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

Xuan, Liang and Jiti, Gao and xiaodong, Gong

The Australian National University, Monash University, University of Canberra

7 January 2021

MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

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

Xuan Liang, Jiti Gao and Xiaodong Gong

April 2021

Working Paper 05/21

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

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

Xuan Liang[†], Jiti Gao[⋆] and Xiaodong Gong[‡]

The Australian National University[†], Monash University[⋆],

The University of Canberra[‡], Australia

Abstract

This paper considers a semiparametric spatial autoregressive panel data model with fixed effects with time–varying coefficients. The time–varying coefficients are allowed to follow an unknown function of time while the other parameters are assumed to be constants. We propose a "local linear concentrated quasi–maximum likelihood estimation" method to obtain consistent estimators for the spatial autoregressive coefficient, the variance of the error term and the nonparametric time–varying coefficients. We show that the estimators of the parametric components converge at the rate of $\sqrt{NT}$, and those of the nonparametric time–varying coefficients converge at the rate of $\sqrt{NTh}$. Monte Carlo simulations are conducted to illustrate the finite sample performance of our proposed method. We apply our method to study the spatial influences and the time–varying spillover effects in the wage level among 159 Chinese cities.

Key Words: Concentrated quasi–maximum likelihood estimation, local linear estimation, time–varying coefficient.

JEL Classifications: C21, C23

1 Introduction

Panel data analysis has been used widely 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). In classical panel data models, we normally assume independence among different units for the errors. Even though some dependence assumptions can be made in the error term, no clear cross–sectional dependence structure can be modeled in pure panel data models.

Spatial econometric models, which are designed to model spatial interactions, have 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, it has become an active research area in spatial econometrics. One issue with spatial econometric models is that the spatial lag term is endogenous. Various estimation methods have been

The first and the second authors acknowledge the Australian Research Council Discovery Grants Program for its financial support under Grant Numbers: DP150101012 & DP170104421.

proposed to deal with this issue, e.g., 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 logical concepts and details of spatial econometrics can be found in classic spatial econometrics books, e.g., Anselin et al. (2013) and LeSage and Pace (2009). As more temporal data becomes available, spatial panel data models have received considerable attentions. Spatial panel data models with SAR disturbances have been considered in Baltagi et al. (2003) and Kapoor et al. (2007). Fingleton (2008) studied 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 models with individual fixed effects. More recent studies on spatial dynamic panel data models can be found in Yu et al. (2008), Lee and Yu (2014) and Li (2017), etc.

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 semi–parametric spatial econometric models have been studied in the literature. Su and Jin (2010) consider a partially linear SAR model. Su (2012) proposes an SAR model with a nonparametric regressor term. Functional–coefficient SAR models are also studied in Sun (2016) and Malikov and Sun (2017). The former mentioned studies are about cross–sectional data. In terms of nonparametric and semi–parametric panel data models in spatial econometric, Zhang and Shen (2015) consider a partially linear SAR panel data model with functional coefficients and random effects while Sun and Malikov (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 coefficients in these functional–coefficient spatial models are mostly permitted to be unknown smooth functions of 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 some discussion in Cai 2007; Silvapulle et al. 2017). The reason behind could be due to changes in the economic structure or environment, policy reform, or technology development, etc. To accommodate such cases, time–varying coefficient models have been well studied in the existing panel data setting, where the coefficients of the regressors were allowed to be unknown smooth functions of time (Li et al. (2011), Chen et al. (2012) and Robinson (2012)). One advantage of the time–varying coefficient model is that the time variable can be self–explanatory and naturally capture the nonlinear time variation in the coefficients. To our knowledge, the time–varying coefficient model and its estimation has not been well studied in spatial econometrics.

In this paper we propose a semiparametric time–varying coefficient spatial panel data model with fixed effects for large N and T. Specifically, the spatial lag term in the model is assumed to be parametric while the regressor coefficients vary with time, specified as nonparametric functions of time. In addition, regressors can be trending non–stationary. To get consistent estimators for both parametric parameters and nonparametric time–varying components, we propose a "local linear concentrated quasi–maximum likelihood estimation" (LLQML) method. When time–varying coefficients are constant and regressors are stationary, our model reduces to a classical spatial autoregressive panel data model which is fully parametric and has been considered in Lee and Yu (2010). Our model only allows the coefficients of the explanatory variables to be time–varying. A more general model with a nonparametric spatial lag term would be less restrictive since the spatial dependence would be likely to change over time as well. However, as the spatial lag term is endogenous, it is very difficult to estimate such a fully nonparametric model with classical nonparametric techniques. Nevertheless, we would like to study such general models in our future work.

Our contributions in this paper are then 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 feature, as it combines the strengths from different models, including the strong identification of panel data models, the clear interpretation of cross–sectional dependence in spatial models, and the model flexibility of time–varying coefficient models. In the existing literature of spatial econometrics, the regressors are often assumed to be non–stochastic (see, e.g., Lee and Yu 2010, Su and Jin 2010). We relax such assumptions in the theoretical derivations so that the regressors can be trending non–stationary, which renders our model and 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 unknown parameters and 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 for the proposed estimator.

(iii) We evaluate the finite–sample performance of our proposed model under several scenarios. We find our estimator produces robust and consistent estimates, not only for the time–varying feature or non–stationary covariates, but also for time–invariant or 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 time–varying effects of factors on labour compensation in urban China over 1995–2009, a period which 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 paper is organized as follows. Section 2 discusses the model setting and the estimation procedure. Section 3 lays out the assumption. Asymptotic theory of the proposed estimator is 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 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 paper takes the following form:

$$Y_{it} = \rho_0 \sum_{j \neq i} w_{ij} Y_{jt} + X_{it}^\top \boldsymbol{\beta}_{0,t} + \alpha_{0,i} + e_{it}, \quad t = 1, \cdots, T, \quad i = 1, \cdots, N, \tag{2.1}$$

where Y_{it} is the response of location i at time t; $X_{it} = (X_{it1}, \cdots, X_{itd})^\top$ is a d-dimensional vector with the corresponding d-dimensional time–varying coefficient vector function $\boldsymbol{\beta}_{0,t} = (\beta_{0,t1}, \cdots, \beta_{0,td})^\top$; $\alpha_{0,i}$ reflects the unobserved individual fixed effect; w_{ij} describes the spatial weight of observation j to i, which can be a decreasing function of spatial distance between i and j; the scalar parameter ρ_0 measures the strength of spatial dependence; the error component is e_{it} with mean zero and variance σ_0^2; T and N are the time length and the number of spatial units, respectively. In this model, the term $\rho_0 w_{ij} Y_{jt}$ captures the spatial interaction and $X_{it}^\top \boldsymbol{\beta}_{0,t}$ measures the covariate effects over time.

When $\boldsymbol{\beta}_{0,t}$ does not vary over time, it reduces to a vector of constants. Model (2.1) becomes the traditional spatial autoregressive panel data model as discussed in Lee and Yu (2010). If only some components of $\boldsymbol{\beta}_{0,t}$ change over time, model (2.1) gives a partially time–varying spatial panel data model, meaning that a few covariates have effects changing over time while the effects of other covariates stay constant. In this paper, we assume that $\boldsymbol{\beta}_{0,t}$ is fully nonparametric and follows the following specification:

$$\boldsymbol{\beta}_{0,t} = \boldsymbol{\beta}_0(\tau_t), \quad t = 1, \cdots, T, \tag{2.2}$$

where $\boldsymbol{\beta}_0(\cdot)$ is a d-dimensional vector of unknown smooth functions defined on $\mathbb{R}^d$ and $\tau_t = t/T \in (0, 1]$. The same specification is used in Li et al. (2011) and Chen et al. (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 $\boldsymbol{\beta}_0(\tau_t)$ when the constant 1 is included in the regressor X_{it}, the individual fixed effects are assumed to satisfy $\sum_{i=1}^N \alpha_{0,i} = 0$. Such condition is standard in the literature, e.g., Su and Ullah (2006) and Chen et al. (2012). The detailed justification of the identification issue is discussed in Appendix A.1.

Let $\mathbf{0}_n$ and $\mathbf{1}_n$ be the vectors with n elements of zeros and ones, respectively. Denote $0_{m_1\times m_2}$ as an $m_1 \times m_2$ matrix with all zero elements and I_m as the m-dimensional identity matrix. Define an $N\times N$ spatial weight matrix $W = (w_{ij})_{N\times N}$ with zero diagonal elements, i.e., $w_{ii} = 0$, an $N\times(N-1)$ matrix $D_0 = (-\mathbf{1}_{N-1}, I_{N-1})^\top$. A clear matrix form of (2.1) can be written as

$$Y_t \quad = \quad \rho_0 W Y_t + X_t \boldsymbol{\beta}_0(\tau_t) + D_0\boldsymbol{\alpha}_0 + \mathbf{e}_t, \quad t = 1, \cdots, T, \tag{2.3}$$

where $Y_t = (Y_{1t}, \cdots, Y_{Nt})^\top$, $X_t = (X_{1t}, \cdots, X_{Nt})^\top$, $\boldsymbol{\alpha}_0 = (\alpha_{0,2}, \cdots, \alpha_{0,N})^\top$ and $\mathbf{e}_t = (e_{1t}, \cdots, e_{Nt})^\top$. Define an $N \times N$ matrix $S_N(\rho) = I_N - \rho W$. Model (2.3) can further be written as

$$S_N(\rho_0)Y_t \quad = \quad X_t\boldsymbol{\beta}_0(\tau_t) + D_0\boldsymbol{\alpha}_0 + \mathbf{e}_t. \tag{2.4}$$

In (2.4), 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 new response variable as if ρ_0 were known. The goal is to construct consistent estimators for the unknown parameters: the spatial coefficient ρ_0 and the variance σ_0^2, and the unknown time–varying coefficient function $\boldsymbol{\beta}_0(\tau)$.

2.2 Estimation

The joint quasi log-likelihood function of model (2.4) can be written as

$$\log\left(L_{N,T}(\rho, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\beta}(\tau))\right) = -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|S_N(\rho)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} U_N^\top U_N, \tag{2.5}$$

where $U_N = S_N(\rho)Y_t - D_0\boldsymbol{\alpha} - X_t\boldsymbol{\beta}(\tau_t)$. If $\boldsymbol{\beta}(\tau)$ is a vector of constants, the model becomes fully parametric so that the traditional QML method based on (2.5) can be used to estimate parameters (see Lee (2004) and Lee and Yu (2010) for more details). In the presence of the nonparametric time–varying component $\boldsymbol{\beta}(\tau)$ in (2.5), 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 $\boldsymbol{\beta}(\tau)$ for fixed ρ and $\boldsymbol{\alpha}$ by the weighted local likelihood or equivalently the local linear kernel method and denote it as $\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau)$; (ii) Plug in $\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau)$ into (2.5), and obtain the QML estimators $\widehat{\rho}$, $\widehat{\sigma}^2$ and $\widehat{\boldsymbol{\alpha}}$. With ρ and $\boldsymbol{\alpha}$ estimated, the estimator of $\boldsymbol{\beta}(\tau)$ can then be updated by $\widehat{\boldsymbol{\beta}}_{\widehat{\rho},\widehat{\boldsymbol{\alpha}}}(\tau)$. To be more specific:

Step one:

For given values of ρ and $\boldsymbol{\alpha}$, we adopt the weighted/local likelihood approach of Fan and Gijbels (1996) in this step to estimate $\boldsymbol{\beta}(\tau)$.

Let $K(\cdot)$ and h be the kernel function and the smoothing bandwidth, respectively. Assuming that $\boldsymbol{\beta}(\cdot)$ has continuous derivatives of up to the second order, applying Taylor expansion we have $\boldsymbol{\beta}(\tau_t) = \boldsymbol{\beta}(\tau) + \boldsymbol{\beta}'(\tau)(\tau_t - \tau) + O\left((\tau_t - \tau)^2\right)$. where $\boldsymbol{\beta}'(\cdot)$ is the first derivative of $\boldsymbol{\beta}(\cdot)$ and $\tau \in (0, 1]$. We also have that $X_t\boldsymbol{\beta}(\tau_t) \approx X_t\boldsymbol{\beta}(\tau) + \left(\frac{\tau_t-\tau}{h} X_t\right) h\boldsymbol{\beta}'(\tau)$. The weighted/local log-likelihood function can be written as

$$Q(\mathbf{a}, \mathbf{b}) = \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right)\left(-\frac{N}{2}\log(2\pi\sigma^2) + \log|S_N(\rho)|\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right)\widetilde{U}_N^\top \widetilde{U}_N, \quad (2.6)$$

where $\widetilde{U}_N = S_N(\rho)Y_t - D_0\boldsymbol{\alpha} - X_t\mathbf{a} - \left(\frac{\tau_t-\tau}{h} X_t\right)\mathbf{b}$. For given values of ρ, α and σ^2, the maximizer of (2.6) can be obtained equivalently by minimizing the following weighted loss function $L(\mathbf{a}, \mathbf{b})$ with respect to $(\mathbf{a}^\top, \mathbf{b}^\top)^\top$

$$L(\mathbf{a}, \mathbf{b}) \;=\; \sum_{t=1}^{T} K\left(\frac{\tau_t - \tau}{h}\right)\left\{S_N(\rho)Y_t - D_0\boldsymbol{\alpha} - X_t\mathbf{a} - \frac{\tau_t - \tau}{h}X_t\mathbf{b}\right\}^\top \left\{S_N(\rho)Y_t - D_0\boldsymbol{\alpha} - X_t\mathbf{a} - \frac{\tau_t - \tau}{h}X_t\mathbf{b}\right\}.$$

Define an NT-dimensional vector $Y = (Y_1^\top, \cdots, Y_T^\top)^\top$ and an $NT \times NT$ matrix $S_{N,T}(\rho) = I_T \otimes S_N(\rho)$, where $\otimes$ denotes the Kronecker product. Denote also an NT-dimensional vector $Y^*(\rho) = S_{N,T}(\rho)Y$ and an $NT \times (N-1)$ matrix $D = \mathbf{1}_T \otimes D_0$. Function $L(\mathbf{a}, \mathbf{b})$ can be re–written as

$$L(\mathbf{a}, \mathbf{b}) \;=\; \left\{Y^*(\rho) - D\boldsymbol{\alpha} - M(\tau)(\mathbf{a}^\top, \mathbf{b}^\top)^\top\right\}^\top \Omega(\tau)\left\{Y^*(\rho) - D\boldsymbol{\alpha} - M(\tau)(\mathbf{a}^\top, \mathbf{b}^\top)^\top\right\},$$

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

$$M(\tau) = \begin{pmatrix} X_1 & \frac{\tau_1-\tau}{h}X_1 \\ \vdots & \vdots \\ X_T & \frac{\tau_T-\tau}{h}X_T \end{pmatrix} \quad \text{and} \quad \Omega(\tau) = \begin{pmatrix} K\left(\frac{\tau_1-\tau}{h}\right)I_N & & \\ & \ddots & \\ & & K\left(\frac{\tau_T-\tau}{h}\right)I_N \end{pmatrix},$$

respectively. The estimators of $\boldsymbol{\beta}(\tau)$ and $h\boldsymbol{\beta}'(\tau)$ for given $(\rho, \boldsymbol{\alpha})$ are then represented by

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau) \\ h\widehat{\boldsymbol{\beta}}'_{\rho,\boldsymbol{\alpha}}(\tau) \end{pmatrix} = \arg\min_{(\mathbf{a}^\top,\mathbf{b}^\top)^\top} L(\mathbf{a}, \mathbf{b}) = \left\{M^\top(\tau)\Omega(\tau)M(\tau)\right\}^{-1} M^\top(\tau)\Omega(\tau)\left\{Y^*(\rho) - D\boldsymbol{\alpha}\right\}.$$

Denoting a $d \times NT$ matrix $\Phi(\tau) = (I_d, 0_{d\times d})\left\{M^\top(\tau)\Omega(\tau)M(\tau)\right\}^{-1} M^\top(\tau)\Omega(\tau)$, the estimator of time–varying coefficient $\boldsymbol{\beta}_0(\cdot)$ can be expressed by

$$\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau) \;=\; \Phi(\tau)\{Y^*(\rho) - D\boldsymbol{\alpha}\}. \quad (2.7)$$

Step Two:

In this step, we plug in $\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau)$ into the original log-likelihood (2.5) and estimate ρ_0 and σ_0^2 by maximizing the quasi log-likelihood function:

$$\begin{aligned}
\log L_{N,T}(\rho,\sigma^2,\boldsymbol{\alpha}) &= -\frac{NT}{2}\log(2\pi\sigma^2)+T\log|S_N(\rho)| \\
&\quad -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left\{S_N(\rho)Y_t-X_t\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau_t)-D_0\boldsymbol{\alpha}\right\}^\top\left\{S_N(\rho)Y_t-X_t\widehat{\boldsymbol{\beta}}_{\rho,\boldsymbol{\alpha}}(\tau_t)-D_0\boldsymbol{\alpha}\right\} \\
&= -\frac{NT}{2}\log(2\pi\sigma^2)+T\log|S_N(\rho)|-\frac{1}{2\sigma^2}\left\{\tilde{Y}(\rho)-\tilde{D}\boldsymbol{\alpha}\right\}^\top\left\{\tilde{Y}(\rho)-\tilde{D}\boldsymbol{\alpha}\right\}, \qquad (2.8)
\end{aligned}$$

where $\tilde{Y}(\rho)=(I_{NT}-S)Y^*(\rho)$ and $\tilde{D}=(I_{NT}-S)D$ are the smoothing versions of $Y^*(\rho)$ and D by $NT\times NT$ matrix $S=\tilde{X}\tilde{\Phi}$, in which the $NT\times dT$ matrix $\tilde{X}$ is a diagonal block matrix with the $N\times d$ matrix X_t being its t-th diagonal block, and $dT\times NT$ matrix $\tilde{\Phi}=(\Phi(\tau_1)^\top,\cdots,\Phi(\tau_T)^\top)^\top$. Taking the derivative of (2.8) with respect to $\boldsymbol{\alpha}$ and setting it to be zero, we have

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

Define an $NT\times NT$ matrix $Q_{N,T}=I_{NT}-\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top$. Plugging $\widehat{\boldsymbol{\alpha}}(\rho)$ into (2.8) leads to

$$\log L_{N,T}(\rho,\sigma^2) = -\frac{NT}{2}\log(2\pi\sigma^2)+T\log|S_N(\rho)|-\frac{1}{2\sigma^2}\tilde{Y}^\top(\rho)Q_{N,T}\tilde{Y}(\rho). \qquad (2.9)$$

Then, taking the derivative of (2.9) with respect to σ^2 and equating it to zero, we have the estimator of σ^2 as the following function of ρ:

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

Replacing σ^2 with $\widehat{\sigma}^2(\rho)$ in (2.9), we obtain the concentrated quasi log-likelihood function:

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

Therefore, we estimate the parameters $\boldsymbol{\theta}_0=(\rho_0,\sigma_0^2)^\top$ and $\boldsymbol{\alpha}_0$ by $\widehat{\boldsymbol{\theta}}=(\widehat{\rho},\widehat{\sigma}^2)^\top$ and $\widehat{\boldsymbol{\alpha}}$ as follows:

$$\widehat{\rho}=\max_{\rho}\log L_{N,T}(\rho),\quad \widehat{\sigma}^2=\frac{1}{NT}\tilde{Y}^\top(\widehat{\rho})Q_{N,T}\tilde{Y}(\widehat{\rho}),\quad \widehat{\boldsymbol{\alpha}}=(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top\tilde{Y}(\widehat{\rho}).$$

Finally, the updated estimator of $\boldsymbol{\beta}_0(\tau)$ is obtained by plugging $\widehat{\rho}$ and $\widehat{\boldsymbol{\alpha}}$ into (2.7):

$$\widehat{\boldsymbol{\beta}}(\tau) = \Phi(\tau)\{Y^*(\widehat{\rho})-D\widehat{\boldsymbol{\alpha}}\}. \qquad (2.10)$$

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 $\|\boldsymbol{a}\|_s = (\sum_{i=1}^n |a_i|^s)^{1/s}$ as the s-norm ($s \geq 1$) for any generic vector $\boldsymbol{a} = (a_1, \cdots, a_n)^\top$. For any generic $m \times m$ matrix $A = (a_{ij})_{m\times m}$, define the diagonal vector of A as $\text{diag}(A) = (a_{11}, \cdots, a_{mm})^\top$, $\|A\|_1 = \max_{1\leq j\leq m} \sum_{i=1}^m |a_{ij}|$ and $\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^m |a_{ij}|$ as the 1-norm and ∞-norm, respectively.

Assumption 1. *Let d-dimensional vector $X_{it} = g(\tau_t) + \mathbf{v}_{it}$ contain a deterministic time trend part $g(\tau) = (g_1(\tau), \cdots, g_d(\tau))^\top$ and a random component $\mathbf{v}_{it} = (v_{it1}, \cdots, v_{itd})^\top$.*

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

(ii) Denote $\mathbf{v}_t = (\mathbf{v}_{1t}, \cdots, \mathbf{v}_{Nt})^\top$. Suppose that $\{\mathbf{v}_t, t \geq 1\}$ is a strictly stationary sequence with mean zero and α-mixing with mixing coefficient $\alpha_{\text{mix,N}}(t)$, and that there exists a function $\alpha_{\text{mix}}(t)$ and a constant δ such that $\alpha_{\text{mix,N}}(t) \leq \alpha_{\text{mix}}(t)$ and $\sum_{t=1}^\infty \alpha_{\text{mix}}(t)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$.

(iii) Let $\{\mathbf{v}_{it}, i \geq 1, t \geq 1\}$ be identically distributed in index i. In addition, we assume $\mathrm{E}|v_{itk}|^{4+\delta} < \infty$ for $k = 1, \cdots, d$ and let $\mathrm{E}(\mathbf{v}_{it}\mathbf{v}_{it}^\top) = \Sigma_{\mathbf{v}} = (\sigma_{\mathbf{v}}^{(k_1,k_2)})_{d\times d}$ where $\sigma_{\mathbf{v}}^{(k_1,k_2)} = E(v_{itk_1}v_{itk_2})$.

