point locations $\mathcal{T}$ and assume that $t_i^* \in \mathcal{T}$ for all $i \in [k]$. Let all assumptions in Theorem 2 hold, with L replaced by $|\mathcal{T}|$. For model (5.8), let $\{\widehat{\phi}_{j,l}\}_{j\in[p],l\in[|\mathcal{T}|]}$ denote the adaptive LASSO solution for $\{\phi^*_{j,l}\}_{j\in[p],l\in[|\mathcal{T}|]}$ in*

$$\mathbf{y}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{p}\Big(\phi^*_{j,0} + \sum_{l=1}^{|\mathcal{T}|} z_{j,l,t}\phi^*_{j,l}\Big)\mathbf{W}_j\mathbf{y}_t + \mathbf{X}_t\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t,$$

where $z_{j,l,t} := \mathbb{1}\{t \le t_l\}$ with t_l being the l-th element of $\mathcal{T}$ for $j \in [p], l \in \big[|\mathcal{T}|\big]$. Write $\widehat{\mathcal{T}} := \{l \neq 0 : \widehat{\phi}_{j,l} \neq 0$ for some $j \in [p]\}$. Then, $\widehat{k} := |\widehat{\mathcal{T}}|$ estimates k consistently and, for every $i \in [\widehat{k}]$, the i-th smallest element in $\widehat{\mathcal{T}}$ estimates t_i^* consistently.

Remark 1 *Suppose the true number of changes is of order L satisfying Assumption (R10), but throughout Section 5.2, the size of the set $\mathcal{T}$ is restricted by (R10). This means that applying our procedure to a single set $\mathcal{T}$ containing many candidate change point locations may violate this assumption. A similar concern arises with the order of L in threshold spatial autoregressive models in Section 5.1.*

To address this in practice, we resort to a divide-and-conquer scheme to work with smaller candidate sets. Consider (5.6) for instance. We first form some subsets $\mathcal{T}_1, \mathcal{T}_2, \ldots$ such that $\mathcal{T} = \cup_j \mathcal{T}_j$ and each $\mathcal{T}_j$ satisfies (R10) (with L replaced by $|\mathcal{T}_j|$). Note that these subsets can overlap. On each subset, we run the estimation algorithm and identify all potential change locations within the subset. Then, we aggregate all those locations into a set $\widetilde{\mathcal{T}}$, which can be shown, using Corollary 2 on each $\mathcal{T}_j$, to satisfy (R10) (with L replaced by $|\widetilde{\mathcal{T}}|$). Finally, we estimate (5.7) with $\mathcal{T}$ replaced by $\widetilde{\mathcal{T}}$ to determine the change point location. Simulations in the online supplement confirms the effectiveness of this scheme.

6. NUMERICAL STUDIES

6.1. Simulations

We conduct simulations to demonstrate the performance of our estimators. For the general setting:

$$\mathbf{y}_t=\Big\{\mathbf{I}_d-\big(0.2+0.2z_{1,1,t}+0z_{1,2,t}\big)\mathbf{W}_1-\big(0+0z_{2,1,t}+0.3z_{2,2,t}\big)\mathbf{W}_2\Big\}^{-1}\big(\boldsymbol{\mu}^*+\mathbf{X}_t\boldsymbol{\beta}^*+\boldsymbol{\epsilon}_t\big), \qquad (6.1)$$

where $\boldsymbol{\mu}^*$ and $\boldsymbol{\beta}^*$ are vectors of 1's, $\mathbf{W}_1$ is generated to have two neighbors ahead and two behind as in Kelejian and Prucha (1998), and $\mathbf{W}_2$ is a contiguity matrix with off-diagonal entries being i.i.d. Bernoulli(0.2). The true parameter is $\boldsymbol{\phi}^*=(0.2,0.2,0,0,0,0.3)^\top$. The disturbance $\boldsymbol{\epsilon}_t$ is jointly Gaussian with its variance-covariance matrix having 1 on the diagonal and each upper triangular entries 0.1 with probability 0.2 and 0 otherwise. For any row of $\mathbf{W}_1$ or $\mathbf{W}_2$ with row sum exceeding one, we divide each entry by the L_1 norm of the row. We use independent standard normal random variables for the dynamic variables $\{z_{1,1,t}\}$, $\{z_{1,2,t}\}$, $\{z_{2,1,t}\}$ and $\{z_{2,2,t}\}$. The covariate matrix $\mathbf{X}_t$ has three columns, with each entry generated as independent standard normal, except that the third column is endogenous by adding $0.5\boldsymbol{\epsilon}_t$. Let $\mathbf{X}_{\text{exo},t}$ be $\mathbf{X}_t$ with the disturbance part removed, then the instruments can be set as $\mathbf{B}_t=\big[\mathbf{X}_{\text{exo},t},\mathbf{W}_1\mathbf{X}_{\text{exo},t},\mathbf{W}_2\mathbf{X}_{\text{exo},t}\big]$. The tuning parameter for the adaptive LASSO is selected by minimizing the BIC in (4.1).

