Semiparametric Spatial Autoregressive Panel Data Model with Fixed Effects and Time-Varying Coefficients

Xuan Liang*, Jiti Gao*, and Xiaodong Gong*

*Research School of Finance, Actuarial Studies and Statistics, The Australian National University, Canberra, Australia; **Monash University, Melbourne, Australia; ***University of Canberra and IZA, Canberra, Australia

ABSTRACT
This article considers a semiparametric spatial autoregressive (SAR) panel data model with fixed effects and time-varying coefficients. The time-varying coefficients are allowed to follow unknown functions of time, while the other parameters are assumed to be unknown constants. We propose a local linear quasi-maximum likelihood estimation method to obtain consistent estimators for the SAR coefficient, the variance of the error term, and the nonparametric time-varying coefficients. The asymptotic properties of the proposed estimators are also established. Monte Carlo simulations are conducted to evaluate the finite sample performance of our proposed method. We apply the proposed model to study labor compensation in Chinese cities. The results show significant spatial dependence among cities and the impacts of capital, investment, and the economy's structure on labor compensation change over time.

ARTICLE HISTORY
Received November 2019
Accepted September 2021

KEYWORDS
Concentrated quasi-maximum likelihood estimation; Local linear estimation; Time-varying coefficient

1. Introduction
Panel data analysis has been widely used in many fields of social sciences as it usually enables strong identification and increases estimation efficiency. A comprehensive review about these methodologies can be found in Arellano (2003), Baltagi (2008), and Hsiao (2014). To model spatial dependence, spatial econometric models have drawn a lot of attention as they provided a way to model the cross-sectional dependence with a clear structure and intuitive interpretations.

A class of spatial autoregressive (SAR) models was first proposed in Cliff and Ord (1973). Since then, spatial econometrics has become an active research area in econometrics. One issue with spatial econometric models is that the spatial lag term is endogenous. Various estimation methods have been proposed to deal with this issue, for example, the instrumental variable (IV) method (Kelejian and Prucha 1998), the generalized method of moments (GMM) framework (Kelejian and Prucha 1999) and the quasi-maximum likelihood (QML) method (Lee 2004). More details of spatial econometrics can be found in classic spatial econometrics books, for example, Anselin, Florax, and Rey (2004) and LeSage and Pace (2009). As more temporal data became available, spatial panel data models have received considerable attentions. Spatial panel data models with SAR disturbances are considered in Baltagi, Song, and Koh (2003) and Kapoor, Kelejian, and Prucha (2007). Fingleton (2008) studies a spatial panel data model with a SAR-dependent variable and a spatial moving average-disturbance. Lee and Yu (2010) focus on a spatial panel model with individual fixed effects. More recent studies on spatial dynamic panel data models can be found in Yu, De Jong, and Lee (2008), Lee and Yu (2014) and Li (2017).

A common feature of the aforementioned models is that they are fully parametric with a linear form in regressors, which may lead to model misspecification. To enhance model flexibility, nonparametric and semiparametric spatial econometric models have been studied in the literature. Su and Jin (2010) considered a partially linear SAR model. Su (2012) proposed an SAR model with a nonparametric regressor term. Functional-coefficient SAR models are also studied in Sun (2016) and Malikov and Sun (2017). These studies are all about cross-sectional data. In terms of nonparametric and semi-parametric spatial econometric models for panel data, Zhang and Shen (2015) considered a partially linear SAR panel data model with functional coefficients and random effects. Sun and Malikov’s (2018) study a functional-coefficient SAR panel data model with fixed effects. It is worth noting that they focus on the case of large N and finite T. In addition, the functional coefficients in these spatial models are mostly permitted to be unknown smooth functions of some exogenous variables. Sometimes, finding such appropriate exogenous variables in practice is challenging.

It has been noted that especially when the time span of data is long, coefficients of covariates are likely to change over time in many real examples (see discussions in Cai 2007; Silvapulle et al. 2017). The reasons for such changes may include changes in the economic structure or environment, policy reform, or technology development. To accommodate such cases, time-varying coefficient models have been studied extensively in the
Panel data setting, where the regressor coefficients are allowed to be unknown smooth functions of time (e.g., Li, Chen, and Gao 2011; Chen, Gao, and Li 2012; Robinson 2012). One advantage of time-varying coefficient models is that the time variable can be self-explanatory and naturally captures the nonlinear time variation in the coefficients. To the best of our knowledge, time-varying coefficient models have not been studied in spatial econometrics.

In this article, we propose a semiparametric time-varying coefficient spatial panel data model with fixed effects for large $N$ and large $T$. Specifically, the coefficient of the spatial lag term in the model is assumed to be constant over time, while the regressor coefficients vary with time. In addition, the regressors can be trending nonstationary. To obtain consistent estimators for both the parametric components and nonparametric time-varying coefficients, we propose a local linear concentrated quasi-maximum likelihood estimation (LLQML) method. When the time-varying coefficients are constant and the regressors are stationary, our model reduces to a fully parametric SAR panel data model that has been considered in Lee and Yu (2010). Our model allows only the coefficients of the explanatory variables to be time-varying. A more general setting with a nonparametric coefficient for the spatial lag term would be less restrictive. The model studied in Sun and Malikov (2018) has a functional-coefficient for the spatial lag term but for fixed $T$. We would like to leave more general models for future research.

Our contributions in this article are summarized as follows.

(i) We propose a semiparametric time-varying coefficient spatial panel data model. This model is suitable for panel data with spatial interaction and time-varying features, as it combines the strengths from different models, including strong identification of panel data models, clear interpretation of cross-sectional dependence in spatial econometric models, and the flexibility of time-varying coefficient models. In the existing literature of spatial econometrics, the regressors are often assumed to be nonstochastic (see, e.g., Lee and Yu 2010; Su and Jin 2010). We relax such assumptions so that the regressors can be trending non-stationary, which renders our model and method of estimation more general and practically useful.

(ii) Since the model consists of both unknown parametric and nonparametric components, we propose the LLQML method to consistently estimate the parameters and the unknown time-varying functions by incorporating the local linear estimation (Fan and Gijbels 1996) into the QML estimation. We also establish the consistency and asymptotic normality of the proposed estimators.

(iii) We evaluate the finite-sample performance of our proposed model under several scenarios. We find that our estimators are consistent and robust against different model specifications, including not only time-varying models or non-stationary covariates, but also time-invariant models and stationary covariates. The results also show that if the time-varying coefficients are misspecified as constants, it would lead to severely inconsistent estimation.

(iv) As an empirical application of our model, we analyze the time-varying effects of factors on labor compensation in urban China over 1995–2009, a period that has seen continuous reforms and dramatic changes in the economy. Consistent with our conjecture, the estimated effects show quite strong time-varying features.

The rest of the article is organized as follows. Section 2 discusses the model setting and the estimation procedure. Section 3 lays out the assumptions. Asymptotic properties of the proposed estimators are established in Section 4. We report the results of Monte Carlo simulations and of the empirical application in Sections 5 and 6, respectively. In Section 7, we conclude. Appendix A provides the justification of the identification condition and then gives the proofs of the main theorems. Technical lemmas and their proofs as well as additional numerical results are given in Appendices B–D of the supplementary material.

2. Model Setting and Estimation

2.1. Model

The model we consider in this article takes the following form:

$$Y_{it} = \rho_0 \sum_{j \neq i} w_{ij} Y_{jt} + X_{it}^\top \beta_{0,t} + \alpha_{0,i} + e_{it},$$

$$t = 1, \ldots, T, \quad i = 1, \ldots, N,$$

where $Y_{it}$ is the response of location $i$ at time $t$, $X_{it} = (X_{i1}, \ldots, X_{itd})^\top$ is the $d$-dimensional covariate, $\beta_{0,t} = (\beta_{0,1,t}, \ldots, \beta_{0,d,t})^\top$ is the corresponding $d$-dimensional time-varying coefficient, $\alpha_{0,i}$ reflects the unobserved individual fixed effect, the error component $e_{it}$ has mean zero and variance $\sigma^2_0$, and $T$ and $N$ are the time length and the number of spatial units, respectively. In this model, $w_{ij}$ describes the spatial weight of location $j$ to $i$, which can be a decreasing function of spatial distance between $i$ and $j$. The scalar parameter $\rho_0$ is a measure of the strength of spatial dependence. Hence, the term $\rho_0 w_{ij} Y_{jt}$ captures the spatial interaction and $X_{it}^\top \beta_{0,t}$ measures the covariate effects over time.

When $\beta_{0,t}$ does not vary over time, it reduces to a vector of constants. In this case, model (1) reduces to the traditional SAR panel data model discussed in Lee and Yu (2010). If only some components of $\beta_{0,t}$ change over time, model (1) takes the form of a partially time-varying spatial panel data model, which means that a few covariates have effects changing over time while the effects of other covariates stay constant. In this article, we assume that $\beta_{0,t}$ is fully nonparametric and follows the specification

$$\beta_{0,t} = \beta_0(t), \quad t = 1, \ldots, T,$$

where $\beta_0(\cdot)$ is a $d$-dimensional vector of unknown smooth functions defined on $\mathbb{R}^d$ and $T = t/T \in (0, 1]$. The same specification is used in Li, Chen, and Gao (2011) and Chen, Gao, and Li (2012). The reason to rescale time onto the interval $(0, 1]$ is for convenience when estimating the model with the kernel method.

For the purpose of identifying $\beta_0(t)$ when the constant 1 is included in $X_{it}$, the individual fixed effects are assumed to satisfy $\sum_{i=1}^N \alpha_{0,i} = 0$. Such a condition is standard in the literature, for example, Su and Ullah (2006) and Chen, Gao, and Li (2012) with its justification provided in Appendix A.1.
Let $\mathbf{0}_n$ and $\mathbf{1}_n$ be the vectors with $n$ elements of zeros and ones, respectively. Let $\mathbf{0}_{m_1 \times m_2}$ denote an $m_1 \times m_2$ matrix with all zero elements and let $I_m$ denote an $m$-dimensional identity matrix. Define an $N \times N$ spatial weight matrix $W = (w_{ij})_{N \times N}$ with zero diagonal elements, that is, $w_{ii} = 0$, an $N \times (N - 1)$ matrix $D_0 = (-\mathbf{1}_{N-1}, \mathbf{I}_{N-1})^T$. A clear matrix form of model (1) can be written as

$$Y_t = \rho_0 W Y_t + X_t \beta_0(\tau_t) + D_0 \alpha_0 + \epsilon_t, \quad t = 1, \ldots, T, \quad (3)$$

where $Y_t = (Y_{1t}, \ldots, Y_{Nt})^T$, $X_t = (X_{1t}, \ldots, X_{Nt})^T$, $\alpha_0 = (\alpha_{01}, \ldots, \alpha_{0N})^T$ and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})^T$. Define an $N \times N$ matrix $S_N(\rho) = I_N - \rho W$. Model (3) can further be written as

$$S_N(\rho_0) Y_t = X_t \beta_0(\tau_t) + D_0 \alpha_0 + \epsilon_t. \quad (4)$$

In (4), we move the spatial lag term ($\rho_0 W Y_t$) to the left side so that $S_N(\rho_0) Y_t$ would be regarded as the response variable as if $\rho_0$ were known. The goal is to construct consistent estimators for the spatial coefficient $\rho_0$, the variance $\sigma^2$, and the unknown time-varying coefficient function $\beta_0(\tau)$.  