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

Assumption 1(i) assumes that the time trend $g(\tau)$ is continuous, which is a standard assumption to model the trend in X_{it}. With this structure, the regressors can be either stationary or non–stationary over time. Specially, if $g(\tau_t)$ reduces to a constant vector, it covers the case with stationary X_{it}. Otherwise, X_{it} is generally non–stationary. By assuming this, we take the non–stationarity of X_{it} into account when we derive the theoretical properties of the estimators. The reason why $g(\tau)$ is defined over $(0, 1]$ is to scale the time domain to a bounded set, for the same reason as for $\boldsymbol{\beta}_0(\tau)$. Note that $g(\tau)$ here can be further generalized to allow for an individual time trend $g_i(\tau)$. To make theoretical derivations less complicated, we consider the homogeneous trend. The trend $g(\tau)$ can be estimated by $\widehat{g}(\tau) = \frac{1}{N}\sum_{i=1}^N \widehat{g}_i(\tau)$, where $\widehat{g}_i(\tau) = \frac{\sum_{t=1}^T K(\frac{\tau_t-\tau}{h})X_{it}}{\sum_{t=1}^T K(\frac{\tau_t-\tau}{h})}$.

To allow for serial dependence in $\{\mathbf{v}_t\}$, we impose the stationarity and α-mixingness in Assumption 1(ii) on $\mathbf{v}_t$ (see, e.g., examples and discussions in Fan and Yao 2008; Gao 2007). Since $\mathbf{v}_t$ is a high dimensional vector depending on N, we need to assume that there exists an upper bound $\alpha_{\text{mix}}(t)$. Similar assumptions can be found in Chen et al. (2012). Moreover, $\sum_{t=1}^\infty \alpha_{\text{mix}}(t)^{\delta/(4+\delta)} < \infty$ is commonly used in the literature; see, e.g., Dou et al. (2016). This assumption is weaker than the exponentially decaying α-mixing coefficient $\alpha_{\text{mix}}(t) = c_\alpha\psi^t$ for $0 < c_\alpha < \infty$ and $0 < \psi < 1$; see, e.g., Chen et al. (2012, 2019).

It is worth noting that we only assume $\{\mathbf{v}_{it}, i \geq 1, t \geq 1\}$ to be identically distributed in index i, which is weaker than the i.i.d. assumption for covariates in Sun and Malikov (2018). This also means the cross–sectional dependence for $\mathbf{v}_{it}$ across index i can be allowed as long as the mixing condition for $\mathbf{v}_t = (\mathbf{v}_{1t}, \cdots, \mathbf{v}_{Nt})^\top$ in Assumption 1(ii) is satisfied.

Meanwhile, it is allowed the constant 1 term to be included in X_{it}. When $g_1(\tau_t)$ reduces to constant 1 and v_{it1} degenerates to $v_{it1} \equiv 0$, $X_{it1} \equiv 1$ is exactly the constant 1 term.

Assumption 2. *The error term* $\{\mathbf{e}_t = (e_{1t}, \cdots, e_{Nt})^\top : t \geq 1\}$ *is a stationary process such that*

(i) for some $\delta > 0$, $\sup_{i\geq 1} \mathrm{E}(|e_{it}|^{4+\delta}) < \infty$;

(ii) $\mathrm{E}(\mathbf{e}_t|\mathcal{E}_{t-1}) = \mathbf{0}_N$ *and* $\mathrm{E}(\mathbf{e}_t\mathbf{e}_t^\top|\mathcal{E}_{t-1}) = \sigma_0^2 I_N$, *where* $\mathcal{E}_{t-1} = \mathcal{F}_V \vee \sigma\langle \mathbf{e}_1, \cdots, \mathbf{e}_{t-1}\rangle$, *is the* σ*-field generated by* $\mathcal{F}_V \cup \langle \mathbf{e}_1, \cdots, \mathbf{e}_{t-1}\rangle$ *and* $\mathcal{F}_V = \sigma\langle\{\mathbf{v}_{it} : i \geq 1, t \geq 1\}\rangle$ *is the* σ*-field generated by* $\{\mathbf{v}_{it} : i \geq 1, t \geq 1\}$;

(iii) Given $\mathcal{E}_{t-1}$, $\widetilde{\mathbf{e}}_i = (e_{i1}, \cdots, e_{iT})^\top$ *is a vector of conditionally independent random errors with* $\mathrm{E}(e_{it}^j|\mathcal{E}_{t-1}) = \mathrm{E}(e_{it}^j) = m_j \in \mathbb{R}$ *for* $j = 3$ *and* 4.

Remark: Assumption 2 summarizes the conditions on the error term. Assumption 2 (ii) implies $\mathrm{E}(e_{it}|\mathcal{F}_V) = 0$, indicating X_{it} is strictly exogenous. Sun and Malikov (2018) also consider the exogenous covariates. A sufficient condition for the conditional independence of $\widetilde{\mathbf{e}}_i$ in Assumption 2 (iii) is that e_{it} are independent in both i and t (e.g., see Assumption 2 of Yu et al. 2008) and $\{e_{it}\}$ is independent of $\mathcal{F}_V$. 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, e.g., the proof of Theorem 2 in Appendix A.2. Further, this technique of the proof can be adapted to model (2.3) if a cross–sectional dependent random structure is specified. Specifically, we still impose Assumption 2 but we replace $\mathbf{e}_t$ in model (2.3) by a cross–sectional dependent random error $\boldsymbol{\varepsilon}_t = L\mathbf{e}_t$, where L is a non–stochastic matrix and $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top|\mathcal{E}_{t-1}) = \sigma_0^2 LL^\top$ 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 below), similar theoretical results can be established but more complicated derivations are involved.

Assumption 3. *(i) The kernel function* $K(\cdot)$ *is a continuous and symmetric probability density function with compact support.*

(ii) The bandwidth is assumed to satisfy $h \to 0$ *as* $\min(N, T) \to \infty$, $Th \to \infty$ *and* $NTh^8 \to 0$.

Remark: Assumption 3 first imposes the conditions on the kernel function used in estimation, which is common in the literature; see, e.g., Chen et al. (2012). Conditions on the bandwidth h along with T and N are also considered in Assumption 3; see similar conditions in Assumption A5 of Chen et al. (2012).

Assumption 4. W *is a non–stochastic spatial weight matrix with zero diagonals and is uniformly bounded in both row and column sums in absolute value (for short, UB), i.e.,* $\sup_{n\geq 1} \|W\|_1 < \infty$ *and* $\sup_{n\geq 1} \|W\|_\infty < \infty$.

Assumption 5. *$S_N(\rho)$ is invertible for all $\rho \in \triangle$, where $\triangle$ is a compact interval with the true value ρ_0 as an interior point. Also, $S_N(\rho)$ and $S_N^{-1}(\rho)$ are both UB, uniformly in $\rho \in \triangle$.*

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 ρ. 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 $\boldsymbol{\beta}_0(\cdot)$ has continuous derivatives of up to the second order.*

Assumption 7. *The fixed effects satisfy that $\|D_0\boldsymbol{\alpha}_0\|_1 < \infty$.*

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, e.g., Condition 2.1 of Li and Racine (2007), Assumption 2.7 of Gao (2007) and Assumption A3 of Chen et al. (2012). Assumption 7 guarantees the uniform boundedness of the sum of absolute fixed effects.

To proceed, we need to introduce the following notation. Let $S_N = S_N(\rho_0)$, $S_{N,T} = S_{N,T}(\rho_0)$, $G_N(\rho) = WS_N^{-1}(\rho)$, $G_N = G_N(\rho_0)$, $G_{N,T} = I_T \otimes G_N$, $P_{N,T} = (I_{NT} - S)^\top Q_{N,T}(I_{NT} - S)$ and $R_{N,T} = G_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0)$ where $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_0^\top(\tau_1), \cdots, \boldsymbol{\beta}_0^\top(\tau_T)\right)^\top$.

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

Remark: Assumption 8 is a condition for the identification of ρ_0, which is similar to Assumption 8 in Lee (2004), Assumption 4 in Lee and Yu (2010), Assumption 7 in Su and Jin (2010). This assumption requires implicitly that after removing the time trend, the generated regressor $R_{N,T}$ and the original regressor $X_{N,T} = (X_{11}, \cdots, X_{NT})^\top$ ($NT \times d$ matrix) are not asymptotically multicollinear. To check the suitability of this assumption in practice, their correlation coefficients or variance inflation factors (VIF) can be used to determine if there exist any multicollinearity problems.

4 Asymptotic Properties

Asymptotic consistency of $\widehat{\boldsymbol{\theta}} = (\widehat{\rho}, \widehat{\sigma}^2)^\top$ to $\boldsymbol{\theta}_0 = (\rho_0, \sigma_0^2)^\top$ is established in Theorem 1. The asymptotic distributions of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\beta}}(\tau)$ are provided in Theorems 2 and 3. The proofs of these theorems are given in Appendix A.2.

Theorem 1. *Under Assumptions 1-8, $\boldsymbol{\theta}_0$ is globally identifiable and $\widehat{\boldsymbol{\theta}}$ is consistent to $\boldsymbol{\theta}_0$.*

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

Theorem 2. *Under Assumptions 1-8, as $T\to\infty$ and $N\to\infty$ simultaneously, then*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}_2, \Sigma_{\boldsymbol{\theta}_0}^{-1} + \Sigma_{\boldsymbol{\theta}_0}^{-1}\Omega_{\boldsymbol{\theta}_0}\Sigma_{\boldsymbol{\theta}_0}^{-1}\right), \tag{4.1}$$

where $\Omega_{\boldsymbol{\theta}_0} = \lim_{N,T\to\infty}\Omega_{NT,\boldsymbol{\theta}_0}$ with $\Omega_{NT,\boldsymbol{\theta}_0}$ being defined by

$$\Omega_{NT,\boldsymbol{\theta}_0} = \begin{pmatrix} \frac{2m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}G_{N,T}))}{NT\sigma_0^4} + \frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}^2)}{NT\sigma_0^4} & \frac{m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}))}{2NT\sigma_0^6} + \frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}p_{ii})}{2NT\sigma_0^6} \\ \frac{m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}))}{2NT\sigma_0^6} + \frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}p_{ii})}{2NT\sigma_0^6} & \frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}p_{ii}^2)}{4\sigma_0^8 NT} \end{pmatrix},$$

in which p_{ii} and $(gp)_{ii}$ are the i-th main diagonal elements of $P_{N,T}$ and $P_{N,T}G_{N,T}$, respectively, and $\Sigma_{\boldsymbol{\theta}_0} = \begin{pmatrix} \frac{1}{\sigma_0^2}\Psi_{R,R} + c_1 & \frac{c_2}{\sigma_0^2} \\ \frac{c_2}{\sigma_0^2} & \frac{1}{2\sigma_0^4} \end{pmatrix}$ *is positive definite as shown in Lemma C.9.*

Since we use the QML method to estimate $\boldsymbol{\theta}_0$, it relaxes the normality assumption on the error term but it 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 $m_4 = 3\sigma_0^2$, the asymptotic covariance matrix in (4.1) reduces to $\Sigma_{\boldsymbol{\theta}_0}^{-1}$, as shown in the following proposition.

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

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}_2, \Sigma_{\boldsymbol{\theta}_0}^{-1}\right)$$

when $\{e_{it}, i\geq 1, t\geq 1\}$ is independent and identically normally distributed with $\Omega_{\boldsymbol{\theta}_0} = 0_{2\times 2}$ due to $m_3 = 0$ and $m_4 = 3\sigma_0^4$.

Define $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$. Let $\boldsymbol{\beta}_0''(\tau)$ be the second derivative of $\boldsymbol{\beta}_0(\tau)$. An asymptotic distribution for $\widehat{\boldsymbol{\beta}}(\tau)$ is established in the following theorem.

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

$$\sqrt{NTh}\left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau) - b_{\boldsymbol{\beta}}(\tau)h^2 + o_{\mathrm{P}}(h^2)\right) \xrightarrow{d} N\left(\mathbf{0}_d, \sigma_0^2\nu_0\Sigma_X^{-1}(\tau)\right), \tag{4.2}$$

provided that $\Sigma_X(\tau)$ is positive definite for each given τ, where $b_{\boldsymbol{\beta}}(\tau) = \frac{1}{2}\mu_2\boldsymbol{\beta}_0''(\tau)$ and $\Sigma_X(\tau) = g(\tau)g(\tau)^\top + \Sigma_{\mathbf{v}}$.

Thus, the rate of convergence of $\widehat{\boldsymbol{\beta}}(\tau)$ is $\sqrt{NTh}$, which is the fastest possible rate in the nonparametric structure. It is also clear that the covariance matrix is related to $g(\tau)$ since it involves the trend of X_{it}. When X_{it} is stationary, the asymptotic covariance matrix in (4.2) reduces to a constant matrix $\sigma_0^2\nu_0(\mu_X\mu_X^\top+\Sigma_{\mathbf{v}})^{-1}$ where $\mu_X=E(X_{it})$.

One can use the following sample version to estimate the unknown covariance matrices involved: $\widehat{\Sigma}_{\boldsymbol{\theta}_0}=\begin{pmatrix}\frac{1}{\widehat{\sigma}^2}\widehat{\Psi}_{R,R}+\widehat{c}_1 & \frac{\widehat{c}_2}{\widehat{\sigma}^2}\\ \frac{\widehat{c}_2}{\widehat{\sigma}^2} & \frac{1}{2\widehat{\sigma}^4}\end{pmatrix}$, $\widehat{\Omega}_{\boldsymbol{\theta}_0}=\Omega_{NT,\boldsymbol{\theta}_0}$ and $\widehat{\Sigma}_X(\tau)=\widehat{g}(\tau)\widehat{g}(\tau)^\top+\widehat{\Sigma}_v$, where $\widehat{\Psi}_{R,R}=(NT)^{-1}R_{N,T}^\top P_{N,T}R_{N,T}$, $\widehat{c}_1=\mathrm{tr}(G_{N,T}^2+G_{N,T}^\top G_{N,T})/NT$, $\widehat{c}_2=\mathrm{tr}(G_{N,T})/NT$, $\widehat{g}(\tau)=\frac{\sum_{i,t}K\left(\frac{\tau_t-\tau}{h}\right)X_{it}}{\sum_{i,t}K\left(\frac{\tau_t-\tau}{h}\right)}$, $\widehat{\Sigma}_{\mathbf{v}}=(NT)^{-1}\sum_{i,t}\widehat{\mathbf{v}}_{it}\widehat{\mathbf{v}}_{it}^\top$ and $\widehat{\mathbf{v}}_{it}=X_{it}-\widehat{g}(\tau_t)$. The consistency of these sample estimators is shown in Lemma C.10 of Appendix C 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 method 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:

$$Y_t \;=\; \rho_0 WY_t+X_t\boldsymbol{\beta}_0(\tau_t)+D_0\boldsymbol{\alpha}_0+\mathbf{e}_t,\quad t=1,\cdots,T.$$

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, and then following Kelejian and Prucha (1999). We also 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 neighbouring units to each unit. Then, the regressor is set to be $X_{it}=(1,X_{it2})^\top$ where $X_{it2}=g(\tau_t)+v_{it2}$. The component v_{it2} is the i-th element of an N-dimensional vector $\mathbf{v}_t$ generated by $\mathbf{v}_t=0.2\mathbf{v}_{t-1}+N(\mathbf{0}_N,\Sigma^*)$ with $\Sigma^*=(0.5^{|i-j|})_{N\times N}$ for $-99\le t\le T$ and $\mathbf{v}_{-100}=\mathbf{0}_N$. It is obvious that $\{v_{it2}\}$ is both serially and cross–sectionally dependent. The error term e_{it} is independent and identically generated from the distribution of $N(0,1)$ so that $\sigma_0^2=1$. The fixed effects follow $\alpha_{0,i}=T^{-1}\sum_{t=1}^T v_{it2}$ for $i=1,\cdots,N-1$ and $\alpha_{0,1}=-\sum_{i=2}^N\alpha_{0,i}$. The time–varying coefficient vector is set to be $\boldsymbol{\beta}_0(\tau)=(\beta_{0,1}(\tau),\beta_{0,2}(\tau))^\top$ where $\beta_{0,1}(\tau)$ and $\beta_{0,1}(\tau)$ represent the time–varying coefficient associated with the constant 1 and X_{it2} in X_{it}, respectively. Various simulation settings are defined by changing the specification of $g(\tau)$, $\boldsymbol{\beta}_0(\tau)$ and ρ_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) $g(\tau) = 2sin(\pi\tau)$;

- Set II (Setting of $\boldsymbol{\beta}_0(\tau)$): (II-1) $\boldsymbol{\beta}_0(\tau) = (1, 1)^\top$; (II-2) $\boldsymbol{\beta}_0(\tau) = (1, 1 + 2\tau + 2\tau^2)^\top$, (II-3) $\boldsymbol{\beta}_0(\tau) = (1 + 3\tau, 1 + 2\tau + 2\tau^2)^\top$;

- Set III (Setting of spatial coefficient): (IV-1) $\rho_0 = 0.3$, (IV-2) $\rho_0 = 0.7$.

Each of these sets (and combinations of them) will generate data of 1) 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; 2) coefficient $\boldsymbol{\beta}_0(\tau)$ with different time–varying feature (Set II): from Set II-1 to Set II-2, $\boldsymbol{\beta}_0(\tau)$ changes from time–invariant, partially time–varying to fully time-varying respectively; and 3) different spatial autoregressive coefficient or spatial dependence among cross–sectional units (Set III). For each scenario, simulations are conducted on 1000 replications. The Epanechnikov kernel $K(u) = 3/4(1 - u^2)I(|u| \leq 1)$ 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. For short, we call it "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 ρ_0 and σ_0^2 under different settings of $g(\tau)$ and $\boldsymbol{\beta}_0(\tau)$, together with those of Lee–Yu model (with ρ_0 fixed at 0.3). A few comments can be made on the results.

Firstly, our estimates of ρ_0 and σ_0^2 are consistent under all settings as the means and SDs of the bias of ρ_0 and σ_0^2 are getting smaller when either N or T is increasing. It shows the robustness of our model in both the time and cross dimensions.

Secondly, if the data are generated by a time–invariant process (Set II-1), the estimates of ρ_0 and σ_0^2 from Lee–Yu model are consistent with smaller biases compared to ours. It makes senses as a time–invariant spatial panel date model is a special case of our model. However, when the coefficient of the covariate involves time–varying features (Set II-2 and Set II-3) in the data generating process, the estimates of ρ_0 and σ_0^2 from 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 ρ_0 and 1.9 for σ_0^2. 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 Lee–Yu model will lead to inconsistent estimation.

Thirdly, comparing different data generating processes, if the data are generated from a fully time–varying model (Set III-3), our estimates have smallest biases and SDs, followed by a 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

Table 1: Means and standard deviations of bias of $\widehat{\rho}$ ($\rho_0 = 0.3, \sigma_0^2 = 1$).

(a) Our model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0662	-0.0493	-0.0208	-0.0166	-0.0140	-0.0053	-0.0131	-0.0122	-0.0046
		(0.1123)	(0.0837)	(0.0563)	(0.0533)	(0.0429)	(0.0275)	(0.0511)	(0.0425)	(0.0272)
	T=15	-0.0403	-0.0293	-0.0131	-0.0097	-0.0071	-0.0030	-0.0078	-0.0060	-0.0026
		(0.0861)	(0.0678)	(0.0451)	(0.0428)	(0.0333)	(0.0226)	(0.0423)	(0.0332)	(0.0225)
	T=30	-0.0244	-0.0162	-0.0074	-0.0056	-0.0029	-0.0011	-0.0051	-0.0025	-0.0010
		(0.0573)	(0.0465)	(0.0310)	(0.0267)	(0.0224)	(0.0151)	(0.0264)	(0.0224)	(0.0150)
(I-2)	T=10	-0.0662	-0.0493	-0.0208	-0.0099	-0.0095	-0.0028	-0.0085	-0.0084	-0.0022
		(0.1123)	(0.0837)	(0.0563)	(0.0516)	(0.0427)	(0.0273)	(0.0513)	(0.0432)	(0.0275)
	T=15	-0.0403	-0.0293	-0.0131	-0.0055	-0.0040	-0.0009	-0.0051	-0.0033	-0.0003
		(0.0861)	(0.0678)	(0.0451)	(0.0428)	(0.0336)	(0.0228)	(0.0427)	(0.0335)	(0.0230)
	T=30	-0.0244	-0.0162	-0.0074	-0.0032	-0.0009	0.0004	-0.0026	-0.0004	0.0008
		(0.0573)	(0.0465)	(0.0310)	(0.0267)	(0.0224)	(0.0151)	(0.0269)	(0.0225)	(0.0152)
(I-3)	T=10	-0.0663	-0.0500	-0.0222	-0.0223	-0.0199	-0.0126	-0.0201	-0.0183	-0.0109
		(0.1060)	(0.0805)	(0.0547)	(0.0445)	(0.0369)	(0.0248)	(0.0447)	(0.0375)	(0.0252)
	T=15	-0.0409	-0.0301	-0.0144	-0.0162	-0.0148	-0.0111	-0.0145	-0.0129	-0.0086
		(0.0847)	(0.0665)	(0.0435)	(0.0373)	(0.0297)	(0.0207)	(0.0378)	(0.0303)	(0.0212)
	T=30	-0.0234	-0.0170	-0.0082	-0.0127	-0.0111	-0.0085	-0.0112	-0.0087	-0.0061
		(0.0561)	(0.0452)	(0.0302)	(0.0241)	(0.0207)	(0.0149)	(0.0244)	(0.0208)	(0.0145)

(b) Lee–Yu model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0137	-0.0136	-0.0054	0.0768	0.0816	0.0878	0.1899	0.1913	0.2002
		(0.0989)	(0.0761)	(0.0537)	(0.0887)	(0.0757)	(0.0528)	(0.0935)	(0.0799)	(0.0571)
	T=15	-0.0073	-0.0064	-0.0026	0.0897	0.0929	0.0977	0.2018	0.2045	0.2083
		(0.0808)	(0.0648)	(0.0437)	(0.0758)	(0.0616)	(0.0437)	(0.0809)	(0.0644)	(0.0467)
	T=30	-0.0069	-0.0056	-0.0028	0.0982	0.1028	0.1027	0.2104	0.2125	0.2134
		(0.0546)	(0.0452)	(0.0299)	(0.0543)	(0.0423)	(0.0328)	(0.0556)	(0.0452)	(0.0320)
(I-2)	T=10	-0.0137	-0.0136	-0.0054	0.2608	0.2594	0.2634	0.4073	0.4054	0.4059
		(0.0989)	(0.0761)	(0.0536)	(0.0919)	(0.0762)	(0.0541)	(0.0593)	(0.0486)	(0.0338)
	T=15	-0.0074	-0.0064	-0.0026	0.2725	0.2716	0.2713	0.4168	0.4141	0.4121
		(0.0808)	(0.0648)	(0.0437)	(0.0773)	(0.0605)	(0.0440)	(0.0479)	(0.0391)	(0.0280)
	T=30	-0.0069	-0.0056	-0.0028	0.2807	0.2779	0.2755	0.4196	0.4169	0.4147
		(0.0546)	(0.0452)	(0.0299)	(0.0515)	(0.0419)	(0.0293)	(0.0327)	(0.0264)	(0.0194)
(I-3)	T=10	-0.0121	-0.0103	-0.0041	0.1732	0.1756	0.1861	0.3170	0.3191	0.3265
		(0.0876)	(0.0683)	(0.0492)	(0.0784)	(0.0686)	(0.0474)	(0.0692)	(0.0587)	(0.0408)
	T=15	-0.0051	-0.0066	-0.0040	0.1855	0.1865	0.1926	0.3271	0.3290	0.3331
		(0.0742)	(0.0593)	(0.0392)	(0.0652)	(0.0551)	(0.0394)	(0.0586)	(0.0479)	(0.0347)
	T=30	-0.0047	-0.0050	-0.0029	0.1902	0.1944	0.1971	0.3303	0.3340	0.3371
		(0.0501)	(0.0402)	(0.0285)	(0.0476)	(0.0391)	(0.0284)	(0.0428)	(0.0338)	(0.0241)

Table 2: Means and standard deviations of bias of $\widehat{\sigma}^2$ ($\rho_0 = 0.3, \sigma_0^2 = 1$).

