Estimation and model selection in general spatial dynamic panel data models

Baisuo Jin[a,1], Yuehua Wu[b,1], Calyampudi Radhakrishna Rao[c,d,1], and Li Hou[a]

[a]Department of Statistics and Finance, University of Science and Technology of China, Anhui, Hefei, China 230026; [b]Department of Mathematics and Statistics, York University, Toronto, ON M3J1P3, Canada; [c]Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14221-3000; and [d]C. R. Rao Advanced Institute of Mathematics, Statistics and Computer Science, Hyderabad-500046, India

Contributed by Calyampudi Radhakrishna Rao, November 13, 2019 (sent for review October 14, 2019; reviewed by Ching-Kang Ing and Runze Li)

Commonly used methods for estimating parameters of a spatial dynamic panel data model include the two-stage least squares, quasi-maximum likelihood, and generalized moments. In this paper, we present an approach that uses the eigenvalues and eigenvectors of a spatial weight matrix to directly construct consistent least-squares estimators of parameters of a general spatial dynamic panel data model. The proposed methodology is conceptually simple and efficient and can be easily implemented. We show that the proposed parameter estimators are consistent and asymptotically normally distributed under mild conditions. We demonstrate the superior performance of our approach via extensive simulation studies. We also provide a real data example.

spatial dynamic panel data model | spatial–temporal model | least squares | eigendecomposition | consistency

Spatiotemporal data are common in many areas of science and engineering such as environmental science, epidemiology, economics, and sociology. For illustration, data derived from neighboring geographical locations can exhibit spatial dependence, as can data generated by adjacent nodes of a social network. Modeling such data is necessary but challenging. For example, urban crime count data exhibit clear spatiotemporal patterns (1), and important economic variables across space may be found that account for the concentrations of violence movement. If we can effectively model where and when crime occurs, we can launch better preventative measures. As another example, data collected from Sina Weibo, the largest Twitter-like social network in China, can be better modeled if one leverages user-specific covariates and information about the network structure; good modeling allows us to detect key players in the network, and this knowledge can be used to improve targeted marketing (2). More examples can be found in the literature; e.g., ref. 3 used a time–space dynamic panel data model with spatial moving average errors to study the employment levels across 255 NUTS regions of the European Union over the period 2001 to 2012 in an application in geographical economics; ref. 4 introduced a spatiotemporal model that uses information from nearby, recently sold properties in predicting the value of a given property.

In the following, we use urban crime count data for illustration of the above examples. The dataset was previously analyzed in ref. 1 by using a count model combined with a latent Gaussian spatiotemporal state process. It contains the monthly counts of crimes from January 2008 to December 2013 (72 mo) in the 138 census tracts in Pittsburgh, PA. These counts account for Part I and Part II offences, as defined in the *Uniform Crime Reporting* (UCR) handbook of the US Department of Justice (ref. 5, p. 8). Part I offences consisted of 8 categories of serious felonies and Part II offences were classified into 21 categories of nonserious felonies and misdemeanors. Since the numbers of Part I offences and Part II offences are integers, we apply a logarithmic transformation to them. The transformed data are displayed in Fig. 1 *A* and *B*, respectively. It is interesting to know whether Part II offences contribute to the modeling of Part I offences. To account for heterogeneity across census tracts, following ref. 1, the following data have also been collected from the Census 2000 (US Census Bureau and Social Explorer tables in ref. 1) on the 15 socioeconomic variables, which are total population (Tp), population density per square mile (Pd), median income (Mi), dropout rate age 16 to 19 y (Dra), civilian unemployment rate (Cur), poverty rate (Pvr), percentage of total population under 18 y (U18), group quarter proportion (Gqp), percentage of total population that is African-American (Paa), percentage of population with less than a high-school degree (Hdl), percentage of population with a bachelor's degree or higher (Bdh), rental housing units as percentage of occupied housing units (Rhu), percentage of households having been in the same house for more than 1 y (Sh1), percentage of female-headed households (Fhh), and housing units vacancy rate (Hvr).

Note that there are two types of data above: One type of data is not dependent on time while another type of data is time dependent. As the locations of census tracts may play roles in the modeling, the neighboring relationships can be given via an adjacency matrix $\mathrm{A}=(a_{ij})_{138\times 138}$, which was chosen in ref. 1 as the queen contiguity matrix such that $a_{ii}=0$, $a_{ij}=a_{ji}$, $j\neq i$, and $a_{ij}=1$ if the borders of tract i and j share at least one common point and $a_{ij}=0$ otherwise. As Part I offences at the time t may be dependent on the neighboring Part I offences via a weight matrix built on the adjacency matrix, the previous Part I offences,

Significance

Spatial dynamic panel data modeling is widely used in many areas of science and engineering such as environmental science, epidemiology, economics, and sociology. The ordinary least-squares estimation of parameters of a general spatial dynamic panel data model is inconsistent in general because the spatially lagged dependent variables are typically correlated with the error term. Other estimations based on maximum likelihood or generalized moments are rather complicated. In this paper, we propose an efficient, distribution-free least-squares estimation method that utilizes the eigendecomposition of a weight matrix. We also present a model selection procedure based on the proposed method. Our approach is very powerful compared to the well-known instrumental variable techniques. Its applicability is demonstrated via a high-dimensional data example.

Author contributions: B.J., Y.W., and C.R.R. designed research; B.J., Y.W., and C.R.R. performed research; L.H. analyzed data; and B.J., Y.W., C.R.R., and L.H. wrote the paper.

Reviewers: C.-K.I., National Tsing Hua University; and R.L., Pennsylvania State University.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: crr1@psu.edu, wuyh@mathstat.yorku.ca, or jbs@ustc.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1917411117/-/DCSupplemental.

First published February 24, 2020.

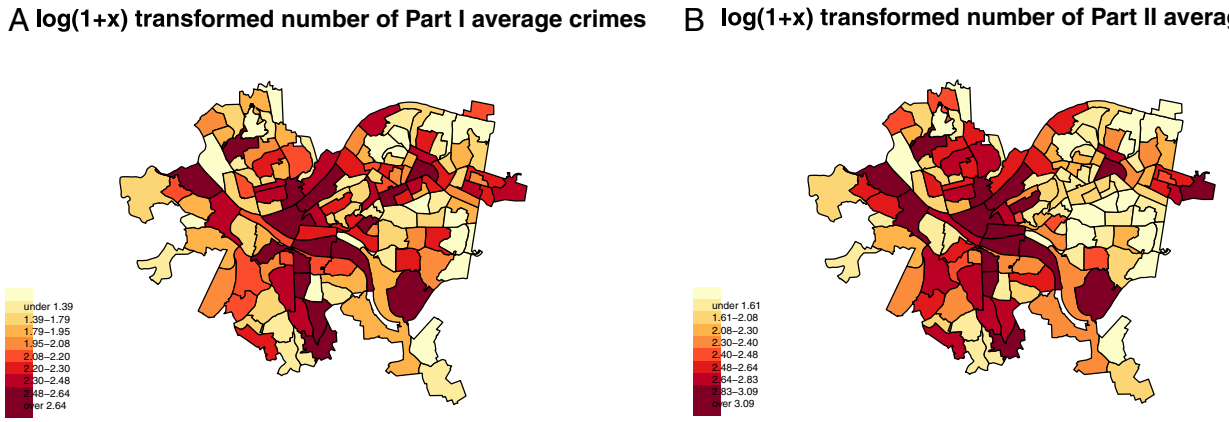

Fig. 1. (*A* and *B*) Time averages of $\log(1+c)$-transformed Part I (*A*) and Part II (*B*) crimes in the 138 Pittsburgh census tracts.