We experiment $d=25,50,75$ and $T=50,100,150$, with each setting repeated 1000 times. Results are presented in Table 1. In there, MSE is the mean squared error; specificity is the proportion of true zeros estimated as zeros; sensitivity is the proportion of nonzeros estimated as nonzeros. The MSE results corroborate the consistency of parameter estimation in Theorem 1 and 3, while the specificity and sensitivity results corroborate the sparsity consistency in Theorem 2. Zeros in $\boldsymbol{\phi}^*$ can be selected with high accuracy, yet the sensitivity results suggest a mild over-identification of zeros. Both increasing the spatial dimension d and time span T improve the performance of our estimators in general, except that when T increases from 100 to 150 and $d=25$, all performance measures slightly deteriorate, which is similarly observed in Table 1 of Lam and Souza (2020). This might suggest the issue of a data set with unbalanced dimensions in practice.

	$T=50$			$T=100$			$T=150$		
	$d=25$	$d=50$	$d=75$	$d=25$	$d=50$	$d=75$	$d=25$	$d=50$	$d=75$
$\widehat{\boldsymbol{\phi}}$ MSE	.002	.001	.002	.000	.000	.000	.003	.001	.000
	(.008)	(.001)	(.001)	(.000)	(.000)	(.000)	(.007)	(.000)	(.000)
$\widehat{\boldsymbol{\beta}}$ MSE	.087	.029	.080	.005	.008	.001	.088	.013	.005
	(.434)	(.011)	(.025)	(.004)	(.002)	(.001)	(.278)	(.005)	(.002)
$\widehat{\boldsymbol{\mu}}$ MSE	.039	.038	.044	.013	.010	.011	.036	.010	.008
	(.124)	(.010)	(.010)	(.006)	(.002)	(.002)	(.083)	(.003)	(.002)
$\widehat{\boldsymbol{\phi}}$ Specificity	.992	1.000	1.000	1.000	1.000	1.000	.994	1.000	1.000
	(.064)	(.000)	(.000)	(.000)	(.000)	(.000)	(.048)	(.000)	(.000)
$\widehat{\boldsymbol{\phi}}$ Sensitivity	.921	.983	.992	.910	.961	.991	.873	.983	.956
	(.133)	(.069)	(.048)	(.132)	(.099)	(.051)	(.148)	(.070)	(.103)

Table 1: Simulation results for the general setting (6.1). The mean and standard deviation (in brackets) of the corresponding error measures over 1000 repetitions are reported.

To better illustrate the asymptotic normality for $\widehat{\boldsymbol{\phi}}$ in Theorem 2, we use the same data generating mechanism as above with $(T,d)=(200,50)$, except that $\boldsymbol{\phi}^*=(0,\text{-}0.5,0.5,0,0,0)^\top$, $\mathbf{X}_t$ is exogenous and $\boldsymbol{\epsilon}_t$ has a diagonal covariance matrix. For ease of presentation, we fix $\widehat{H}=\{2,3\}$ which is the index set of true nonzero parameters. The remaining components of the covariance matrix are estimated according to the last part of Section 4. In particular, Figure 1 displays the histogram of $T^{1/2}(\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_\gamma\widehat{\mathbf{R}}_\beta\widehat{\boldsymbol{\Sigma}}_\beta\widehat{\mathbf{R}}_\beta^\top\widehat{\mathbf{S}}_\gamma^\top\widehat{\mathbf{R}}_{\widehat{H}}^\top)^{-1/2}(\widehat{\boldsymbol{\phi}}_{\widehat{H}}-\boldsymbol{\phi}^*_{\widehat{H}})$. The plots show good normal approximation to the distribution of this quantity, and hence confirm the results in Theorem 2 together with the validity of our proposed covariance estimator. Some discrepancies are present on the tails, potentially due to insufficient dimensions T and d. We leave potential improvements to future studies.

6.2. Change point analysis

We demonstrate in this subsection the performance of our dynamic framework with structural changes as described in Section 5.2, with simulation results for the threshold model in Section 5.1 in-