(a) Our model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.1677	-0.1157	-0.0576	-0.1609	-0.1110	-0.0517	-0.1565	-0.1076	-0.0499
		(0.1349)	(0.1070)	(0.0802)	(0.1332)	(0.1062)	(0.0794)	(0.1310)	(0.1057)	(0.0791)
	T=15	-0.1392	-0.0963	-0.0479	-0.1349	-0.0907	-0.0422	-0.1317	-0.0885	-0.0414
		(0.1058)	(0.0933)	(0.0661)	(0.1052)	(0.0932)	(0.0656)	(0.1042)	(0.0927)	(0.0656)
	T=30	-0.1249	-0.0817	-0.0415	-0.1195	-0.0765	-0.0361	-0.1183	-0.0755	-0.0357
		(0.0759)	(0.0645)	(0.0459)	(0.0751)	(0.0644)	(0.0458)	(0.0753)	(0.0645)	(0.0457)
(I-2)	T=10	-0.1677	-0.1157	-0.0576	-0.1513	-0.1030	-0.0462	-0.1496	-0.1015	-0.0449
		(0.1349)	(0.1070)	(0.0802)	(0.1318)	(0.1065)	(0.0796)	(0.1310)	(0.1068)	(0.0802)
	T=15	-0.1392	-0.0963	-0.0479	-0.1283	-0.0852	-0.0375	-0.1276	-0.0831	-0.0360
		(0.1058)	(0.0933)	(0.0661)	(0.1046)	(0.0929)	(0.0660)	(0.1047)	(0.0935)	(0.0660)
	T=30	-0.1249	-0.0817	-0.0415	-0.1149	-0.0721	-0.0325	-0.1133	-0.0706	-0.0312
		(0.0759)	(0.0645)	(0.0459)	(0.0761)	(0.0650)	(0.0460)	(0.0760)	(0.0654)	(0.0465)
(I-3)	T=10	-0.1662	-0.1149	-0.0574	-0.1417	-0.0945	-0.0387	-0.1423	-0.0965	-0.0430
		(0.1336)	(0.1064)	(0.0802)	(0.1318)	(0.1067)	(0.0802)	(0.1316)	(0.1065)	(0.0795)
	T=15	-0.1388	-0.0953	-0.0473	-0.1174	-0.0756	-0.0308	-0.1203	-0.0794	-0.0363
		(0.1050)	(0.0933)	(0.0657)	(0.1044)	(0.0932)	(0.0654)	(0.1046)	(0.0931)	(0.0655)
	T=30	-0.1239	-0.0810	-0.0410	-0.1066	-0.0648	-0.0273	-0.1100	-0.0696	-0.0324
		(0.0764)	(0.0648)	(0.0460)	(0.0754)	(0.0652)	(0.0467)	(0.0760)	(0.0653)	(0.0460)

(b) Lee–Yu model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0216	-0.0170	-0.0074	1.2871	1.3011	1.2988	1.6606	1.6689	1.6666
		(0.1474)	(0.1153)	(0.0824)	(0.4128)	(0.3395)	(0.2321)	(0.5083)	(0.4168)	(0.2815)
	T=15	-0.0083	-0.0072	-0.0030	1.3137	1.2993	1.2992	1.6963	1.6724	1.6611
		(0.1175)	(0.1003)	(0.0681)	(0.3459)	(0.2689)	(0.1943)	(0.4357)	(0.3400)	(0.2388)
	T=30	-0.0086	-0.0039	-0.0023	1.3162	1.3194	1.3020	1.6955	1.6872	1.6642
		(0.0846)	(0.0692)	(0.0476)	(0.2466)	(0.1838)	(0.1335)	(0.2981)	(0.2317)	(0.1661)
(I-2)	T=10	-0.0216	-0.0170	-0.0074	1.9066	1.9021	1.8800	2.3929	2.3879	2.3492
		(0.1474)	(0.1153)	(0.0824)	(0.5532)	(0.4513)	(0.3006)	(0.6869)	(0.5590)	(0.3731)
	T=15	-0.0083	-0.0072	-0.0030	1.9372	1.8994	1.8689	2.4061	2.3633	2.3205
		(0.1175)	(0.1003)	(0.0681)	(0.4755)	(0.3698)	(0.2546)	(0.5856)	(0.4605)	(0.3123)
	T=30	-0.0086	-0.0039	-0.0023	1.9383	1.9122	1.8717	2.3997	2.3658	2.3162
		(0.0846)	(0.0692)	(0.0476)	(0.3211)	(0.2497)	(0.1778)	(0.3918)	(0.3087)	(0.2191)
(I-3)	T=10	-0.0210	-0.0173	-0.0076	2.0818	2.0406	1.9949	2.9138	2.8141	2.6892
		(0.1477)	(0.1155)	(0.0827)	(0.5574)	(0.4374)	(0.2958)	(0.6650)	(0.5303)	(0.3507)
	T=15	-0.0083	-0.0068	-0.0025	2.1024	2.0328	1.9842	2.9110	2.7826	2.6595
		(0.1171)	(0.1005)	(0.0682)	(0.4473)	(0.3461)	(0.2520)	(0.5418)	(0.4245)	(0.2930)
	T=30	-0.0090	-0.0036	-0.0023	2.1010	2.0478	1.9840	2.9043	2.7783	2.6477
		(0.0844)	(0.0692)	(0.0476)	(0.3224)	(0.2429)	(0.1691)	(0.3835)	(0.2840)	(0.2014)

Table 3: Means and standard deviations of MSE of $\widehat{\boldsymbol{\beta}}(\tau) = (\widehat{\beta}_1(\tau), \widehat{\beta}_2(\tau))^\top$ ($\rho_0 = 0.3, \sigma_0^2 = 1$).

(a) $\widehat{\beta}_1(\tau)$

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	0.0792	0.0480	0.0193	0.0447	0.0296	0.0124	0.0774	0.0531	0.0209
		(0.1034)	(0.0537)	(0.0206)	(0.0455)	(0.0285)	(0.0116)	(0.0887)	(0.0653)	(0.0225)
	T=15	0.0463	0.0287	0.0128	0.0283	0.0176	0.0081	0.0505	0.0307	0.0145
		(0.0587)	(0.0314)	(0.0126)	(0.0283)	(0.0165)	(0.0074)	(0.0597)	(0.0352)	(0.0150)
	T=30	0.0202	0.0124	0.0058	0.0125	0.0077	0.0036	0.0218	0.0140	0.0065
		(0.0205)	(0.0123)	(0.0058)	(0.0107)	(0.0066)	(0.0032)	(0.0237)	(0.0158)	(0.0072)
(I-2)	T=10	0.1983	0.1166	0.0446	0.1287	0.0840	0.0337	0.2105	0.1403	0.0540
		(0.2750)	(0.1498)	(0.0524)	(0.1491)	(0.1015)	(0.0347)	(0.2535)	(0.1922)	(0.0650)
	T=15	0.1116	0.0672	0.0291	0.0826	0.0493	0.0238	0.1365	0.0791	0.0384
		(0.1462)	(0.0792)	(0.0307)	(0.0881)	(0.0548)	(0.0239)	(0.1634)	(0.1010)	(0.0439)
	T=30	0.0480	0.0287	0.0130	0.0366	0.0224	0.0103	0.0581	0.0362	0.0163
		(0.0528)	(0.0301)	(0.0141)	(0.0383)	(0.0223)	(0.0101)	(0.0707)	(0.0428)	(0.0183)
(I-3)	T=10	0.1946	0.1164	0.0443	0.1016	0.0692	0.0285	0.1665	0.1158	0.0477
		(0.3156)	(0.1710)	(0.0594)	(0.1308)	(0.0901)	(0.0333)	(0.2311)	(0.1621)	(0.0607)
	T=15	0.1119	0.0647	0.0276	0.0647	0.0386	0.0194	0.1075	0.0648	0.0331
		(0.1572)	(0.0881)	(0.0313)	(0.0834)	(0.0442)	(0.0207)	(0.1442)	(0.0792)	(0.0387)
	T=30	0.0459	0.0288	0.0125	0.0288	0.0186	0.0089	0.0474	0.0313	0.0151
		(0.0541)	(0.0365)	(0.0146)	(0.0342)	(0.0207)	(0.0106)	(0.0610)	(0.0393)	(0.0191)

(b) $\widehat{\beta}_2(\tau)$

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	0.0419	0.0263	0.0123	0.0482	0.0326	0.0167	0.0467	0.0312	0.0169
		(0.0410)	(0.0263)	(0.0112)	(0.0405)	(0.0266)	(0.0122)	(0.0395)	(0.0253)	(0.0120)
	T=15	0.0289	0.0181	0.0076	0.0352	0.0236	0.0125	0.0341	0.0231	0.0127
		(0.0276)	(0.0182)	(0.0068)	(0.0288)	(0.0192)	(0.0085)	(0.0277)	(0.0183)	(0.0086)
	T=30	0.0139	0.0083	0.0040	0.0181	0.0127	0.0087	0.0178	0.0130	0.0089
		(0.0133)	(0.0076)	(0.0035)	(0.0141)	(0.0090)	(0.0054)	(0.0135)	(0.0091)	(0.0055)
(I-2)	T=10	0.0419	0.0263	0.0123	0.0464	0.0309	0.0168	0.0462	0.0305	0.0172
		(0.0410)	(0.0263)	(0.0112)	(0.0387)	(0.0248)	(0.0120)	(0.0372)	(0.0242)	(0.0118)
	T=15	0.0289	0.0181	0.0076	0.0344	0.0232	0.0124	0.0342	0.0232	0.0128
		(0.0276)	(0.0182)	(0.0068)	(0.0282)	(0.0185)	(0.0084)	(0.0275)	(0.0178)	(0.0083)
	T=30	0.0139	0.0083	0.0040	0.0177	0.0128	0.0084	0.0179	0.0131	0.0089
		(0.0133)	(0.0076)	(0.0035)	(0.0136)	(0.0090)	(0.0054)	(0.0134)	(0.0090)	(0.0056)
(I-3)	T=10	0.0374	0.0238	0.0111	0.0406	0.0287	0.0163	0.0403	0.0283	0.0150
		(0.0388)	(0.0248)	(0.0102)	(0.0327)	(0.0217)	(0.0106)	(0.0313)	(0.0216)	(0.0109)
	T=15	0.0260	0.0161	0.0070	0.0305	0.0215	0.0124	0.0299	0.0206	0.0103
		(0.0254)	(0.0162)	(0.0065)	(0.0235)	(0.0152)	(0.0084)	(0.0236)	(0.0152)	(0.0077)
	T=30	0.0123	0.0074	0.0035	0.0169	0.0132	0.0086	0.0158	0.0112	0.0062
		(0.0123)	(0.0070)	(0.0032)	(0.0122)	(0.0088)	(0.0058)	(0.0120)	(0.0083)	(0.0044)

estimates for ρ_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 estimates 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 coefficient $\beta_{0,1}(\tau)$ and $\beta_{0,2}(\tau)$ are reported, respectively, where for an estimate $\widehat{\beta}_k(\tau)$ $(k = 1, 2)$, the MSE is defined by

$$\mathrm{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left\{\widehat{\beta}_k(\tau_t) - \beta_{0,k}(\tau_t)\right\}^2.$$

The results show that both the means and SDs of MSE for $\widehat{\beta}_k(\tau)$ $(k = 1, 2)$ decrease when either N or T increases, confirming the consistency of our estimation. The results for a different spatial coefficient $\rho_0 = 0.7$ are similar to our benchmark case of $(\rho_0 = 0.3)$ and the corresponding simulation results can be found in Tables D.8-D.10 of Appendix D in 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 labour and is closely related to productivity; see for example Van Biesebroeck (2015) for a literature review and Combes et al. (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 labour 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, e.g., Braid (2002) and Baltagi et al. (2012). Secondly, China has experienced an unprecedented economic growth and change in 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 labour, and how renumeration was determined. The series of reforms in housing, the health system, and education also changed both demand and supply of labour. 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 labour input should be included in the model. The only variable that is available to us is the size of 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 labour 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.

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 yuans, 1994 price)
$log(Asset)$	Log value of total asset (million yuans, 1994 price)
GDP_m	Proportion of GDP by the manufactural sector
GDP_s	Proportion of GDP by the services sector
Emp_{ms}	Proportion of employed persons in the manufactural sector out of the non–agricultural sector

Our data are derived from Statistic Year Books of China (various years), for 1995 to 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 prefectural levels in China. A prefectural city in China means a city that 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 non–stationary. To checked whether the macro–level regressors considered here are trending stationary, we first fit the unknown trend in the regressors with the local linear estimation method. After removing the

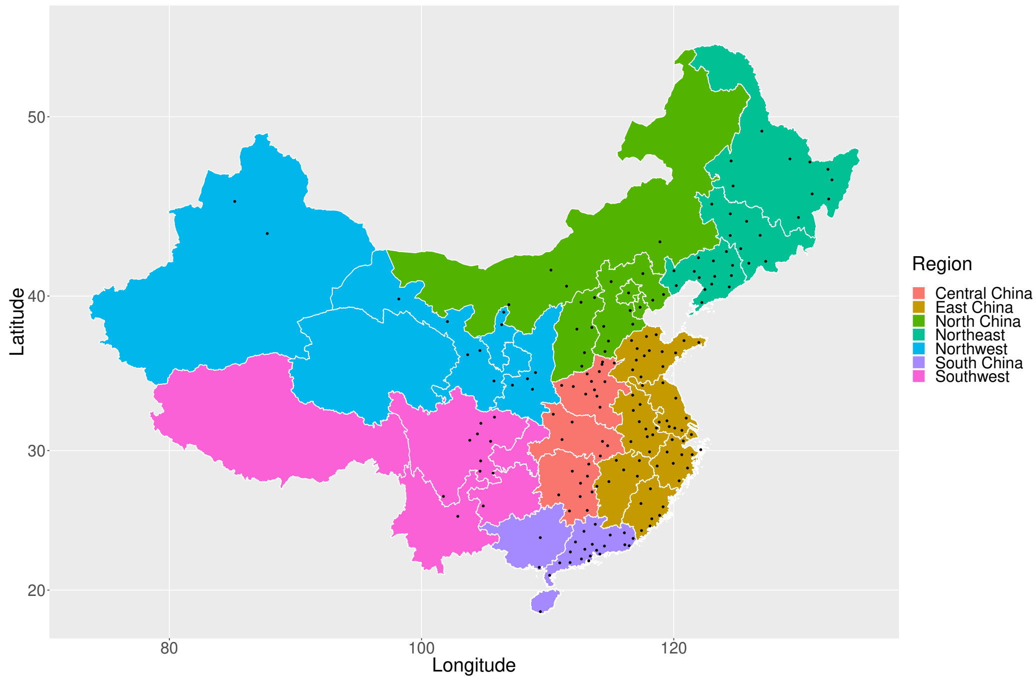

Figure 1: Map of 159 Chinese cities where each color represents one of the following regions from Central China (CC), East China (EC), North China (NC), North East (NE), North West (NW), South China (SC) and South West(SW).

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

Residuals in Regressors	W_{tbar}	p-value	Z_{tbar}	p-value
$log(FDI)$	-17.7893	< 0.0001	-17.9520	< 0.0001
$log(Asset)$	-5.3533	< 0.0001	-5.3726	< 0.0001
GDP_m	-12.4843	< 0.0001	-12.5689	< 0.0001
GDP_s	-12.5017	< 0.0001	-12.6032	< 0.0001
Emp_{ms}	-13.6499	< 0.0001	-13.7323	< 0.0001

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 Equation (4.10) and (4.6) in Im et al. (2003) for the IPS test statistics W_{tbar} and Z_{tbar}, 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 bandwidth choice is crucial for nonparametric kernel estimation. We first estimate the model with the optimal bandwidth ($h_{opt} = 0.4000$) obtained by the leave–one–unit–out cross–validation method, and then compare the results with a number of different bandwidths around it:

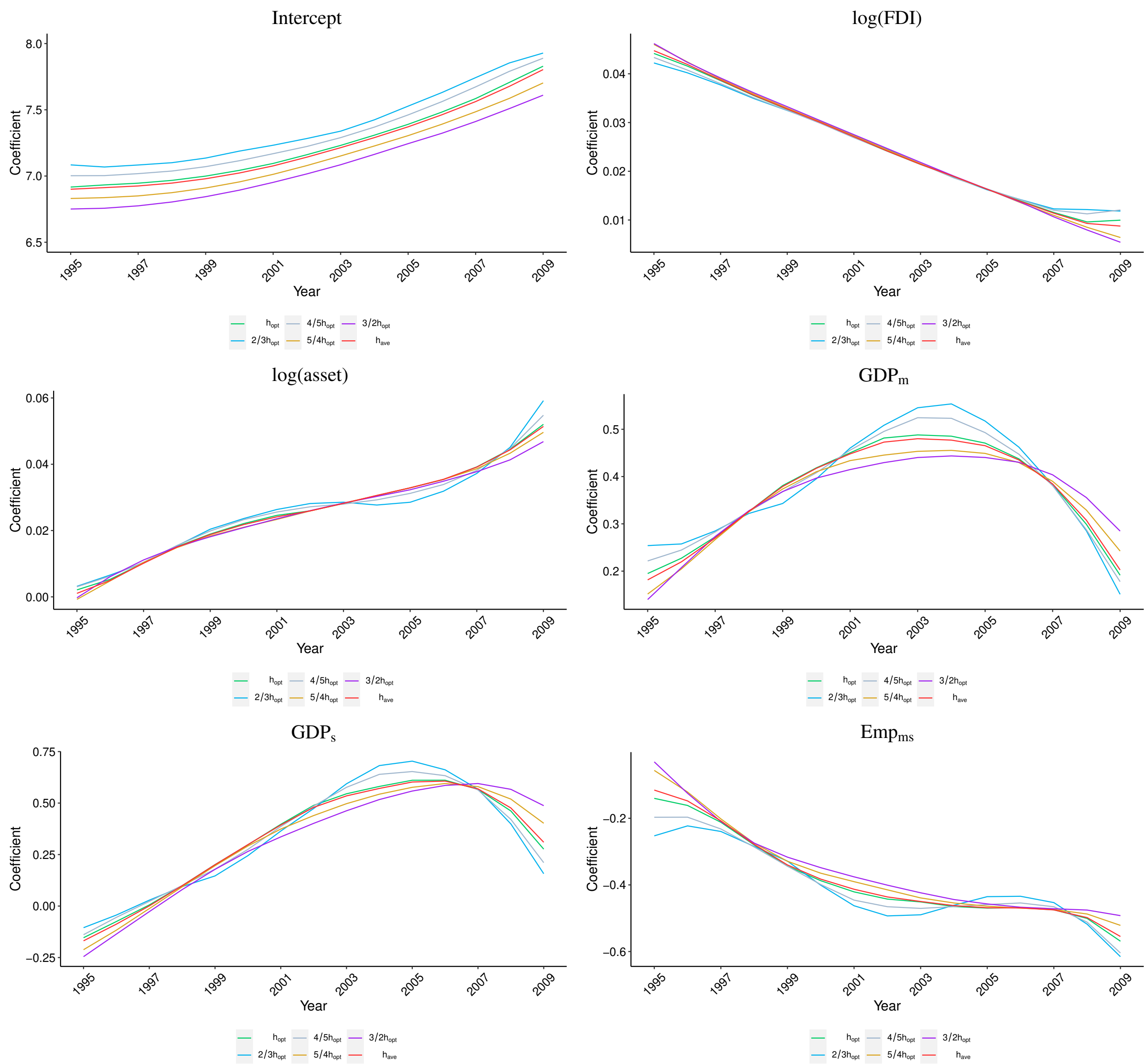

Figure 2: Estimated curves of the time–varying coefficients under different bandwidths for the whole country (159 cities).

$2/3h_{opt}$, $4/5h_{opt}$, $5/4h_{opt}$ and $3/2h_{opt}$. Figure 2 shows that the results under different bandwidths are consistent. Table 6 reports that the estimates of the spatial coefficient ρ_0 and variance σ_0^2, showing that the estimates are quite similar in these specifications. Given the robustness of the results, we decide to use the average bandwidth $h_{ave} = 0.4173$ of those five bandwidths for the rest of the paper.