Part II offences at t, the 15 socioeconomic variables, seasonality, etc., it motivates us to consider a general spatial dynamic panel data model that is conceptually simple, efficient, and easily implemented for analyzing spatiotemporal data.

This general spatial dynamic panel data (GSDPD) model has the form

$$\boldsymbol{y}_t = \rho \mathrm{W}_n \boldsymbol{y}_t + \alpha \mathbf{1}_n + Z\boldsymbol{\gamma}_0 + \mathrm{W}_n Z\boldsymbol{\beta}_0 + \boldsymbol{Z}_t\boldsymbol{\gamma} + \mathrm{W}_n\boldsymbol{Z}_t\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \quad t = 1, \ldots, T, \qquad [1]$$

where n is the size of spatial sites; $\mathbf{1}_n = (1, \ldots, 1)^\top$ is an n-dimensional vector; $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})^\top$ is an n-dimensional vector of observations at time t; W_n is an $n \times n$ spatial weight matrix; Z is an $n \times d_0$ design matrix; $\boldsymbol{Z}_t$ is an $n \times d_1$ time-dependent matrix of predictor variables; $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})^\top$, $\varepsilon_{it}, i = 1, \ldots, n, t = 1, \ldots, T$ are independently and identically distributed (iid) random errors with zero means, variance σ^2, and finite fourth moment μ_4; ρ and α are unknown parameters; and $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_0$, and $\boldsymbol{\beta}$ are unknown parameter vectors. The construction of the model Eq. **1** is displayed in Fig. 2. The above GSDPD model includes the classical spatial autoregressive (SAR) model as its special case, which is Eq. **1** with $T = 1$, $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_0 = \mathbf{0}$, and $\boldsymbol{\gamma} = \boldsymbol{\beta} = \mathbf{0}$. For this SAR model, the ordinary least-squares estimation is inconsistent in general, because the spatially lagged dependent variable is typically correlated with the error term (6, 7). To attain the consistent estimation, ref. 7 proposed maximum-likelihood (ML) estimation, combined with a Newton–Raphson procedure to optimize the objective, to estimate ρ.

A more general special case of the GSDPD model is the general first-order serial and spatial autoregressive distributed lag model considered in ref. 8, which can be represented by Eq. **1** with $\alpha = 0$, $\boldsymbol{\gamma}_0 = \boldsymbol{\beta}_0 = 0$, and $\boldsymbol{Z}_t = (\boldsymbol{y}_{t-1}, \boldsymbol{x}_t, \boldsymbol{x}_{t-1})_{n\times 3}$ and has been widely used in practice. The ML method was employed to achieve the consistent estimation in ref. 8. Another spatial dynamic panel data (SDPD) model with fixed effects considered in ref. 9 is also a special case of the GSDPD model with $\alpha = 0$, $\boldsymbol{\beta}_0 = \mathbf{0}$, $\boldsymbol{\beta} = (\rho_1, \mathbf{0}_{d_1-1}^\top)^\top$, $d_0 = n$, $Z = I_n$, and $\boldsymbol{Z}_t = (\boldsymbol{y}_{t-1}, \boldsymbol{X}_t)$, where I_n denotes the $n \times n$ identity matrix. That work also investigated asymptotic properties of the quasi–maximum-likelihood (QML) estimator of the model.

Computing complexity is high for both ML and QMLE since they both need to compute the determinant of the Jacobian matrix which is a nonlinear function of ρ and hence the computation time increases as n increases. This motivated ref. 10 to propose the generalized method of moments (GMM) to estimate the SDPD model. For the same reason, to estimate the spatial Durbin dynamic panel model, another special case of the SDPD, ref. 11 proposed using a combination of two-stage least-squares (2SLS) and QML approaches. We remark that moment functions and instrumental variables need be selected to use the GMM approach and to compute the 2SLS estimates.

In this paper, we propose an approach that targets directly estimation of a GSDPD model by ordinary least squares (OLS), which require neither an iterative algorithm nor having to select moment functions and instrumental variables (IV). To obtain the consistent estimation, we need only to use the eigendecomposition of a spatial weight matrix. There are several major innovations in our approach. First, the proposed estimation is $\sqrt{n(1+d_1)T}$ consistent and asymptotically normally distributed. Second, the proposed estimates have explicit forms and do not need to be iteratively solved, and thus our method is an easy and efficient one. Third, as the spatial weight matrix is conventionally sparse, computation of its eigenvalues and eigenvectors via the Arnoldi and Lanczos algorithms is very fast. Finally, the proposed method can also be applied to select a model from a broader set of models, which, as demonstrated in our simulation studies, outperforms the model selection methods (12) based on the well-known instrumental variables (13) in terms of estimation accuracy and computational speed.

Our main contributions are summarized as follows:

1) Development of a GSDPD model that accounts for many of the classical models as its special cases. The optimal model can be obtained by performing model selection.

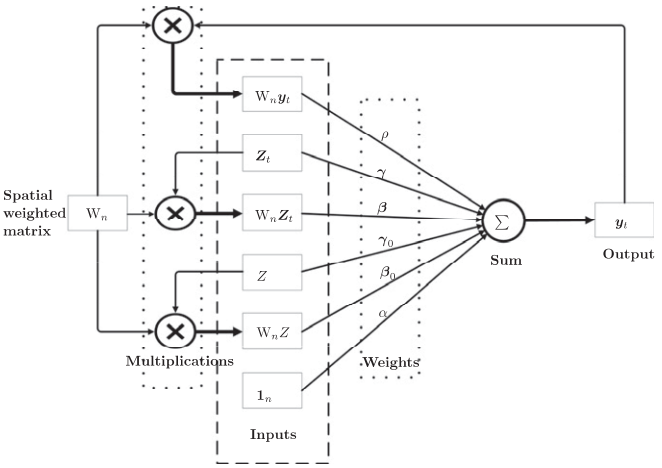

Fig. 2. The construction of the model Eq. **1**.

2) An eigendecomposition-based least-squares (EDLS) estimation method for the general spatial dynamic panel data model in terms of the eigendecomposition of a conventional spatial weight matrix.
3) A model selection method on the general spatial dynamic panel data model based on EDLS.
4) Theoretical analysis of the limiting behavior of both the estimation of EDLS and the proposed model selection method.
5) A real data analysis using the developed method.