2.2. Estimation

The joint quasi log-likelihood function of model (4) can be written as

$$\log L_{N,T}(\rho, \sigma^2, \alpha, \beta(\cdot)) = -\frac{NT}{2} \log(2\pi \sigma^2) + T \log |S_N(\rho)|$$

$$- \frac{1}{2\sigma^2} \sum_{t=1}^T U_t^T U_t, \quad (5)$$

where $U_t = S_N(\rho_0) Y_t - D_0 \alpha \times X_t \beta(\tau_t)$ and $\beta(\cdot)$ is a smooth function on $\mathbb{R}^d$. If $\beta(\tau)$ is a vector of constants, the model (4) becomes fully parametric so that the traditional QML method based on Equation (5) can be used to estimate the parameters (see Lee 2004; Lee and Yu 2010 for more details). In the presence of the nonparametric time-varying component $\beta(\tau)$ in Equation (5), the traditional QML would fail. Motivated by Su and Ullah (2006) and Su and Jin (2010), we propose the LLQML method, which is a two-step procedure: (i) Estimate $\hat{\beta}_0(\tau)$ for fixed $\rho$ and $\alpha$ by the local linear kernel method (Fan and Gijbels 1996) and denote it as $\hat{\beta}_{\rho,\alpha}(\tau)$; (ii) Plug $\hat{\beta}_{\rho,\alpha}(\tau)$ into (5), and obtain the QML estimators $\hat{\rho}, \hat{\sigma}^2$ and $\hat{\alpha}$. With $\rho$ and $\alpha$ estimated, the estimator of $\beta_0(\tau)$ can then be updated by $\hat{\beta}_{\rho,\hat{\alpha}}(\tau)$. To be more specific:

Step 1. Assume that $\beta_0(\cdot)$ has continuous derivatives of up to the second order. Let $K(\cdot)$ and $h$ be the kernel function and the smoothing bandwidth, respectively. Applying the Taylor expansion to $\beta_0(\tau_t)$ at $\tau$ such that $|\tau_t - \tau| = o(1)$, we obtain $\beta_0(\tau_t) = \beta_0(\tau) + \beta_0'(\tau)(\tau_t - \tau) + O((\tau_t - \tau)^2)$, where $\beta_0'(\cdot)$ is the first derivative of $\beta_0(\cdot)$ and $\tau \in [0, 1]$. We also have that $X_t \beta_0(\tau_t) \approx X_t \beta_0(\tau) + h \beta_0''(\tau)$. According to the local linear kernel method, the estimators of $\beta_0(\tau)$ and $h \beta_0''(\tau)$ for given $(\rho, \alpha)$ can be obtained by minimizing the weighted loss function $L(a, b)$ with respect to $(a^T, b^T)^T$, that is,

$$\begin{pmatrix} \hat{\beta}_{\rho,\alpha}(\tau) \\ h \hat{\beta}''_{\rho,\alpha}(\tau) \end{pmatrix} = \arg \min_{(a^T, b^T)^T} L(a, b),$$

where

$$L(a, b) = \sum_{t=1}^T K \left( \frac{\tau_t - \tau}{h} \right) \left[ S_N(\rho) Y_t - D_0 \alpha - X_t \beta_0(\tau) \right]^T$$

$$\times \left[ S_N(\rho) Y_t - D_0 \alpha - X_t \beta_0(\tau) \right].$$

Define an $NT$-dimensional vector $Y = (Y_1^T, \ldots, Y_T^T)^T$ and an $NT \times NT$ matrix $S_N(T, \rho) = I_T \otimes S_N(\rho)$, where $\otimes$ denotes the Kronecker product. Further denote an $NT$-dimensional vector $Y(\rho) = S_N(T, \rho) Y$ and an $NT \times (N - 1)$ matrix $D = I_T \otimes D_0$. Function $L(a, b)$ can be rewritten as

$$L(a, b) = \left\{ Y(\rho) - D \alpha - M(\tau)(a^T, b^T)^T \right\}^T \Omega(\tau) \times \left\{ Y(\rho) - D \alpha - M(\tau)(a^T, b^T)^T \right\},$$

where the $NT \times 2d$ matrix $M(\tau)$ and the $NT \times NT$ matrix $\Omega(\tau)$ are defined as follows:

$$M(\tau) = \begin{pmatrix} X_1 & X_1 \\ \vdots & \vdots \\ X_T & X_T \end{pmatrix}$$

and

$$\Omega(\tau) = \begin{pmatrix} K \left( \frac{\tau - \tau}{h} \right) I_N & \cdots \\ \vdots & \cdots \end{pmatrix}.$$ 

Taking the first derivative of $L(a, b)$ with respect to $(a^T, b^T)^T$ and equating it to zero, we obtain

$$\begin{pmatrix} \hat{\beta}_{\rho,\alpha}(\tau) \\ h \hat{\beta}''_{\rho,\alpha}(\tau) \end{pmatrix} = \left\{ M(\tau) \Omega(\tau) M(\tau) \right\}^{-1} M(\tau) \Omega(\tau) \times \left\{ Y(\rho) - D \alpha \right\}. \quad (6)$$

Denoting a $d \times NT$ matrix $\Phi(\tau) = (I_d, 0_{d \times NT}) \left( M(\tau) \Omega(\tau) M(\tau) \right)^{-1} M(\tau) \Omega(\tau)$, the estimator of the time-varying coefficient $\beta_0(\cdot)$ can be expressed by

$$\hat{\beta}_{\rho,\alpha}(\tau) = \Phi(\tau) \left( Y(\rho) - D \alpha \right).$$

Step 2. In this step, we plug $\hat{\beta}_{\rho,\alpha}(\tau)$ into the original log-likelihood (5) and estimate $\rho_0$ and $\sigma^2$ by maximizing the quasi log-likelihood function:

$$\log L_{N,T}(\rho, \sigma^2, \alpha)$$

$$= -\frac{NT}{2} \log(2\pi \sigma^2) + T \log |S_N(\rho)|$$

$$- \frac{1}{2\sigma^2} \sum_{t=1}^T \left[ S_N(\rho) Y_t - X_t \hat{\beta}_{\rho,\alpha}(\tau_t) - D_0 \alpha \right]^T$$

$$\times \left[ S_N(\rho) Y_t - X_t \hat{\beta}_{\rho,\alpha}(\tau_t) - D_0 \alpha \right]$$

$$= -\frac{NT}{2} \log(2\pi \sigma^2) + T \log |S_N(\rho)|$$

$$- \frac{1}{2\sigma^2} \sum_{t=1}^T \left[ \hat{Y}(\rho) - \hat{\alpha} \right]^T \left[ \hat{Y}(\rho) - \hat{\alpha} \right],$$

where $\hat{Y}(\rho) = (I_{NT} - S)Y(\rho)$ and $\hat{D} = (I_{NT} - S)D$ are the smoothing versions of $Y(\rho)$ and $D$ by the $NT \times NT$ matrix $S = \hat{X} \Phi$, in which the $NT \times dT$ matrix $\hat{X}$ is a diagonal block.
matrix with $X_t$ being its $r$th diagonal block, and $dT \times NT$ matrix $\tilde{\Phi} = (\phi(t_1))^{-1}, \ldots, (\phi(t_T))^{-1})^{-1}$. Taking the derivative of (7) with respect to $\alpha$ and setting it to be zero, we have

$$\tilde{\alpha}(\rho) = (\tilde{D}^T\tilde{D})^{-1}\tilde{D}^T\tilde{Y}(\rho).$$

Define an $NT \times NT$ matrix $Q_{N,T} = I_{NT} - \tilde{D}(\tilde{D}^T\tilde{D})^{-1}\tilde{D}^T$. Plugging $\tilde{\alpha}(\rho)$ into (7) leads to

$$\log L_{N,T}(\rho, \sigma^2) = -\frac{NT}{2}\log(2\pi \sigma^2) + T\log|\Sigma_{N}(\rho)| - \frac{1}{2\sigma^2}\tilde{Y}(\rho)Q_{N,T}\tilde{Y}(\rho).$$

(8)

Then, taking the derivative of Equation (8) with respect to $\sigma^2$ and equating it to zero, we have the estimator of $\sigma^2_0$ as the following function of $\rho$:

$$\hat{\sigma}^2(\rho) = \frac{1}{NT}\tilde{Y}(\rho)Q_{N,T}\tilde{Y}(\rho).$$

Replacing $\sigma^2$ with $\hat{\sigma}^2(\rho)$ in Equation (8), we obtain the concentrated quasi log-likelihood function:

$$\log L_{N,T}(\rho) = -\frac{NT}{2}\left\{\log(2\pi) + 1\right\} - \frac{NT}{2}\log\left\{\frac{1}{NT}\tilde{Y}(\rho)Q_{N,T}\tilde{Y}(\rho)\right\} + T\log|\Sigma_{N}(\rho)|.$$

Therefore, we estimate the parameters $\theta_0 = (\rho_0, \sigma^2_0)^T$ and $\alpha_0$ by $\hat{\theta} = (\hat{\rho}, \hat{\sigma}^2_0)^T$ and $\tilde{\alpha}$ as follows:

$$\hat{\rho} = \arg\max_{\rho} \log L_{N,T}(\rho), \quad \hat{\sigma}^2 = \frac{1}{NT}\tilde{Y}(\hat{\rho})Q_{N,T}\tilde{Y}(\hat{\rho}),$$

$$\tilde{\alpha} = (\tilde{D}^T\tilde{D})^{-1}\tilde{D}^T\tilde{Y}(\hat{\rho}).$$

Finally, the updated estimator of $\beta_0(\tau)$ is obtained by plugging $\hat{\rho}$ and $\tilde{\alpha}$ into Equation (6):

$$\tilde{\beta}(\tau) = \Phi(\tau)(\tilde{Y}^*(\hat{\rho}) - D\tilde{\alpha}).$$

(9)

In order to establish asymptotic properties for the proposed estimators, we need to introduce the following assumptions.

### 3. Model Assumptions

In this section, we lay out the assumptions for our model. Denote $||a||_s = \left(\sum_{n=1}^{\infty} |a|^s\right)^{\frac{1}{s}}$ as the $s$-norm ($s \geq 1$) for any generic vector $a = (a_1, \ldots, a_n)^T$. For any generic $m \times m$ matrix $A = (a_{ij})_{m \times m}$, define the diagonal vector of $A$ as $\text{diag}(A) = (a_{11}, \ldots, a_{nn})^T$, $||A|| = \max |a_{ij}|$ and $||A||_\infty = \max |a_{ij}|$ as the row sum norm and column sum norm, respectively.

**Assumption 1.** Let the $d$-dimensional vector $X_{it} = g(t) + \epsilon_{it}$ contain a deterministic time trend $g(t) = \left(g_1(t), \ldots, g_d(t) \right)^T$ and a random component $\epsilon_{it} = (\epsilon_{it1}, \ldots, \epsilon_{itd})^T$. If

(i) Suppose that $g(t)$ is a continuous function for any $0 < t \leq 1$.

(ii) Denote $\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tn})^T$. Suppose that $\{\epsilon_t, t \geq 1\}$ is a strictly stationary sequence with mean zero and $\alpha$-mixing with mixing coefficient $\alpha_{\text{mix}}(n)$, and that there exists a function $\alpha_{\text{mix}}(n)$ and a constant $\delta$ such that $\alpha_{\text{mix}}(n) \leq \alpha_{\text{mix}}(n)$ and $\sum_{n=1}^{\infty} \alpha_{\text{mix}}(n)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$.

(iii) Let $v_{it}$ be identically distributed in index $i$ for any given $t$. In addition, we assume $E|\epsilon_{itk}|^{4+\delta} < \infty$ for $k = 1, \ldots, d$ and let $E(\epsilon_{itk}) = 0$ for $k = 1, \ldots, d$ and $\sigma_{\epsilon_{itk}}^2 = E(\epsilon_{itk}^2)$.