Table 6: Estimates of parameters under different bandwidths.

	h_{opt}	$2/3h_{opt}$	$4/5h_{opt}$	$5/4h_{opt}$	$3/2h_{opt}$
ρ_0	0.1214 (0.0185)	0.1052 (0.0185)	0.1126 (0.0185)	0.1332 (0.0186)	0.1418 (0.0186)
σ_0^2	0.0062 (0.0002)	0.0061 (0.0002)	0.0062 (0.0002)	0.0064 (0.0002)	0.0064 (0.0002)

We estimate the model for China as a whole and then for East China (Figure 3). Table 7 reports

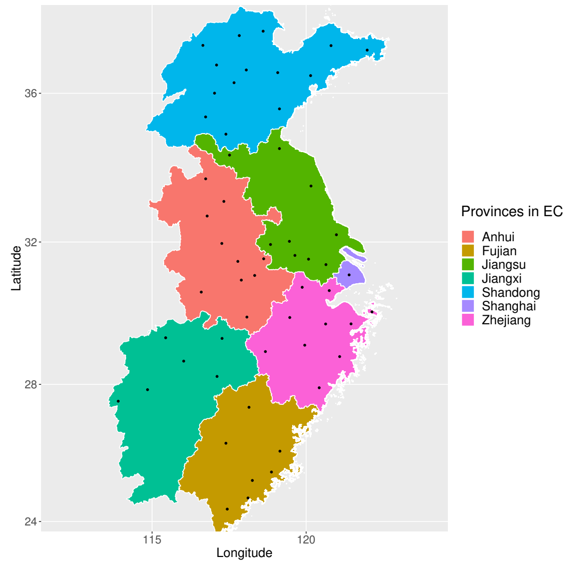

Figure 3: Map of 56 cities in Region EC where each color represents different province.

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
ρ_0	0.1240*** (0.0185)	0.2239*** (0.0242)
σ_0^2	0.0063*** (0.0002)	0.0044*** (0.0002)

the estimates of the spatial coefficient ρ_0, σ_0^2 and time average $\boldsymbol{\beta}_0(\tau)$ for the whole country and for East China, respectively. The significance of the spatial coefficient estimate reflects the spatial dependence and confirms the existences 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 percent. If capital increases by one percent, real wage would increase by 0.026 percent. The estimates also show that economic structure affects wages as well. For example, if the share of manufactural or service sector increases by one percent, the average real wage would increase by 0.365 or 0.326 percent, respectively, but the increase in wage would be 0.380 percent higher if the size of service sector is one percent larger relative to the manufactural sector. Comparison between the whole country and East China illustrates that the spatial dependence is much bigger in EC than the whole country. It makes sense as the economic

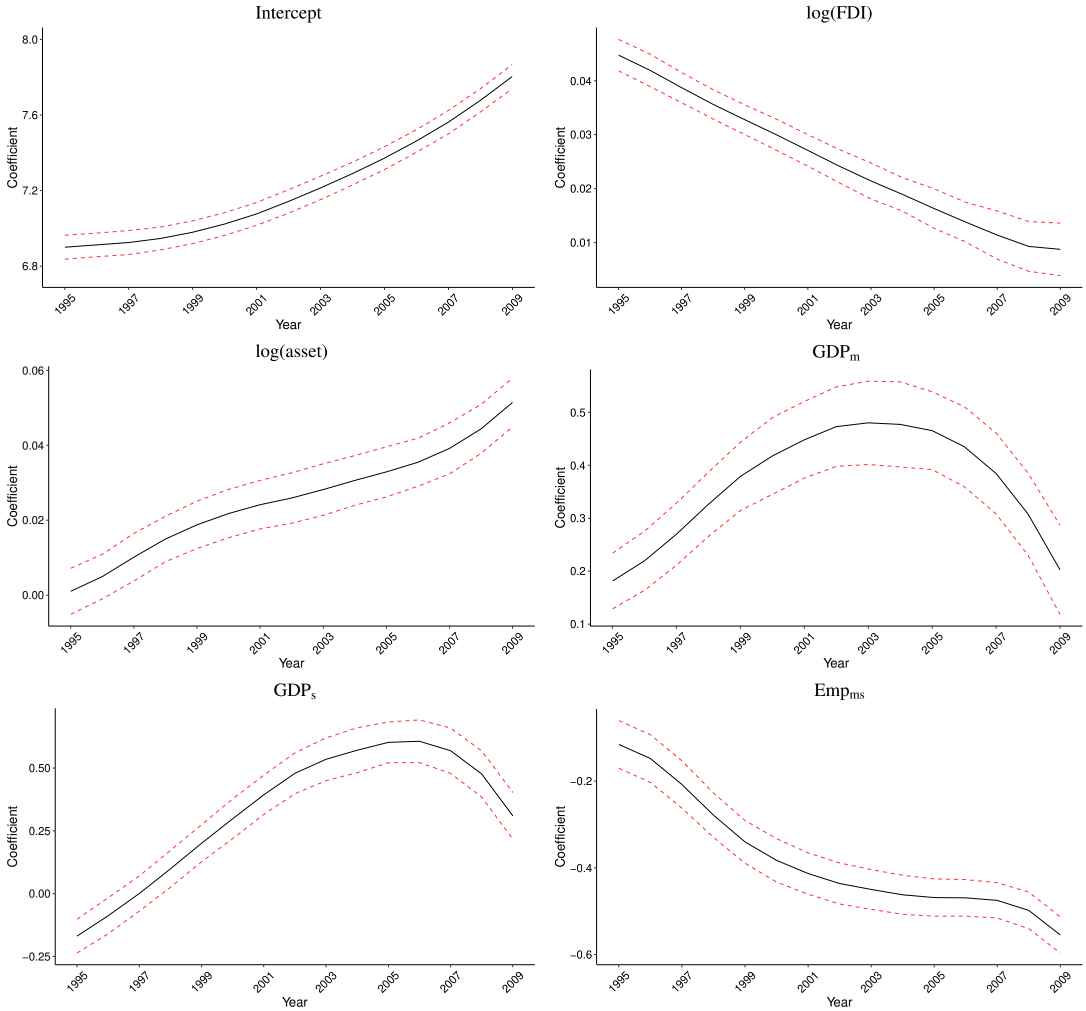

Figure 4: Estimated curves and 95% confidence bands of time–varying coefficients for the whole country (159 cities) with bandwidth $h_{ave} = 0.4173$.

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 to the local economy is not as large as for less developed parts of the country. The average effect of capital and the share of manufactural in EC are larger than the whole country. This is because it is the most dense region in the country and its manufactural sectors are more developed.

Figure 4 displays the time–varying coefficient curves for each variable 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-hows, which also push up the labour demand, but as the domestic economy catches up, the impact of FDI on the labour market becomes less important. Meanwhile the impact of capital on the wage level kept increasing over time. This could be because as the

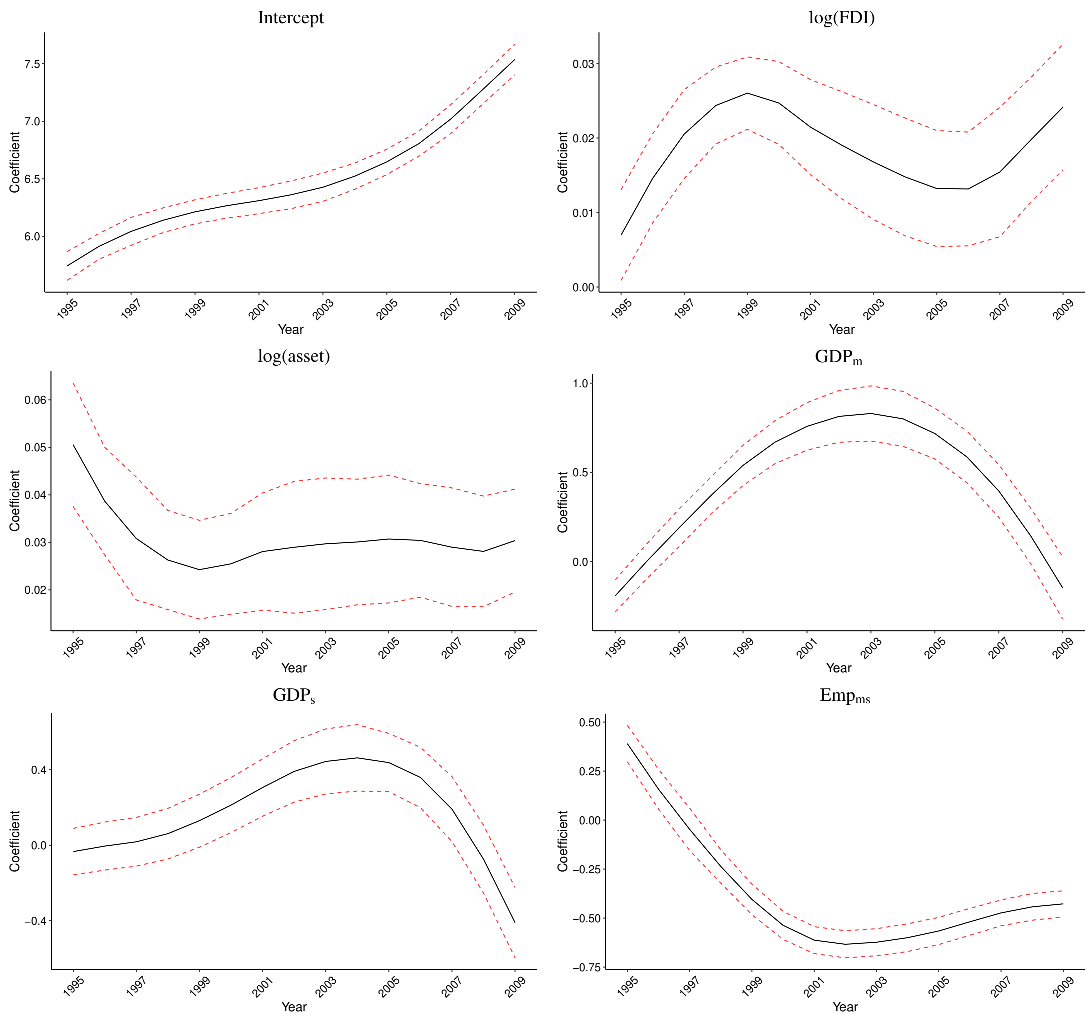

Figure 5: Estimated curves and 95% confidence bands of time–varying coefficients for Region of East China with bandwidth $h_{ave} = 0.3477$.

economy becomes less labour intensive, as the capital level increases, the demand for labour 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 appear quite strong in EC as well. The results imply if a time-invariant model were used, the impacts of these variables would be estimated with biases.

In addition, we have conducted model diagnostics. To save space, we just report the results of the model for the whole country since the results of model for 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 have carried out the Box-Pierce test (see Box and Pierce 1970) on the estimated residuals for each city. It is worth noting that we are interested in a set of hypotheses

$$\mathrm{H}_{i,0}: \text{the estimated residuals for each city } i \text{ are white noise.} \quad \text{v.s. } \mathrm{H}_{i,1}: \text{otherwise.} \tag{6.1}$$

for cities $i = 1, \cdots, N$. 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, we apply the Benjamini and Hochberg (1995)'s multiple testing procedure that controls the false discovery rate (FDR) of (6.1) 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 (6.1) 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 residuals. Moreover, Assumption 8 is also valid due to small VIF (less than 3) shown in Table D.11 of the supplementary material. All the aforementioned diagnostics results support the validity of our regression model.

7 Conclusion

We have considered a semiparametric spatial autoregressive panel data model with fixed effects. This model is designed particularly for situations where covariate effects on the dependent variables change over time so that they follow unknown functions of 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 quasi–maximum likelihood 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 is 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 coefficient is misspecified to be constant, using the standard time–invariant spatial panel data model would lead to inconsistent estimation while our proposed model is always consistent and robust.

We have also applied the proposed model to study labour 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 labour compensation have changed over time. The results also imply that for a fast changing economies such as China, many important economic parameters may not be consistently estimated with a time-invariant model.

Acknowledgements

The authors would like to thank the Co–Editor, the Associate Editor and two referees for their constructive comments and suggestions on an earlier version of this submission. The authors also acknowledge seminar participants from several seminars for their comments and suggestions.

References

Anselin, L., Florax, R., and Rey, S. J. (2013). *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer Science & Business Media.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.

Baltagi, B. H., Blien, U., and Wolf, K. (2012). A dynamic spatial panel data approach to the german wage curve. *Economic Modelling*, 29(1):12–21.

Baltagi, B. H., Song, S. H., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117(1):123–150.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526.

Braid, R. M. (2002). The spatial effects of wage or property tax differentials, and local government choice between tax instruments. *Journal of Urban Economics*, 51(3):429–445.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *the Annals of Probability*, 1(1):19–42.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chen, J., Li, D., and Linton, O. (2019). A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables. *Journal of Econometrics*, 212(1):155–176s.

Chung, K. L. (2001). *A Course in Probability Theory*. Academic Press.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation, Monographs in Spatial Environmental Systems Analysis*. London: Pion Limited.

Combes, P.-P., Démurger, S., and Li, S. (2017). Productivity gains from agglomeration and migration in the people's republic of china between 2002 and 2013. *Asian Development Review*, 34(2):184–200.

Dou, B., Parrella, M. L., and Yao, Q. (2016). Generalized yule–walker estimation for spatio-temporal models with unknown diagonal coefficients. *Journal of Econometrics*, 194(2):369–382.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and Their Applications*, 84(2):313–342.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall/CRC.

Fan, J. and Yao, Q. (2008). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Science & Business Media.

Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis*, 3(1):27–44.

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, London.

Gao, Y. and Li, K. (2013). Nonparametric estimation of fixed effects panel data models. *Journal of Nonparametric Statistics*, 25(3):679–693.

Hsiao, C. (2014). *Analysis of Panel Data*. Cambridge University Press.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of econometrics*, 115(1):53–74.

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

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

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

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

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

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

Lee, L.-F. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180(2):174–197.

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

Li, D., Chen, J., and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *Econometrics Journal*, 14(3):387–408.

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

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

Lin, Z. and Bai, Z. (2011). *Probability inequalities*. Springer Science & Business Media.

Malikov, E. and Sun, Y. (2017). Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *Journal of Econometrics*, 199(1):12–34.

Miller Jr, R. G. (1966). *Simultaneous Statistical Inference*. Springer.

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, 169(1):4–14.

Seber, G. A. F. (2007). *A Matrix Handbook for Statisticians*. Wiley-Interscience.

Silvapulle, 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.

Su, L. (2012). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics*, 167(2):543–560.

Su, L. and Jin, S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics*, 157(1):18–33.

Su, L. and Ullah, A. (2006). Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters*, 92(1):75–81.

Sun, Y. (2016). Functional-coefficient spatial autoregressive models with nonparametric spatial weights. *Journal of Econometrics*, 195(1):134–153.

Sun, Y. and Malikov, E. (2018). Estimation and inference in functional-coefficient spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 203(2):359–378.

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

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

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

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.

Appendix A

A.1 Justification of Identification Condition $\sum_{i=1}^{N} \alpha_{0,i} = 0$

Considering the specification of $\boldsymbol{\beta}_{0,t} = \boldsymbol{\beta}_0(\tau_t)$ in (2.2), our model (2.1) becomes

$$Y_{it} = \rho_0 \sum_{j \neq i} w_{ij} Y_{jt} + X_{it}^\top \boldsymbol{\beta}_0(\tau_t) + \alpha_{0,i} + e_{it}, \quad t = 1, \cdots, T, \quad i = 1, \cdots, N.$$

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

Without loss of generality, let $X_{it} = (X_{it1}, X_{it,-1}^\top)^\top$, $\boldsymbol{\beta}_0(\tau_t) = (\beta_{0,1}(\tau_t), \boldsymbol{\beta}_{0,-1}(\tau_t)^\top)^\top$ where $X_{it1} = 1$ and $X_{it,-1} = (X_{it2}, \cdots, X_{itd})^\top$ and $\boldsymbol{\beta}_{0,-1}(\tau_t) = (\beta_{0,2}(\tau_t), \cdots, \beta_{0,d}(\tau_t))^\top$. Then, our model becomes

$$Y_{it} = \rho_0 \sum_{j \neq i} w_{ij} Y_{jt} + \beta_{0,1}(\tau_t) + X_{it,-1}^\top \boldsymbol{\beta}_{0,-1}(\tau_t) + \alpha_{0,i} + e_{it}, \quad t = 1, \cdots, T, \quad i = 1, \cdots, N.$$

Let $Y_{\cdot t} = N^{-1} \sum_{i=1}^N Y_{it}$, $W_{\cdot t} = N^{-1} \sum_{i=1}^N \sum_{j \neq i} w_{ij} Y_{jt}$, $X_{\cdot t,-1} = N^{-1} \sum_{i=1}^N X_{it,-1}$, $\overline{\alpha} = N^{-1} \sum_{i=1}^N \alpha_{0,i}$ and $e_{\cdot t} = N^{-1} \sum_{i=1}^N e_{it}$. We then have

$$Y_{\cdot t} = \rho_0 W_{\cdot t} + \beta_{0,1}(\tau_t) + X_{\cdot t,-1}^\top \boldsymbol{\beta}_{0,-1}(\tau_t) + \overline{\alpha} + e_{\cdot t}, \tag{A.1}$$

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

A.2 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 $Q_{N,T}(\rho) = \max_{\sigma^2} \mathrm{E}\{\log L_{N,T}(\boldsymbol{\theta})\}$, where $\boldsymbol{\theta} = (\rho, \sigma^2)$. In order to show the consistency of $\widehat{\boldsymbol{\theta}}$, it suffices to show

$$\frac{1}{NT}\{\log L_{N,T}(\rho) - Q_{N,T}(\rho)\} \xrightarrow{\mathrm{P}} 0 \text{ uniformly on } \triangle, \tag{A.2}$$

and the uniqueness identification condition that

$$\limsup_{N,T \to \infty} \max_{\rho \in N_\epsilon^c(\rho_0)} \frac{1}{NT}\{Q_{N,T}(\rho) - Q_{N,T}(\rho_0)\} < 0 \text{ for any } \epsilon > 0, \tag{A.3}$$

by using White (1996) and Lee (2004), where $N_\epsilon^c(\rho_0)$ is the complement of an open neighbourhood of ρ_0 on $\triangle$ of diameter ϵ.

(1) Proof of (A.2). Observe that $Q_{N,T}(\rho) = -\frac{NT}{2}\{\log(2\pi)+1\} - \frac{NT}{2}\log\{\sigma^{2*}(\rho)\} + T\log|S_N(\rho)|$, where

$$\sigma^{2*}(\rho) = (NT)^{-1}\mathrm{E}\{\tilde{Y}^\top(\rho)Q_{N,T}\tilde{Y}(\rho)\} = (NT)^{-1}\mathrm{E}\left\{Y^{*\top}(\rho)P_{N,T}Y^*(\rho)\right\}.$$

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

$$\begin{aligned}
&\sigma^{2*}(\rho)\frac{1}{NT}(\rho_0-\rho)^2\mathrm{E}(R_{N,T}^\top P_{N,T}R_{N,T}) + \frac{2}{NT}(\rho_0-\rho)\mathrm{E}((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}R_{N,T}) \\
&+\frac{1}{NT}\mathrm{E}((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)) + \frac{\sigma_0^2}{NT}\mathrm{tr}\{S_{N,T}^{-1\top}S_{N,T}^\top(\rho)\mathrm{E}(P_{N,T})S_{N,T}(\rho)S_{N,T}^{-1}\}.
\end{aligned} \tag{A.4}$$

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

$$\begin{aligned}
\widehat{\sigma}^2(\rho) &= \frac{1}{NT}Y^{*\top}(\rho)P_{N,T}Y^*(\rho) \\
&= \frac{1}{NT}(\rho_0-\rho)^2R_{N,T}^\top P_{N,T}R_{N,T} + \frac{2}{NT}(\rho_0-\rho)(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}R_{N,T} + \frac{1}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0) \\
&\quad+\frac{2}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e} + \frac{2(\rho_0-\rho)}{NT}R_{N,T}^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e} \\
&\quad+\frac{1}{NT}\mathbf{e}^\top S_{N,T}^{-1\top}S_{N,T}(\rho)^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}.
\end{aligned} \tag{A.5}$$

To show (A.2), it is sufficient to show that

$$\widehat{\sigma}^2(\rho) - \sigma^{2*}(\rho) = o_{\mathrm{P}}(1), \text{ uniformly on } \triangle. \tag{A.6}$$

According to Lemma B.2, we know

$$\begin{aligned}
\frac{1}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0) = o_{\mathrm{P}}(1), &\quad \frac{1}{NT}\mathrm{E}((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)) = o(1), \\
\frac{1}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}R_{N,T} = o_{\mathrm{P}}(1), &\quad \frac{1}{NT}\mathrm{E}((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}R_{N,T}) = o(1), \\
\frac{1}{NT}R_{N,T}^\top P_{N,T}R_{N,T} = \frac{1}{NT}&\mathrm{E}(R_{N,T}^\top P_{N,T}R_{N,T}) + o_{\mathrm{P}}(1) \to \Psi_{R,R}.
\end{aligned}$$

By (A.4) and (A.5),

$$\begin{aligned}
\widehat{\sigma}^2(\rho) - \sigma^{2*}(\rho) &= \frac{2}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e} + \frac{2(\rho_0-\rho)}{NT}R_{N,T}^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e} \\
&\quad+\frac{1}{NT}\mathbf{e}^\top S_{N,T}^{-1\top}S_{N,T}(\rho)^\top P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e} - \sigma^2(\rho) + o_{\mathrm{P}}(1) \\
&:= 2H_1(\rho) + 2(\rho_0-\rho)H_2(\rho) + H_3(\rho) - \sigma^2(\rho) + o_{\mathrm{P}}(1),
\end{aligned}$$

where $\sigma^2(\rho) = (NT)^{-1}\sigma_0^2\mathrm{tr}\{S_{N,T}^{-1\top}S_{N,T}^\top(\rho)\mathrm{E}(P_{N,T})S_{N,T}(\rho)S_{N,T}^{-1}\}$.

According to Lemma B.3, we get (A.6) so that (A.2) holds.