Expanded Eigendecomposition-Based Least-Squares Estimation Procedure

The Methodology. Throughout this paper, we consider the scenario where all diagonal elements of W_n are zeros. Denote the eigenvalues and eigenvectors of $\mathrm{W}_n^\top$ by $\lambda_{i,n}$ and $\boldsymbol{\eta}_{i,n}$; i.e., $\mathrm{W}_n^\top \boldsymbol{\eta}_{i,n} = \lambda_{i,n}\boldsymbol{\eta}_{i,n}$, $i = 1, \ldots, n$. We further restrict that $\lambda_{i,n}$, $i = 1, \ldots, n$, are real and are not all equal. This restriction can be justified by the following two common settings. In the first setting, W_n is a symmetric matrix. In the second setting, $\mathrm{W}_n = D\mathrm{A}$ where one of A and D is a symmetric matrix and the other is a positive definite matrix.

Denote the true values of the regression coefficients of the model Eq. **1** as $\boldsymbol{\theta}^o = (\rho^o, \alpha^o, (\boldsymbol{\gamma}_0^o)^\top, (\boldsymbol{\beta}_0^o)^\top, (\boldsymbol{\gamma}^o)^\top, (\boldsymbol{\beta}^o)^\top)^\top = (\theta_1^o, \ldots, \theta_p^o)^\top$, where $p = 2 + 2d_0 + 2d_1$. Throughout the rest of this paper, the superscript "o" is suppressed to simplify notation. We propose the following expanded eigendecomposition-based least-squares estimation (EDLS+) procedure:

Step 1. Left multiply both sides of the model Eq. **1** by $\boldsymbol{\eta}_{i,n}^\top$; i.e.,

$$\boldsymbol{\eta}_{i,n}^\top \boldsymbol{y}_t = \frac{1}{1-\rho\lambda_{i,n}}\Big[\boldsymbol{\eta}_{i,n}^\top \mathbf{1}_n \alpha + \boldsymbol{\eta}_{i,n}^\top Z(\boldsymbol{\gamma}_0 + \lambda_{i,n}\boldsymbol{\beta}_0) + \boldsymbol{\eta}_{i,n}^\top \boldsymbol{Z}_t(\boldsymbol{\gamma} + \lambda_{i,n}\boldsymbol{\beta}) + \boldsymbol{\eta}_{i,n}^\top \boldsymbol{\varepsilon}_t\Big]. \quad [2]$$

Denote $y_{i,t}^* = \boldsymbol{\eta}_{i,n}^\top \boldsymbol{y}_t$, $\boldsymbol{z}_{i,t}^* = (1, \boldsymbol{\eta}_{i,n}^\top \boldsymbol{Z}_t)^\top$, $\boldsymbol{\beta}_i^* = \big(\boldsymbol{\eta}_{i,n}^\top \mathbf{1}_n \alpha + \boldsymbol{\eta}_{i,n}^\top Z(\boldsymbol{\gamma}_0 + \lambda_{i,n}\boldsymbol{\beta}_0), (\boldsymbol{\gamma} + \lambda_{i,n}\boldsymbol{\beta})^\top\big)^\top/(1-\rho\lambda_{i,n})$, and $\varepsilon_{i,t}^* = \boldsymbol{\eta}_{i,n}^\top \boldsymbol{\varepsilon}_t/(1-\rho\lambda_{i,n})$. Eq. **2** can be written as

$$y_{i,t}^* = (\boldsymbol{z}_{i,t}^*)^\top \boldsymbol{\beta}_i^* + \varepsilon_{i,t}^*, i = 1, \ldots, n, \ t = 1, \ldots, T, \quad [3]$$

where $\boldsymbol{z}_{i,t}^* = (z_{i,t,1}, \ldots, z_{i,t,d_1+1})^\top$ with $z_{i,t,1} = 1$. By Eq. **3**, compute the least-squares (LS) estimate of $\boldsymbol{\beta}_i^*$ as follows:

$$\hat{\boldsymbol{\beta}}_i^* = \left(\sum_{t=1}^T \boldsymbol{z}_{i,t}^*(\boldsymbol{z}_{i,t}^*)^\top\right)^{-1} \sum_{t=1}^T \boldsymbol{z}_{i,t}^* y_{i,t}^*. \quad [4]$$

Step 2. Find the Cholesky decomposition $\boldsymbol{\Gamma}_i$, a lower triangular matrix, such that $\boldsymbol{\Gamma}_i^\top \boldsymbol{\Gamma}_i = \frac{1}{T}\sum_{t=1}^T \boldsymbol{z}_{i,t}^*(\boldsymbol{z}_{i,t}^*)^\top$. Define $\boldsymbol{\varsigma}_{i,T} = \sqrt{T}(1-\rho\lambda_{i,n})\boldsymbol{\Gamma}_i(\hat{\boldsymbol{\beta}}_i^* - \boldsymbol{\beta}_i^*)$. Thus,

$$\boldsymbol{\Gamma}_i\hat{\boldsymbol{\beta}}_i^* = \rho\lambda_{i,n}\boldsymbol{\Gamma}_i\hat{\boldsymbol{\beta}}_i^* + (1-\rho\lambda_{i,n})\boldsymbol{\Gamma}_i\boldsymbol{\beta}_i^* + \frac{\boldsymbol{\varsigma}_{i,T}}{\sqrt{T}} = \boldsymbol{U}_i\boldsymbol{\theta}^o + \frac{\boldsymbol{\varsigma}_{i,T}}{\sqrt{T}}, \quad i = 1, \ldots, n, \quad [5]$$

where $\boldsymbol{U}_i = \boldsymbol{\Gamma}_i\left(\lambda_{i,n}\hat{\boldsymbol{\beta}}_i^*, \boldsymbol{B}_i\right)$ and

$$\boldsymbol{B}_i = \begin{pmatrix} \boldsymbol{\eta}_{i,n}^\top \mathbf{1}_n & \boldsymbol{\eta}_{i,n}^\top Z & \lambda_{i,n}\boldsymbol{\eta}_{i,n}^\top Z & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{d_1} & \lambda_{i,n}\boldsymbol{I}_{d_1} \end{pmatrix}_{(1+d_1)\times(p-1)}.$$

Eq. **5** can be rewritten in matrix form as

$$\boldsymbol{v} = \boldsymbol{U}\boldsymbol{\theta}^o + \frac{\boldsymbol{\varsigma}_T}{\sqrt{T}}, \quad [6]$$

where $\boldsymbol{v} = (\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_n^\top)^\top$ with $\boldsymbol{v}_i = \boldsymbol{\Gamma}_i\hat{\boldsymbol{\beta}}_i^*$, $\boldsymbol{U} = (\boldsymbol{U}_1^\top, \ldots, \boldsymbol{U}_n^\top)^\top = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p)$, and $\boldsymbol{\varsigma}_T = (\boldsymbol{\varsigma}_{1,T}^\top, \ldots, \boldsymbol{\varsigma}_{n,T}^\top)^\top$.

Step 3.