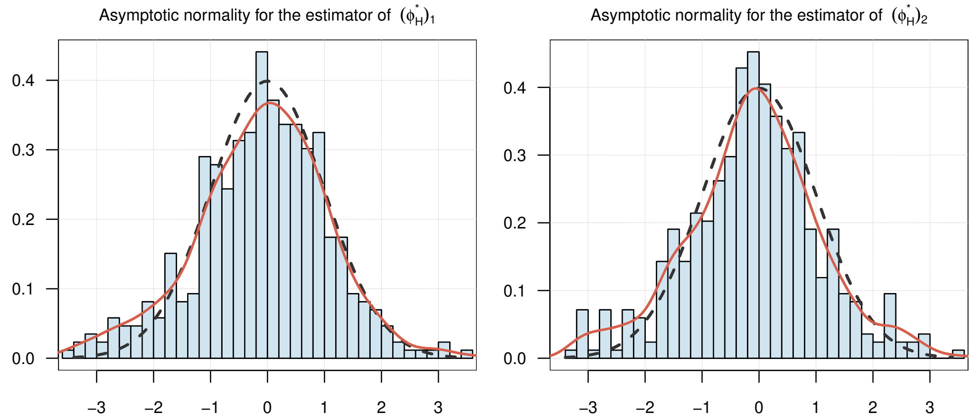

Figure 1: Histogram of $T^{1/2}(\widehat{\mathbf{R}}_{\widehat{H}}\widehat{\mathbf{S}}_{\gamma}\widehat{\mathbf{R}}_{\beta}\widehat{\mathbf{\Sigma}}_{\beta}\widehat{\mathbf{R}}_{\beta}^{\top}\widehat{\mathbf{S}}_{\gamma}^{\top}\widehat{\mathbf{R}}_{\widehat{H}}^{\top})^{-1/2}(\widehat{\boldsymbol{\phi}}_{\widehat{H}}-\boldsymbol{\phi}^*_{\widehat{H}})$ for $(T,d)=(200,50)$, shown for the first coordinate (left panel) and the second coordinate (right panel). The red curves are the empirical density, and the black dotted curves are the density for $\mathcal{N}(0,1)$.

cluded in the online supplement. We first consider a spatial autoregressive model with a single change:

$$\mathbf{y}_t=\Big\{\mathbf{I}_d-0.3\cdot\mathbb{1}\{t\leq 30\}\mathbf{W}_1-0.3\cdot\mathbb{1}\{t>30\}\mathbf{W}_2\Big\}^{-1}\big(\boldsymbol{\mu}^*+\mathbf{X}_t\boldsymbol{\beta}^*+\boldsymbol{\epsilon}_t\big), \tag{6.2}$$

where $\{\mathbf{W}_1,\mathbf{W}_2,\boldsymbol{\mu}^*,\mathbf{X}_t,\boldsymbol{\beta}^*\}$ are constructed in the same way as in (6.1). We generate $\boldsymbol{\epsilon}_t$ as i.i.d. $\mathcal{N}(0,1)$ and t_6, respectively, to demonstrate the robustness of our estimators under heavy-tailed noise. This data generating mechanism (6.2) represents a change on the true spatial weight matrix at time $t=30$ from $0.3\mathbf{W}_1$ to $0.3\mathbf{W}_2$. We then fit a model

$$\mathbf{y}_t=\Bigg\{\mathbf{I}_d-\sum_{l=1}^{\lfloor T/\Delta\rfloor-1}z_{1,l,t}\phi^*_{1,l}\mathbf{W}_1-\sum_{l=1}^{\lfloor T/\Delta\rfloor-1}z_{2,l,t}\phi^*_{2,l}\mathbf{W}_2\Bigg\}^{-1}\big(\boldsymbol{\mu}^*+\mathbf{X}_t\boldsymbol{\beta}^*+\boldsymbol{\epsilon}_t\big),\text{ where} \tag{6.3}$$

$z_{1,l,t}=\mathbb{1}\{t\leq t_l\}$, $z_{2,l,t}=\mathbb{1}\{t>t_l\}$, $t_l=\Delta\cdot l$.

We consider a grid of candidate change locations, spaced at intervals of $\Delta=5$. From (6.2), the true value of every $(\phi^*_{1,l},\phi^*_{2,l})$ is $(0,0)$ except the one corresponding to $t_l=30$. Table 2 presents the results for each (T,d) setting specified therein, together with the results under a stronger change signal, where the true spatial weight matrix switches from $0.5\mathbf{W}_1$ to $0.5\mathbf{W}_2$.