**Remark:** **Assumption 1** is a list of assumptions about the $d$-dimensional explanatory variable $X_{it}$.

**Assumption 2.** The error term $\{\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tn})^T : t \geq 1\}$ is a stationary process such that

(i) for some $\delta > 0$, $\sup_{|\delta|} E(\epsilon_{itk})^{4+\delta} < \infty$;

(ii) $E(\epsilon_{itk}) = 0_N$ and $E(\epsilon_{itk}^*|\mathcal{F}_{t-1}) = \sigma_{\epsilon_{itk}}^2 I_n$, where $\mathcal{F}_{t-1} = \mathcal{F}_{t-1}^\vee \vee \sigma(\epsilon_{11}, \ldots, \epsilon_{itk-1})$, is the $\sigma$-field generated by $\mathcal{F}_{t} \cup \{\epsilon_{itk} : i \geq 1, t \geq 1\}$; and $\mathcal{F}_{t}^\vee \sigma(\{\epsilon_{itk} : i \geq 1, t \geq 1\})$ is the $\sigma$-field generated by $\{\epsilon_{itk} : i \geq 1, t \geq 1\}$;

(iii) Given $\mathcal{F}_{t-1} = \tilde{\epsilon}_t = (\epsilon_{it1}, \ldots, \epsilon_{itd})^T$ is a vector of conditionally independent random errors with $E(\tilde{\epsilon}_t|\mathcal{F}_{t-1}) = E(\tilde{\epsilon}_t) = m_{\epsilon}$, almost surely for $j = 3$ and 4.
Remark: Assumption 2 summarizes the conditions on the error term. Assumption 2(ii) ensures that the first and second conditional moments of $e_i$ given $E_{-1}$ do not depend on the information generated by $\{v_t; i \geq 1, t \geq 1\}$. In particular, Assumption 2(ii) implies $E(e_i|F_{-1}) = E(E(e_i|E_{-1})|F_{-1}) = 0$. Note that if $\{e_i\}$ and $\{v_t\}$ are independent of each other, this result naturally holds for the mean zero error vector $e_i$. Assumption 2(iii) assumes the conditional independence of $e_{it}$ (i = 1, . . . , N) given $E_{t-1}$ as well as its conditional moments. A sufficient condition for Assumption 2(ii) and (iii) is that $e_i$ and $v_t$ are independent in both i and t (e.g., see Assumption 2 of Yu, De Jong, and Lee 2008) and $\{e_i\}$ is independent of $F_{-1}$. It is worth noting that the conditional independence of $e_{it}$ in Assumption 2(iii) along with Assumption 2(ii) can help form a martingale difference array in both i and t in the theoretical derivations; see, for example, the proof of Theorem 2 in Appendix A.2. Further, this technique of the proof can be adapted to model (3) if a cross-sectional dependent random structure is specified. For example, we can still impose Assumption 2 but we replace $e_i$ in model (3) by a cross-sectional dependent random error $e_{it} = \epsilon_{it}$, where $L$ is a non-stochastic matrix and $E(e_{it}e_{it}^T|E_{t-1}) = \sigma^2_{LL}I_N$ can measure the cross-sectional dependence. If we assume that $L$ is uniformly bounded in both row and column sums in absolute value (analogously to Assumption 4), similar theoretical results can be established but more technical derivations are involved.

Assumption 3. The kernel function $K(\cdot)$ is a continuous and symmetric probability density function with compact support. The bandwidth is assumed to satisfy $h \rightarrow 0$ as $min(N, T) \rightarrow \infty$, $Th \rightarrow \infty$ and $NTth^8 \rightarrow 0$.

Remark: Assumption 3 first imposes the conditions on the kernel function, which is common in the literature; see, for example, Chen, Gao, and Li (2012). Conditions on the bandwidth $h$ along with $T$ and $N$ are also considered; see similar conditions in Assumption A5 of Chen, Gao, and Li (2012).

Assumption 4. To be more specific on the dimension of the $N \times N$ spatial weight matrix $W$, we denote $W$ as $W_N$ in this assumption. We assume $W_N$ is a non-stochastic matrix with zero diagonals and is uniformly bounded in both row sum norm and column sum norm (for short, UB), that is, $sup_{N \geq 1} \| W_N \|_1 < \infty$ and $sup_{N \geq 1} \| W_N \|_{\infty} < \infty$.

Assumption 5. $S_N(\rho)$ is invertible for all $\rho \in \Delta$, where $\Delta$ is a compact interval with the true value $\rho_0$ as an interior point. Also, $S_N(\rho)$ and $S_N^{-1}(\rho)$ are both UB, uniformly in $\rho \in \Delta$.

Remark: Assumptions 4 and 5 are standard assumptions originated from Kelejian and Prucha (1998, 2001) and also used in Lee (2004). When $W$ is row-normalized, a compact subset of ($-1, 1$) has often been taken as the parameter space for $\rho$. The UB conditions limit the spatial correlation to a manageable degree. To save space, we refer readers to Kelejian and Prucha (2001) for more discussions.

Assumption 6. The time-varying coefficient $\beta_0(\cdot)$ has continuous derivatives of up to the second order.

Assumption 7. The fixed effects satisfy that $\|D_0\alpha_0\|_1 < \infty$.

Remark: Assumption 6 is a mild condition on the smoothness of the functions which is required by the local linear fitting procedure. Such an assumption is common for nonparametric estimation methods, for example, Li and Racine (2007, condit. 2.1), Gao (2007, assump. 2.7) and Assumption A3 of Chen, Gao, and Li (2012). Assumption 7 guarantees the uniform boundedness of the sum of absolute fixed effects.

To proceed, we need to introduce the following notations. Let $S_N = S_N(\rho_0)$, $S_{N,T} = S_{N,T}(\rho_0)$, $G_N(\rho) = W_N^{-1}(\rho)$, $G_N = G_N(\rho_0)$, $G_{N,T} = G_{N,T}(\rho_0)$, $P_N = (NT - S)^TQ_N(NT - S)$ and $R_{N,T} = G_{N,T}e_0 + D\sigma_0$ where $\beta_0 = (\beta_0(\tau_1), \ldots, \beta_0(\tau_T))^T$.

Assumption 8. $\Psi_{RR} = \lim_{NT \rightarrow \infty} \frac{1}{NT}E(R_{N,T}^T P_{N,T} R_{N,T}) > 0$.

Remark: Assumption 8 is a condition for the identification of $\rho_0$, which is similar to Lee (2004, assump. 8), Lee and Yu (2010, assump. 4), and Su and Jin (2010, assump. 7). To interpret Assumption 8, we rewrite $\Psi_{RR} = \lim_{NT \rightarrow \infty} \frac{1}{NT} E(R_{N,T}^T Q_{N,T} R_{N,T}^T)$, where $R_{N,T} = (NT - S)R_{N,T}$. Similar to the explanation of Assumption 7 in Su and Jin (2010), $R_{N,T}$ can be interpreted as the remainder of $R_{N,T}$ deviated from the time-varying projection onto $X^\tau \hat{\theta}(\tau)$. In practice, one may use the estimate of $\Psi_{RR}$ to check this assumption.

4. Asymptotic Properties

Asymptotic consistency of $\hat{\theta} = (\hat{\rho}, \hat{\sigma}^2)^T$ to $\theta_0 = (\rho_0, \sigma_0^2)^T$ is established in Theorem 1. The asymptotic distributions of $\hat{\theta}$ and $\hat{\theta}(\tau)$ are provided in Theorems 2 and 3. The proofs of these theorems are given in Appendix A.2.

Theorem 1. Under Assumptions 1–8, $\theta_0$ is globally identifiable and $\hat{\theta}$ is consistent to $\theta_0$.

Denote $c_1 = \lim_{NT \rightarrow \infty} \text{tr}(G^2_{N,T} + G^T_{N,T}G_{N,T})/NT$, $c_2 = \lim_{NT \rightarrow \infty} \text{tr}(G_{N,T}/NT)$ where the existence proofs of the limits are shown in Lemma C.7 of the supplementary material.

Theorem 2. Under Assumptions 1–8, as $T \rightarrow \infty$ and $N \rightarrow \infty$ simultaneously, then

$$\sqrt{NT}(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0_2, \Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0} \Omega_{NT\theta_0} \Sigma^{-1}_{\theta_0}),$$

(10)

where $\Omega_{\theta_0} = \lim_{NT \rightarrow \infty} \Omega_{NT\theta_0}$ with $\Omega_{NT\theta_0}$ being defined by

$$
\Omega_{NT\theta_0} = \frac{2m_3E(R_{N,T}^T P_{N,T} \text{diag}(P_{N,T} G_{N,T}))}{m_2E(R_{N,T}^T P_{N,T} \text{diag}(P_{N,T}))} + \frac{(m_4 - 3m_3^2)E(\sum_{i=1}^{NT} (g_i p_i))}{2NTm_2^2} + \frac{2m_2E(R_{N,T}^T P_{N,T} \text{diag}(P_{N,T}))}{m_2E(R_{N,T}^T P_{N,T} \text{diag}(P_{N,T}))} + \frac{(m_4 - 3m_3^2)E(\sum_{i=1}^{NT} (g_i p_i))}{2NTm_2^2}.
$$
in which $p_i$ and $(gp)_i$ are the $i$th main diagonal elements of $P_{N,T}$ and $P_{N,T}G_{N,T}$, respectively, and $\Sigma_{\theta_0} = \begin{pmatrix} \frac{1}{\sigma^2_0} \psi_{RR} + c_1 \frac{\sigma_2^2}{\sigma^2_0} & \frac{c_1}{\sigma_0} \\ \frac{c_1}{\sigma_0} & \frac{2c_1}{\sigma_0} \end{pmatrix}$ is positive definite as shown in Lemma C.9.

Since we use the QML method to estimate $\theta_0$, it relaxes the normality assumption on the error term but adds an additional term to the variance that is a function of the error term's third and fourth moments. If the third and fourth moments are satisfied with $m_3 = 0$ and $m_4 = 3\sigma^2_0$, the asymptotic covariance matrix in Equation (10) reduces to $\Sigma_{\theta_0}^{-1}$, as shown in the following proposition.

**Proposition 1.** Let Assumptions 1–8 hold. Then as $T \to \infty$ and $N \to \infty$ simultaneously

$$\sqrt{NT}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \Sigma^{-1}\right)$$

when $\{e_{it}, i \geq 1, t \geq 1\}$ is independent and identically normally distributed with $\Omega_{\theta_0} = 0_{2 \times 2}$ due to $m_3 = 0$ and $m_4 = 3\sigma^2_0$.

Define $\mu_1 = \int u'K(u)du$ and $v_1 = \int u'K^2(u)du$. Let $\beta_0''(\tau)$ be the second derivative of $\beta_0(\tau)$. The asymptotic distribution of $\hat{\beta}(\tau)$ is established in the following theorem.

**Theorem 3.** Let Assumptions 1–8 hold. As $T \to \infty$ and $N \to \infty$ simultaneously, we have

$$\sqrt{NT}(\hat{\beta}(\tau) - \beta_0(\tau) - \beta(\tau)h^2 + o_p(h^2)) \xrightarrow{d} N\left(0_2, \sigma^2_0^{-1}\Sigma_{X}(\tau)\right),$$

(11)

provided that $\Sigma_{X}(\tau)$ is positive definite for each given $\tau$ in $(0, 1)$, where $\beta(\tau) = \frac{1}{\sigma^2_0} \psi_{RR}''(\tau)$ and $\Sigma_{X}(\tau) = g(\tau)g(\tau)\tau^2 + \nu v$.

Thus, the rate of convergence of $\hat{\beta}(\tau)$ is $\sqrt{NT}$, which is standard in the nonparametric estimation. It is also clear that the covariance matrix is related to $g(\tau)$ since it involves the trend of $X_t$. When $X_t$ is stationary, the asymptotic covariance matrix in (11) reduces to a constant matrix $\sigma^2_0 v_0(\mu \mu' + \Sigma_\nu)$, where $\mu_X = E(X_{it})$.

One can use the following sample version to estimate the unknown covariance matrices involved: $\hat{\Sigma}_{\theta_0} = \begin{pmatrix} \frac{1}{\sigma^2_0} \hat{\psi}_{RR} + \hat{c}_1 \frac{\sigma_2^2}{\sigma^2_0} & \frac{\hat{c}_1}{\sigma_0} \\ \frac{\hat{c}_1}{\sigma_0} & \frac{2\hat{c}_1}{\sigma_0} \end{pmatrix}$, $\hat{\Omega}_{\theta_0} = \Omega_{NT,\theta_0}$ and $\hat{\Sigma}_{X}(\tau) = \hat{g}(\tau)\hat{g}(\tau)\tau^2 + \hat{\nu}_v$, where $\hat{\psi}_{RR} = (NT)^{-1}R_{N,T}P_{N,T}R_{N,T}$, $\hat{c}_1 = \text{tr}(G_{N,T}^2 + C_{N,T}G_{N,T})/NT$, $\hat{c}_2 = \text{tr}(G_{N,T}^2)/NT$, $\hat{g}(\tau) = \sum_{l=1}^L K(\frac{m-1}{h}u_l)\hat{\gamma}_l$, $\hat{\nu}_v = (NT)^{-1} \sum_{l=1}^L \hat{\gamma}_l'\hat{\nu}_l^2$ and $\hat{\nu}_l = X_{it} - \hat{g}(\tau_l)$. The consistency of these sample estimators is shown in Lemma C.10 in the supplementary material.