3a. If U is of full rank, an estimate of $\boldsymbol{\theta}^o$ is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{EDLS}} = (\boldsymbol{U}^\top\boldsymbol{U})^{-1}\boldsymbol{U}^\top\boldsymbol{v}. \quad [7]$$

3b. Assume that some elements of $\boldsymbol{\theta}^o$ are zeros. An estimate of $\boldsymbol{\theta}^o$ can be obtained by the penalized model selection method

$$\hat{\boldsymbol{\theta}}_{\mathrm{EDLS+}} = \arg\min_{\boldsymbol{\theta}}\left\{\frac{1}{n(1+d_1)}\|\boldsymbol{v} - \boldsymbol{U}\boldsymbol{\theta}\|^2 + \sum_{j=1}^p p_{\zeta,\gamma}(|\theta_i|)\right\}, \quad [8]$$

where $\zeta > 0$, $\gamma > 0$, and the penalty function $p_{\zeta,\gamma}(|x|)$ satisfies the following conditions:

$$p_{\zeta,\gamma}(0) = 0, \ p'_{\zeta,\gamma}(x) = 0 \text{ if } x > \gamma\zeta \text{ and } p'_{\zeta,\gamma}(0) = \zeta. \quad [9]$$

The above conditions are satisfied by the following two penalty functions among others. One is the smoothly clipped absolute deviation (SCAD) penalty defined in ref. 14,

$$p_{\zeta,\gamma}(x) = \zeta x I_{[0,\,\zeta]}(x) + \frac{\gamma\zeta x - 0.5(x^2+\zeta^2)}{\gamma-1} I_{(\zeta,\,\gamma\zeta]}(x) + \frac{\zeta^2(\gamma^2-1)}{2(\gamma-1)} I_{(\gamma\zeta,\,\infty)}(x), \quad x \in [0,\,\infty),$$

and the other one is the minimax concave penalty (MCP) given in ref. 15,

$$p_{\zeta,\gamma}(x) = \zeta x - \frac{x^2}{2\gamma} I_{[0,\,\gamma\zeta]}(x) + \frac{1}{2}\gamma\zeta^2 I_{(\gamma\zeta,\,\infty)}(x), \quad x \in [0,\,\infty).$$

Remark 1: Denote $\mathcal{G} = \{j : \theta_j^o \neq 0, j = 1, \ldots, p\}$ and $\boldsymbol{U}_\mathcal{G} = (\boldsymbol{u}_j, j \in \mathcal{G})$, $\boldsymbol{\theta}_\mathcal{G}^o = (\theta_j^o, j \in \mathcal{G})^\top$. If $\mathcal{G}$ is known and $U_\mathcal{G}$ is of full rank, $\boldsymbol{\theta}^o$ can be estimated by $\hat{\boldsymbol{\theta}}^o$, the oracle estimator, such that $\hat{\boldsymbol{\theta}}_\mathcal{G}^o = (\boldsymbol{U}_\mathcal{G}^\top\boldsymbol{U}_\mathcal{G})^{-1}\boldsymbol{U}_\mathcal{G}^\top\boldsymbol{v}$, and $\hat{\boldsymbol{\theta}}_j^o = 0$ for $j \notin \mathcal{G}$.

Theoretical Justification. Denote the smallest and largest eigenvalues of a matrix Ψ by $\lambda_{\min}(\Psi)$ and $\lambda_{\max}(\Psi)$, respectively.

Denote $\Pi_n = (\boldsymbol{\eta}_{1,n}, \ldots, \boldsymbol{\eta}_{n,n})_{n\times n}$ and

$$\boldsymbol{\Psi}_t = \begin{pmatrix} (\boldsymbol{\Gamma}_1^\top)^{-1}\boldsymbol{z}_{1,t}^* & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Gamma}_2^\top)^{-1}\boldsymbol{z}_{2,t}^* & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & (\boldsymbol{\Gamma}_n^\top)^{-1}\boldsymbol{z}_{n,t}^* \end{pmatrix}_{n(1+d_1)\times n}.$$

Define $\boldsymbol{U}_i^* = \boldsymbol{\Gamma}_i(\lambda_{i,n}\boldsymbol{\beta}_i^*, \boldsymbol{B}_i)$, $\boldsymbol{U}^* = \big((\boldsymbol{U}_1^*)^\top, \ldots, (\boldsymbol{U}_n^*)^\top\big)^\top = (\boldsymbol{u}_1^*, \ldots, \boldsymbol{u}_p^*)$, and $\boldsymbol{U}_\mathcal{G}^* = (\boldsymbol{u}_j^*, j \in \mathcal{G})$. We have the following lemma, which is needed for proving *Theorem 1*.

Lemma 1. *Assume that* $\sqrt{n(1+d_1)/T} \to c$ *with* $0 \le c < \infty$ *and that there exist two positive definite matrices* $\Sigma_\mathcal{G}$ *and* Σ_ς *such that* $(\boldsymbol{U}_\mathcal{G}^*)^\top\boldsymbol{U}_\mathcal{G}^*/(n(1+d_1)) \to_p \Sigma_\mathcal{G}$, $(\boldsymbol{U}_\mathcal{G}^*)^\top\left[\sum_{t=1}^T \boldsymbol{\Psi}_t\Pi_n^\top\Pi_n\boldsymbol{\Psi}_t^\top/T\right]\boldsymbol{U}_\mathcal{G}^*/[n(1+d_1)] \to_p \Sigma_\varsigma$. *If* $\sup_{i,n}|\lambda_{i,n}/(1-\rho\lambda_{i,n})| < w$ *for some finite* $w > 0$, *and* $(1/n)\sum_{i=1}^n \lambda_{i,n}/(1-\rho\lambda_{i,n}) \to m_\rho$ *with* $|m_\rho| < w$ *as* $n \to \infty$, *under the assumptions of SI Appendix, Lemma S-1.1, we have that for* $\Sigma_\theta = \Sigma_\mathcal{G}^{-1}\Sigma_\varsigma\Sigma_\mathcal{G}^{-1}$,

1) *if* $1 \notin \mathcal{G}$, $\sqrt{n(1+d_1)\,T}\left(\hat{\boldsymbol{\theta}}_\mathcal{G}^o - \boldsymbol{\theta}_\mathcal{G}^o\right) \to_d N\left(\mathbf{0}, \sigma^2\Sigma_\theta\right)$;
2) *if* $1 \in \mathcal{G}$,

$$\sqrt{n(1+d_1)T}\left(\hat{\boldsymbol{\theta}}^o_{\mathcal{G}}-\boldsymbol{\theta}^o_{\mathcal{G}}\right)\to_d N\left(c\Sigma^{-1}_{\mathcal{G}}\left(m_\rho\sigma^2,\mathbf{0}\right)^\top,\sigma^2\Sigma_\theta\right).$$

Lemma 1 implies that the asymptotic bias tends to zero at the rate $O(1/T)$ if $1\in\mathcal{G}$. It is noted that the asymptotic bias of the MLE or GMM estimator of the autoregressive parameter in the autoregressive panel data model with random effects tends to zero at the same rate (theorem 1-3 of ref. 16) and so does the asymptotic bias of the QML estimators of the SDPD model with fixed effects (theorem 3 of ref. 9). The proof of *Lemma 1* is given in *SI Appendix*.