Weak Change Signal ($0.3\mathbf{W}_1$ to $0.3\mathbf{W}_2$)								
$\boldsymbol{\epsilon}_t$	i.i.d. $\mathcal{N}(0,1)$				i.i.d. t_6			
(T,d)	(50,25)	(50,50)	(100,50)	(100,75)	(50,25)	(50,50)	(100,50)	(100,75)
$\widehat{\boldsymbol{\phi}}$ MSE	.008	.004	.004	.003	.010	.004	.004	.003
	(.007)	(.003)	(.002)	(.002)	(.009)	(.003)	(.003)	(.002)
$\widehat{\boldsymbol{\phi}}$ True-Unique	.611	.965	.668	.938	.559	.944	.647	.898
	(.488)	(.184)	(.472)	(.242)	(.497)	(.231)	(.479)	(.303)
Strong Change Signal ($0.5\mathbf{W}_1$ to $0.5\mathbf{W}_2$)								
$\boldsymbol{\epsilon}_t$	i.i.d. $\mathcal{N}(0,1)$				i.i.d. t_6			
(T,d)	(50,25)	(50,50)	(100,50)	(100,75)	(50,25)	(50,50)	(100,50)	(100,75)
$\widehat{\boldsymbol{\phi}}$ MSE	.006	.001	.003	.001	.007	.001	.004	.001
	(.011)	(.001)	(.006)	(.001)	(.012)	(.003)	(.007)	(.002)
$\widehat{\boldsymbol{\phi}}$ True-Unique	.826	.994	.846	.990	.759	.984	.811	.972
	(.379)	(.077)	(.361)	(.100)	(.428)	(.126)	(.392)	(.166)

Table 2: Simulation results for the model (6.3) with either a weak or strong change signal. $\widehat{\boldsymbol{\phi}}$ *True-Unique* is defined to be 1 if the only nonzero pair $(\phi^*_{1,l},\phi^*_{2,l})$ corresponds to $t_l=30$. The mean and standard deviation (in brackets) of the corresponding error measures over 500 repetitions are reported.

Results for both weak and strong change signals display similar patterns. The performance of change detection is slightly affected by heavy-tailed noise and weaker signals, but remains effective overall. Detection accuracy improves with an increasing spatial dimension d. However, a larger T appears to undermine the $\widehat{\boldsymbol{\phi}}$ True-Unique measure, primarily due to the greater number of dynamic variables $z_{1,l,t}$ and $z_{2,l,t}$ included in the setting.

Comparison with Li (2018)

To further assess the performance of our estimator and compare both the computational cost and accuracy of our method with that in Li (2018), we consider

$$\mathbf{y}_t=\Big\{\mathbf{I}_d-\big(-0.3+\varrho\cdot\mathbb{1}\{t>30\}\big)\mathbf{W}_1\Big\}^{-1}\big(\boldsymbol{\mu}^*+\mathbf{X}_t\boldsymbol{\beta}^*+\boldsymbol{\epsilon}_t\big), \tag{6.4}$$

with $T=100$ and other parameters the same as in (6.2). This data generating process represents a single change on the spatial weight matrix at $t=30$ from $-0.3\mathbf{W}_1$ to $(-0.3+\varrho)\mathbf{W}_1$. For simplicity, we only examine the change point candidates in $\{10,20,...,90\}$, and compare three estimators:

1. AL: The adaptive LASSO estimator with dynamic variables $\{z_{t,0,1},...,z_{t,9,1}\}$, where $z_{t,0,1}=1$ and $z_{t,k,1}=\mathbb{1}\{t>10\cdot k\}$ for $k\in[9]$ and $t\in[100]$.
2. AL-DC: The adaptive LASSO estimator using the divide-and-conquer scheme in Remark 1, with three subsets $\{z_{t,0,1},z_{t,1,1},z_{t,2,1},z_{t,3,1}\}$, $\{z_{t,0,1},z_{t,4,1},z_{t,5,1},z_{t,6,1}\}$, and $\{z_{t,0,1},z_{t,7,1},z_{t,8,1},z_{t,9,1}\}$, where all dynamic variables are the same as in method 1 above. The final estimator is selected from the three results with the smallest estimation error.
3. QMLE (Li, 2018): The estimation is performed for each candidate in $\{10,20,...,90\}$ and the final estimator is the one with the smallest estimation error within this set.

Table 3 reports the results for $T=100$, $d=25,50$ with $\varrho=0.5$, where all simulations are repeated 500 times, and Table 4 presents the results for a smaller change of $\varrho=0.1$ at $t=30$. Note that specificity and sensitivity are not reported here, as QMLE leverages prior information of a single change. We use the estimation error $\sum_{t=1}^{T}\|\widehat{\boldsymbol{\epsilon}}_t\|^2/(Td)$ to quantify overall performance.

$\boldsymbol{\epsilon}_t$	i.i.d. $\mathcal{N}(0,1)$			i.i.d. t_6		
$d=25$	AL	AL-DC	QMLE	AL	AL-DC	QMLE
$\widehat{\phi}$ MSE	.027	**.016**	.044	.032	**.019**	.042
	(.027)	(.018)	(.008)	(.031)	(.023)	(.005)
Estimation Error	1.01	**1.00**	1.11	1.51	**1.49**	1.61
	(.035)	(.030)	(.038)	(.069)	(.067)	(.070)
Run Time (s)	**1.43**	1.82	5.43	**1.47**	1.89	5.58
	(.091)	(.081)	(.357)	(.084)	(.086)	(.496)
$d=50$	AL	AL-DC	QMLE	AL	AL-DC	QMLE
$\widehat{\phi}$ MSE	.008	**.005**	.036	.010	**.007**	.031
	(.007)	(.005)	(.013)	(.009)	(.007)	(.017)
Estimation Error	.995	**.992**	1.07	1.49	**1.48**	1.56
	(.020)	(.020)	(.047)	(.049)	(.048)	(.073)
Run Time (s)	**6.75**	8.47	40.9	**6.74**	8.43	41.7
	(.136)	(.142)	(1.25)	(.114)	(.090)	(.648)

Table 3: Simulation results for model (6.4) with $\varrho=0.5$ using methods: AL, AL-DC, and QMLE. The mean and standard deviation (in brackets) of the measures over 500 repetitions are reported.