### 5. Monte Carlo Simulations

We now conduct a number of simulations to evaluate the finite sample performance and the robustness of our proposed model and estimation under a rich set of scenarios, which are different in stationarity of the covariates, variation in time of the coefficients, and the degree of spatial dependence.

The simulated data are generated from the following model:

$$Y_t = \rho_0 W Y_t + X_t \beta_0(\tau_t) + D_t \alpha_0 + \epsilon_t, \quad t = 1, \ldots, T.$$ 

The data generating process for our simulation is summarized below. First, the spatial matrix $W$ in the data generating process is chosen as a "q step head and q step behind" spatial weights matrix as in Kelejian and Prucha (1999) with $q = 2$ in this section. The procedure is as follows: all the units are arranged in a circle and each unit is affected only by the $q$ units immediately before it and immediately after it with the weight being 1. Then following Kelejian and Prucha (1999), we normalize the spatial weights matrix by letting the sum of each row equal to 1 so that it generates an equal weight influence from all the neighboring units to each unit. Then, the regressor is set to be $X_{it} = (1, X_{it2})^T$ where $X_{it2} = g(\tau_t) + v_{it2}$ with $v_{it2}$ is the $i$-th element of an $N$-dimensional vector $v_i$ generated by $v_i = 0.2v_{i-1} + N(0_{Nt}, \Sigma^*)$ with $\Sigma^* = (0.5^{i-j})_{N, N}$ for $-99 \leq t \leq T$ and $v_{i,-100} = 0_N$. It is obvious that $v_{it2}$ is both serially and cross-sectionally dependent. The error term $\epsilon_{it}$ is independent and identically distributed from the generation of $N(0, 1)$ so that $\sigma^2_{\epsilon} = 1$. The fixed effects follow $\alpha_{it} = T^{-1} \sum_{t=1}^T v_{it2}$ for $i = 1, \ldots, N$ and $\alpha_{i,1} = -\sum_{t=2}^N \alpha_{it}$. The time-varying coefficient vector is set to be $\beta_0(\tau_t) = (\beta_{01}(\tau_t), \beta_{02}(\tau_t))^T$ where $\beta_{01}(\tau_t)$ and $\beta_{02}(\tau_t)$ represent the time-varying coefficients associated with the constant 1 and $X_{it2}$ in $X_{it}$, respectively. Various simulation settings are defined by choosing the specification of $g(\tau)$, $\beta_0(\tau)$ and $\rho_0$. Specifically, we consider the following scenarios:

- **Set I** (Setting of $g(\tau)$): (I-1) $g(\tau) = 0$; (I-2) $g(\tau) = 1$ and (I-3) $g(\tau) = 2 \sin(\pi \tau)$;
- **Set II** (Setting of $\beta_0(\tau)$): (II-1) $\beta_0(\tau) = (1, 1)^T$; (II-2) $\beta_0(\tau) = (1 + 3\tau, 1 + 2\tau + 2\tau^2)^T$; (II-3) $\beta_0(\tau) = (1 + 3\tau, 1 + 2\tau + 2\tau^2)^T$;
- **Set III** (Setting of spatial coefficient): (III-1) $\rho_0 = 0.3$, (III-2) $\rho_0 = 0.7$.

Each of these sets (and combinations of them) will generate data of (i) covariates of different stationarity (Set I): in Sets I-1 and I-2 $X_{it2}$ is stationary and in Set I-3 $X_{it2}$ is non-stationary; (ii) coefficient $\beta_0(\tau)$ with different time-varying features (Set II): from Set II-1 to Set II-3, $\beta_0(\tau)$ changes from time-invariant, partially time-varying to fully time-varying respectively; and (iii) different SAR coefficient or spatial dependence among cross-sectional units (Set III). For each scenario, simulations are conducted on 1000 replications. The Epanechnikov kernel $K(u) = 3/4(1-u^2)$ if $|u| \leq 1$ is used where $I(\cdot)$ is the indicator function. The bandwidth is selected through a leave-one-unit-out cross-validation method explained in Appendix A.3.

The simulated data are first estimated by our proposed model and estimation method, and then estimated by a standard time-invariant spatial panel data model considered in Lee and Yu (2010) and their proposed estimation. In short, we call it the "Lee-Yu model". Tables 1 and 2 report the means and standard deviations (SDs) (in parentheses) of the bias for the estimates of our model for $\rho_0$ and $\sigma^2_\epsilon$ under different settings of $g(\tau)$ and
\(\beta_0(\tau)\), together with those of the Lee-Yu model (with \(\rho_0\) fixed at 0.3, that is, Set III-1). A few comments can be made on the results.

First, our estimates of \(\rho_0\) and \(\sigma_\tau^2\) are consistent under all settings as the means and SDs of the bias of \(\rho_0\) and \(\sigma_\tau^2\) are getting smaller when either \(N\) or \(T\) is increasing. It shows the robustness of our model against different settings on the covariates and time-varying coefficients.

Second, if the data are generated by a time-invariant process (Set II-1), the estimates of \(\rho_0\) and \(\sigma_\tau^2\) from the Lee-Yu model are consistent with smaller biases compared to ours. It makes sense as a time-invariant spatial panel data model is a special case of our model. However, when the coefficients of the covariates have time-varying features (Set II-2 and Set II-3) in the data generating process, the estimates of \(\rho_0\) and \(\sigma_\tau^2\) from the Lee-Yu model are not consistent and exhibit large biases. For example, under the combination of Set I-2 and Set II-2, the biases are around 0.27 for \(\rho_0\) and 1.9 for \(\sigma_\tau^2\). When there are more coefficients having time-varying features, (e.g., from Set II-2 to II-3), the biases become larger. These findings confirm that when the time-varying model is misspecified as a time-invariant model, following the estimation of the Lee-Yu model will lead to inconsistent estimates.

Third, comparing different data-generating processes, if the data are generated from a fully time-varying model (Set II-3), our estimates have the smallest biases and SDs, followed by a partially linear model (Set II-2) and then a time-invariant model (Set II-1) given the setting of \(X_{it2}\). For example, when \(N = 15\), \(T = 15\) and \(X_{it2}\) follows Set I-3, the means and SDs of biases of our estimates for \(\rho_0\) are \(-0.0301 (0.665), -0.0148 (0.0297), -0.0086 (0.0212)\), respectively, from Set II-1 to Set II-3.

We also evaluate our model by examining the finite-sample performance of the estimators for the two time-varying coefficients \(\beta_{0,1}(\tau)\) and \(\beta_{0,2}(\tau)\). In Table 3, the means and SDs (in parentheses) of the mean squared errors (MSEs) of the estimates of the time-varying coefficients \(\beta_{0,1}(\tau)\) and \(\beta_{0,2}(\tau)\) are reported, where for an estimator \(\hat{\beta}_k(\tau)\) \((k = 1, 2)\), the MSE is defined by

\[
\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{\beta}_k(\tau_t) - \beta_{0,k}(\tau_t) \right)^2.
\]

The results show that both the means and SDs of MSE for \(\hat{\beta}_k(\tau)\) \((k = 1, 2)\) decrease when either \(N\) or \(T\) increases, confirming the
consistency of our estimators. The results for a different spatial coefficient $\rho_0 = 0.7$ (Set III-2) are similar to our benchmark case of Set III-1 ($\rho_0 = 0.3$), and the corresponding simulation results can be found in Tables D.1-D.3 of the supplementary material.

### 6. Empirical Application

As a case study, we apply our model to analyze the level of real wages in 159 Chinese cities over the period between 1995 and 2009. The level of real wages measures the demand for labor and is closely related to productivity; see for example Van Biesebroeck (2015) for a literature review and Combes, Démurger, and Li (2017) for another example. We believe that this is an ideal empirical application for our proposed model for the following two reasons. First, China's vast internal urban labor markets are inter-linked. The wage level in each city is not only determined by the characteristics and/or performance of itself, but also depends on those around it. The spatial effects of wages are also discussed in the literature; see, for example, Braid (2002) and Baltagi, Blien, and Wolf (2012). Second, China has experienced unprecedented economic growth and change in the economic structure during the last four decades or so. During the period, the organization of the economy, the environment in which economic agents operated, and perhaps even the agents themselves have changed dramatically. For example, the reforms of State-owned Firm started in 1997 changed the ownership of most of the firms, the way they are managed, the productivity of labor, and how remuneration was determined. The series of reforms in housing, the health system, and education also changed both demand and supply of labor. In other words, it is likely that the economy has experienced "structural changes" over the years and that many of the key relations between economic variables may not remain constant over time.

In this case study, we explain the logarithm of the average wage level of a city by a number of variables, including capital (measured by asset), investment (measured by FDI), and the economic structure of the city (measured by the proportion of industries and sectors). Ideally, a variable reflecting the level of labor input should be included in the model. The only variable that is available to us is the size of the population in each city. The variable appears to be highly correlated with the time trend--the correlation coefficient between the average (log) population
size of the cities and the time trend is about 0.98. Thus, we chose not to include the variable. The impact of labor input is absorbed by the coefficient of the time trend when the constant term is included in the time-varying model. Table 4 provides the definitions of these variables. Our model captures spatial inter-dependence and potential change in the effects of these explanatory variables.

Our data are derived from Statistic Year Books of China (various years), for 1995–2009 (T = 15) and cover 159 (N) cities, including four cities like Beijing, Shanghai, Tianjin, and Chongqing directly administrated by the central government, and other 155 cities at or above prefectoral levels in China. A prefectural city in China means a city that is directly controlled by provincial governments. The geographical location of these cities can be found in Figure 1. Following the convention, we divide China into seven regions: East China (EC), South China (SC), Southwest (SW), North China (NC), Northwest China (NW), Central China (CC), and Northeast China (NE). The densities of cities in these regions are quite different that EC has 56 cities, almost one-third of the whole country while it is very sparse in the western region, reflecting the uneven distribution of cities. We use highway distances between pairs of cities in kilometers to measure the spatial distances between cities, as they can reflect economic distances between cities. These data are collected using the service provided by Google Map Services. We specify W as the inverse of highway distances between cities, standardized by its maximum eigenvalue.

As described in the assumptions of Section 3, regressors are allowed to be trending stationary. To check whether the macro-level regressors considered here are trending stationary, we first

---

**Table 3.** Means and standard deviations of MSE of $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau), \hat{\beta}_2(\tau))^\top$ ($\rho_0 = 0.3, \sigma^2 = 1$).

|     | (I-1)       | (I-2)       | (I-3)       |
|-----|-------------|-------------|-------------|
| T   | N=10        | N=15        | N=30        |
|     | MSE         | MSE         | MSE         |
|     | Standard    | Standard    | Standard    |
|     | deviation   | deviation   | deviation   |

|     | (II-1)       | (II-2)       | (II-3)       |
|-----|-------------|-------------|-------------|
| T   | N=10        | N=15        | N=30        |
|     | MSE         | MSE         | MSE         |
|     | Standard    | Standard    | Standard    |
|     | deviation   | deviation   | deviation   |

---

**Table 4.** Variable definitions.

| Dependent variable (Y) | Definition |
|------------------------|------------|
| log(wage)              | Log value of average wage per worker (1994 price) |

| Independent variables (X) | Definition |
|---------------------------|------------|
| log(FDI)                  | Log value of FDI (10 thousand yuan, 1994 price) |
| log(Asset)                | Log value of total asset (million yuan, 1994 price) |
| GDP<sub>m</sub>           | Proportion of GDP by the manufacturing sector |
| GDP<sub<s</sub>          | Proportion of GDP by the service sector |
| Emp<sub>s</sub>           | Proportion of employed persons in the manufacturing sector |
|                         | Proportion of employed persons in the nonagricultural sector |
Figure 1. Map of 159 Chinese cities where each color represents one of the following regions from Central China (CC), East China (EC), North China (NC), North East (NE), North West (NW), South China (SC) and South West (SW).