(2) Proof of (A.3). Consider an auxiliary SAR panel process: $Y_t = \rho_0WY_t + \mathbf{e}_t$ where $\mathbf{e}_t \sim N(\mathbf{0}_N, \sigma_0^2I_N)$ and $t = 1,\cdots,T$. Denote the log-likelihood of this model as $\log L_{N,T}^a(\rho,\sigma^2)$. Let $\tilde{\sigma}^2(\rho) =$

$(NT)^{-1}\sigma_0^2 \mathrm{tr}\{S_{N,T}^{-1\top} S_{N,T}^\top(\rho) S_{N,T}(\rho) S_{N,T}^{-1}\}$. It can be verified that

$$\max_{\sigma^2} \mathrm{E}_a\{\log L_{N,T}^a(\rho,\sigma^2)\} = -\frac{NT}{2}\{\log(2\pi)+1\} - \frac{NT}{2}\log\{\tilde\sigma^2(\rho)\} + T\log|S_N(\rho)| := \tilde Q_{N,T}(\rho),$$

where E_a denotes the expectation under this auxiliary model. Hence, for any $\rho \in \triangle$, we have $\tilde Q_{N,T}(\rho) \le \max_{\rho,\sigma^2} \mathrm{E}_a\{\log L_{N,T}^a(\rho,\sigma^2)\} = \mathrm{E}_a\{\log L_{N,T}^a(\rho_0,\sigma_0^2)\} = \tilde Q_{N,T}(\rho_0)$ so that

$$\frac{1}{NT}\{\tilde Q_{N,T}(\rho) - \tilde Q_{N,T}(\rho_0)\} \le 0 \quad \text{uniformly on } \triangle\,.$$

The term in (A.3) can be rewritten as

$$\frac{1}{NT}\{Q_{N,T}(\rho) - Q_{N,T}(\rho_0)\} = \frac{1}{NT}\{\tilde Q_{N,T}(\rho) - \tilde Q_{N,T}(\rho_0)\} + \frac{1}{2}H_4(\rho) + \frac{1}{2}H_5,$$

where $H_4(\rho) = \log\{\tilde\sigma^2(\rho)\} - \log\{\sigma^{2*}(\rho)\}$ and $H_5 = \log\{\sigma^{2*}(\rho_0)\} - \log\{\tilde\sigma^2(\rho_0)\}$.

By Lemma C.5, we have $(NT)^{-1}\mathrm{tr}(\mathrm{E}(P_{N,T})) = (NT)^{-1}\mathrm{tr}(I_{NT}) = 1$ so that

$$\sigma^{2*}(\rho_0) - \tilde\sigma^2(\rho_0) = \frac{1}{NT}\mathrm{E}((\tilde X\tilde{\boldsymbol\beta}_0)^\top P_{N,T}(\tilde X\tilde{\boldsymbol\beta}_0)) + \sigma_0^2\left\{\frac{\mathrm{tr}(\mathrm{E}(P_{N,T}))}{NT} - 1\right\} = o(1).$$

Accordingly, we obtain $H_5 = o(1)$.

To show $H_4(\rho) \le 0$ uniformly on $\triangle$, it suffices to show $\tilde\sigma^2(\rho) - \sigma^{2*}(\rho) \le 0$ uniformly on $\triangle$. Observe that $\tilde\sigma^2(\rho) - \sigma^{2*}(\rho) = \frac{\sigma_0^2}{NT}H_4^1(\rho) - \frac{1}{NT}H_4^2(\rho)$, where

$$\begin{aligned}\frac{1}{NT}H_4^1(\rho) &= 1 - \frac{\mathrm{tr}(\mathrm{E}(P_{N,T}))}{NT} + 2(\rho_0-\rho)\frac{\mathrm{tr}(G_{NT}) - \mathrm{tr}(\mathrm{E}(P_{N,T})G_{NT})}{NT}\\ &\quad + (\rho_0-\rho)^2\frac{\mathrm{tr}(G_{NT}^\top G_{NT}) - \mathrm{tr}(G_{NT}^\top \mathrm{E}(P_{N,T})G_{NT})}{NT}\end{aligned}$$

and $H_4^2(\rho) = \mathrm{E}\left(\{\tilde X\tilde{\boldsymbol\beta}_0 + (\rho_0-\rho)R_{N,T}\}^\top P_{N,T}\{\tilde X\tilde{\boldsymbol\beta}_0 + (\rho_0-\rho)R_{N,T}\}\right)$.

By Lemma C.5, we get $(NT)^{-1}H_4^1(\rho) = o_p(1)$ uniformly on $\triangle$. Since $P_{N,T} = (I_{NT} - \tilde X\tilde\Phi)^\top Q^\top Q(I_{NT} - \tilde X\tilde\Phi)$ and $\tilde X\tilde{\boldsymbol\beta}_0 + (\rho_0-\rho)R_{N,T}$ is not a zero vector, $H_4^2(\rho)$ can be written as $\mathrm{E}(x^\top Ax)$ with a positive semidefinite matrix A. Thus, $(NT)^{-1}H_4^2(\rho) \le 0$ uniformly on $\triangle$ so that $H_4(\rho) \le 0$ holds uniformly on $\triangle$.

Hence, we have $\lim_{N,T\to\infty}\sup\max_{\rho\in N_\epsilon^c(\rho_0)} \frac{1}{NT}\{Q_{N,T}(\rho) - Q_{N,T}(\rho_0)\} \le 0$ for any $\epsilon > 0$.

If the identification uniqueness condition was not satisfied, without loss of generality, there would exist a sequence ρ_n converging to $\rho^* \ne \rho_0$ such that $\lim_{N,T\to\infty}\sup\max_{\rho\in N_\epsilon^c(\rho_0)} \frac{1}{NT}\{Q_{N,T}(\rho_n) - Q_{N,T}(\rho^*)\} = 0$. This will follow if (a) $\lim_{N,T\to\infty}\{\tilde\sigma^2(\rho^*) - \sigma^{2*}(\rho^*)\} = 0$ and (b) $\lim_{N,T\to\infty}(NT)^{-1}[\tilde Q_{N,T}(\rho^*) - \tilde Q_{N,T}(\rho_0)] = 0$. Since $\lim_{N,T\to\infty}\{\tilde\sigma^2(\rho) - \sigma^{2*}(\rho)\} = \Psi_{R,R} > 0$ 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 $\widehat{\boldsymbol\theta}$. According to Taylor expansion of the first order condition of (2.9), we obtain

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = -\left\{\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}^{-1}\frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+o_{\mathrm{P}}(1),$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ and hence it converges to $\boldsymbol{\theta}_0$ in probability by Theorem 1. If

$$\frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\boldsymbol{\theta}_0}+\Omega_{\boldsymbol{\theta}_0}), \tag{A.7}$$

$$-\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}-\Sigma_{\boldsymbol{\theta}_0} = o_{\mathrm{P}}(1), \tag{A.8}$$

$$\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}-\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = o_{\mathrm{P}}(1) \quad \text{uniformly in} \quad \tilde{\boldsymbol{\theta}}, \tag{A.9}$$

all hold, the asymptotic normality follows: $\sqrt{NT}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\boldsymbol{\theta}_0}^{-1}+\Sigma_{\boldsymbol{\theta}_0}^{-1}\Omega_{\boldsymbol{\theta}_0}\Sigma_{\boldsymbol{\theta}_0}^{-1})$.

(1) Proof of (A.7). First, we can get the first order derivative of $\log L_{N,T}(\boldsymbol{\theta})$ as follows

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} &= \begin{pmatrix} \frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\rho} \\ \frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\sigma^2}\end{pmatrix} \\
&= \frac{1}{\sqrt{NT}}\begin{pmatrix} \frac{R_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^2}+\frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}-T\sigma_0^2\mathrm{tr}(G_N)}{\sigma_0^2}+\frac{R_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0}{\sigma_0^2}+\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}G_{N,T}\mathbf{e}}{\sigma_0^2} \\ \frac{\mathbf{e}^\top P_{N,T}\mathbf{e}-NT\sigma_0^2}{2\sigma_0^4}+\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)+2(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\mathbf{e}}{2\sigma_0^4}\end{pmatrix} \\
&= \frac{1}{\sqrt{NT}}\begin{pmatrix} \frac{1}{\sigma_0^2}\left\{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}-\sigma_0^2\mathrm{tr}(G_{N,T})\right\}+\frac{1}{\sigma_0^2}R_{N,T}^\top P_{N,T}\mathbf{e} \\ \frac{1}{2\sigma_0^4}(\mathbf{e}^\top P_{N,T}\mathbf{e}-NT\sigma_0^2)\end{pmatrix}+o_{\mathrm{P}}(1)
\end{aligned}$$

due to

$$\frac{1}{\sqrt{NT}}R_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 = o_{\mathrm{P}}(1), \quad \frac{1}{\sqrt{NT}}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0) = o_{\mathrm{P}}(1),$$

$$\frac{1}{\sqrt{NT}}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\mathbf{e} = o_{\mathrm{P}}(1), \quad \frac{1}{\sqrt{NT}}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}G_{N,T}\mathbf{e} = o_{\mathrm{P}}(1),$$

in Lemma B.2. By Lemma C.5, we have

$$\frac{1}{\sqrt{NT}}\mathrm{tr}(P_{N,T}-I_{NT}) = o_{\mathrm{P}}(1), \frac{1}{\sqrt{NT}}\mathrm{tr}(P_{N,T}G_{N,T}-G_{N,T}) = o_{\mathrm{P}}(1),$$

so that

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\frac{\partial \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} &= \frac{1}{\sqrt{NT}}\begin{pmatrix} \frac{1}{\sigma_0^2}\left\{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}-\sigma_0^2\mathrm{tr}(G_{N,T}^\top P_{N,T})\right\}+\frac{1}{\sigma_0^2}R_{N,T}^\top P_{N,T}\mathbf{e} \\ \frac{1}{2\sigma_0^4}\{\mathbf{e}^\top P_{N,T}\mathbf{e}-\sigma_0^2\mathrm{tr}(P_{N,T})\}\end{pmatrix}+o_{\mathrm{P}}(1) \\
&= \frac{1}{\sqrt{NT}}\{\mathcal{Q}-\mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}+o_{\mathrm{P}}(1),
\end{aligned}$$

where $\mathcal{Q} = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e} + \frac{1}{\sigma_0^2} R_{N,T}^\top P_{N,T}\mathbf{e} \\ \frac{1}{2\sigma_0^4}\mathbf{e}^\top P_{N,T}\mathbf{e} \end{pmatrix}$.

Next, we show the asymptotic normality of $(NT)^{-1/2}\{\mathcal{Q} - \mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}$. By the Cramér-Wold device, it is sufficient to derive the asymptotic normality of $(NT)^{-1/2}(a_1, a_2)\{\mathcal{Q} - \mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}$, where $(a_1, a_2)^\top$ is any given two-dimensional constant vector. Note that

$$(a_1, a_2)\mathcal{Q} = \mathbf{e}^\top \left(\frac{a_1}{2\sigma_0^2} G_{N,T}^\top P_{N,T} + \frac{a_1}{2\sigma_0^2} P_{N,T} G_{N,T} + \frac{a_2}{2\sigma_0^4} P_{N,T} \right) \mathbf{e} + \frac{a_1}{\sigma_0^2} R_{N,T}^\top P_{N,T}\mathbf{e} =: \mathbf{e}^\top \mathcal{A}_{N,T}\mathbf{e} + \mathbf{b}_{N,T}^\top \mathbf{e},$$

which is a linear quadratic form of $\mathbf{e} = (e_{11}, \cdots, e_{N1}, \cdots, e_{1T}, \cdots, e_{NT})^\top$. Here $\mathcal{A}_{N,T} = \frac{a_1}{2\sigma_0^2} G_{N,T}^\top P_{N,T} + \frac{a_1}{2\sigma_0^2} P_{N,T} G_{N,T} + \frac{a_2}{2\sigma_0^4} P_{N,T}$ and $\mathbf{b}_{N,T}^\top = \frac{a_1}{\sigma_0^2} R_{N,T}^\top P_{N,T}$ are complicated functions of $\{X_t, t = 1, \cdots, T\}$ based on the definitions of $P_{N,T}$ and $R_{N,T}$ given before Assumption 8. Consequently, $\mathcal{A}_{N,T}$ and $\mathbf{b}_{N,T}^\top$ are both random and hence different from the fixed setting in the central limit theorem for linear quadratic forms of Theorem 1 in Kelejian and Prucha (2001).

In order to adapt the proof of Theorem 1 in Kelejian and Prucha (2001) in our case, the conditions for $\mathbf{e}$ 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 $\mathbf{e}^\top \mathcal{A}_{N,T}\mathbf{e} + \mathbf{b}_{N,T}^\top \mathbf{e}$.

To construct the martingale difference array, let $j = N(t-1) + i$ for $1 \le i \le N$ and $1 \le t \le T$ and $e_{it} = e^{(j)}$ for $j = 1, \cdots, NT$. Consequently, we get $\mathbf{e} = (e_{11}, \cdots, e_{N1}, \cdots, e_{1T}, \cdots, e_{NT}) = (e^{(1)}, \cdots, e^{(N)}, \cdots, e^{N(T-1)+1}, \cdots, e^{(NT)})^\top$. Let

$$\mathcal{F}_{j-1} = \mathcal{F}_V \vee \sigma\langle e^{(1)}, \cdots, e^{(j-1)} \rangle$$

for $j \ge 1$ where $\mathcal{F}_{j-1} := \mathcal{F}_V$ if $j = 1$, and $\mathcal{F}_V \vee \sigma\langle e^{(1)}, \cdots, e^{(j-1)} \rangle$ is the σ-field generated by $\mathcal{F}_V \cup \sigma\langle e^{(1)}, \cdots, e^{(j-1)} \rangle$. Due to the conditional independence of e_{it} and $\{e_{1t}, \cdots, e_{i-1,t}\}$ given $\mathcal{E}_{t-1}$, we obtain $\mathrm{E}(e^{(j)}|\mathcal{F}_{j-1}) = \mathrm{E}(e_{it}|\mathcal{E}_{t-1}, e_{1t}, \cdots, e_{i-1,t}) = \mathrm{E}(e_{it}|\mathcal{E}_{t-1}) = 0$. Thus $\{(e^{(j)}, \mathcal{F}_j) : 1 \le j \le NT\}$ 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 $\mathbf{e}^\top \mathcal{A}_{N,T}\mathbf{e} + \mathbf{b}_{N,T}^\top \mathbf{e}$, we let $(A)_{j_1 j_2}$ be the (j_1, j_2)-th element of matrix A and $(\mathbf{b})_j$ be the j-th element of vector $\mathbf{b}$. Note that both $\mathcal{A}_{N,T}, \mathbf{b}_{N,T} \in \mathcal{F}_V$ because they are functions of $\{X_t, t = 1, \cdots, T\}$ and also $\mathcal{A}_{N,T}$ is symmetric. Hence we obtain $\mathbf{e}^\top \mathcal{A}_{N,T}\mathbf{e} + \mathbf{b}_{N,T}^\top \mathbf{e} - \mathrm{E}(\mathbf{e}^\top \mathcal{A}_{N,T}\mathbf{e} + \mathbf{b}_{N,T}^\top \mathbf{e}|\mathcal{F}_V) = \sum_{j=1}^{NT} \mathcal{Z}_j$, where

$$\mathcal{Z}_j = (\mathcal{A}_{N,T})_{jj} \left\{ e^{(j)2} - \sigma_0^2 \right\} + 2e^{(j)} \sum_{j_1=1}^{j-1} (\mathcal{A}_{N,T})_{jj_1} e^{(j_1)} + (\mathbf{b}_{N,T})_j e^{(j)}.$$

Since $\mathcal{A}_{N,T}, \mathbf{b}_{N,T} \in \mathcal{F}_V$, we then have $\mathrm{E}(\mathcal{Z}_j|\mathcal{F}_{j-1}) = 0$. This implies that $\{(\mathcal{Z}_j, \mathcal{F}_j) : 1 \le j \le NT\}$ 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

$$\frac{1}{\sqrt{NT}}(a_1,a_2)\{\mathcal{Q}-\mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}=\frac{1}{\sqrt{NT}}\sum_{j=1}^{NT}\mathcal{Z}_j\xrightarrow{d}N\left(0,(a_1,a_2)\Sigma_{\mathcal{Q}}(a_1,a_2)^\top\right),$$

where $\Sigma_{\mathcal{Q}}=\lim_{N,T\to\infty}(NT)^{-1}\mathrm{E}\,\mathrm{Cov}(Q|\mathcal{F}_V)=\lim_{N,T\to\infty}(\Sigma_{NT,\boldsymbol{\theta}_0}+\Omega_{NT,\boldsymbol{\theta}_0})$,

$$\begin{aligned}\Sigma_{NT,\boldsymbol{\theta}_0} &= \begin{pmatrix}\frac{\mathrm{E}(R_{N,T}^\top P_{N,T}^2R_{N,T})}{NT\sigma_0^2}+\frac{\mathrm{E}(\mathrm{tr}(P_{N,T}G_{N,T}P_{N,T}G_{N,T}+P_{N,T}^2G_{N,T}G_{N,T}^\top))}{NT} & \frac{\mathrm{E}(\mathrm{tr}(G_{N,T}^\top P_{N,T}^2))}{NT\sigma_0^2}\\ \frac{\mathrm{E}(\mathrm{tr}(G_{N,T}^\top P_{N,T}^2))}{NT\sigma_0^2} & \frac{\mathrm{E}(\mathrm{tr}(P_{N,T}^2))}{2\sigma_0^4NT}\end{pmatrix}\\ &= \begin{pmatrix}\frac{\mathrm{E}(R_{N,T}^\top P_{N,T}R_{N,T})}{NT\sigma_0^2}+\frac{\mathrm{tr}(G_{N,T}^2)+\mathrm{tr}(G_{N,T}G_{N,T}^\top)}{NT} & \frac{\mathrm{tr}(G_{N,T})}{NT\sigma_0^2}\\ \frac{\mathrm{tr}(G_{N,T})}{NT\sigma_0^2} & \frac{1}{2\sigma_0^4}\end{pmatrix}+o(1),\end{aligned}$$

and

$$\Omega_{NT,\boldsymbol{\theta}_0}$$
$$=\begin{pmatrix}\frac{2m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}G_{N,T}))}{NT\sigma_0^4}+\frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}^2)}{NT\sigma_0^4} & \frac{m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}))}{2NT\sigma_0^6}+\frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}p_{ii})}{2NT\sigma_0^6}\\ \frac{m_3\mathrm{E}(R_{N,T}^\top P_{N,T}\mathrm{diag}(P_{N,T}))}{2NT\sigma_0^6}+\frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}(gp)_{ii}p_{ii})}{2NT\sigma_0^6} & \frac{(m_4-3\sigma_0^4)\mathrm{E}(\sum_{i=1}^{NT}p_{ii}^2)}{4\sigma_0^8NT}\end{pmatrix},$$

where p_{ii} and $(gp)_{ii}$ are the i-th main diagonal elements of $P_{N,T}$ and $G_{N,T}^\top P_{N,T}$, respectively. As

$$\frac{\mathrm{E}(R_{N,T}^\top PR_{N,T})}{NT}\to\Phi_{R,R},\quad \lim_{N,T\to\infty}\frac{\mathrm{tr}(G_{N,T}^2)+\mathrm{tr}(G_{N,T}^\top G_{N,T})}{NT}=c_1,\quad \frac{\mathrm{tr}(G_{N,T})}{NT}=c_2,$$

we have $\lim_{N,T\to\infty}\Sigma_{NT,\boldsymbol{\theta}_0}=\Sigma_{\boldsymbol{\theta}_0}=\begin{pmatrix}\frac{1}{\sigma_0^2}\Phi_{R,R}+c_1 & \frac{c_2}{\sigma_0^2}\\ \frac{c_2}{\sigma_0^2} & \frac{1}{2\sigma_0^4}\end{pmatrix}$. Therefore, we obtain

$$\frac{1}{\sqrt{NT}}\frac{\partial\log L_{N,T}(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\xrightarrow{d}N(0,\Sigma_{\boldsymbol{\theta}_0}+\Omega_{\boldsymbol{\theta}_0}),$$

where $\Omega_{\boldsymbol{\theta}_0}=\lim_{N,T\to\infty}\Omega_{NT,\boldsymbol{\theta}_0}$.

(2) Proof of (A.8). The second derivative can be obtained as follows

$$\begin{aligned}\frac{\partial^2\log L_{N,T}(\boldsymbol{\theta})}{\partial\rho^2} &= -T\mathrm{tr}\left\{G_N^2(\rho)\right\}-\frac{1}{\sigma^2}\left\{(I_T\otimes W)Y\right\}^\top P_{N,T}\left\{(I_T\otimes W)Y\right\},\\ \frac{\partial^2\log L_{N,T}(\boldsymbol{\theta})}{\partial\rho\partial\sigma^2} &= -\frac{\left\{(I_T\otimes W)Y\right\}^\top P_{N,T}\left\{S_{N,T}(\rho)Y\right\}}{\sigma^4},\\ \frac{\partial^2\log L_{N,T}(\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2} &= \frac{NT}{2\sigma^4}-\frac{\left\{S_{N,T}(\rho)Y\right\}^\top Q_{N,T}\left\{S_{N,T}(\rho)Y\right\}}{\sigma^6}.\end{aligned}\tag{A.10}$$

By some calculations, we have

$$
\begin{aligned}
\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial \rho^2} &= -\frac{\operatorname{tr}(G_{N,T}^2)}{NT} - \frac{R_{N,T}^\top P_{N,T} R_{N,T} + 2R_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e}}{NT\sigma_0^2},\\
\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial \rho \partial \sigma^2} &= -\frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 + R_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0}{NT\sigma_0^4},\\
\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial \sigma^2 \partial \sigma^2} &= \frac{1}{2\sigma_0^4} - \frac{\mathbf{e}^\top P_{N,T}\mathbf{e} + (\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 + 2(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\mathbf{e}}{NT\sigma_0^6}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
-\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} &= \frac{1}{NT}\begin{pmatrix} \frac{2R_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e}}{\sigma_0^2} & \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} \\ \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} & \frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{\sigma_0^6} \end{pmatrix}\\
&\quad + \begin{pmatrix} \frac{\operatorname{tr}(G_{N,T}^2)}{NT} + \frac{R_{N,T}^\top P_{N,T} R_{N,T}}{NT\sigma_0^2} & 0 \\ 0 & -\frac{1}{2\sigma_0^4} \end{pmatrix} + o_{\mathrm{P}}(1).
\end{aligned}
$$

By Lemma B.2, we have

$$
\begin{aligned}
&\frac{1}{NT}\begin{pmatrix} \frac{2R_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e}}{\sigma_0^2} & \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} \\ \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} & \frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{\sigma_0^6} \end{pmatrix}\\
&= \mathrm{E}\left\{\frac{1}{NT}\begin{pmatrix} \frac{2R_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T} G_{N,T}\mathbf{e}}{\sigma_0^2} & \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} \\ \frac{R_{N,T}^\top P_{N,T}\mathbf{e} + \mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sigma_0^4} & \frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{\sigma_0^6} \end{pmatrix}\right\} + o_P(1)\\
&= \begin{pmatrix} \frac{\operatorname{tr}(G_{N,T}^\top \mathrm{E}(P_{N,T}) G_{N,T})}{NT} & \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}) G_{N,T})}{NT\sigma_0^2} \\ \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}) G_{N,T})}{NT\sigma_0^2} & \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}))}{NT\sigma_0^4} \end{pmatrix} + o_P(1).
\end{aligned}
$$

According to Lemma C.5, we can get

$$
\begin{aligned}
-\frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} &= \begin{pmatrix} \frac{\operatorname{tr}(G_{N,T}^\top \mathrm{E}(P_{N,T}) G_{N,T}) + \operatorname{tr}(G_{N,T}^2)}{NT} + \frac{\mathrm{E}(R_{N,T}^\top P_{N,T} R_{N,T})}{NT\sigma_0^2} & \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}) G_{N,T})}{NT\sigma_0^2} \\ \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}) G_{N,T})}{NT\sigma_0^2} & \frac{\operatorname{tr}(\mathrm{E}(P_{N,T}))}{NT\sigma_0^4} - \frac{1}{2\sigma_0^4} \end{pmatrix}\\
&= \begin{pmatrix} \frac{\operatorname{tr}(G_{N,T}^2) + \operatorname{tr}(G_{N,T}^\top G_{N,T})}{NT} + \frac{\mathrm{E}(R_{N,T}^\top P_{N,T} R_{N,T})}{NT\sigma_0^2} & \frac{\operatorname{tr}(G_{N,T})}{NT\sigma_0^2} \\ \frac{\operatorname{tr}(G_{N,T})}{NT\sigma_0^2} & \frac{1}{2\sigma_0^4} \end{pmatrix} + o(1) = \Sigma_{\boldsymbol{\theta}_0} + o(1).
\end{aligned}
$$

Thus, (A.8) has been proved.