Even though $\hat{\boldsymbol{\theta}}_{\text{EDLS+}}$ is not unique in general, the following theorem shows that the oracle estimator $\hat{\boldsymbol{\theta}}^o$ is a solution of Eq. **8** in probability:

Theorem 1. *Suppose that the conditions of Lemma 1 and conditions in Eq.* **9** *hold. If* $\zeta\to 0$, $\sqrt{n(1+d_1)T}\zeta\to\infty$, *and* $\min\{|\theta^o_j|,j\in\mathcal{G}\}>c_0$ *for a finite* $c_0>0$, *then* $\hat{\boldsymbol{\theta}}^o$ *is a solution of Eq.* **8** *in probability.*

The proof of *Theorem 1* is given in *SI Appendix*. Denote $S_n=\frac{1}{n(1+d_1)}\boldsymbol{U}^\top\boldsymbol{U}$ and let ϱ_n be the smallest eigenvalue of S_n. If $\varrho_n+\min_{u>0}\{p''_{\zeta,\gamma}(u)\}>0$, then $Q(\boldsymbol{\theta})=\|\boldsymbol{v}-\boldsymbol{U}\boldsymbol{\theta}\|^2+\sum_{j=1}^p p_{\zeta,\gamma}(|\theta_i|)$ is strictly convex and hence $\hat{\boldsymbol{\theta}}_{\text{EDLS+}}$ is uniquely characterized by the Karush–Kuhn–Tucker (KKT) conditions. Note that $\varrho_n=0$ if $p>n(1+d_1)$. Thus under the condition that $p\le n(1+d_1)$, the following corollary is an immediate consequence of *Theorem 1*:

Corollary 1. *Suppose that the conditions of Theorem 1 are satisfied. Assume that* $p\le n(1+d_1)$ *and* $\varrho_n\to_p c^*$, *where* $c^*>-\min_{u>0}p''_{\zeta,\gamma}(u)$ *is a positive constant. Then*

$$P(A_n)\to 1,\ P(\hat{\boldsymbol{\theta}}_{EDLS+}\neq\hat{\boldsymbol{\theta}}^o|A_n)\to 0,\ \text{as } n\to\infty,\ T\to\infty,$$

where $A_n=\{\varrho_n+\min_{u>0}p''_{\zeta,\gamma}(u)>0\}$.

By *Corollary 1*, the difference between $\hat{\boldsymbol{\theta}}_{\text{EDLS+}}$ and the oracle least-squares estimator $\hat{\boldsymbol{\theta}}^o$ tends to zero in probability, which implies that the proposed EDLS+ procedure is consistent.

Data Examples

Simulations. Let $\mathrm{A}=(a_{ij})_{n\times n}$ be an adjacency matrix such that $a_{ii}=0$, $a_{ij}=a_{ji}$, $j\neq i$. We generate a_{ij}, $i>j$, using Bernoulli distribution $B(1,10/n)$. We define the weight matrix $\mathrm{W}_n=D^{-1}\mathrm{A}$, where $D=\mathrm{diag}(\sum_{j=1}^n a_{1j},\sum_{j=1}^n a_{2j},\ldots,\sum_{j=1}^n a_{nj})$, and each row sum of W_n is scaled to one.

In the model Eq. **1**, let Z be an $n\times 2$ matrix and $\boldsymbol{Z}_t=(\boldsymbol{y}_{t-1},\boldsymbol{X}_t,\boldsymbol{X}_{t-1})$, where Z are generated from the multivariate normal distribution with zero mean vector and covariance matrix I_2, $\boldsymbol{X}_t$ is a $n\times 2$ dimension matrix, $\boldsymbol{X}_t$ are generated from the multivariate normal distribution with zero mean vector and covariance matrix given by $\Sigma_0=(c_{ij})_{2\times 2}$ with $c_{ij}=0.5^{|i-j|}$, and the error terms ε_t are iid from normal distribution $N(0,1)$ or t distribution $t(3)$. Denote $\boldsymbol{\theta}=(\rho,\alpha,\boldsymbol{\gamma}_0^\top,\boldsymbol{\beta}_0^\top,\boldsymbol{\gamma}^\top,\boldsymbol{\beta}^\top)^\top$, whose true value is $\boldsymbol{\theta}^o=(0.2,0.5,0,0,-1.5,2.5,0.3,0,0,0,0,0.5,-1,2,0,0)^\top_{16\times 1}$

Table 1. Performance of EDLS+ and the IV method for estimating θ

		N(0,1)				*t*(3)			
Method	(*n*, *T*)	ρ	MSE	CR	ICR	ρ	MSE	CR	ICR
EDLS+									
Oracle	(50,50)	0.205	0.037	8.000	0.000	0.200	0.039	8.000	0.000
LASSO		0.248	0.094	2.949	0.000	0.246	0.095	2.981	0.002
MCP		0.199	0.060	6.658	0.111	0.200	0.060	6.665	0.097
SCAD		0.198	0.060	6.711	0.112	0.198	0.060	6.693	0.102
OGA+		0.142	0.099	7.945	0.395	0.146	0.094	7.943	0.369
Oracle	(50,100)	0.203	0.018	8.000	0.000	0.200	0.018	8.000	0.000
LASSO		0.231	0.045	2.969	0.000	0.229	0.047	2.869	0.002
MCP		0.203	0.020	7.613	0.011	0.201	0.022	7.590	0.012
SCAD		0.204	0.020	7.603	0.010	0.201	0.022	7.582	0.013
OGA+		0.202	0.021	7.967	0.022	0.199	0.022	7.961	0.030
Oracle	(100,50)	0.210	0.019	8.000	0.000	0.207	0.019	8.000	0.000
LASSO		0.241	0.053	3.165	0.000	0.238	0.050	3.128	0.000
MCP		0.214	0.021	7.690	0.008	0.210	0.021	7.669	0.007
SCAD		0.214	0.021	7.687	0.008	0.210	0.021	7.669	0.007
OGA+		0.213	0.021	7.985	0.013	0.208	0.022	7.981	0.021
IV									
Oracle	(50,50)	0.332	0.104	8.000	0.000	0.327	0.102	8.000	0.000
LASSO		0.379	0.231	3.126	0.001	0.374	0.227	3.091	0.005
MCP		0.377	0.236	7.208	0.429	0.367	0.217	7.107	0.374
SCAD		0.378	0.240	7.207	0.445	0.367	0.218	7.085	0.384
OGA+		0.431	0.479	7.899	1.212	0.425	0.470	7.904	1.185
Oracle	(50,100)	0.330	0.077	8.000	0.000	0.329	0.076	8.000	0.000
LASSO		0.370	0.157	2.951	0.000	0.368	0.155	2.894	0.001
MCP		0.335	0.083	7.414	0.001	0.335	0.084	7.382	0.005
SCAD		0.334	0.082	7.349	0.001	0.334	0.083	7.326	0.005
OGA+		0.471	0.458	7.913	1.000	0.471	0.459	7.918	1.003
Oracle	(100,50)	0.280	0.038	8.000	0.000	0.275	0.036	8.000	0.000
LASSO		0.317	0.097	3.241	0.000	0.312	0.093	3.136	0.002
MCP		0.285	0.041	7.744	0.000	0.280	0.040	7.770	0.003
SCAD		0.285	0.041	7.678	0.000	0.280	0.040	7.693	0.003
OGA+		0.399	0.409	7.967	1.113	0.389	0.410	7.972	1.136