Our adaptive LASSO estimator offers significant computational advantages over QMLE while achieving comparable or superior estimation accuracy. Among all three methods, the adaptive

ϵ_t	i.i.d. $\mathcal{N}(0,1)$			i.i.d. t_6		
$d=25$	AL	AL-DC	QMLE	AL	AL-DC	QMLE
$\widehat{\phi}$ MSE	.033	**.005**	.007	.046	.007	**.003**
	(.033)	(.007)	(.017)	(.044)	(.010)	(.004)
Estimation Error	1.01	**.992**	1.01	1.51	**1.48**	1.50
	(.034)	(.028)	(.049)	(.071)	(.066)	(.070)
Run Time (s)	**1.42**	1.81	5.97	**1.44**	1.84	6.05
	(.081)	(.084)	(.405)	(.078)	(.077)	(.518)
$d=50$	AL	AL-DC	QMLE	AL	AL-DC	QMLE
$\widehat{\phi}$ MSE	.010	.003	**.002**	.014	.004	**.002**
	(.010)	(.003)	(.003)	(.014)	(.004)	(.004)
Estimation Error	.993	**.990**	1.00	1.49	**1.48**	1.49
	(.020)	(.019)	(.034)	(.050)	(.050)	(.053)
Run Time (s)	**6.73**	8.41	39.9	**6.72**	8.40	41.7
	(.105)	(.091)	(1.08)	(.098)	(.078)	(2.34)

Table 4: Simulation results for model (6.4) with $\varrho=0.1$ using methods: AL, AL-DC, and QMLE. The mean and standard deviation (in brackets) of the measures over 500 repetitions are reported.

LASSO estimator without using divide-and-conquer has the shortest run time across all settings; however, its performance deteriorates as the signal weakens. Although the QMLE method is naturally expected to perform well in low signal strength settings due to its likelihood-maximizing nature, it requires an optimization step at each candidate location to maximize the log-likelihood, resulting in high computational costs. We employ the BFGS method for optimization in the simulations and note that the statistical performance of the procedure could possibly be improved, albeit at the expense of even higher computational costs. In contrast, our adaptive LASSO estimator with the divide-and-conquer scheme achieves the best balance between computational efficiency and statistical accuracy.

6.3. A real data example: stock returns

We now use our proposed model to analyze the stock return spillover effects of the largest $d=50$ firms traded on the New York Stock Exchange (NYSE) in 2024 and compare with a selection of benchmark models. The data set comprises 251 days of log-returns, and we use the Fama–French three factors[3], namely the size factor (Rm-Rf), the small-minus-big factor (SMB), and the high-minus-low book-to-market equity factor (HML), as our explanatory variables[4]. Strictly speaking, the Fama–French factors are not exogenous to individual NYSE stock returns, but the issue of endogeneity is fortunately mild. Thus, we treat the factors as exogenous for the convenience of this analysis, leaving further scope for the possible specification of instrumental variables in practice.

We construct three candidate spatial weight matrices: the inverse distance between firm headquarters ($\mathbf{W}_1$), firms belonging to the same sectors according to the GICS classifications ($\mathbf{W}_2$), and firms belonging to the same sub-industries ($\mathbf{W}_3$). If the sum of any row exceeds one, each entry in that row is divided by the L_1 norm of the row. For a thorough comparison, we examine both in-sample and out-of-sample performance for each model. Specifically, the first 100 timestamps are used for the in-sample study, while a rolling window of $T=50$ on the remaining data is used for the out-of-sample study. We consider five measures of performance: (1) in-sample error; (2) in-sample run time; (3) out-of-sample one-step ahead prediction error; (4) out-of-sample two-step ahead prediction error; and (5) annualized Sharpe ratio based on one-step ahead prediction[5].