(3) Proof of (A.9). By (A.10), we note that $1/\sigma^2$ appears either in linear, quadratic or cubic from in all the elements of the second derivative of $\log L_{N,T}(\boldsymbol{\theta})$ and ρ only appears in linear form in $\partial^2 \log L_{N,T}(\boldsymbol{\theta})/\partial\rho\partial\sigma^2$ and $\partial^2 \log L_{N,T}(\boldsymbol{\theta})/\partial\sigma^2\partial\sigma^2$. According to the convergence of $\tilde{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ in probability, it is easy to show (A.9) holds for all elements but the second derivative of $\log L_{N,T}(\boldsymbol{\theta})$ with respect to ρ. Since $G_N(\rho) = WS_N(\rho)$,

we have $\mathrm{tr}\left\{G_N^2(\tilde\rho)\right\} = \mathrm{tr}\left(G_N^2\right) + 2\mathrm{tr}\left\{G_N^3(\tilde\rho^*)\right\}(\tilde\rho^* - \rho_0)$ for some $\tilde\rho^*$ between $\tilde\rho$ and ρ_0 by Assumption 5 and mean value theorem. Consequently, we have

$$\begin{aligned} & \frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\tilde{\boldsymbol\theta})}{\partial\rho^2} - \frac{1}{NT}\frac{\partial^2 \log L_{N,T}(\boldsymbol\theta_0)}{\partial\rho^2} \\ = \quad & -2\frac{\mathrm{tr}\left\{\mathrm{G}_\mathrm{N}^3(\tilde\rho^*)\right\}}{N}(\tilde\rho^* - \rho_0) + \left(\frac{1}{\sigma_0^2} - \frac{1}{\tilde\sigma^2}\right)\frac{\{(I_T\otimes W)Y\}^\top P_{N,T}\{(I_T\otimes W)Y\}}{NT}. \end{aligned} \tag{A.11}$$

Since $G_N(\rho)$ is UB, $N^{-1}\mathrm{tr}\left\{\mathrm{G}_\mathrm{N}^3(\tilde\rho^*)\right\}$ is bounded, implying the first term (A.11) is $o_\mathrm{P}(1)$. The second term is also $o_\mathrm{P}(1)$ since $(NT)^{-1}\{(I_T\otimes W)Y\}^\top P_{N,T}\{(I_T\otimes W)Y\}$ which can be shown by Lemma B.2. Consequently, (A.9) holds. Hence, we have finished the whole proof.

Proof of Theorem 3. As we know $S_{N,T}(\widehat\rho) = S_{N,T} + (\rho_0 - \widehat\rho)I_T\otimes W$ as well as the definition of $\widehat{\boldsymbol\beta}(\tau)$ in (2.10), we have

$$\begin{aligned} & \widehat{\boldsymbol\beta}(\tau) - \boldsymbol\beta_0(\tau) = \Phi(\tau)\{Y^*(\widehat\rho) - D\widehat{\boldsymbol\alpha}\} - \boldsymbol\beta_0(\tau) = \Phi(\tau)\{Y^*(\widehat\rho) - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top\tilde Y(\widehat\rho)\} - \boldsymbol\beta_0(\tau) \\ & = \Phi(\tau)\{I_{NT} - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top(I_{NT} - S)\}\tilde X\tilde{\boldsymbol\beta}_0 - \boldsymbol\beta_0(\tau) + \Phi(\tau)\{I_{NT} - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top(I_{NT} - S)\}D\boldsymbol\alpha \\ & \qquad + \Phi(\tau)\{I_{NT} - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top(I_{NT} - S)\}\mathbf{e} \\ & \qquad + (\rho_0 - \widehat\rho)\Phi(\tau)\{I_{NT} - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top(I_{NT} - S)\}G_{N,T}(\tilde X\tilde{\boldsymbol\beta}_0 + D\boldsymbol\alpha + \mathbf{e}) \\ & \qquad := \Xi_{N,T}(1) + \Xi_{N,T}(2) + \Xi_{N,T}(3) + \Xi_{N,T}(4). \end{aligned} \tag{A.12}$$

Due to $\tilde D = (I_{NT} - S)D$, we get

$$\Xi_{N,T}(2) = \Phi(\tau)\{D - D(\tilde D^\top\tilde D)^{-1}\tilde D^\top\tilde D\}\boldsymbol\alpha = \mathbf{0}_d. \tag{A.13}$$

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

$$\Xi_{N,T}(1) = \frac{1}{2}\mu_2\boldsymbol\beta_0''(\tau)h^2 + o_\mathrm{P}(h^2). \tag{A.14}$$

Also, Lemma B.5 shows that

$$\sqrt{NTh}\,\Xi_{N,T}(3) \xrightarrow{d} N(\mathbf{0}_d, \nu_0\sigma_0^2\Sigma_X^{-1}(\tau)). \tag{A.15}$$

From Theorem 2, we have

$$\sqrt{NTh}\,\Xi_{N,T}(4) = \sqrt{NTh}\,O_\mathrm{P}(\rho_0 - \widehat\rho) = o_\mathrm{P}(1). \tag{A.16}$$

By (A.12)-(A.16), we have accomplished the proof.

A.3 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 the paper, we adopt the leave–one–unit–out cross–validation method to choose optimal bandwidth. Such method is also used in Li et al. (2011) and Chen et al. (2012). The initial value $\tilde{\rho}$ is obtained from the parametric spatial panel data model Lee and Yu (2010).

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

$$h_{opt} = \arg\min_h \left[Z(\tilde{\rho}) - B\left(X, \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-)}\right)\right]^\top M^{*\top} M^* \left[Z(\tilde{\rho}) - B\left(X, \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-)}\right)\right], \tag{A.17}$$

where $M^* = I_{NT} - T^{-1}(i_T i_T') \otimes I_N$ is used to delete the unobserved fixed effect due to $M^* D = 0_{NT\times(N-1)}$, $B(X, \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-)}) = \left(\left[X_{11}^\top \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-1)}(\tau_1)\right]^\top, \cdots, \left[X_{N1}^\top \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-N)}(\tau_1)\right]^\top, \cdots, \left[X_{1T}^\top \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-1)}(\tau_T)\right]^\top, \cdots, \left[X_{NT}^\top \widehat{\boldsymbol{\beta}}_{\tilde{\rho}}^{(-N)}(\tau_T)\right]^\top\right)^\top$.

Appendix B: Main Lemmas and Their Proofs

Let

$$\lambda_\mu = \begin{pmatrix} \mu_2(\tau) \\ \mu_3(\tau) \end{pmatrix}, \Lambda_\mu(\tau) = \begin{pmatrix} \mu_0(\tau) & \mu_1(\tau) \\ \mu_1(\tau) & \mu_2(\tau) \end{pmatrix}, \Lambda_\nu(\tau) = \begin{pmatrix} \nu_0(\tau) & \nu_1(\tau) \\ \nu_1(\tau) & \nu_2(\tau) \end{pmatrix}, N(\tau) = \begin{pmatrix} \left(\frac{1-\tau T}{Th}\right)^2 X_1 \\ \vdots \\ \left(\frac{T-\tau T}{Th}\right)^2 X_T \end{pmatrix},$$

where $\mu_\ell(\tau) = \mu_\ell = \int_{-\infty}^{\infty} u^\ell K(u)\mathrm{d}u$ and $\nu_\ell(\tau) = \nu_\ell = \int_{-\infty}^{\infty} u^\ell K^2(u)\mathrm{d}u$ if $0 < \tau < 1$. and $\mu_\ell(\tau) = \int_{-\infty}^{0} u^\ell K(u)\mathrm{d}u$ and $\nu_\ell(\tau) = \int_{-\infty}^{0} u^\ell K^2(u)\mathrm{d}u$ if $\tau = 1$.

Define $\Delta_M(\tau) = \Lambda_\mu(\tau) \otimes \Sigma_X(\tau)$, $\Delta_{M,N}(\tau) = \lambda_\mu \otimes \Sigma_X(\tau)$,

$$\widehat{\Delta}_M(\tau) = \frac{M^\top(\tau)\Omega(\tau)M(\tau)}{NTh}, \quad \widehat{\Delta}_{M,N}(\tau) = \frac{M^\top(\tau)\Omega(\tau)N(\tau)}{NTh} \quad \text{and} \quad \widehat{\Delta}_{M,\mathbf{e}}(\tau) = \frac{M^\top(\tau)\Omega(\tau)\mathbf{e}}{NTh}.$$

Lemma B.1. *Suppose Assumptions 1, 3 and 6 hold. When $N, T \to \infty$,*

$$\Xi_{N,T}(1,1) = \Phi(\tau)\tilde{X}\tilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0(\tau) = \frac{1}{2}\mu_2 \boldsymbol{\beta}_0''(\tau)h^2 + o_P(h^2) \quad \textit{for} \quad 0 < \tau \leq 1.$$

Proof. According to the Taylor Expansion and some transformations, we can write $\Xi_{N,T}(1,1)$ as the sum of four terms. Observe that

$$
\begin{aligned}
& \Phi(\tau)\tilde{X}\tilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0(\tau) = (I_d, 0_{d\times d})\widehat{\Delta}_M^{-1}(\tau)\widehat{\Delta}_{M,N}(\tau)\left(\frac{1}{2}\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2)\right) \\
& = (I_d, 0_{d\times d})\Delta_M^{-1}(\tau)\Delta_{M,N}(\tau)\left(\frac{1}{2}\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2)\right) + (I_d, 0_{d\times d})\left(\widehat{\Delta}_M^{-1}(\tau) - \Delta_M^{-1}(\tau)\right)\Delta_{M,N}(\tau) \\
& \times\left(\frac{1}{2}\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2)\right) \\
& +(I_d, 0_{d\times d})\Delta_M^{-1}(\tau)\left(\widehat{\Delta}_{M,N}(\tau) - \Delta_{M,N}(\tau)\right)\left(\frac{1}{2}\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2)\right) \\
& +(I_d, 0_{d\times d})\left(\widehat{\Delta}_M^{-1}(\tau) - \Delta_M^{-1}(\tau)\right)\left(\widehat{\Delta}_{M,N}(\tau) - \Delta_{M,N}(\tau)\right)\left(\frac{1}{2}\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2)\right) \\
:= \; & \Xi_{N,T}(1,1,1) + \Xi_{N,T}(1,1,2) + \Xi_{N,T}(1,1,3) + \Xi_{N,T}(1,1,4).
\end{aligned}
$$

Simple algebra yields that

$$
\Xi_{N,T}(1,1,1) = \frac{1}{2}\mu_2\boldsymbol{\beta}_0''(\tau)h^2 + o(h^2).
$$

According to the convergence of $\widehat{\Delta}_M^{-1}(\tau)$ and $\widehat{\Delta}_{M,N}(\tau)$ in Lemma C.2, we can get $\Xi_{N,T}(1,1,\ell) = o_{\mathrm{P}}(h^2)$, for $\ell = 2,3,4$. So the proof of this lemma has been completed.

□

Lemma B.2. *Under Assumptions 1-8, as $N, T \to \infty$, we have*

(1) $\frac{1}{\sqrt{NT}}A_{NT}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 = o_{\mathrm{P}}(1)$ *for* $A_{NT} = \tilde{X}\tilde{\boldsymbol{\beta}}_0$, $G_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0$, $R_{N,T}$, $\mathbf{e}$ *and* $G_{N,T}\mathbf{e}$.

(2) $\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} = O_{\mathrm{P}}(1)$, $\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top G_{N,T}^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} = O_{\mathrm{P}}(1)$ *and* $\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top G_{N,T}^\top P_{N,T} G_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} = O_{\mathrm{P}}(1)$.

Proof. (1) By Lemma B.1, it is obvious that

$$
X_{it}^\top\boldsymbol{\beta}_0(\tau_t) - X_{it}^\top\Phi(\tau_t)\tilde{X}\tilde{\boldsymbol{\beta}}_0 \;=\; -\frac{1}{2}\mu_2 X_{it}^\top\boldsymbol{\beta}_0''(\tau_t)h^2 + o_{\mathrm{P}}(h^2).
$$

Define $\boldsymbol{\beta}_0^{''} = (\boldsymbol{\beta}_0^{''}(\tau_1)^\top, \cdots, \boldsymbol{\beta}_0^{''}(\tau_T)^\top)^\top$. Then,

$$
(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0 = \begin{pmatrix} X_{11}^\top\boldsymbol{\beta}(\tau_1) - X_{11}^\top\Phi(\tau_1)\tilde{X}\tilde{\boldsymbol{\beta}}_0 \\ \vdots \\ X_{NT}^\top\boldsymbol{\beta}(\tau_t) - X_{NT}^\top\Phi(\tau_t)\tilde{X}\tilde{\boldsymbol{\beta}}_0 \end{pmatrix} = -\frac{1}{2}\mu_2 h^2\tilde{X}\tilde{\boldsymbol{\beta}}_0^{''} + o_{\mathrm{P}}(h^2).
$$

It can be shown that the element in $\tilde{X}\tilde{\boldsymbol{\beta}}_0^{''} = O_{\mathrm{P}}(1)$ by Assumptions 1 and 6 so that we have

$$
\begin{aligned}
\|(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_\infty &= \left\|-\frac{1}{2}\mu_2 h^2\tilde{X}\tilde{\boldsymbol{\beta}}_0^{''} + o(h^2)\right\|_\infty = O_{\mathrm{P}}(h^2) \\
\text{and } \|(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 &= \left\|-\frac{1}{2}\mu_2 h^2\tilde{X}\tilde{\boldsymbol{\beta}}_0^{''} + o(h^2)\right\|_2 = O_{\mathrm{P}}(\sqrt{NT}h^2).
\end{aligned}
$$

By Lemma C.7, $\|Q_{NT}\|_2 = O_{\mathrm{P}}(1)$ holds so that

$$\|Q_{NT}(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 \quad = \quad O_{\mathrm{P}}(\sqrt{NT}h^2).$$

Due to $NTh^8 \to 0$ in Assumption 3, we have

$$\frac{1}{\sqrt{NT}}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 \quad = \quad \frac{1}{\sqrt{NT}}\|Q_{N,T}(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2^2 = O_{\mathrm{P}}(\sqrt{NT}h^4) = o_{\mathrm{P}}(1).$$

Similarly, we can show that

$$\|Q_{N,T}(I_{NT} - S)G_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 = O_{\mathrm{P}}(\sqrt{NT}h^2)$$

and

$$\|Q_{N,T}(I_{NT} - S)R_{N,T}\|_2 = O_{\mathrm{P}}(\sqrt{NT}h^2).$$

Then,

$$\begin{aligned}\frac{1}{\sqrt{NT}}(G_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 &= \frac{1}{\sqrt{NT}}\|Q_{N,T}(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2\|Q_{N,T}(I_{NT} - S)G_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 \\ &= O(\sqrt{NT}h^4) = o_{\mathrm{P}}(1),\end{aligned}$$

$$\begin{aligned}\frac{1}{\sqrt{NT}}R_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 &= \frac{1}{\sqrt{NT}}\|Q_{N,T}(I_{NT} - S)\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2\|Q_{N,T}(I_{NT} - S)R_{N,T}\|_2 \\ &= O(\sqrt{NT}h^4) = o_{\mathrm{P}}(1).\end{aligned}$$

Thus, we have shown that

$$\frac{1}{\sqrt{NT}}A_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0 = o_{\mathrm{P}}(1)$$

holds for $A_{NT} = \tilde{X}\tilde{\boldsymbol{\beta}}_0$, $G_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0$ and $R_{N,T}$.

As we have known $\mathrm{E}\left((NT)^{-1/2}A_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) = 0$ for $A_{N,T} = \mathbf{e}$ and $G_{N,T}\mathbf{e}$, to prove the lemma it suffices to show that

$$\mathrm{Var}\left(\frac{1}{\sqrt{NT}}A_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) = o(1).$$

It can be show that $\|P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 = O_{\mathrm{P}}(\sqrt{NT}h^2)$, $\|G_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2 = O_{\mathrm{P}}(\sqrt{NT}h^2)$ and $\mathrm{Var}(\mathbf{e}|\mathcal{F}_V) = \sigma_0^2 I_{NT}$ so that we have

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{\sqrt{NT}}\mathbf{e}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) &= \frac{1}{NT}\mathrm{E}\left((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}\mathrm{Var}(\mathbf{e}|\mathcal{F}_V)P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) \\ &= \frac{\sigma_0^2}{NT}\mathrm{E}\left((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}^2\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) = \frac{\sigma_0^2}{NT}\mathrm{E}\left(\|P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2^2\right) = o(h^4)\end{aligned}$$

and

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{\sqrt{NT}}\mathbf{e}^\top G_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right) &= \frac{1}{NT}\mathrm{E}\left((\tilde{X}\tilde{\boldsymbol{\beta}}_0)^\top P_{N,T}G_{N,T}\operatorname{Var}(\mathbf{e}|\mathcal{F}_V)G_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\right)\\
&= \frac{\sigma_0^2}{NT}\mathrm{E}\left(\|G_{N,T}^\top P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0\|_2^2\right) = o(h^4).
\end{aligned}$$

Hence, we have completed the proof.

(2) According to the results in (1), we obtain

$$\begin{aligned}
\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} &= \frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{NT} + o_{\mathrm{P}}(1),\\
\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top G_{N,T}^\top P_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} &= \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{NT} + \frac{R_{N,T}^\top P_{N,T}\mathbf{e}}{NT} + o_{\mathrm{P}}(1),\\
\frac{(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})^\top G_{N,T}^\top P_{N,T}G_{N,T}(\tilde{X}\tilde{\boldsymbol{\beta}}_0 + D\boldsymbol{\alpha}_0 + \mathbf{e})}{NT} &= \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}G_{N,T}\mathbf{e}}{NT} + \frac{R_{N,T}^\top P_{N,T}R_{N,T}}{NT}\\
&\quad + \frac{2R_{N,T}^\top P_{N,T}G_{N,T}\mathbf{e}}{NT}. \qquad \text{(B.1)}
\end{aligned}$$

In the proof of Theorem 2, we have shown that $(NT)^{-1/2}\{\mathcal{Q} - \mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}$ converges in probability to normal distribution, where $\mathcal{Q} = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e} + \frac{1}{\sigma_0^2}R_{N,T}^\top P_{N,T}\mathbf{e} \\ \frac{1}{2\sigma_0^4}\mathbf{e}^\top P_{N,T}\mathbf{e} \end{pmatrix}$. It means that

$$\frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{\sqrt{NT}} - \frac{\operatorname{tr}(P_{N,T})}{\sqrt{NT}}\sigma_0^2 = O_{\mathrm{P}}(1), \qquad \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{\sqrt{NT}} - \frac{\operatorname{tr}(G_{N,T}^\top P_{N,T})}{\sqrt{NT}}\sigma_0^2 = O_{\mathrm{P}}(1).$$

Then,

$$\frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{NT} = \frac{\operatorname{tr}(P_{N,T})}{NT}\sigma_0^2 + o_{\mathrm{P}}(1), \qquad \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{NT} = \frac{\operatorname{tr}(G_{N,T}^\top P_{N,T})}{NT}\sigma_0^2 + o_{\mathrm{P}}(1).$$

Similarly, we can show that

$$\frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}G_{N,T}\mathbf{e}}{NT} = \frac{\operatorname{tr}(G_{N,T}^\top P_{N,T}G_{N,T})}{NT}\sigma_0^2 + o_{\mathrm{P}}(1).$$

By Lemma C.5, we know

$$\begin{aligned}
\frac{tr(P_{N,T})}{NT} &= \frac{\operatorname{tr}(I_{NT})}{NT} + o_{\mathrm{P}}(1) = O_p(1), \quad \frac{\operatorname{tr}(G_{N,T}^\top P_{N,T})}{NT} = \frac{tr(G_{N,T})}{NT} + o_{\mathrm{P}}(1) = O_p(1),\\
\frac{\operatorname{tr}(G_{N,T}^\top P_{N,T}G_{N,T})}{NT} &= \frac{\operatorname{tr}(G_{N,T}^\top G_{N,T})}{NT} + o_{\mathrm{P}}(1) = O_p(1).
\end{aligned}$$

Then, we obtain

$$\frac{\mathbf{e}^\top P_{N,T}\mathbf{e}}{NT} = O_{\mathrm{P}}(1), \quad \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}\mathbf{e}}{NT} = O_{\mathrm{P}}(1), \quad \frac{\mathbf{e}^\top G_{N,T}^\top P_{N,T}G_{N,T}\mathbf{e}}{NT} = O_{\mathrm{P}}(1).$$

The variances of the other terms, like $(NT)^{-1}R_{N,T}^{\top}P_{N,T}\mathbf{e}$ and $(NT)^{-1}R_{N,T}^{\top}P_{N,T}G_{NT}\mathbf{e}$ and $(NT)^{-1}R_{N,T}^{\top}P_{N,T}R_{N,T}$ can be shown to be $o(1)$ so that

$$\frac{R_{N,T}^{\top}P_{N,T}\mathbf{e}}{NT}=\mathrm{E}\left(\frac{R_{N,T}^{\top}P_{N,T}\mathbf{e}}{NT}\right)+o_{\mathrm{P}}(1)=o_{\mathrm{P}}(1),\quad \frac{R_{N,T}^{\top}P_{N,T}G_{N,T}\mathbf{e}}{NT}=\mathrm{E}\left(\frac{R_{N,T}^{\top}P_{N,T}G_{N,T}\mathbf{e}}{NT}\right)+o_{\mathrm{P}}(1)=o_{\mathrm{P}}(1),$$

and

$$\frac{R_{N,T}^{\top}P_{N,T}R_{N,T}}{NT}=\mathrm{E}\left(\frac{R_{N,T}^{\top}P_{N,T}R_{N,T}}{NT}\right)+o_{\mathrm{P}}(1)\to\Psi_{R,R}.$$

The last equation is obtained by Assumption 8.