Table 2. Time consumptions of EDLS+ and the IV method

	EDLS+					IV				
(n, T)	(50,100)	(100,50)	(100,100)	(100,200)	(200,100)	(50,100)	(100,50)	(100,100)	(100,200)	(200,100)
LASSO	0.036	0.071	0.071	0.076	0.177	0.520	0.515	2.072	9.198	9.847
MCP	0.069	0.105	0.122	0.135	0.211	0.554	0.547	2.090	9.084	9.802
SCAD	0.089	0.124	0.146	0.165	0.233	0.572	0.565	2.122	9.420	10.172
OGA+	0.069	0.144	0.146	0.145	0.356	4.329	4.158	16.701	73.901	76.564

The entries are the average running times in seconds based on 100 Monte Carlo replications. All computations are performed on the same computer [Intel(R) Core(TM) i7-8700 processor, 4.27 GHz, 12 M caches, 8 GB memory].

in which there are eight nonzero coefficients and $\mathcal{G} = \{1, 2, 5, 6, 7, 12, 13, 14\}$. The sample size (n, T) is chosen respectively as $(50, 50)$, $(50, 100)$, and $(100, 50)$.

In addition to the EDLS+ using both SCAD and MCP penalty functions, some other estimation methods are also considered for estimating $\boldsymbol{\theta}^o$, including the oracle estimator (as $\mathcal{G}$ is known in simulations), the Least Absolute Shrinkage and Selection Operator (LASSO) estimator (17), and the estimator obtained by using the orthogonal greedy algorithm (OGA) plus high-dimensional Hannan–Quinn criterion (HDHQ) plus trimming (TRIM) (OGA+HDHQ+TRIM) (18) that is simplified as OGA+ in this paper. It is noted that OGA is a forward stepwise regression method, HDHQ is used to choose a set of regressors along the OGA path by minimizing HDHQ, and TRIM is to exclude irrelevant variables. We select the tuning parameters in the LASSO, SCAD, and MCP penalty functions by the Bayesian information criterion (BIC).

We also compare the proposed procedure with the IV method, which is a natural generalization of the method introduced in ref. 12 with $T = 1$,

$$\hat{\boldsymbol{\theta}}_{IV} = \arg\min_{\boldsymbol{\theta}} \frac{1}{nT} \sum_{t=1}^{T} \|\boldsymbol{y}_t - \boldsymbol{Z}_t^* \boldsymbol{\theta}\|^2 + \sum_{j=0}^{p} p_{\zeta,\gamma}(|\theta_j|),$$

where $\boldsymbol{Z}_t^* = (\mathrm{H}_t(\mathrm{H}_t^\top \mathrm{H}_t)^{-1}\mathrm{H}_t^\top \mathrm{W}_n \boldsymbol{y}_t, \boldsymbol{X}_t^*)$, $\mathrm{H}_t = \mathrm{W}_n(I - \hat{\rho}\mathrm{W}_n)^{-1}\boldsymbol{X}_t^*$ is an instrumental variable, $\boldsymbol{X}_t^* = (\mathbf{1}_n, Z, \mathrm{W}_n Z, \boldsymbol{Z}_t, \mathrm{W}_n \boldsymbol{Z}_t)$, and $\hat{\rho}$ is estimated by directly using the least-squares method in the model Eq. **1**.

We perform 1,000 Monte Carlo simulations. We report the mean-squared errors (MSE), the average numbers of zero coefficients which are correctly estimated to be zero (CR), and the average numbers of nonzero coefficients that are erroneously set to zero (ICR) for estimating $\boldsymbol{\theta}$, where the MSE is calculated as $MSE(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{1,000} \|\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}^o\|^2 / 1{,}000$. The simulation results are reported in Tables 1 and 2. We can see from Tables 1 and 2 that

1) EDLS+ outperforms the IV method;
2) EDLS+ is much faster than the IV method;
3) The larger the n and T, the better is the performance of all of the methods; and
4) Neither the normal distribution nor the t distribution of the random error has significantly influenced the performance of both methods.

A Real Data Analysis. We go back to the example of modeling Part I offences based on urban crime count data discussed in the beginning of this paper. The detailed logarithmic transformations of Part I offences $c_{it}^{(1)}$ and Part II offences $c_{it}^{(2)}$ in census tract $i \in [1, 138]$ in month $t \in [1, 72]$ are respectively $y_{it}^{(j)} = \log(1 + c_{it}^{(j)})$, $j = 1, 2$. The time plot of the average $\bar{y}_t^{(j)} = \sum_{i=1}^{138} y_{i,t}^{(j)} / 138$ is shown in Fig. 3*A*. The partial autocorrelation functions (PACF) of $\bar{y}_t^{(j)}, j = 1, 2$ are plotted in Fig. 3*B*. From Fig. 3 *A* and *B*, we can assume that $\bar{y}_t^{(1)}$ has a period of 12 mo, and $\bar{y}_t^{(1)}$ at lag 1 is correlated with $\bar{y}_t^{(2)}$. In Fig. 3 *C* and *D*, we display

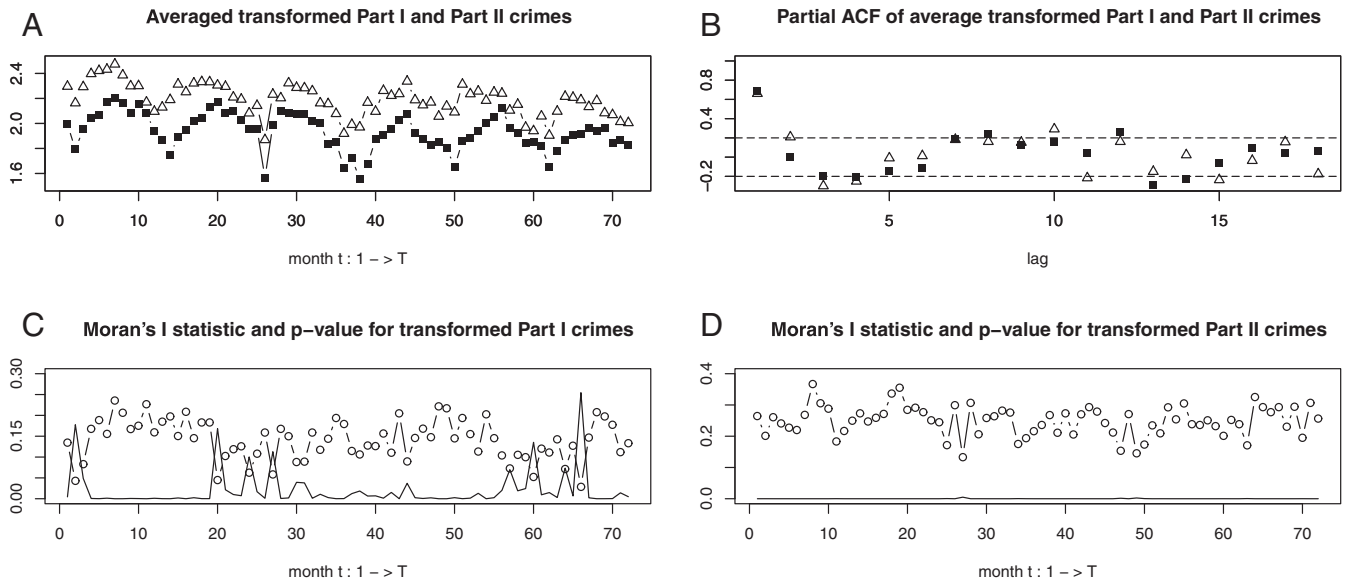

Fig. 3. (*A*) Time plot of transformed Part I crimes (solid squares) and Part II crimes (open triangles) averaged across census tracts. (*B*) Partial autocorrelations averaged across census tracts. (*C*) Time plot of Moran's I (diamonds) and *P* value (solid line) for transformed Part I crimes. (*D*) Time plot of Moran's I (diamonds) and *P* value (solid line) for transformed Part II crimes. ACF, autocorrelation function.