Inspired by the key feature of volatility clustering in financial data (Chai et al., 2020), we apply our framework in a threshold setting with the threshold variable z_t computed as the average of squared log-returns over $[t-\tau, t-1]$. This can be seen as a sparse threshold autoregressive panel data model. We focus on the case with $\tau=5$, as the model interpretations remain similar for other choices of τ. Moreover, for clearer interpretation, we use only the median of the threshold variable, 0.000197, as the candidate threshold value. For our in-sample study, the model, denoted as DSAR-1, can be written as

$$\textit{log-return}_t = \boldsymbol{\mu}^* + \sum_{j=1}^{3}\Big(\phi^*_{j,0} + \phi^*_{j,1}\mathbb{1}\{z_t \geq 0.000197\}\Big)\mathbf{W}_j\,\textit{log-return}_t + (\textit{Rm-Rf}_t, \textit{SMB}_t, \textit{HML}_t)\boldsymbol{\beta}^* + \boldsymbol{\epsilon}_t.$$

[3]For the log-return at time t, computed as the logarithm of the ratio of prices at t and $t-1$, its factors are taken at time $t-1$.

[4]See more details at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[5]For the annualized Sharpe ratio: at each timestamp, the portfolio position is set as the predicted return based on one-step prediction; the portfolio profit is the average of predicted return times actual return; the annualized Sharpe ratio is calculated as $\sqrt{255}$ times the ratio of sample mean to sample standard deviation of the portfolio profit.

We also consider a simple variant, DSAR-2, in which $\mathbf{W}_2$ is the only spatial weight matrix in the model, and the 0.2, 0.4, 0.6, and 0.8 empirical quantiles are used to construct the dynamic variables. Specifically, the spatial correlation for $\mathbf{W}_2$ is

$$\phi_{2,0}^{*}+\phi_{2,1}^{*}\mathbb{1}\{z_t\geq 0.000166\}+\phi_{2,2}^{*}\mathbb{1}\{z_t\geq 0.000186\}+\phi_{2,3}^{*}\mathbb{1}\{z_t\geq 0.000213\}+\phi_{2,4}^{*}\mathbb{1}\{z_t\geq 0.000236\}.$$

The estimated parameters for our two models DSAR-1 and DSAR-2 are presented in Table 5. The nonzero coefficients suggest the varying spillover effects in response to recent overall market fluctuations. For DSAR-1, it is also worth noting that both coefficients $\widehat{\phi}_{1,0}$ and $\widehat{\phi}_{1,1}$ for $\mathbf{W}_1$ are nonzero. While The co-movement of stock returns based on headquarter distances might seem unconventional, this finding is consistent with Pirinsky and Wang (2006), who attribute such a phenomenon to the distinct trading behaviors of local residents. The estimated coefficients in DSAR-2 reveal three phases of sector spillover effect, which can be interpreted as an increasing co-movement of equity prices within the same sector as the market becomes more active.

	$\widehat{\phi}_{1,0}$	$\widehat{\phi}_{1,1}$	$\widehat{\phi}_{2,0}$	$\widehat{\phi}_{2,1}$	$\widehat{\phi}_{3,0}$	$\widehat{\phi}_{3,1}$	$\widehat{\beta}_{Rm\text{-}Rf}$	$\widehat{\beta}_{SMB}$	$\widehat{\beta}_{HML}$	BIC
DSAR-1	.015	-1.12	-.025	.000	.000	.118	.001	-.000	.001	-3.77
	(.000)	(.000)	(.000)			(.000)				
	$\widehat{\phi}_{2,0}$	$\widehat{\phi}_{2,1}$	$\widehat{\phi}_{2,2}$	$\widehat{\phi}_{2,3}$	$\widehat{\phi}_{2,4}$		$\widehat{\beta}_{Rm\text{-}Rf}$	$\widehat{\beta}_{SMB}$	$\widehat{\beta}_{HML}$	BIC
DSAR-2	.000	.024	.000	.116	.000		.000	-.000	.001	-4.63
		(.000)		(.000)						

Table 5: Estimated coefficients for models DSAR-1 and DSAR-2, with standard errors (in brackets) computed as described in the final part of Section 4. $\widehat{\beta}_{Rm\text{-}Rf}$, $\widehat{\beta}_{SMB}$ and $\widehat{\beta}_{HML}$ denote the estimates of $\boldsymbol{\beta}^{*}$ corresponding to $Rm\text{-}Rf_t$, SMB_t and HML_t, respectively.

We compare our models, DSAR-1 and DSAR-2, with several existing ones: Lam and Souza (2020), which also allows for several spatial weight candidates to be specified but consider only time-invariant weights; Li (2018), which utilizes QMLE within a single change point framework searching potential change locations among $\{0.2T, 0.4T, 0.6T, 0.8T\}$, with each of the three spatial weight matrices experimented on separately; Li and Lin (2024), adapted to panel data, where a threshold spatial autoregressive model is estimated using QMLE, with the aforementioned z_t as the threshold variable and candidates for the threshold value taken from the $\{0.2,0.4,0.6,0.8\}$ empirical quantiles of $\{z_t\}$; and Liang et al. (2022) where a time-varying regression coefficient and semiparametric spatial weights are considered, estimated by the method described therein, again experimenting with each spatial weight matrix respectively. We present the performance results of all those models in Table 6.