Table 5. IPS unit root test statistics and p-values for the regressor residuals.

| Residuals in regressors | $W_{\text{bar}}$ | p-value | $Z_{\text{bar}}$ | p-value |
|-------------------------|------------------|---------|------------------|---------|
| $\log(\text{FDI})$     | -17.7893         | <0.0001 | -17.9520         | <0.0001 |
| $\log(\text{Asset})$   | -5.3533          | <0.0001 | -5.3726          | <0.0001 |
| $\text{GDP}_m$         | -12.4843         | <0.0001 | -12.5689         | <0.0001 |
| $\text{GDP}_s$         | -12.5017         | <0.0001 | -12.6032         | <0.0001 |
| $\text{Emp}_{ms}$      | -13.6499         | <0.0001 | -13.7323         | <0.0001 |

fit the unknown trend in the regressors with the local linear estimation method. After removing the fitted trend, we obtain the residuals in regressors. Then, we conduct the Im-Pesaran-Shin (IPS) panel unit root tests on these residuals. Refer to Im, Pesaran, and Shin (2003, eqs. (4.10) and (4.6)) for the IPS test statistics $W_{\text{bar}}$ and $Z_{\text{bar}}$, respectively. According to the test statistics and p-values in Table 5, the null hypotheses of panel unit root for these variables are all rejected, indicating that the assumption of the trend stationary regressors is valid for our dataset.

It is known that the choice of bandwidth is important for a nonparametric kernel estimation. We first estimate the model with the optimal bandwidth ($h_{\text{opt}} = 0.4000$) obtained by the leave-one-unit-out cross-validation method, and then compare the results with a number of different bandwidths around it: $2/3h_{\text{opt}}$, $4/5h_{\text{opt}}$, $5/4h_{\text{opt}}$, and $3/2h_{\text{opt}}$. Figure 2 shows that the results under different bandwidths are consistent. Table 6 reports that the estimates of the spatial coefficient $\rho_0$ and variance $\sigma_0^2$, which shows that the estimates are quite similar in these specifications. Given the robustness of the results, we decide to use the average bandwidth $h_{\text{ave}} = 0.4173$ of those five bandwidths for the rest of the article.

We estimate the model for China as a whole and then for East China (Figure 3). Table 7 reports the estimates of the spatial coefficient $\rho_0$, $\sigma_0^2$ and the time average of $\beta_0(\tau)$ for the whole country and for East China, respectively. The significance of the spatial coefficient estimate reflects the spatial dependence and confirms the existence of spillover effects between cities. From the table we can see that, over the years, FDI contributed positively to the wage level on average, if FDI increases one percent, the average level of real wage would increase by 0.025%. If capital increases by one percent, the average real wage would increase by 0.026%. The estimates also show that economic structure affects wages as well. For example, if the share of the manufacturer or service sector increases by one percent, the average real wage would increase by 0.365% or 0.326%, respectively, but the increase in wage would be 0.380% higher if the size of the service sector is one percent larger relative to the manufacturer sector. Comparison between the whole country and East China illustrates that the spatial dependence is much bigger in EC than in the whole country. It makes sense as the economic connection is spatially stronger in small regions. The impact of FDI on wage is smaller in EC. This is likely due to a smaller difference in economic development between EC and the rest of the country so that the additional effect of FDI on the local economy is not as large as for less developed parts of the country. The average effect of capital and the share of manufacturer in EC are larger than the whole country. This is because it is the densest region in the country and its manufacturer sectors are more developed.

Figure 4 displays the time-varying coefficient curves for all variables with their 95% confidence bands for the whole country. It shows the parameters of the explanatory variables evolve clearly over time. For example, the impact of FDI on the wage level has been decreasing over time. This can be explained by the fact that in the early stage of reform, foreign investment brought advanced technology and management know-how, which also push up the labor demand, but as the domestic economy catches up, the impact of FDI on the labor market becomes less important. Meanwhile, the impact of capital on the wage level kept increasing over time. This could be because as
the economy becomes less labor-intensive, as the capital level increases, the demand for work also increases accordingly. The effects of economic structure on wage have also changed over time. These findings confirm that indeed the relations between economic variables changed dramatically during such a period of fast development. Figure 5 shows the time-varying features of these variables and the intercept appears quite strong in EC as well. The results imply that if a time-invariant model were used, the impacts of these variables would be estimated with biases.

In addition, we conduct model diagnostics. To save space, we just report the results of the model for the whole country since the results of the model for the EC region are quite similar. To check the stationarity of residuals, we implement IPS unit root test. Two test statistics are $W_{tbar} = -7.3741$ and $Z_{tbar} = -7.4154$ with $p$-values less than 0.0001. So we reject the null hypothesis of panel unit root and conclude the residuals are stationary. To further check whether there is a serial correlation, we carry out the Box-Pierce test (see Box and Pierce 1970) on the estimated residuals for each city. It is worth noting that we are

---

**Table 6.** Estimates of parameters under different bandwidths.

| $h_{opt}$ | $h_{opt}/2$ | $h_{opt}/4$ | $h_{opt}/5$ | $h_{opt}/4$ | $h_{opt}/2$ |
|-----------|-------------|-------------|-------------|-------------|-------------|
| $\rho_0$  | 0.1214 (0.0185) | 0.1052 (0.0185) | 0.1126 (0.0185) | 0.1132 (0.0186) | 0.1418 (0.0186) |
| $\sigma_2$ | 0.0062 (0.0002) | 0.0061 (0.0002) | 0.0062 (0.0002) | 0.0064 (0.0002) | 0.0064 (0.0002) |

---

**Figure 2.** Estimated curves of the time-varying coefficients under different bandwidths for the whole country (159 cities).

**Figure 3.** Map of 56 cities in Region EC where each color represents different province.
interested in a set of hypotheses

$H_{i,0} :$ the estimated residuals for each city $i$ are white noise.

v.s. $H_{i,1} :$ otherwise. (12)

for cities $i = 1, \ldots, N$. Accordingly, for each of the hypotheses above, we can calculate the Box-Pierce test statistic and its $p$-value. To identify whether or not there exists a significant test,

Table 7. Estimation results of semiparametric spatial autoregressive panel data model (the covariate coefficient estimates are calculated by the average over time).

|                | Whole country | East China |
|----------------|--------------|-----------|
| Intercept      | 7.2190       | 6.4822    |
| log(FDI)       | 0.0250       | 0.0183    |
| log(Asset)     | 0.0256       | 0.0308    |
| $GDP_m$        | 0.3645       | 0.4319    |
| $GDP_s$        | 0.3256       | 0.1663    |
| $Emp_{ms}$     | $-0.3799$    | $-0.3717$ |
| $\rho$         | 0.1240*** (0.0185) | 0.2239*** (0.0242) |
| $\sigma^2$     | 0.0063*** (0.0002) | 0.0044*** (0.0002) |

we apply the Benjamini and Hochberg (1995)'s multiple testing procedure that controls the false discovery rate (FDR) of (12) at the rate of 0.05. We also apply the Bonferroni correction method (see Miller Jr 1966) that controls the familywise error rate (FWER) of (12) at the rate of 0.05. Both of these two multiple testing procedures show that the null hypotheses for all cities cannot be rejected, which means there is no serial correlation in the residuals. All the aforementioned diagnostics results support the validity of our regression model.

7. Conclusion

We have considered a semiparametric SAR panel data model with fixed effects. This model is designed particularly for situations where covariate effects on the dependent variable change over time. The spatial dependence structure between units is assumed to be time-invariant presented by a parametric spatial lag term. To consistently estimate both the parametric
and nonparametric components, we have proposed a local linear concentrated QML estimation method. Asymptotic properties for the estimators have been derived with parametric $\sqrt{NT}$ and nonparametric $\sqrt{NTh}$ rate of convergence, respectively, when both the cross-sectional size $N$ and the time length $T$ go to infinity.

The finite-sample performance of our model has been evaluated and compared with those from a time-invariant spatial panel data model using Monte Carlo simulations. The results showed that when the time-varying coefficients are misspecified to be constants, using the standard time-invariant spatial panel data model would lead to inconsistent estimates while our proposed model is always consistent and robust.

We have also applied the proposed model to study labor compensation in Chinese cities. Our results have illustrated that as China became more developed, the impacts of capital, investment, and the structure of the economy on labor compensation have changed over time. The results also imply that for a fast-changing economy such as China, many important economic parameters may not be consistently estimated with a time-invariant model.

**Acknowledgments**

The authors are grateful for the editor, the associate editor and anonymous referees for their constructive comments and suggestions on earlier versions of this submission. The authors also acknowledge seminar participants for their comments and suggestions. This project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

**Funding**

Xuan Liang and Jiti Gao thank to the Australian Research Council Discovery Grants Program under grant numbers: DP150101012 & DP170104421 for its financial support. Xuan Liang also acknowledges the financial support of the ANU RSEFAS Cross-Disciplinary Grant.
Appendix A: Justification of Identification Condition

\[ \sum_{i=1}^{N} \alpha_{i,j} = 0 \]

Considering the specification of \( \beta_{0,j} = \beta_{0}(\tau_i) \) in Equation (2), our model (1) becomes

\[
Y_{it} = \rho_{0} \sum_{j \neq i} w_{ij} Y_{jt} + X_{it}^{\top} \beta_{0}(\tau_i) + \alpha_{i,j} + \epsilon_{it},
\]

\[ t = 1, \ldots, T, \quad i = 1, \ldots, N. \]

where the constant 1 is included in the regressor \( X_{it} \).

Without loss of generality, let \( X_{it} = (X_{it1}, X_{it,-1}^{\top}) \), \( \beta_{0}(\tau_i) = (\beta_{0,1}(\tau_i), \beta_{0,-1}(\tau_i))^{\top} \) where \( X_{it1} = 1 \) and \( X_{it,-1} = (X_{it2}, \ldots, X_{itd})^{\top} \) and \( \beta_{0,-1}(\tau_i) = (\beta_{0,2}(\tau_i), \ldots, \beta_{0,d}(\tau_i))^{\top} \). Then, our model becomes

\[
Y_{it} = \rho_{0} \sum_{j \neq i} w_{ij} Y_{jt} + \beta_{0,1}(\tau_i) + X_{it,-1}^{\top} \beta_{0,-1}(\tau_i) + \alpha_{i,j} + \epsilon_{it},
\]

\[ t = 1, \ldots, T, \quad i = 1, \ldots, N. \]

Let \( Y_{t} = N^{-1} \sum_{i=1}^{N} Y_{it}, \; W_{t} = N^{-1} \sum_{i=1}^{N} \sum_{j \neq i} w_{ij} Y_{jt}, \; X_{t,-1} = N^{-1} \sum_{i=1}^{N} X_{it,-1}, \; \overline{\sigma} = N^{-1} \sum_{i=1}^{N} \alpha_{i,j} \) and \( e_{t} = N^{-1} \sum_{i=1}^{N} \epsilon_{it} \). We then have

\[
Y_{t} = \rho_{0} W_{t} + \beta_{0,1}(\tau_i) + X_{t,-1}^{\top} \beta_{0,-1}(\tau_i) + \overline{\sigma} + e_{t}, \quad (A1)
\]

Without imposing the assumption \( \overline{\sigma} = 0 \), model (A1) implies that there is an identification issue with the identifiability and then estimation of \( \beta_{0,1}(\tau_i) \). Hence, in order to identify \( \beta_{0,1}(\tau_i) \) we require the condition \( \sum_{i=1}^{N} \alpha_{i,j} = 0 \). In fact, the advantage of this identification condition is to allow us to estimate \( \beta_{0,1}(\tau_i) \) as a smooth time-varying trend effect in contrast to the fixed effects structure. Therefore, we believe that our model is more flexible and applicable.

Appendix B: Proofs of Theorems

Proof of Theorem 1. Even though we have the nonparametric terms in our model, the idea for proving the consistency of the parametric estimators and the identification can be adopted from Lee (2004). Define \( Q_{N,T}(\rho) = \max_{\theta} E\{\log L_{N,T}(\theta)\} \), where \( \theta = (\rho, \sigma^2) \). In order to show the consistency of \( \theta \), it suffices to show

\[
\frac{1}{N T} \log L_{N,T}(\rho) - Q_{N,T}(\rho) \overset{p}{\underset{N \to \infty}{\to}} 0 \quad \text{uniformly on } \Delta, \quad (A2)
\]

and the uniqueness identification condition that

\[
\limsup_{N \to \infty} \max_{\rho \in N_{\epsilon}(\rho_0)} \frac{1}{N T} \{Q_{N,T}(\rho) - Q_{N,T}(\rho_0)\} < 0 \quad \text{for any } \epsilon > 0, \quad (A3)
\]

by using White (1996) and Lee (2004), where \( N_{\epsilon}(\rho_0) \) is the complement of an open neighborhood of \( \rho_0 \) on \( \Delta \) of diameter \( \epsilon \).