Table 3. Estimates of the regression coefficients by EDLS+

	Without $\boldsymbol{y}_t^{(2)}$				With $\boldsymbol{y}_t^{(2)}$			
Regressors	LASSO+	MCP	SCAD	OGA+	LASSO+	MCP	SCAD	OGA+
$W_n\boldsymbol{y}_t^{(1)}$	1.68e-01**	2.71e-01***	2.82e-01***		5.84e-02*			
$\boldsymbol{y}_{t-1}^{(1)}$	5.71e-01***	5.81e-01***	6.79e-01***	6.59e-01***	4.12e-01***	4.22e-01***	4.14e-01***	4.48e-01***
Tp	1.47e-04***	1.52e-04***	9.02e-05***	1.26e-04***	4.44e-05***	3.41e-05***	4.37e-05***	
Dp	−8.26e-06*	−9.86e-06*	−5.53e-07		−5.83e-06*	−2.94e-06	−5.69e-06	
Mi	−5.96e-07	−1.02e-06 .	−1.01e-06		−1.94e-06***	−2.17e-06**	−1.95e-06*	−1.46e-06***
Gqp	−5.64e-01***	−5.49e-01***		−4.25e-01***				
Bdh					4.96e-01***	5.47e-01***	5.07e-01***	6.38e-01***
Rhu	4.43e-01***	4.68e-01***		6.05e-01***	1.77e-01**		1.85e-01***	
Hvr	4.06e-01**							
W_nTp	−8.32e-05**	−1.01e-04***	−7.47e-05**		−3.52e-05 .	−4.38e-05*	−3.52e-05.	
W_nDp	3.04e-06	3.20e-07	3.23e-06		8.34e-06 .	1.29e-05**	9.90e-06*	
W_nMi	2.93e-06*	9.77e-07	1.39e-06		4.27e-07	8.64e-07	6.21e-07	
W_nDra				2.74e-01***				
W_nGqp	3.58e-01*	5.57e-01**						
W_nBdh	−3.50e-01*							
W_nRhu	8.14e-02							
$\boldsymbol{y}_t^{(2)}$					3.68e-01***	3.78e-01***	3.67e-01***	4.21e-01***
$W_n\boldsymbol{y}_t^{(2)}$						5.85e-02**	3.98e-02*	
$\cos(\frac{2\pi t}{12})\mathbf{1}_n$	−6.61e-02***	−5.51e-02***		−8.10e-02***	−3.69e-02**	−3.39e-02*	−3.81e-02**	
$\sin(\frac{2\pi t}{12})\mathbf{1}_n$					−3.17e-02*	−3.49e-02*	−3.65e-02**	
R^2	0.9795	0.9789	0.9745	0.9762	0.9856	0.9853	0.9856	0.9842
σ	0.1455	0.1472	0.1614	0.1556	0.1087	0.1097	0.1088	0.1130
P value	0.0025	0.0026	0.0012	0.0012	0.2818	0.1228	0.1905	0.3945
AIC	−1.3482	−1.3397	−1.2647	−1.2823	−1.4531	−1.4426	−1.4486	−1.4259
BIC	−1.3372	−1.3309	−1.2589	−1.2779	−1.4436	−1.4337	−1.4390	−1.4229

Significance values: ***P = 0.001, **P = 0.01, *P = 0.05. P value is computed by a two-sided Kolmogorov–Smirnov test where null hypothesis is that the residuals are normality.

Moran's I statistics (19) and P values under the null hypothesis of no spatial correlation between $y_{i,t}^{(1)}$ and $y_{i,t}^{(2)}$ at each time t, which are calculated using the spatial weight matrix $\mathrm{W}_n = D^{-1}\mathrm{A}$, where $\mathrm{A} = (a_{ij})$ is the queen contiguity matrix chosen by following ref. 1, and $D = \mathrm{diag}(\sum_{j=1}^n a_{1j}, \sum_{j=1}^n a_{2j}, \ldots, \sum_{j=1}^n a_{nj})$ so that each row sum of W_n is scaled to one. By these two plots, it can be seen that $\{y_{it}^{(1)}\}$ and $\{y_{it}^{(2)}\}$ are clearly spatially correlated.

Denote $\boldsymbol{y}_{t-1}^{(1)} = (y_{1,t-1}^{(1)}, \ldots, y_{138,t-1}^{(1)})^\top$ and $\boldsymbol{y}_t^{(2)} = (y_{1,t}^{(2)}, \ldots, y_{138,t}^{(2)})^\top$. To model $\boldsymbol{y}_t^{(1)} = (y_{1,t}^{(1)}, \ldots, y_{138,t}^{(1)})^\top$ by a GSDPD model, we let

$$\boldsymbol{Z}_t = \left(\boldsymbol{y}_{t-1}^{(1)}, \boldsymbol{y}_t^{(2)}, \cos(2\pi t/12)\mathbf{1}_{138\times 1}, \sin(2\pi t/12)\mathbf{1}_{138\times 1}\right)_{138\times 4},$$

$Z_{138\times 15}$ be the 15 socioeconomic variables over the 138 census tracts and the weight matrix be W_n given above. As there are 13 socioeconomic variables having a total of 83 missing values for 13 census tracts, we impute them by the medians of the corresponding socioeconomic variables. Thus, for this GSDPD model, $d_0 = 15$, $d_1 = 4$ so that the number of regression coefficients is equal to $p = 2 + 2d_0 + 2d_1 = 40$, which implies that it is necessary to perform model selection for this model.

To find a strong confirmation of the "broken-windows" phenomenon (20), we compare the differences in modeling of $\{y_{i,t}^{(1)}\}$ without or with $\{y_{i,t}^{(2)}\}$. For the former one, we accordingly replace $\boldsymbol{Z}_t$ by $\boldsymbol{Z}_t^*$ by deleting the second column of $\boldsymbol{Z}_t$.