Model	In-sample Error*	Run Time (s)	One-step Error*	Two-step Error*	Sharpe Ratio
DSAR-1	4.15	3.75	**2.47**	**2.42**	.545
DSAR-2	1.87	**3.24**	6.68	4.68	**.695**
Lam & Souza	1.94	1,722	3.11	3.30	-.654
Li-$\mathbf{W}_1$	154	20.3	603	1,630	-3.25
Li-$\mathbf{W}_3$	2.30	18.3	9,407	23,134	-1.57
Li & Lin-$\mathbf{W}_1$	1,407	21.9	712	1,447	-1.48
Li & Lin-$\mathbf{W}_3$	401	21.4	24,561	793,957	-1.60
Liang et al.-$\mathbf{W}_1$	1.74	2,087	2.92	2.78	-1.74
Liang et al.-$\mathbf{W}_2$	**1.66**	1,968	2.97	2.82	-.830
Liang et al.-$\mathbf{W}_3$	1.76	1,502	3.00	2.84	-.747

*All errors are multiplied by 10^4

Table 6: Model performance on the return data. Lam & Souza refers to the estimator in Lam and Souza (2020), Li-$\mathbf{W}_j$ denotes that in Li (2018) using $\mathbf{W}_j$; Li & Lin-$\mathbf{W}_j$ denotes that adapted from Li and Lin (2024) using $\mathbf{W}_j$; and Liang et al.-$\mathbf{W}_j$ refers to the estimator in Liang et al. (2022).

We have excluded both Li (2018) and Li and Lin (2024) with spatial weight matrix $\mathbf{W}_2$, as their in-sample errors explode. This may be due to (near) non-stationarity, as large errors are also observed in Li-$\mathbf{W}_1$, Li-$\mathbf{W}_3$, Li & Lin-$\mathbf{W}_1$, and Li & Lin-$\mathbf{W}_3$, suggesting that more sophisticated spatial weight configurations may be necessary for their models to perform adequately.

Overall, both DSAR-1 and DSAR-2 perform very well across all performance measures. In particular, DSAR-1, which uses three different spatial weight matrices, achieves the best predictive power, while DSAR-2 attains smaller in-sample errors. Notably, DSAR-2 achieves the highest Sharpe ratios among all models. We observe that Lam and Souza (2020) and Liang et al. (2022) can achieve comparable or slightly better performance. However, both approaches come with much higher computational complexity, being over 500 times slower than our model fitting. This again highlights the advantage of our proposed approach: achieving strong statistical performance with low computational cost.

7. CONCLUSIONS

In this work, we advance spatial modeling by introducing a framework that enables data-driven selection of time-varying spatial structures, addressing a critical gap in the spatial econometrics literature. By integrating adaptive penalization with instrumental variables, our framework provides practitioners with a principled approach for resolving uncertainty in the specification of spatial weight matrices while detecting structural breaks. Empirical studies demonstrate that our method achieves strong statistical performance at a substantially lower computational cost compared with existing models.

A promising direction for future research is to extend our framework to incorporate time-varying regression coefficients or even spatial autoregressive terms using lagged panels, which would further enhance flexibility. In applied econometrics, our method can provide new insights into evolving spatial systems, with applications ranging from financial contagion to climate risk propagation.

SUPPLEMENTARY MATERIAL

Online supplement. Contains additional mathematical details for Section 2.2, extra simulation results, an additional real data analysis, auxiliary results and all the proofs.

DISCLOSURE STATEMENT

The authors report there are no competing interests to declare.

References

Altieri, L., Scott, E. M., Cocchi, D., and Illian, J. B. (2015), "A changepoint analysis of spatio-temporal point processes," *Spatial Statistics*, 14, 197–207, spatio-Temporal Stochastic Modelling of Environmental Hazards.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Barigozzi, M. and Trapani, L. (2020), "Sequential Testing for Structural Stability in Approximate Factor Models," *Stochastic Processes and their Applications*, 130, 5149–5187.

Beenstock, M. and Felsenstein, D. (2012), “Nonparametric Estimation of the Spatial Connectivity Matrix Using Spatial Panel Data,” *Geographical Analysis*, 44, 386–397.

Bhattacharjee, A. and Jensen-Butler, C. (2013), “Estimation of the Spatial Weights Matrix under Structural Constraints,” *Regional Science and Urban Economics*, 43, 617–634.

Catania, L. and Billé, A. G. (2017), “Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances,” *Journal of Applied Econometrics*, 32, 1178–1196.

Chai, S., Zhang, Z., Du, M., and Jiang, L. (2020), “Volatility Similarity and Spillover Effects in G20 Stock Market Comovements: An ICA-Based ARMA-APARCH-M Approach,” *Complexity*, 2020, 8872307.