Therefore, we show that the right sides of (B.1) are $O_{\mathrm{P}}(1)$ which completes the proof.

□

Lemma B.3. *Under Assumptions 1-8, as $N,T\to\infty$, we have*

(1) $H_1(\rho)=\frac{1}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}=o_{\mathrm{P}}(1)$ uniformly on $\triangle$.

(2) $H_2(\rho)=\frac{1}{NT}R_{N,T}^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}=o_{\mathrm{P}}(1)$ uniformly on $\triangle$.

(3) $H_3(\rho)=\frac{1}{NT}\mathbf{e}^{\top}S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}=\sigma^2(\rho)+o_{\mathrm{P}}(1)$ uniformly on $\triangle$,
where $\sigma^2(\rho)=(NT)^{-1}\sigma_0^2\mathrm{tr}\{S_{N,T}^{-1\top}S_{N,T}^{\top}(\rho)\mathrm{E}(P_{N,T})S_{N,T}(\rho)S_{N,T}^{-1}\}$.

Proof. (1) By the defination of $S_{N,T}(\rho)$, we have $S_{N,T}(\rho)S_{N,T}^{-1}=\{S_{N,T}+(\rho_0-\rho)W\}S_{N,T}^{-1}=I_{NT}-(\rho_0-\rho)G_{N,T}$. Accordingly, we get

$$H_1(\rho)=\frac{1}{NT}(\tilde{X}\tilde{\boldsymbol{\beta}}_0)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}=\frac{1}{NT}\mathbf{e}^{\top}P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}_0+(\rho_0-\rho)\frac{1}{NT}(G_{N,T}\mathbf{e})^{\top}P_{N,T}\tilde{X}\tilde{\boldsymbol{\beta}}.$$

Since $H_1(\rho)$ is a linear form of ρ, by Lemma B.2 we can get $H_1(\rho)=O_{\mathrm{P}}(\frac{1}{\sqrt{NT}})=o_{\mathrm{P}}(1)$ uniformly on $\triangle$.

(2) Similarly, $H_2(\rho)$ can be written as

$$H_2(\rho)\quad=\quad\frac{1}{NT}R_{N,T}^{\top}P_{N,T}\mathbf{e}+(\rho_0-\rho)\frac{1}{NT}R_{N,T}^{\top}P_{N,T}G_{N,T}\mathbf{e}=o_{\mathrm{P}}(1).$$

As we have shown that $(NT)^{-1}R_{N,T}^{\top}P_{N,T}\mathbf{e}=o_{\mathrm{P}}(1)$ and $(NT)^{-1}R_{N,T}^{\top}P_{N,T}G_{N,T}\mathbf{e}=o_{\mathrm{P}}(1)$ in Lemma B.2 (2), $H_2(\rho)=o_{\mathrm{P}}(1)$ uniformly on $\triangle$ due to the linear form of ρ.

(3) Similarly to the proof that $(NT)^{-1/2}\{\mathcal{Q}-\mathrm{E}(\mathcal{Q}|\mathcal{F}_V)\}$ converges in probability to the normal distribution in Theorem 2, we can show that the asymptotic distribution of the following term is normal distribution as well so that

$$\frac{1}{\sqrt{NT}}\mathbf{e}^{\top}S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\mathbf{e}-\frac{\sigma_0^2}{\sqrt{NT}}\mathrm{tr}(S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1})=O_{\mathrm{P}}(1).$$

Then,

$$\begin{aligned}H_3(\rho)\quad&=\quad\frac{\sigma_0^2}{NT}\mathrm{tr}(S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top}P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1})+o_{\mathrm{P}}(1)\\&=\quad\frac{\sigma_0^2}{NT}\mathrm{tr}(S_{N,T}(\rho)S_{N,T}^{-1}S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top})+o_{\mathrm{P}}(1),\end{aligned}$$

where the last line is by the property of trace and Lemma C.5. Also, we get

$$\begin{aligned}\sigma^2(\rho) &= \frac{\sigma_0^2}{NT}\mathrm{E}(\mathrm{tr}\{S_{N,T}^{-1\top}S_{N,T}^{\top}(\rho)P_{N,T}S_{N,T}(\rho)S_{N,T}^{-1}\}) \\ &= \frac{\sigma_0^2}{NT}\mathrm{E}(\mathrm{tr}(S_{N,T}(\rho)S_{N,T}^{-1}S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top}+o_{\mathrm{P}}(1)) \\ &= \frac{\sigma_0^2}{NT}\mathrm{tr}(S_{N,T}(\rho)S_{N,T}^{-1}S_{N,T}^{-1\top}S_{N,T}(\rho)^{\top})+o(1).\end{aligned}$$

Therefore, $H_3(\rho)=\sigma^2(\rho)+o_{\mathrm{P}}(1)$ uniformly on $\triangle$.

□

Lemma B.4. *Suppose Assumptions 1, 3 and 6 hold. As N, $T\to\infty$,*

$$\|\Xi_{N,T}(1,2)\|_2=o_{\mathrm{P}}(h^2).$$

Proof. Using a similar proof to that of Lemma 4.2 in Su and Ullah (2006) and Lemma A.5 in Gao and Li (2013), we can show

$$(\tilde{D}^{\top}\tilde{D})^{-1}=(D^{\top}D)^{-1}\left\{1+O_{\mathrm{P}}\left(h^2+\frac{1}{\sqrt{NTh}}\right)\right\},$$

so that the leading term of $\Xi_{N,T}(1,2)$ is

$$\Xi_{N,T}(1,2,1) = \Phi(\tau)D(D^{\top}D)^{-1}D^{\top}(I_{NT}-S)^{\top}(I_{NT}-S)\tilde{X}\tilde{\boldsymbol{\beta}}_0.$$

Define $\Upsilon=S^{\top}+S-S^{\top}S$. Then, Υ is a $NT\times NT$ matrix with the (s_1,s_2)-th $N\times N$ block matrix being

$$\Upsilon_{s_1,s_2} = \{X_{s_2}\Phi_{s_1}(\tau_{s_2})\}^{\top}+X_{s_1}\Phi_{s_2}(\tau_{s_1})-\sum_{t=1}^{T}\{X_t\Phi_{s_1}(\tau_t)\}^{\top}X_t\Phi_{s_2}(\tau_t).$$

Thus,

$$(I_{NT}-S)^{\top}(I_{NT}-S)\tilde{X}\tilde{\boldsymbol{\beta}}_0=\tilde{X}\tilde{\boldsymbol{\beta}}_0-\Upsilon\tilde{X}\tilde{\boldsymbol{\beta}}_0=\begin{pmatrix}X_1\boldsymbol{\beta}(\tau_1)-\sum_{s_1=1}^{T}\Upsilon_{1,s_1}X_{s_1}\boldsymbol{\beta}(\tau_{s_1})\\ \vdots \\ X_T\boldsymbol{\beta}(\tau_T)-\sum_{s_1=1}^{T}\Upsilon_{T,s_1}X_{s_1}\boldsymbol{\beta}(\tau_{s_1})\end{pmatrix}:=\begin{pmatrix}\zeta_1\\ \vdots\\ \zeta_T\end{pmatrix},$$

where $\zeta_t=X_t\boldsymbol{\beta}(\tau_t)-\sum_{s_1=1}^{T}\Upsilon_{t,s_1}X_{s_1}\boldsymbol{\beta}(\tau_{s_1})$ which is shown to be $X_t\left(o(h^2)\right)$ in Lemma C.3.

Define $\mathbb{K}_s(\tau)=\left(K\left(\frac{\tau_s-\tau}{h}\right)I_d,\frac{\tau_s-\tau}{h}K\left(\frac{\tau_s-\tau}{h}\right)I_d\right)^{\top}$. Since $D(D^{\top}D)^{-1}D^{\top}=\frac{1}{T}\mathbf{1}_T\mathbf{1}_T^{\top}\otimes\left(I_N-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^{\top}\right)$, we have

$$
\begin{aligned}
& \Xi_{N,T}(1,2,1) \\
= \;& \Phi(\tau)\left\{\frac{1}{T}\mathbf{1}_T\mathbf{1}_T^\top \otimes \left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\right)\right\}(I_{NT}-S)^\top(I_{NT}-S)\tilde{X}\tilde{\boldsymbol{\beta}}_0 \\
= \;& (I_d, 0_{d\times d})\left\{\frac{M^\top(\tau)\Omega(\tau)M(\tau)}{NTh}\right\}^{-1}\frac{1}{NTh}M^\top(\tau)\Omega(\tau)\left[\mathbf{1}_T \otimes \left\{\frac{1}{T}\sum_{t=1}^{T}\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\right)\zeta_t\right\}\right] \\
= \;& \frac{o(h^2)}{NT^2h}(I_d, 0_{d\times d})\widehat{\Delta}_M(\tau)^{-1}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s - \frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\} \\
= \;& \frac{o(h^2)}{NT^2h}(I_d, 0_{d\times d})\left\{\widehat{\Delta}_M(\tau)^{-1} - \Delta_M(\tau)^{-1}\right\}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s - \frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\} \\
& \frac{o(h^2)}{NT^2h}(I_d, 0_{d\times d})\Delta_M(\tau)^{-1}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s - \frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\}.
\end{aligned}
$$

By Lemma C.2 and Lemma C.4, we can get $\|\Xi_{N,T}(1,2,1)\|_2 = o_{\mathrm{P}}(h^2)$.

□

Lemma B.5. *Suppose Assumptions 1, 2, 3 and 6 hold. As $N, T \to \infty$,*

$$\sqrt{NTh}\,\Xi_{N,T}(3) \xrightarrow{d} N(\mathbf{0}_d, \nu_0\sigma_0^2\Sigma_X^{-1}(\tau)).$$

Proof. By definition, we can get

$$
\begin{aligned}
\sqrt{NTh}\,\Xi_{N,T}(3) &= \sqrt{NTh}\,\Phi(\tau)\{I_{N,T} - D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\}\mathbf{e} \\
&= \sqrt{NTh}\,\Phi(\tau)\mathbf{e} + \sqrt{NTh}\,\Phi(\tau)D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\mathbf{e} \\
&:= \Xi_{N,T}(3,1) + \Xi_{N,T}(3,2).
\end{aligned}
$$

(1) First, we show $\Xi_{N,T}(3,1) \xrightarrow{d} N\left(\mathbf{0}_d, \nu_0\sigma_0^2\Sigma_X^{-1}(\tau)\right)$.

By definition, we obtain

$$
\begin{aligned}
\Xi_{N,T}(3,1) &= \sqrt{NTh}(I_d, 0_{d\times d})\widehat{\Delta}_M^{-1}(\tau)\widehat{\Delta}_{M,\mathbf{e}}(\tau) \\
&= \sqrt{NTh}(I_d, 0_{d\times d})\Delta_M^{-1}(\tau)\widehat{\Delta}_{M,\mathbf{e}}(\tau) + \sqrt{NTh}(I_d, 0_{d\times d})\left(\widehat{\Delta}_M^{-1}(\tau) - \Delta_M^{-1}(\tau)\right)\widehat{\Delta}_{M,\mathbf{e}}(\tau) \\
&:= \Xi_{N,T}(3,1,1) + \Xi_{N,T}(3,1,2).
\end{aligned}
$$

As long as we can show

$$\sqrt{NTh}\,\widehat{\Delta}_{M,\mathbf{e}} \xrightarrow{d} N\left(\mathbf{0}_{2d}, \sigma_0^2\Lambda_\mu(\tau)\otimes\Sigma_X(\tau)\right), \tag{B.2}$$

by the Cramér-Rao device we get $\Xi_{N,T}(3,1,1) \xrightarrow{d} N\left(\mathbf{0}_d, \nu_0\sigma_0^2\Sigma_X^{-1}(\tau)\right)$, and together with Lemma C.2 we get $\|\Xi_{N,T}(3,1,2)\|_2 = o_{\mathrm{P}}(1)$. Therefore, we obtain

$$\Xi_{N,T}(3,1) \xrightarrow{d} N\left(\mathbf{0}_d, \nu_0\sigma_0^2\Sigma_X^{-1}(\tau)\right).$$

Next, we show the proof of (B.2). Define $Z_t = \frac{1}{\sqrt{NTh}}\begin{pmatrix} K\left(\frac{t-\tau T}{Th}\right) X_t^\top \mathbf{e}_t \\ \left(\frac{t-\tau T}{Th}\right) K\left(\frac{t-\tau T}{Th}\right) X_t^\top \mathbf{e}_t \end{pmatrix}$. By Assumption 2 and $X_t \in \mathcal{F}_V$, we obtain $\mathrm{E}(Z_t|\mathcal{E}_{t-1}) = \mathbf{0}_{2d}$, and hence $\{Z_t, \mathcal{E}_t\}$ also forms a martingale difference array. Note that $\sqrt{NTh}\widehat{\Delta}_{M,\mathbf{e}}(\tau) = \sum_{t=1}^T Z_t$. According to the central limit theorem for martingale difference arrays, to show (B.2) we only need to check for any $\epsilon > 0$ and $2d\times 1$ element bounded vector $\mathbf{a}$, it holds that

$$\sum_{t=1}^T \mathrm{E}\left\{\left|\mathbf{a}^\top Z_t Z_t^\top \mathbf{a}\right| I\left(\left|\mathbf{a}^\top Z_t\right| > \epsilon\right) |\mathcal{E}_{t-1}\right\} = o_{\mathrm{P}}(1) \tag{B.3}$$

and

$$\sum_{t=1}^T \mathrm{E}\left(\mathbf{a}^\top Z_t Z_t^\top \mathbf{a}|\mathcal{E}_{t-1}\right) = \sigma_0^2\mathbf{a}^\top\left(\Lambda_\mu(\tau)\otimes\Sigma_X\right)\mathbf{a} + o_{\mathrm{P}}(1). \tag{B.4}$$

To prove (B.3), it is sufficient to prove $\sum_{t=1}^T \mathrm{E}\left\{\left|\mathbf{a}^\top Z_t\right|^{2+\delta/2}\right\} = o(1)$ for some $\delta > 0$.

Without loss of generality, we assume $d = 1$ so that $X_{it} = g(\tau_t) + v_{it}$ where v_{it} is univariate. Define $\mathbf{a} = (a_0, a_1)^\top$ and $K_t^\ell(\tau) = \left(\frac{\tau_t-\tau}{h}\right)^\ell K\left(\frac{\tau_t-\tau}{h}\right)$. We have

$$\begin{aligned}
&\sum_{t=1}^T \mathrm{E}\left\{\left|\mathbf{a}^\top Z_t\right|^{2+\delta/2}\right\} \\
= &\ \frac{1}{(NTh)^{1+\delta/4}}\sum_{t=1}^T \mathrm{E}\left\{\left|K_t^0(\tau)\sum_{i=1}^N(a_0X_{it}e_{it}) + K_t^1(\tau)\sum_{i=1}^N(a_1X_{it}e_{it})\right|^{2+\delta/2}\right\} \\
= &\ \frac{1}{(NTh)^{1+\delta/4}}\sum_{t=1}^T \mathrm{E}\left\{\left|(K_t^0(\tau)a_0 + K_t^1(\tau)a_1)g(\tau_t)\left(\sum_{i=1}^N e_{it}\right) + (K_t^0(\tau)a_0 + K_t^0(\tau)a_1)\left(\sum_{i=1}^N v_{it}e_{it}\right)\right|^{2+\delta/2}\right\} \\
\le &\ \frac{C}{(NTh)^{1+\delta/4}}\sum_{t=1}^T |K_t^0(\tau) + K_t^1(\tau)|^{2+\delta/2}\left[\left\{\mathrm{E}\left|\sum_{i=1}^N e_{it}\right|^{2+\delta/2}\right\}^{1/(2+\delta/2)} + \left\{\mathrm{E}\left|\sum_{i=1}^N v_{it}e_{it}\right|^{2+\delta/2}\right\}^{1/(2+\delta/2)}\right]^{2+\delta/2} \\
= &\ \frac{C}{N^{1+\delta/4}(Th)^{\delta/4}}\left\{\frac{1}{Th}\sum_{t=1}^T |K_t^0(\tau) + K_t^1(\tau)|^{2+\delta/2}\right\} O\left(N^{1+\delta/4}\right) = O\left(\frac{1}{(Th)^{\delta/2}}\right) = o(1)
\end{aligned}$$

where the last line is due to

$$\mathrm{E}\left|\sum_{i=1}^N e_{it}\right|^{2+\delta/2} = O\left(N^{1+\delta/4}\right) \quad \text{and} \quad \mathrm{E}\left|\sum_{i=1}^N v_{it}e_{it}\right|^{2+\delta/2} = O\left(N^{1+\delta/4}\right), \tag{B.5}$$

which are to be shown later and $(Th)^{-1}\sum_{t=1}^T |K_t^0(\tau) + K_t^1(\tau)|^{2+\delta/2} = O(1)$ is because of the definition of

Riemann integral uniformly for $0 < \tau < 1$. Similar techniques is used in the proof of (B.7) in Chen et al. (2012, p. 83).

Next we show (B.5). For given N, we get

$$\mathrm{E}\left|\sum_{i=1}^{N} e_{it}\right|^{2+\delta/2} < \infty \quad \text{and} \quad \mathrm{E}\left|\sum_{i=1}^{N} v_{it} e_{it}\right|^{2+\delta/2} < \infty$$

due to the Minkowski's inequality and

$$\mathrm{E}\left|e_{it}\right|^{2+\delta/2} \leq \sqrt{\mathrm{E}\left|e_{it}\right|^{4+\delta}} < \infty, \quad \mathrm{E}\left|v_{it} e_{it}\right|^{2+\delta/2} \leq \sqrt{\mathrm{E}\left|v_{it}\right|^{4+\delta}} \sqrt{\mathrm{E}\left|e_{it}\right|^{4+\delta}} < \infty,$$

which is from Hölder's inequality. It means the moments of $\mathrm{E}\left|e_{it}\right|^{2+\delta/2}$ and $\mathrm{E}\left|v_{it} e_{it}\right|^{2+\delta/2}$ exist.

Next, we derive the order of $\mathrm{E}\left|\sum_{i=1}^{N} e_{it}\right|^{2+\delta/2}$ and $\mathrm{E}\left|\sum_{i=1}^{N} v_{it} e_{it}\right|^{2+\delta/2}$. For given t, we define

$$\mathscr{F}_{i-1,t} = \mathcal{E}_{t-1} \vee \sigma\langle e_{1t}, \cdots, e_{i-1,t}\rangle$$

for $i \geq 1$ where $\mathscr{F}_{i-1,t} := \mathcal{E}_{t-1}$ if $i = 1$, and $\mathcal{E}_{t-1} \vee \sigma\langle e_{1t}, \cdots, e_{i-1,t}\rangle$ is the σ-field generated by $\mathcal{E}_{t-1} \cup \sigma\langle e_{1t}, \cdots, e_{i-1,t}\rangle$. Due to the conditional independence of e_{it} and $\{e_{1t}, \cdots, e_{i-1,t}\}$ given $\mathcal{E}_{t-1}$ based on Assumption 2, we obtain $\mathrm{E}(e_{it}|\mathscr{F}_{i-1,t}) = \mathrm{E}(e_{it}|\mathcal{E}_{t-1} \vee \sigma\langle e_{1t}, \cdots, e_{i-1,t}\rangle) = \mathrm{E}(e_{it}|\mathcal{E}_{t-1}) = 0$ which is in fact from the property of the conditional independence (e.g., Theorem 9.2.1 in Chung (2001), p. 322). Thus $\{(e_{it}, \mathscr{F}_{it}) : 1 \leq i \leq N\}$ forms a martingale difference array. Using the property that $v_{it} \in \mathcal{F}_V$, we further obtain $\mathrm{E}(v_{it} e_{it}|\mathscr{F}_{i-1,t}) = 0$, and hence $\{(v_{it} e_{it}, \mathscr{F}_{it}) : 1 \leq i \leq N\}$ also forms a martingale difference array. This result, together with Lin and Bai (2011) and Burkholder (1973), leads to

$$\mathrm{E}\left|\sum_{i=1}^{N} e_{it}\right|^{2+\delta/2} \leq CN^{1+\delta/4} \quad \text{and} \quad \mathrm{E}\left|\sum_{i=1}^{N} v_{it} e_{it}\right|^{2+\delta/2} \leq CN^{1+\delta/4}.$$

which are both uniformly in t. Hence, we have shown (B.3) holds.

To prove (B.4), we only need to show $\sum_{t=1}^{T} \mathrm{E}\left(Z_t Z_t^\top | \mathcal{E}_{t-1}\right) = \sigma_0^2 \Lambda_\mu(\tau) \otimes \Sigma_X + o_{\mathrm{P}}(1)$. As

$$\sum_{t=1}^{T} Z_t Z_t^\top = \frac{1}{NTh} \sum_{t=1}^{T} \begin{pmatrix} \left(K_t^0(\tau)\right)^2 X_t^\top \mathbf{e}_t \mathbf{e}_t^\top X_t & \left(K_t^1(\tau)\right)^2 X_t^\top \mathbf{e}_t \mathbf{e}_t^\top X_t \\ \left(K_t^1(\tau)\right)^2 X_t^\top \mathbf{e}_t \mathbf{e}_t^\top X_t & \left(K_t^2(\tau)\right)^2 X_t^\top \mathbf{e}_t \mathbf{e}_t^\top X_t \end{pmatrix},$$

and for each element we can show

$$\begin{aligned} &\frac{1}{NTh} \sum_{t=1}^{T} \left(K_t^\ell(\tau)\right)^2 \mathrm{E}\left(X_t^\top \mathbf{e}_t \mathbf{e}_t^\top X_t | \mathcal{E}_{t-1}\right) = \frac{1}{NTh} \sum_{t=1}^{T} \left(K_t^\ell(\tau)\right)^2 X_t^\top \mathrm{E}\left(\mathbf{e}_t \mathbf{e}_t^\top | \mathcal{E}_{t-1}\right) X_t \\ &= \sigma_0^2 \frac{1}{NTh} \sum_{t=1}^{T} \left(K_t^\ell(\tau)\right)^2 X_t^\top X_t = \sigma_0^2 \Sigma_X(\tau) \nu_\ell(\tau) + o_{\mathrm{P}}(1), \end{aligned}$$

where the last equality can be shown by the same approach as in Lemma C.2 (1). Hence, we have shown

(B.4) holds.