(1) Proof of Equation (A2). Observe that \( Q_{N,T}(\rho) = -\frac{N T}{2} \log(2\pi) + \frac{N T}{2} \log(\sigma^2(\rho)) + T \log(\text{Det} S_{N,T}(\rho)) \), where \( \sigma^2(\rho) = (NT)^{-1} \frac{1}{N T} E\{Y^{\top}(\rho) Q_{N,T}(\rho)\} = (NT)^{-1} E\{Y^{\top}(\rho) P_{N,T} Y(\rho)\}. \)

Due to \( P_{N,T} D = 0_{N,T,N-1}, \; P_{N,T} = P_{N,T}^{\top} + \rho_0 I_T \otimes W \), we can rewrite \( \sigma^2(\rho) \) as

\[
\sigma^2(\rho) = \frac{2}{N T}(\rho_0 - \rho)^2 E(R_{N,T}^{\top} P_{N,T} R_{N,T}) + \frac{2}{N T}(\rho_0 - \rho) E((\tilde{X}_0)^{\top} P_{N,T} R_{N,T}) + \frac{1}{N T} E((\tilde{X}_0)^{\top} P_{N,T} \tilde{X}_0) + \frac{\sigma_T^2}{N T} \text{tr}[S_{N,T}^{-1} H_{N,T}(\rho) P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1}], \quad (A4)
\]

Then, \( (NT)^{-1} \{\log L_{N,T}(\rho) - Q_{N,T}(\rho)\} = -(1/2) \left[ \log(\hat{\sigma}^2(\rho)) - \log(\sigma^2(\rho)) \right] \), where

\[
\hat{\sigma}^2(\rho) = \frac{1}{NT} E(Y^{\top}(\rho) P_{N,T} Y(\rho)) = \frac{1}{NT} \rho_0^2 + \frac{2}{NT} \rho_0 \tilde{X}_0^{\top} P_{N,T} \tilde{X}_0 + \frac{1}{NT} \text{tr}[S_{N,T}^{-1} H_{N,T}(\rho) P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1}], \quad (A5)
\]

To show Equation (A2), it is sufficient to show that

\[
\hat{\sigma}^2(\rho) - \sigma^2(\rho) = o_p(1), \quad \text{uniformly on } \Delta. \quad (A6)
\]

According to Lemma B.2, we know

\[
\frac{1}{NT}(\tilde{X}_0)^{\top} P_{N,T} \tilde{X}_0 = o_p(1), \quad \frac{1}{NT} E(\tilde{X}_0^{\top} P_{N,T} \tilde{X}_0) = o(1), \quad \frac{1}{NT}(\tilde{X}_0)^{\top} P_{N,T} R_{N,T} = o_p(1), \quad \frac{1}{NT} E(\tilde{X}_0^{\top} P_{N,T} R_{N,T}) = o(1),
\]

\[
\frac{1}{NT} R_{N,T}^{\top} P_{N,T} R_{N,T} = \frac{1}{NT} E(R_{N,T}^{\top} P_{N,T} R_{N,T}) + op(1) \to \Psi_{R,R}.
\]

By Equations (A4) and (A5),

\[
\hat{\sigma}^2(\rho) - \sigma^2(\rho) = \frac{2}{NT}(\tilde{X}_0)^{\top} P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1} \tilde{X}_0 + \frac{2}{NT} \rho_0 \tilde{X}_0^{\top} P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1} \tilde{X}_0 + \frac{1}{NT} \text{tr}[S_{N,T}^{-1} H_{N,T}(\rho) P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1}] - \sigma^2(\rho) + op(1),
\]

\[
\leq 2H_2(\rho) + 2(\rho_0 - \rho) H_2(\rho) + H_3(\rho) - \sigma^2(\rho) + op(1),
\]

where \( \sigma^2(\rho) = (NT)^{-1} \frac{1}{N T} \text{tr}[S_{N,T}^{-1} H_{N,T}(\rho) P_{N,T} S_{N,T}(\rho) S_{N,T}^{-1}] \).

According to Lemma B.3, we get Equation (A6) so that Equation (A2) holds.

(2) Proof of Equation (A3). Consider an arbitrary SAR panel process: \( Y_t = \rho_0 W_{it} + \epsilon_{it} \), where \( \epsilon_{it} \sim N(0, \sigma^2_{\epsilon1 N}) \) and \( i = 1, \ldots, T \). Denote the log-likelihood of this model as \( L_{N,T}(\rho, \sigma^2) \). Let \( \tilde{\sigma}^2(\rho) = (NT)^{-1} \frac{1}{N T} \text{tr}[S_{N,T}^{-1} H_{N,T}(\rho) S_{N,T}(\rho) S_{N,T}^{-1}] \). It can be verified that

\[
\max E_{\theta}[\log L_{N,T}(\theta, \sigma^2)] = -\frac{NT}{2} \log(2\pi) + \frac{NT}{2} \log(\hat{\sigma}^2(\rho)) + T \log(S_{N,T}(\rho)) = \hat{Q}_{N,T}(\rho),
\]

\[
\hat{Q}_{N,T}(\rho) = -\frac{NT}{2} \log(2\pi) + \frac{NT}{2} \log(\hat{\sigma}^2(\rho)) + T \log(S_{N,T}(\rho)) \overset{p}{\underset{N \to \infty}{\to}} 0 \quad \text{uniformly on } \Delta.
\]

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
where \( E_a \) denotes the expectation under this auxiliary model. Hence, for any \( \rho \in \Delta \), we have \( \tilde{Q}_{N,T} (\rho) \leq \max_{\rho, \sigma^2} E_a [\log L_{N,T} (\rho, \sigma^2)] = E_a [\log L_{N,T} (\rho_0, \sigma_0^2)] = \tilde{Q}_{N,T} (\rho_0) \) so that
\[
\frac{1}{N T} \sum_{\rho} (\tilde{Q}_{N,T} (\rho) - \tilde{Q}_{N,T} (\rho_0)) \leq 0 \quad \text{uniformly on } \Delta.
\]
The term in Equation (A3) can be rewritten as follows:
\[
\frac{1}{N T} \sum_{\rho} (\tilde{Q}_{N,T}(\rho) - Q_{N,T}(\rho_0)) = \frac{1}{N T} \sum_{\rho} (\tilde{Q}_{N,T}(\rho) - \tilde{Q}_{N,T}(\rho_0)) + \frac{1}{2} H_4(\rho) + \frac{1}{2} H_5,
\]
where \( H_4(\rho) = \log(\sigma^2(\rho)) - \log(\sigma^2(\rho_0)) \) and \( H_5 = \log(\sigma^2(\rho_0)) - \log(\sigma^2(\rho_0)). \)

By Lemma C.5, we have \((NT)^{-1} \text{tr}(E(P_{N,T})) = (NT)^{-1} \text{tr}(N_T) = 1\) so that
\[
\sigma^2(\rho_0) - \tilde{\sigma}^2(\rho_0) = \frac{1}{NT} E((\tilde{X}_0)^\top P_{N,T}(\tilde{X}_0)) + \tilde{\sigma}_0^2 \left\{ \frac{\text{tr}(E(P_{N,T}))}{NT} - 1 \right\} = o(1).
\]
Accordingly, we obtain \( H_5 = o(1) \).

To show \( H_4(\rho) \leq 0 \) uniformly on \( \Delta \), it suffices to show \( \tilde{\sigma}^2(\rho) - \sigma^2(\rho) \geq 0 \) uniformly on \( \Delta \). Observe that \( \tilde{\sigma}^2(\rho) - \sigma^2(\rho) = \frac{\sigma^2}{N T} H_4(\rho) - \frac{1}{N T} H_5(\rho) \), where
\[
\frac{1}{N T} H_4(\rho) = 1 - \frac{\text{tr}(E(P_{N,T}))}{N T} + 2(\rho_0 - \rho) \frac{\text{tr}(G_{N,T}) - \text{tr}(E(P_{N,T})G_{N,T})}{N T} + (\rho_0 - \rho)^2 \frac{\text{tr}(G_{N,T}^2 G_{N,T}) - \text{tr}(G_{N,T}^2 E(P_{N,T})G_{N,T})}{N T}
\]
and \( H_5(\rho) = \frac{1}{N T} \left\{ \text{tr}(E(P_{N,T})G_{N,T}) - \text{tr}(E(P_{N,T})G_{N,T}) \right\} \).

By Lemma C.5, we get \((NT)^{-1} H_4(\rho) = o(1)\) uniformly on \( \Delta \). Since \( P_{N,T} = (I_{N_T} - \tilde{X}_0^\top Q_{N_T} Q(I_{N_T} - \tilde{X}_0^\top) \) and \( \tilde{X}_0^\top \sigma_0^2(\rho_0) \), \( \tilde{X}_0^\top \sigma_0^2(\rho_0) \) is not a zero vector, \( H_5^2(\rho) \) can be written as \( E(\tilde{X}_0^\top A) \) with a positive semidefinite matrix \( A \). Thus, \((NT)^{-1} H_5^2(\rho) \leq 0 \) uniformly on \( \Delta \) so that \( H_5(\rho) \leq 0 \) holds uniformly on \( \Delta \).

Hence, we have \( \lim_{N_T \to \infty} \sup_{\rho \in N_T(\rho_0)} \frac{1}{N T} [Q_{N,T}(\rho) - Q_{N,T}(\rho_0)] \leq 0 \) for any \( \epsilon > 0 \).

If the identification uniqueness condition was not satisfied, without loss of generality, there would exist a sequence \( \rho_n \) converging to \( \rho^* \neq \rho_0 \) such that \( \lim_{N_T \to \infty} \sup_{\rho \in N_T(\rho_0)} \frac{1}{N T} [Q_{N,T}(\rho) - Q_{N,T}(\rho_0)] = 0 \). This will follow if (a) \( \lim_{N_T \to \infty} \sup_{\rho \in N_T(\rho_0)} [\tilde{\sigma}^2(\rho^*) - \sigma^2(\rho_0)] = 0 \) and (b) \( \lim_{N_T \to \infty} \sup_{\rho \in N_T(\rho_0)} [Q_{N,T}(\rho^*) - Q_{N,T}(\rho_0)] = 0 \). Since \( \lim_{N_T \to \infty} \sup_{\rho \in N_T(\rho_0)} [\tilde{\sigma}^2(\rho) - \sigma^2(\rho)] = \Psi_{KK} > 0 \) assumed in Assumption 8, (a) contradicts with this assumption. Hence, we have accomplished the proof.

**Proof of Theorem 2.** In this section, we provide the proof of the asymptotic distribution of \( \hat{\theta} \). According to the Taylor expansion of the first order condition of Equation (8), we obtain
\[
\sqrt{N T} (\hat{\theta} - \theta_0) \sim - \frac{1}{\sqrt{N T}} \frac{\partial^2 \log L_{N,T}(\theta)}{\partial \theta^2}^{-1} \left[ \frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta)}{\partial \theta} \right]
\]
where \( \sqrt{N T} (\hat{\theta} - \theta_0) \) lies between \( \hat{\theta} \) and \( \theta_0 \) and hence it converges to \( \theta_0 \) in probability by Theorem 1. If
\[
\frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta_0)}{\partial \theta} \frac{d}{d \theta} \to N(0_2, \Sigma_{\theta_0} + \Omega_{\theta_0}) \quad \text{as} \ \Delta \to \infty
\]
then all hold, the asymptotic normality follows: \( \sqrt{N T} (\hat{\theta} - \theta_0) \to N(0_2, \Sigma_{\theta_0} + \Omega_{\theta_0}) \).