Since the bias in estimating $\boldsymbol{\theta}$ by LASSO is large compared to others, we modify this approach by first using LASSO for performing model selection and then using OLS to estimate the regression coefficients in the selected model, which we denote by LASSO+. We report the modeling results by EDLS+ in Table 3. In Table 3, it can be seen that in terms of R^2, residual SE σ, the P value of the Kolmogorov–Smirnov (KS) normality test,

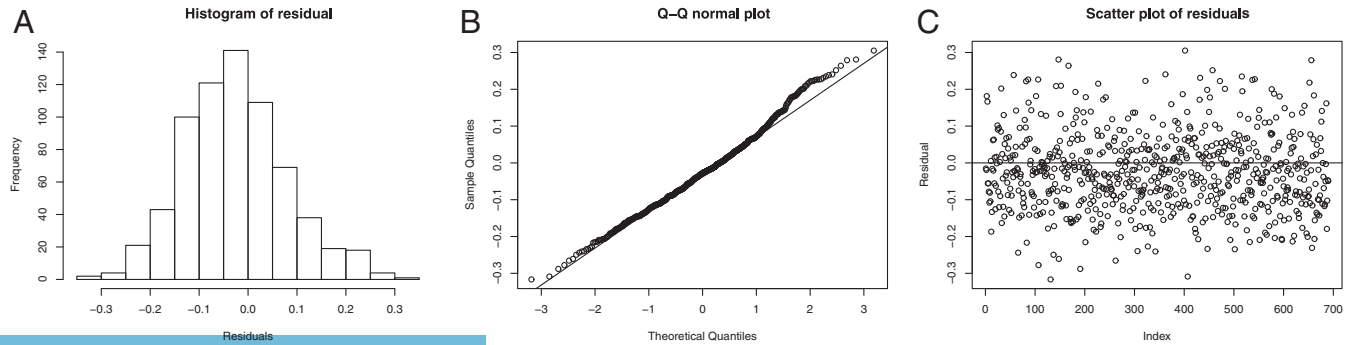

Fig. 4. (*A–C*) Residual analysis of LASSO+ with $\boldsymbol{y}_t^{(2)}$.

the Akaike information criterion (AIC), and the BIC strongly support including $\boldsymbol{y}_t^{(2)}$, as it is shown to have a significant and positive effect on $\boldsymbol{y}_t^{(1)}$. From Table 3, it can also be observed that the spatial lag also has a positive effect on $\boldsymbol{y}_t^{(1)}$. All these observations are fully in accordance with the broken-windows theory (20). In addition, Table 3 shows that LASSO+ with $\boldsymbol{y}_t^{(2)}$ performs the best as it has not only the smallest residual SE σ, AIC, and BIC values but also the largest R^2. The residual analysis of LASSO+ with $\boldsymbol{y}_t^{(2)}$ is displayed in Fig. 4, which indicates that the distribution of residuals is approximately normally distributed.

A further examination of Table 3 shows that both $\cos\left(\frac{2\pi t}{12}\right)\mathbf{1}_n$ and $\sin\left(\frac{2\pi t}{12}\right)\mathbf{1}_n$ are significant, which implies that $\boldsymbol{y}_t^{(1)}$ is periodic with a period of 12 mo. It can also be observed that the population size (Tp) has a significantly positive effect on $\{\boldsymbol{y}_{it}^{(1)}\}$ while both population density per square mile (Pd) and median income (Mi) have significantly negative effects on it. These results are in agreement with those reported in the literature suggesting that concentrations of violence typically occur in disadvantaged communities and regions with a large population size (21, 22). Finally, Table 3 reveals that the percentage of population with a bachelor's degree or higher (Bdh) has a significantly positive effect on $\{\boldsymbol{y}_{it}^{(1)}\}$, which is in line with the result of ref. 1.

ACKNOWLEDGMENTS. We thank Dr. Zhidong Bai for his helpful suggestions. B.J.'s research is partially supported by the National Natural Science Foundation (Grants 71873128, 11571337, 71631006, 71921001). Y.W.'s research is partially supported by the Natural Sciences and Engineering Research Council of Canada (Grant RGPIN-2017-05720).

1. R. Liesenfeld, J. F. Richard, J. Vogler, Likelihood-based inference and prediction in spatio-temporal panel count models for urban crimes. *J. Appl. Econom.* **32**, 600–620 (2017).
2. X. Zhu, R. Pan, G. Li, Y. Liu, H. Wang, Network vector autoregression. *Ann. Stat.* **45**, 1096–1123 (2017).
3. B. H. Baltagi, B. Fingleton, A. Pirotte, A time-space dynamic panel data model with spatial moving average errors. *Reg. Sci. Urban Econ.* **76**, 13–31 (2019).
4. R. K. Pace, R. Barry, J. M. Clapp, M. Rodriquez, Spatiotemporal autoregressive models of neighborhood effects. *J. Real Estate Finance Econ.* **17**, 15–33 (1998).
5. US Federal Bureau Of Investigation, *UCR: Uniform Crime Reporting Handbook* (US Department of Justice, Federal Bureau of Investigation, Washington, DC, 2004). https://lccn.loc.gov/2004483104. Accessed 12 February 2020.
6. P. Whittle, On stationary processes in the plane. *Biometrika*, **41**, 434–449 (1954).
7. K. Ord, Estimation methods for models of spatial interaction. *J. Am. Stat. Assoc.* **70**, 120–126 (1975).
8. J. P. Elhorst, Dynamic models in space and time. *Geogr. Anal.* **33**, 119–140 (2001).
9. J. Yu, R. De Jong, L. F. Lee, Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *J. Econom.* **146**, 118–134 (2008).
10. L. F. Lee, J. Yu, Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *J. Econom.* **180**, 174–197 (2014).
11. L. F. Lee, J. Yu, Identification of spatial Durbin panel models. *J. Appl. Econom.* **31**, 133–162 (2016).
12. T. Xie, R. Cao, J. Du, Variable selection for spatial autoregressive models with a diverging number of parameters. *Stat. Pap.*, 10.1007/s00362-018-0984-2 (2018).
13. T. W. Anderson, C. Hsiao, Estimation of dynamic models with error components. *J. Am. Stat. Assoc.* **76**, 598–606 (1981).
14. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **96**, 1348–1360 (2001).
15. C. H. Zhang, Nearly unbiased variable selection under minimax concave penalty. *Ann. Stat.* **38**, 894–942 (2010).
16. J. Alvarez, M. Arellano, The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* **71**, 1121–1159 (2003).
17. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
18. C. K. Ing, T. L. Lai, A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Stat. Sin.* **21**, 1473–1513 (2011).
19. P. A. Moran, The interpretation of statistical maps. *J. R. Stat. Soc. B* **10**, 243–251 (1948).
20. J. Q. Wilson, G. L. Kelling, Broken windows. *Atl. Mon.* **249**, 29–38 (1982).
21. R. D. Baller, L. Anselin, S. F. Messner, G. Deane, D. F. Hawkins, Structural covariates of US county homicide rates: Incorporating spatial effects. *Criminology* **39**, 561–588 (2001).
22. M. Helbich, J. Jokar Arsanjani, Spatial eigenvector filtering for spatiotemporal crime mapping and spatial crime analysis. *Cartogr. Geogr. Inf. Sci.* **42**, 134–148 (2015).