Chang, H.-Y., Song, X., and Yu, J. (2025), “Trending Time-Varying Coefficient Spatial Panel Data Models,” *Journal of Business & Economic Statistics*, 43, 191–203.

Deng, Y. (2018), “Estimation for the Spatial Autoregressive Threshold Model,” *Economics Letters*, 171, 172–175.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), “Least Angle Regression,” *The Annals of Statistics*, 32, 407 – 499.

Giordano, F., Niglio, M., and Parrella, M. L. (2024), “Testing Spatial Dynamic Panel Data Models with Heterogeneous Spatial and Regression Coefficients,” *Journal of Time Series Analysis*, 45, 771–799.

Hansen, B. E. (1999), “Threshold effects in non-dynamic panels: Estimation, testing, and inference,” *Journal of Econometrics*, 93, 345–368.

Higgins, A. and Martellosio, F. (2023), “Shrinkage Estimation of Network Spillovers with Factor Structured Errors,” *Journal of Econometrics*, 233, 66–87.

Hollaway, M. J. and Killick, R. (2024), “Detection of spatiotemporal changepoints: a generalised additive model approach,” *Statistics and Computing*, 34.

Hong, H., Ju, G., Li, Q., and Yan, K. X. (2024), “Varying-coefficient Spatial Dynamic Panel Data Models with Fixed Effects: Theory and Application,” *Journal of Econometrics*, 245, 105883.

Kelejian, H. and Prucha, I. (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, 99–121.

Lam, C. and Souza, P. C. (2020), "Estimation and Selection of Spatial Weight Matrix in a Spatial Lag Model," *Journal of Business & Economic Statistics*, 38, 693–710.

LeSage, J. and Pace, R. (2009), *Introduction to Spatial Econometrics*, Boca Raton (USA): Chapman-Hall.

Li, K. (2018), "Spatial Panel Data Models with Structural Change," MPRA Paper 85388, University Library of Munich, Germany.

Li, K. and Lin, W. (2024), "Threshold Spatial Autoregressive Model," *Journal of Econometrics*, 244, 105841.

Liang, X., Gao, J., and Gong, X. (2022), "Semiparametric Spatial Autoregressive Panel Data Model with Fixed Effects and Time-Varying Coefficients," *Journal of Business & Economic Statistics*, 40, 1784–1802.

Liu, W., Xiao, H., and Wu, W. B. (2013), "Probability and Moment Inequalities under Dependence," *Statistica Sinica*, 23, 1257–1272.

Meng, X. and Yang, Z. (2023), "Threshold Spatial Panel Data Models with Fixed Effects," Technical report, Working Paper. Singapore Singapore Management University.

Moradi, M., Cronie, O., Pérez-Goya, U., and Mateu, J. (2023), "Hierarchical Spatio-Temporal Change-Point Detection," *The American Statistician*, 77, 390–400.

Pinkse, J., Slade, M. E., and Brett, C. (2002), "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, 70, 1111–1153, doi: 10.1111/1468-0262.00320.

Pirinsky, C. and Wang, Q. (2006), "Does Corporate Headquarters Location Matter for Stock Returns?" *The Journal of Finance*, 61, 1991–2015.

Sengupta, A. (2017), "Testing for a Structural Break in a Spatial Panel Model," *Econometrics*, 5.

Shao, X. (2010), "Nonstationarity-extended Whittle Estimation," *Econometric Theory*, 26, 1060–1087.

Sun, Y. (2016), "Functional-coefficient Spatial Autoregressive Models with Nonparametric Spatial Weights," *Journal of Econometrics*, 195, 134–153.

Sun, Y. and Malikov, E. (2018), "Estimation and Inference in Functional-coefficient Spatial Autoregressive Panel Data Models with Fixed Effects," *Journal of Econometrics*, 203, 359–378.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.

Tong, H. (1978), "On a Threshold Model in Pattern Recognition and Signal Processing," *Pattern recognition and signal processing*, 575–586.

Wang, H. and Leng, C. (2007), "Unified LASSO Estimation by Least Squares Approximation," *Journal of the American Statistical Association*, 102, 1039–1048.

Wu, W. B. (2005), "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Sciences*, 102, 14150–14154.

Wu, Z., Jiang, W., and Xu, X. (2024), "Applications of Functional Dependence to Spatial Econometrics," *Econometric Theory*, 1–36.

Zhang, Y. and Shen, D. (2015), "Estimation of Semi-parametric Varying-coefficient Spatial Panel Data Models with Random-effects," *Journal of Statistical Planning and Inference*, 159, 64–80.

Zhao, P. and Yu, B. (2006), "On Model Selection Consistency of Lasso," *Journal of Machine Learning Research*, 7, 2541–2563.

Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429.