(1) Proof of Equation (A7). First, we can get the first-order derivative of \( \log L_{N,T}(\theta) \) as follows:
\[
\frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta_0)}{\partial \theta} = \left( \frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta)}{\partial \theta} \right)_{\theta_0} = \left( \frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta)}{\partial \theta} \right)_{\theta_0} = \frac{1}{\sqrt{N T}} \left[ \frac{R_{N,T} P_{N,T} \tilde{X}_0}{\sigma_0^2} + \frac{e^\top G_{N,T} P_{N,T} \tilde{X}_0 \text{tr}(G) + \sigma_0^2}{\sigma_0^2} + \frac{(\tilde{X}_0)^\top P_{N,T} \tilde{X}_0}{\sigma_0^2} \right]
\]
due to
\[
\frac{1}{\sqrt{N T}} R_{N,T} P_{N,T} \tilde{X}_0 = o(1), \quad \frac{1}{\sqrt{N T}} (\tilde{X}_0)^\top P_{N,T} \tilde{X}_0 = o(1), \quad \frac{1}{\sqrt{N T}} \left( \frac{1}{\sqrt{N T}} \frac{\partial \log L_{N,T}(\theta)}{\partial \theta} \right) = o(1),
\]
\[
\frac{1}{\sqrt{N T}} (\tilde{X}_0)^\top P_{N,T} e = o(1), \quad \frac{1}{\sqrt{N T}} (\tilde{X}_0)^\top P_{N,T} G_{N,T} e = o(1),
\]
in Lemma B.2. By Lemma C.5, we have
\[
\frac{1}{\sqrt{N T}} (\tilde{X}_0)^\top P_{N,T} \tilde{X}_0 = o(1), \quad \frac{1}{\sqrt{N T}} \text{tr}(P_{N,T} - I_{N_T}) = o(1), \quad \frac{1}{\sqrt{N T}} \text{tr}(P_{N,T} G_{N,T} - G_{N,T}) = o(1),
\]

1798 X. LIANG, J. GAO, AND X. GONG
so that
\[
\frac{1}{\sqrt{NT}} \frac{\partial \log L_{NT}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{NT}} \left( \frac{1}{\sigma_0^2} \left( G_{N,T}^\top P_{N,T} \varepsilon - \frac{3}{2} \varepsilon (G_{N,T}^\top P_{N,T}) + \frac{1}{\sigma_0^2} R_{N,T}^\top P_{N,T} \varepsilon \right) + \phi_1(1) \right) + \phi_2(1).
\]

where 
\[
\mathbb{Q} = \left( \frac{1}{\sigma_0^2} \mathbb{E} \left[ \frac{1}{\sigma_0^2} \left( G_{N,T}^\top P_{N,T} \varepsilon + \frac{1}{\sigma_0^2} R_{N,T}^\top P_{N,T} \varepsilon \right) \right] \right)\].

Next, we show the asymptotic normality of \((NT)^{-1/2}(\mathbb{Q} - \mathbb{E}(\mathbb{Q}|F_V))\). By the Cramer-Wold device, it is sufficient to derive the asymptotic normality of \((NT)^{-1/2}(a_1, a_2) (\mathbb{Q} - \mathbb{E}(\mathbb{Q}|F_V))\), where \((a_1, a_2)^T\) is any given two-dimensional constant vector. Note that
\[
(a_1, a_2) \mathbb{Q} = e^\top \left( \frac{a_1}{\sigma_0^2} G_{N,T}^\top P_{N,T} + \frac{a_1}{\sigma_0^2} R_{N,T}^\top P_{N,T} \right) e + \frac{a_2}{\sigma_0^2} R_{N,T}^\top P_{N,T} e = e^\top A_{N,T} e + b_{N,T}^\top e,
\]
which is a linear quadratic form of \(e = (e_1, \ldots, e_N, e_1, \ldots, e_N)^T\). Here \(A_{N,T} = \frac{a_1}{\sigma_0^2} G_{N,T}^\top P_{N,T} + \frac{a_1}{\sigma_0^2} R_{N,T}^\top P_{N,T} + \frac{a_2}{\sigma_0^2} R_{N,T}^\top P_{N,T}\) and \(b_{N,T} = \frac{a_2}{\sigma_0^2} R_{N,T}^\top P_{N,T}\) are complicated functions of \((X_t, t = 1, \ldots, T)\) based on the definitions of \(P_{N,T}\) and \(R_{N,T}\) given before Assumption 8. Consequently, \(A_{N,T}\) and \(b_{N,T}\) are both random and hence different from the fixed setting in the central limit theorem for linear quadratic forms in Kelejian and Prucha (2001, theor. 1).

In order to adapt the proof of Theorem 1 in Kelejian and Prucha (2001) in our case, the conditions for \(e\) in Assumption 2 can help us establish a martingale difference array. Accordingly, the central limit theorem for martingale difference arrays used in the proof of Theorem 1 in Kelejian and Prucha (2001) can be applied to derive the asymptotic normality of the linear quadratic form \(e^\top A_{N,T} e + b_{N,T}^\top e\).

To construct the martingale difference array, let \(J = N(t - 1) + 1\) for \(1 \leq i \leq N\) and \(1 \leq t \leq T\) and \(e_t = e^{(j)}\) for \(j = 1, \ldots, NT\). Consequently, we get \(e = (e_1, \ldots, e_N, e_1, \ldots, e_N)^T = (e^{(1)}, \ldots, e^{(N)}, e^{(N)}, \ldots, e^{(T-1)}, \ldots, e^{(j)}_T)^T\). Let
\[
F_{j-1} = F_V \lor \sigma(e^{(1)}, \ldots, e^{(j-1)}),
\]
for \(j \geq 1\) where \(F_{j-1} := F_V\) if \(j = 1\), and \(F_V \lor \sigma(e^{(1)}, \ldots, e^{(j-1)}))\) is the \(\sigma\)-field generated by \(F_V \lor \sigma(e^{(1)}, \ldots, e^{(j-1)}))\). Due to the conditional independence of \(e_t\) and \((e_{t+1}, \ldots, e_{t+T-1})\) given \(E_t\), we obtain \(E(e_t|F_{j-1}) = E(e_t|E_{t-1}, e_1, \ldots, e_{t-1}) = E(e_t|E_{t-1}) = 0\). Thus, \(\{e^{(j)}, F_V\} : 1 \leq j \leq NT\) forms a martingale difference array.

To make use of the above constructed martingale difference array to derive the asymptotic normality of the linear quadratic form \(e^\top A_{N,T} e + b_{N,T}^\top e\), we let \((A_{N,T})_{ij}\) be the \((j, j)\)-th element of matrix \(A\) and \(b_j\) be the \(j\)-th element of vector \(b\). Note that both \(A_{N,T}, b_{N,T} \in F_V\) because they are functions of \((X_t, t = 1, \ldots, T)\) and also \(A_{N,T}\) is symmetric. Hence we obtain \(e^\top A_{N,T} e + b_{N,T}^\top e = (e^\top A_{N,T} e + b_{N,T}^\top e) F_V = \sum_{j=1}^{NT} Z_j\), where
\[
Z_j = (A_{N,T})_{jj} \left( \frac{e^{(j)}_2 - \sigma_0^2}{\sigma_0^2} + e^{(j)}_1 (A_{N,T})_{jj} + (b_{N,T})_j e^{(j)}_1 \right).
\]

Since \(A_{N,T}, b_{N,T} \in F_V\), we then have \(E(Z_j|F_{j-1}) = 0\). This implies that \(\{Z_j, F_j\} : 1 \leq j \leq NT\) forms a martingale difference array. At this stage we can apply the central limit theorem for martingale difference arrays similarly to the proof of Theorem 1 in Kelejian and Prucha (2001), and obtain that
\[
\frac{1}{\sqrt{NT}} (a_1, a_2) \mathbb{Q} - \mathbb{E}(\mathbb{Q}|F_V) = \frac{1}{\sqrt{NT}} \sum_{j=1}^{NT} Z_j \overset{d}{\to} N \left(0, (a_1, a_2)^\top \Sigma_Q (a_1, a_2)^\top \right),
\]
where \(\Sigma_Q = \lim_{N, T \to \infty} (NT)^{-1} E \text{Cov}(Q|F_V) = \lim_{N, T \to \infty} (\Sigma_{NT, \theta_0} + \Omega_{NT, \theta_0})\),

and
\[
\Sigma_{NT, \theta_0} = \left( \begin{array}{cc}
\frac{E(R_{N,T}^\top P_{N,T} \text{diag}(P_{N,T}))}{N \tau_0^2} & \frac{E(tr(P_{N,T}G_{N,T}^\top P_{N,T}G_{N,T} + P_{N,T}G_{N,T}^\top G_{N,T}^\top P_{N,T}))}{NT \sigma_0^2} \\
\frac{E(tr(G_{N,T}^\top P_{N,T}))}{N \tau_0^2} & \frac{E(tr(G_{N,T}^\top G_{N,T}))}{NT \sigma_0^2}
\end{array} \right) + o(1),
\]
and
\[
\Omega_{NT, \theta_0} = \left( \begin{array}{c}
\frac{2m_3 E(R_{N,T}^\top P_{N,T} \text{diag}(P_{N,T}))}{2NT \sigma_0^6} \\
\frac{m_3 E(R_{N,T}^\top P_{N,T} \text{diag}(P_{N,T}))}{2NT \sigma_0^6} + \frac{(m_3 - 3m_4) E(\sum_{i=1}^{NT} (gp)_i^2)}{2NT \sigma_0^6} + \frac{E(R_{N,T}^\top P_{N,T} \text{diag}(P_{N,T}))}{2NT \sigma_0^6} + \frac{(m_3 - 3m_4) E(\sum_{i=1}^{NT} (gp)_i^2)}{2NT \sigma_0^6}
\end{array} \right).
\]

where \(p_{ij}\) and \((gp)_i\) are the \(i\)-th main diagonal elements of \(P_{N,T}\) and \(G_{N,T}^\top P_{N,T}\), respectively. As
\[
\lim_{N,T \to \infty} \frac{\text{tr}(G_{N,T}^\top P_{N,T}) + \text{tr}(G_{N,T}^\top G_{N,T})}{NT} = c_1,
\]
and
\[
\lim_{N,T \to \infty} \frac{\text{tr}(G_{N,T}^\top G_{N,T})}{NT} = c_2.
\]

we have \(\lim_{N,T \to \infty} \Sigma_{NT, \theta_0} = \Sigma_{\theta_0} = \left( \begin{array}{ccc}
\frac{1}{\sigma_0^2} & \frac{1}{\sigma_0^2} & \frac{1}{\sigma_0^2} \\
\frac{m_2}{\sigma_0^2} & \frac{m_2}{\sigma_0^2} & \frac{m_2}{\sigma_0^2} \\
\frac{m_3}{\sigma_0^2} & \frac{m_3}{\sigma_0^2} & \frac{m_3}{\sigma_0^2}
\end{array} \right) \).

Therefore, we obtain
\[
\frac{1}{\sqrt{NT}} \frac{\partial \log L_{NT}(\theta_0)}{\partial \theta} d \to N(0, \Sigma_{\theta_0} + \Omega_{\theta_0}),
\]
where \(\Omega_{\theta_0} = \lim_{N,T \to \infty} \Omega_{NT, \theta_0}\).
(2) Proof of (A8). The second derivative can be obtained as follows:

\[
\frac{\partial^2 \log L_{N,T}(\theta)}{\partial \theta^2} = -\frac{\text{tr}(G_N^2(\rho))}{NT} - \frac{1}{\sigma^2} \left( \frac{1}{\sigma^2} \right) \left( (I_T \otimes W)Y \right)^\top P_{N,T} \left( (I_T \otimes W)Y \right),
\]

\[
\frac{\partial^2 \log L_{N,T}(\theta)}{\partial \rho \partial \sigma^2} = -\frac{\text{tr}(G_N^2(\rho))}{\sigma^4} - \frac{1}{\sigma^4} \left( (I_T \otimes W)Y \right)^\top P_{N,T} \left\{ S_{N,T}(\rho) \right\},
\]

\[
\frac{\partial^2 \log L_{N,T}(\theta)}{\partial \sigma^2 \partial \sigma^2} = \frac{NT}{2\sigma^4} - \frac{\left\{ S_{N,T}(\rho) \right\}^\top Q_{N,T} \left\{ S_{N,T}(\rho) \right\}}{\sigma^6}.
\]

By some calculations, we have

\[
\frac{1}{NT} \frac{\partial^2 \log L_{N,T}(\theta_0)}{\partial \theta^2} = -\frac{\text{tr}(G_{N}^2(\rho))}{NT} - \frac{\text{tr}(G_{N}^2(\rho))}{NT\sigma_0^2} + 2R_{N,T}^\top P_{N,T} G_{N,T} e + e^\top G_{N,T}^\top P_{N,T} G_{N,T} e
\]

\[
+ \frac{R_{N,T}^\top P_{N,T} e}{NT\sigma_0^4} - \frac{\text{tr}(G_{N}^2(\rho))}{NT\sigma_0^2} + e^\top P_{N,T} e \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0} + \frac{1}{2\sigma_0^2} \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0}.
\]

Thus,

\[
-\frac{1}{NT} \frac{\partial^2 \log L_{N,T}(\theta_0)}{\partial \theta^2} = \frac{1}{NT} \begin{pmatrix}
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho))
\end{pmatrix} \begin{pmatrix}
\sigma_0^2 \\
\sigma_0^2 \\
\sigma_0^2
\end{pmatrix} + \frac{1}{NT\sigma_0^4} \begin{pmatrix}
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho))
\end{pmatrix} \begin{pmatrix}
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho))
\end{pmatrix} + \frac{1}{2\sigma_0^2} \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0}
\]

By Lemma B.2, we have

\[
\frac{1}{NT} \begin{pmatrix}
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho))
\end{pmatrix} \begin{pmatrix}
\sigma_0^2 \\
\sigma_0^2 \\
\sigma_0^2
\end{pmatrix} = \frac{\text{tr}(G_{N}^2(\rho))}{NT\sigma_0^4} \begin{pmatrix}
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho)) \\
\text{tr}(G_{N}^2(\rho))
\end{pmatrix} + \frac{1}{2\sigma_0^2} \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0}
\]

According to Lemma C.5, we can get

\[
-\frac{1}{NT} \frac{\partial^2 \log L_{N,T}(\theta_0)}{\partial \theta^2} = \frac{\text{tr}(G_{N}^2(\rho))}{NT\sigma_0^2} \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0} + \frac{1}{2\sigma_0^2} \bigg|_{\theta = \theta_0} \bigg|_{\rho = \rho_0}
\]

Thus, Equation (A8) has been proved.

(3) Proof of Equation (A9). By Equation (A10), we note that $1/\sigma^2$ appears either in linear, quadratic or cubic from in all the elements of the second derivative of $\log L_{N,T}(\theta)$ and $\rho$ only appears in linear form in $\partial^2 \log L_{N,T}(\theta)/\partial \rho \partial \sigma^2$ and $\partial^2 \log L_{N,T}(\theta)/\partial \sigma^2 \partial \sigma^2$. According to the convergence of $\hat{\theta}$ to $\theta_0$ in probability, it is easy to show Equation (A9) holds for all elements but the second derivative of $\log L_{N,T}(\theta)$ with respect to $\rho$. Since $G_N(\rho) = W\Sigma_N(\rho)$, we have $\text{tr} \left[ G_N^2(\rho) \right] = \text{tr} \left[ G_N^2(\rho^*) \right] + 2\text{tr} \left[ G_N^2(\rho^*) \right] (\hat{\rho}^* - \rho_0)$ for some $\hat{\rho}^*$ between $\hat{\rho}$ and $\rho_0$ by Assumption 5 and the mean value theorem. Consequently, we have
1 \frac{\sigma^2 \log \text{log}_{N,T}(\hat{\theta})}{\sigma^2} - 1 \frac{\sigma^2 \log \text{log}_{N,T}(\theta_0)}{\sigma^2} = -2 \frac{\text{tr} \left[ G_N(\hat{\rho}^*) \right]}{N} (\hat{\rho}^* - \rho_0) \nonumber \nonumber + \left( \frac{1}{\sigma^2} - \frac{1}{\sigma^2} \right) \frac{[(I_T \otimes W)Y]^T P_{N,T} (I_T \otimes W)Y}{NT} . \tag{A11}

Since $G_N(\rho)$ is UB, $N^{-1} \text{tr} \left[ G_N(\hat{\rho}^*) \right]$ is bounded, implying the first term (A11) is $o(p_1)$. The second term is also $o(p_1)$ since $(NT)^{-1} ((I_T \otimes W)Y)^T P_{N,T} (I_T \otimes W)Y$ which can be shown by Lemma B.2. Consequently, Equation (A9) holds. Hence, we have finished the whole proof.

**Proof of Theorem 3.** As we know $S_{N,T}(\hat{\rho}) = S_{N,T} + (\rho_0 - \hat{\rho}) I_T \otimes W$ as well as the definition of $\hat{\beta}(\tau)$ in Equation (9), we have

\[
\hat{\beta}(\tau) - \beta_0(\tau) = \Phi(\tau)(Y^*(\rho) - D \hat{D}^T \hat{D} - \hat{D}^T \hat{Y}(\rho)) - \beta_0(\tau) = \Phi(\tau)(I_{NT} - D \hat{D}^T \hat{D} - \hat{D}^T \hat{Y}(\rho - S)X) \hat{\beta}_0 - \beta_0(\tau) + \Phi(\tau)(I_{NT} - D \hat{D}^T \hat{D} - \hat{D}^T \hat{Y}(\rho) - S) \Delta \alpha + \Phi(\tau)(I_{NT} - D \hat{D}^T \hat{D} - \hat{D}^T \hat{Y}(\rho - S) \beta_0 + \Delta \alpha + \epsilon)
\]

\[
= \mathcal{Z}_{N,T}(1) + \mathcal{Z}_{N,T}(2) + 2 \mathcal{Z}_{N,T}(3) + \mathcal{Z}_{N,T}(4). \tag{A12}
\]

Due to $\hat{D} = (I_T - S)\hat{D}$, we get

\[
\mathcal{Z}_{N,T}(2) = \Phi(\tau)(D - D \hat{D}^T \hat{D} - \hat{D}^T \hat{D}) \alpha = 0_d. \tag{A13}
\]

We may rewrite $\mathcal{Z}_{N,T}(1)$ as $\mathcal{Z}_{N,T}(1) = \left\{ \Phi(\tau)X \tilde{\beta}_0 - \beta_0(\tau) \right\} - \Phi(\tau)D(\hat{D}^T \hat{D} - \hat{D}^T \hat{D}) (I_{NT} - S)X \tilde{\beta}_0 := \mathcal{Z}_{N,T}(1) - \mathcal{Z}_{N,T}(1,2)$. Then, according to Lemmas B.1 and B.4, we know

\[
\mathcal{Z}_{N,T}(1) = \frac{1}{2} \mu^2 \beta_0^*(\tau) h^2 + o_p(h^2). \tag{A14}
\]

Also, Lemma B.5 shows that

\[
\sqrt{NT} \mathcal{Z}_{N,T}(3) \xrightarrow{d} N(0_d, \nu_0 \sigma_0^2 \Sigma^{-1}(\tau)). \tag{A15}
\]

From **Theorem 2**, we have

\[
\sqrt{NT} \mathcal{Z}_{N,T}(4) = \sqrt{NT} \text{op}(\rho_0 - \hat{\rho}) = o_p(1). \tag{A16}
\]

By Equations (A12)–(A16), we have accomplished the proof.

**Appendix C: Optimal Bandwidth Selection**

Due to the existence of individual fixed effects, the traditional cross-validation method may not provide satisfactory results in panel data when selecting the optimal bandwidth. Hence, throughout this article, we adopt the leave-one-unit-out cross-validation method to choose optimal bandwidth. Such method is also used in Li, Chen, and Gao (2011) and Chen, Gao, and Li (2012). The initial value $\hat{\rho}$ is obtained from the parametric spatial panel data model Lee and Yu (2010).

The idea is that firstly, use $(N - 1)T$ observations among all data except for the $i$th unit $(X_i^T, Y_i) \leq t \leq T$ as the training data to obtain the estimate of $\beta(\tau)$, which is denoted as $\hat{\beta}_{(i)}(\tau)$ for each $1 \leq i \leq N$. The optimal bandwidth is the one that minimizes a weighted squared prediction error of the form

\[
h_{\text{opt}} = \arg \min_h \left\{ \frac{1}{M^*T} \left[ Z(\hat{\rho}) - B \left( X, \hat{\beta}_{(i)}^*(-) \right) \right]^T \right\}.
\]

where $M^* = INT - T^{-1} (IT^T) \otimes IN$ is used to delete the unobserved fixed effect due to $M^*D = 0_{NT \times (N-1)}$. $B(\hat{\beta}_{(i)}^*(-) = \left\{ \left( X_1 \hat{\beta}_{(i)}^{-1}(\tau_1) \right)^T, \ldots, \left( X_N \hat{\beta}_{(i)}^{-1}(\tau_1) \right)^T \right\}$, $\ldots$, $\left\{ \left( X_1 \hat{\beta}_{(i)}^{-1}(\tau_T) \right)^T, \ldots, \left( X_N \hat{\beta}_{(i)}^{-1}(\tau_T) \right)^T \right\}$. $\ldots$
Kelejian, H. H., and Prucha, I. R. (1999), “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, 40, 509–533. [1784,1789]

Kelejian, H. H., and Prucha, I. R. (2001), “On the Asymptotic Distribution of the Moran I Test Statistic With Applications,” *Journal of Econometrics*, 104, 219–257. [1788,1799]

Lee, L.-F. (2004), “Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models,” *Econometrica*, 72, 1899–1925. [1784,1786,1788,1797]

Lee, L.-F., and Yu, J. (2010), “Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects,” *Journal of Econometrics*, 154, 165–185. [1784,1785,1786,1788,1789,1801]

——— (2014), “Efficient GMM Estimation of Spatial Dynamic Panel Data Models With Fixed Effects,” *Journal of Econometrics*, 180, 174–197. [1784]

LeSage, J., and Pace, R. K. (2009), *Introduction to Spatial Econometrics*, London: Chapman and Hall/CRC. [1784]

Li, D., Chen, J., and Gao, J. (2011), “Non-Parametric Time-Varying Coefficient Panel Data Models With Fixed Effects,” *Econometrics Journal*, 14, 387–408. [1785,1801]

Li, K. (2017), “Fixed-Effects Dynamic Spatial Panel Data Models and Impulse Response Analysis,” *Journal of Econometrics*, 198, 102–121. [1784]

Li, Q., and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, NJ: Princeton University Press. [1788]

Malikov, E., and Sun, Y. (2017), “Semiparametric Estimation and Testing of Smooth Coefficient Spatial Autoregressive Models,” *Journal of Econometrics*, 199, 12–34. [1784]

Miller Jr, R. G. (1966), *Simultaneous Statistical Inference*, New York: Springer. [1795]

Miller, P. M. (2012), “Nonparametric Trending Regression With Cross-Sectional Dependence,” *Journal of Econometrics*, 169, 4–14. [1785]

Rahav, P., Smyth, R., Zhang, X., and Fenech, J.-P. (2017), “Nonparametric Panel Data Model for Crude Oil and Stock Market Prices in Net Oil Importing Countries,” *Energy Economics*, 67, 255–267. [1784]

Su, L. (2012), “Semiparametric GMM Estimation of Spatial Autoregressive Models,” *Journal of Econometrics*, 167, 543–560. [1784]

Su, L., and Jin, S. (2010), “Profile Quasi-Maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models,” *Journal of Econometrics*, 157, 18–33. [1784,1785,1786,1788]

Su, L., and Ullah, A. (2006), “Profile Likelihood Estimation of Partially Linear Panel Data Models With Fixed Effects,” *Economics Letters*, 92, 75–81. [1785,1786]

Sun, Y. (2016), “Functional-Coefficient Spatial Autoregressive Models With Nonparametric Spatial Weights,” *Journal of Econometrics*, 195, 134–153. [1784]

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. [1784,1785,1787]

Van Biesebroeck, J. (2015), *How Tight is the Link Between Wages and Productivity?: A Survey of the Literature*. ILO. [1791]

White, H. (1996). *Estimation, Inference and Specification Analysis*, Cambridge: Cambridge University Press. [1797]

Yu, J., De Jong, R., and Lee, L.-F. (2008), “Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data With Fixed Effects When Both n and t Are Large,” *Journal of Econometrics*, 146, 118–134. [1784,1788]

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. [1784]