(2) Next, we show $\Xi_{N,T}(3,2)=o_{\mathrm{P}}(1)$.

By Lemma C.8, we get $\mathrm{E}\left(\Xi_{N,T}(3,2)\right)=\mathrm{E}\left(\sqrt{NTh}\Phi(\tau)D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\mathrm{E}(\mathbf{e}|\mathcal{F}_V)\right)=\mathbf{0}_d$ and

$$
\begin{aligned}
& \mathrm{Var}\left(\Xi_{N,T}(3,2)\right) \\
= & \mathrm{E}\left(\mathrm{Var}(\Xi_{N,T}(3,2)|\mathcal{F}_V)\right)+\mathrm{Var}\left(\mathrm{E}(\Xi_{N,T}(3,2)|\mathcal{F}_V)\right) \\
= & \mathrm{E}\left(\mathrm{Var}(\sqrt{NTh}\Phi(\tau)D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\mathbf{e}|\mathcal{F}_V)\right) \\
= & (NTh)\mathrm{E}\left(\Phi(\tau)D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\mathrm{Var}(\mathbf{e}|\mathcal{F}_V)(I_{NT}-S)^\top\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}D^\top\Phi(\tau)^\top\right) \\
= & \sigma_0^2\mathrm{E}\left(NTh\Phi(\tau)D(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)(I_{NT}-S)^\top\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}D^\top\Phi(\tau)^\top\right) \\
:= & \sigma_0^2\mathrm{E}(\Xi_{N,T}(3,2,1)).
\end{aligned}
$$

Therefore, we need only to show $\left\|\Xi_{N,T}(3,2,1)\right\|_2=o_{\mathrm{P}}(1)$. By definition, we can write $\Xi_{N,T}(3,2,1)$ as

$$
\begin{aligned}
\Xi_{N,T}(3,2,1) = & NTh\Phi(\tau)D\left(D^\top(I_{NT}-\Upsilon)D\right)^{-1}D^\top\Phi(\tau)^\top \\
& -NTh\Phi(\tau)D\left(D^\top(I_{NT}-\Upsilon)D\right)^{-1}D^\top\Upsilon D\left(D^\top(I_{NT}-\Upsilon)D\right)^{-1}D^\top\Phi(\tau)^\top \\
& +NTh\Phi(\tau)D\left(D^\top(I_{NT}-\Upsilon)D\right)^{-1}D^\top\Upsilon\Upsilon D\left(D^\top(I_{NT}-\Upsilon)D\right)^{-1}D^\top\Phi(\tau)^\top \\
:= & \Xi_{N,T}(3,2,1,1)+\Xi_{N,T}(3,2,1,2)+\Xi_{N,T}(3,2,1,3).
\end{aligned}
$$

The leading order of $\Xi_{N,T}(3,2,1,1)$ is

$$
\begin{aligned}
& NTh\Phi(\tau)D\left(D^\top D\right)^{-1}D^\top\Phi(\tau)^\top \\
= & NTh(I_d,0_{d\times d})\left\{\frac{M^\top(\tau)\Omega(\tau)M(\tau)}{NTh}\right\}^{-1}\frac{1}{NTh}M^\top(\tau)\Omega(\tau)\left\{\frac{1}{T}\mathbf{1}_T\mathbf{1}_T^\top\otimes\left(I_N-\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top\right)\right\} \\
& \frac{1}{NTh}\Omega(\tau)M(\tau)\left\{\frac{M^\top(\tau)\Omega(\tau)M(\tau)}{NTh}\right\}^{-1}(I_d,0_{d\times d})^\top \\
= & \frac{1}{NT^2h}(I_d,0_{d\times d})\widehat{\Delta}_M(\tau)^{-1}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s-\frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\mathbb{K}_s(\tau)^\top\right\}\widehat{\Delta}_M(\tau)^{-1}(I_d,0_{d\times d})^\top
\end{aligned}
$$

By Lemma C.2 and Lemma C.4, we get $\left\|NTh\Phi(\tau)D\left(D^\top D\right)^{-1}D^\top\Phi(\tau)^\top\right\|_2=o_{\mathrm{P}}(1)$. Similarly, we can show $\|\Xi_{N,T}(3,2,1,2)\|=o_{\mathrm{P}}(1)$ and $\|\Xi_{N,T}(3,2,1,3)\|=o_{\mathrm{P}}(1)$. Hence, we have completed the proof.

□

Appendix C: Additional Lemmas and Their Proofs

In this appendix, we denote C as a finite constant, which adjusts its values under different scenarios.

Lemma C.1. *Under Assumption 1, for any $1\le t,s\le T$, $1\le k_1,k_2,k_3,k_4\le d$, $1\le i_1,i_2,i_3,i_4\le N$ and some $\delta>0$, we have*

$$|\mathrm{Cov}\left(X_{i_1tk_1}X_{i_2tk_2},X_{i_3sk_3}X_{i_4sk_4}\right)|\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}.$$

Proof. By definition, we get

$$
\begin{aligned}
X_{i_1tk_1}X_{i_2tk_2} &= g_{k_1}(\tau_t)g_{k_2}(\tau_t)+g_{k_2}(\tau_t)v_{i_1tk_1}+g_{k_1}(\tau_t)v_{i_2tk_2}+v_{i_1tk_1}v_{i_2tk_2}\\
\text{and}\quad X_{i_3sk_3}X_{i_4sk_4} &= g_{k_3}(\tau_s)g_{k_4}(\tau_s)+g_{k_4}(\tau_s)v_{i_3sk_3}+g_{k_3}(\tau_s)v_{i_4sk_4}+v_{i_3sk_3}v_{i_4sk_4}
\end{aligned}
$$

so that we obtain

$$
\begin{aligned}
\mathrm{Cov}\left(X_{i_1tk_1}X_{i_2tk_2},X_{i_3sk_3}X_{i_4sk_4}\right) &= g_{k_2}(\tau_t)g_{k_4}(\tau_s)\mathrm{Cov}(v_{i_1tk_1},v_{i_3sk_3})+g_{k_1}(\tau_t)g_{k_4}(\tau_s)\mathrm{Cov}(v_{i_2tk_2},v_{i_3sk_3})+\\
&\quad g_{k_2}(\tau_t)g_{k_3}(\tau_s)\mathrm{Cov}(v_{i_1tk_1},v_{i_4sk_4})+g_{k_1}(\tau_t)g_{k_3}(\tau_s)\mathrm{Cov}(v_{i_2tk_2},v_{i_4sk_4})+\\
&\quad g_{k_4}(\tau_s)\mathrm{Cov}(v_{i_1tk_1}v_{i_2tk_2},v_{i_3sk_3})+g_{k_3}(\tau_s)\mathrm{Cov}(v_{i_1tk_1}v_{i_2tk_2},v_{i_4sk_4})+\\
&\quad g_{k_2}(\tau_t)\mathrm{Cov}(v_{i_1tk_1},v_{i_3sk_3}v_{i_4sk_4})+g_{k_1}(\tau_t)\mathrm{Cov}(v_{i_2tk_2},v_{i_3sk_3}v_{i_4sk_4})+\\
&\quad \mathrm{Cov}(v_{i_1tk_1}v_{i_2tk_2},v_{i_3sk_3}v_{i_4sk_4}):=\sum_{m=1}^{9}\Xi_{cov}(m).
\end{aligned}
$$

Then, we get

$$
\left|\mathrm{Cov}\left(X_{i_1tk_1}X_{i_2tk_2},X_{i_3sk_3}X_{i_4sk_4}\right)\right|\le\sum_{m=1}^{9}\left|\Xi_{cov}(m)\right|.
$$

To complete the proof, it is sufficient to show that $|\Xi_{cov}(m)|\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}$ for $m=1,\cdots,9$. Here we only show the following terms, as the other terms can be obtained similarly.

According to Proposition 2.5 in Fan and Yao (2008), under Assumption 1 we have

$$
\begin{aligned}
|\Xi_{cov}(1)| &\le \max_{1\le t,s\le T,1\le k_1,k_2\le d}|g_{k_1}(\tau_t)||g_{k_2}(\tau_s)|\cdot|\mathrm{Cov}(v_{i_1tk_1},v_{i_3sk_3})|\\
&\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}\left\{\mathrm{E}(|v_{i_1tk_1}|^{2+\delta/2})\right\}^{2/(4+\delta)}\left\{\mathrm{E}(|v_{i_3sk_3}|^{2+\delta/2})\right\}^{2/(4+\delta)}\\
&\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)},
\end{aligned}
$$

where the last inequality is due to $\mathrm{E}(|v_{itk}|^{2+\delta/2})\le\sqrt{\mathrm{E}(|v_{itk}|^{4+\delta})}<\infty$ by Hölder's inequality. Similarly, we have

$$
\begin{aligned}
|\Xi_{cov}(5)| &\le \max_{1\le s\le T,1\le k_4\le d}|g_{k_4}(\tau_s)|\cdot|\mathrm{Cov}(v_{i_1tk_1}v_{i_2tk_2},v_{i_3sk_3})|\\
&\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}\left\{\mathrm{E}(|v_{i_1tk_1}v_{i_2tk_2}|^{2+\delta/2})\right\}^{2/(4+\delta)}\left\{\mathrm{E}(|v_{i_3sk_3}|^{2+\delta/2})\right\}^{2/(4+\delta)}\\
&\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}\left\{\mathrm{E}(|v_{i_1tk_1}|^{4+\delta}\mathrm{E}|v_{i_2tk_2}|^{4+\delta})\right\}^{1/(4+\delta)}\left\{\mathrm{E}(|v_{i_3sk_3}|^{2+\delta/2})\right\}^{2/(4+\delta)}\\
&\le C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)},
\end{aligned}
$$

and

$$
\begin{aligned}
|\Xi_{cov}(9)| &\leq |\mathrm{Cov}(v_{i_1tk_1}v_{i_2tk_2}, v_{i_3sk_3}v_{i_4sk_4})| \\
&\leq C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}\left\{\mathrm{E}(|v_{i_1tk_1}v_{i_2tk_2}|^{2+\delta/2})\right\}^{2/(4+\delta)}\left\{\mathrm{E}(|v_{i_3sk_3}v_{i_4sk_4}|^{2+\delta/2})\right\}^{2/(4+\delta)} \\
&\leq C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}\left\{\mathrm{E}(|v_{i_1tk_1}|^{4+\delta}\mathrm{E}|v_{i_2tk_2}|^{4+\delta})\right\}^{1/(4+\delta)}\left\{\mathrm{E}(|v_{i_3tk_3}|^{4+\delta}\mathrm{E}|v_{i_4tk_4}|^{4+\delta})\right\}^{1/(4+\delta)} \\
&\leq C\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)}.
\end{aligned}
$$

So the proof is accomplished.

□

Lemma C.2. *Suppose Assumption 1 and 3 hold. As $N, T \to \infty$ simultaneously and $0 < \tau \leq 1$, we have*

$$
(1)\left\|\frac{1}{NTh}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^{\ell}K\left(\frac{t-\tau T}{Th}\right)X_t^{\top}X_t - \Sigma_X(\tau)\mu_\ell(\tau)\right\|_2 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right),\ \textit{for } 0 \leq \ell \leq 3,
$$

$$
(2)\left\|\widehat{\Delta}_M(\tau) - \Delta_M(\tau)\right\|_2 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right) = o_{\mathrm{P}}(1),
$$

$$
(3)\left\|\widehat{\Delta}_{M,N}(\tau) - \Delta_{M,N}(\tau)\right\|_2 = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right) = o_{\mathrm{P}}(1).
$$

Proof. (1) Denote $K_t^{\ell}(\tau) = \left(\frac{t-\tau T}{Th}\right)^{\ell}K\left(\frac{t-\tau T}{Th}\right)$, $\Delta_X = \Sigma_X(\tau)\mu_\ell(\tau) = \left(\Delta_X^{(k_1,k_2)}\right)_{d\times d}$ and $\widehat{\Delta}_X = \left(\widehat{\Delta}_X^{(k_1,k_2)}\right)_{d\times d} = \frac{1}{NTh}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^{\ell}K\left(\frac{t-\tau T}{Th}\right)X_t^{\top}X_t$ where $\Delta_X^{(k_1,k_2)}$ and $\widehat{\Delta}_X^{(k_1,k_2)}$ are the (k_1,k_2)-element of these two matrices. Since Δ_X and $\widehat{\Delta}_X$ all have finite dimension, now it suffices to prove for any $1 \leq k_1, k_2 \leq d$, it holds that

$$
\left|\widehat{\Delta}_X^{(k_1,k_2)} - \Delta_X^{(k_1,k_2)}\right| = \left|\widehat{\Delta}_X^{(k_1,k_2)} - \mathrm{E}\left(\widehat{\Delta}_X^{(k_1,k_2)}\right) + \mathrm{E}\left(\widehat{\Delta}_X^{(k_1,k_2)}\right) - \Delta_X^{(k_1,k_2)}\right| = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right). \tag{C.1}
$$

By Assumptions 1 and 3, we get

$$
\begin{aligned}
\mathrm{E}\left(\widehat{\Delta}_X^{(k_1,k_2)}\right) &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}K_t^{\ell}(\tau)\mathrm{E}\left(X_{itk_1}X_{itk_2}\right) \\
&= \frac{1}{Th}\sum_{t=1}^{T}\left\{K_t^{\ell}(\tau)\left(g_{k_1}(\tau_t)g_{k_2}(\tau_t) + \sigma_{\mathbf{v}}^{(k_1,k_2)}\right)\right\} \\
&= \Delta_X^{(k_1,k_2)} + O\left(\frac{1}{Th}\right),
\end{aligned}
$$

where the last term is due to the definition of Riemann integral uniformly for $0 < \tau \leq 1$. Similar techniques is used in the proof of (B.7) in Chen et al. (2012, p. 83). Hence, to prove (C.1), we only need to show that $\left|\widehat{\Delta}_X^{(k_1,k_2)} - \mathrm{E}\left(\widehat{\Delta}_X^{(k_1,k_2)}\right)\right| = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right)$.

As $\widehat{\Delta}_X^{(k_1,k_2)} - \mathrm{E}\widehat{\Delta}_X^{(k_1,k_2)} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Th}\sum_{t=1}^{T}K_t^{\ell}(\tau)\left\{X_{itk_1}X_{itk_2} - \mathrm{E}\left(X_{itk_1}X_{itk_2}\right)\right\}$, it is sufficient to show

$$
\frac{1}{Th}\sum_{t=1}^{T}K_t^{\ell}(\tau)\left\{X_{itk_1}X_{itk_2} - \mathrm{E}\left(X_{itk_1}X_{itk_2}\right)\right\} = O_{\mathrm{P}}\left(\frac{1}{\sqrt{Th}}\right)
$$

for each fixed (i, k_1, k_2). By Lemma C.1, Assumptions 1 and 3, we obtain the order of its variance as follows:

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{Th}\sum_{t=1}^{T} K_t^{\ell}(\tau)\left\{X_{itk_1}X_{itk_2} - \mathrm{E}\left(X_{itk_1}X_{itk_2}\right)\right\}\right) = \frac{1}{(Th)^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_t^{\ell}(\tau)K_s^{\ell}(\tau)\mathrm{Cov}\left(X_{itk_1}X_{itk_2}, X_{isk_1}X_{isk_2}\right) \\
&\leq \frac{1}{(Th)^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_t^{\ell}(\tau)K_s^{\ell}(\tau)\left|\mathrm{Cov}\left(X_{itk_1}X_{itk_2}, X_{isk_1}X_{isk_2}\right)\right| \\
&\leq \frac{C}{(Th)^2}\sum_{t=1}^{T}\sum_{s=1}^{T} K_t^{\ell}(\tau)K_s^{\ell}(\tau)\alpha_{\mathrm{mix}}(|t-s|)^{\delta/(4+\delta)} \\
&\leq \frac{2C}{Th}\left\{\sum_{h=0}^{T-1}\alpha_{\mathrm{mix}}(h)^{\delta/(4+\delta)}\right\}\left\{\frac{1}{Th}\sum_{t=1}^{T}K_t^{\ell}(\tau)\right\} = O\left(\frac{1}{Th}\right).
\end{aligned}$$

Therefore, we have completed the proof of (1).

Proof of (2) and (3):

By definition we have

$$\widehat{\Delta}_M(\tau) = \frac{1}{NTh}\begin{pmatrix} \sum_{t=1}^{T} K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t & \sum_{t=1}^{T} \frac{t-\tau T}{Th} K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t \\ \sum_{t=1}^{T} \frac{t-\tau T}{Th} K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t & \sum_{t=1}^{T} \left(\frac{t-\tau T}{Th}\right)^2 K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t \end{pmatrix}$$

and

$$\widehat{\Delta}_{M,N}(\tau) = \frac{1}{NTh}\begin{pmatrix} \sum_{t=1}^{T} \left(\frac{t-\tau T}{Th}\right)^2 K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t \\ \sum_{t=1}^{T} \left(\frac{t-\tau T}{Th}\right)^3 K\left(\frac{t-\tau T}{Th}\right) X_t^{\top} X_t, \end{pmatrix}$$

which are block matrices with each block having the same convergence rate as (1). Hence, (2) and (3) hold.

□

Lemma C.3. *Under Assumption 1, 3 and 6, $\zeta_t = X_t\boldsymbol{\beta}(\tau_t) - \sum_{s_1=1}^{T}\Upsilon_{t,s_1}X_{s_1}\boldsymbol{\beta}(\tau_{s_1}) = X_t\left(o(h^2)\right)$ as $N, T \to \infty$ where Υ_{t,s_1} is the (t, s_1)-th block matrix of $\Upsilon = S^{\top} + S - S^{\top}S$.*

Proof. By definition, we get

$$\zeta_t = X_t\left(\boldsymbol{\beta}(\tau_t) - \sum_{s_1=1}^{T}\Phi_{s_1}(\tau_t)X_{s_1}\boldsymbol{\beta}(\tau_{s_1})\right) - \sum_{s_1=1}^{T}\left[(X_{s_1}\Phi_t(\tau_{s_1}))^{\top} - \sum_{s_2=1}^{T}\{X_{s_2}\Phi_t(\tau_{s_2})\}^{\top}\ X_{s_2}\Phi_{s_1}(\tau_{s_2})\right]X_{s_1}\boldsymbol{\beta}(\tau_{s_1}).$$

Then, according to Taylor expansion, the first term can be written as

$$
\begin{aligned}
& X_t\left(\boldsymbol{\beta}(\tau_t)-\sum_{s_1=1}^{T}\Phi_{s_1}(\tau_t)X_{s_1}\boldsymbol{\beta}(\tau_{s_1})\right)\\
= \;& X_t\left\{\boldsymbol{\beta}(\tau_t)-(I_d,0_{d\times d})\widehat{\Delta}_M(\tau_t)^{-1}\frac{1}{NTh}\sum_{s_1=1}^{T}\mathbb{K}_{s_1}(\tau_t)X_{s_1}^\top X_{s_1}\left(\boldsymbol{\beta}(\tau_t)+\frac{1}{2}h^2\boldsymbol{\beta}''(\tau_t)+o(h^2)\right)\right\}\\
= \;& X_t\left\{\boldsymbol{\beta}(\tau_t)-(I_d,0_{d\times d})\widehat{\Delta}_M(\tau_t)^{-1}\widehat{\Delta}_M(\tau_t)(I_d,0_{d\times d})^\top\left(\boldsymbol{\beta}(\tau_t)+\frac{1}{2}h^2\boldsymbol{\beta}''(\tau_t)+o(h^2)\right)\right\}\\
= \;& -\frac{1}{2}h^2X_t\boldsymbol{\beta}''(\tau_t)+X_t\left(o(h^2)\right).
\end{aligned}
$$

Similarly, the second term can be shown to be $-\frac{1}{2}h^2X_t\boldsymbol{\beta}''(\tau_t)+X_t\left(o(h^2)\right)$ as well.

Hence, $\zeta_t = X_t\left(o(h^2)\right)$ holds.

□

Lemma C.4. *Under Assumption 1 and 3, we have*

$$
\left\|\frac{1}{NT^2h}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s-\frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\}\right\|_2 = o_{\mathrm{P}}(1)
$$

as $N,T\to\infty$, *where* $\mathbb{K}_t(\tau)=\left(K\left(\frac{\tau_t-\tau}{h}\right)I_d,\frac{\tau_t-\tau}{h}K\left(\frac{\tau_t-\tau}{h}\right)I_d\right)^\top$.

Proof. By definition, we have a $2d\times d$ matrix

$$
\begin{aligned}
\frac{1}{NT^2h}\sum_{t,s=1}^{T}\left\{\mathbb{K}_t(\tau)\left(\mathbf{v}_t^\top\mathbf{v}_s-\frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\} &= \begin{pmatrix}\frac{1}{NT^2h}\sum_{t,s=1}^{T}\left\{K\left(\frac{\tau_t-\tau}{h}\right)\left(\mathbf{v}_t^\top\mathbf{v}_s-\frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\}\\ \frac{1}{NT^2h}\sum_{t,s=1}^{T}\left\{\frac{\tau_t-\tau}{h}K\left(\frac{\tau_t-\tau}{h}\right)\left(\mathbf{v}_t^\top\mathbf{v}_s-\frac{1}{N}\mathbf{v}_t^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{v}_s\right)\right\}\end{pmatrix}\\
:= \begin{pmatrix}\widehat{\Delta}_{\mathbf{v}}^{(0)}\\ \widehat{\Delta}_{\mathbf{v}}^{(1)}\end{pmatrix}.
\end{aligned}
$$

So we only need to prove $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2)}=o_{\mathrm{P}}(1)$, which is the (k_1,k_2)-th element of $\widehat{\Delta}_{\mathbf{v}}^{(\ell)}, \ell=0,1$. By simple algebra, we can get

$$
\begin{aligned}
\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2)} &= \frac{1}{NT^2h}\sum_{t,s=1}^{T}\sum_{i=1}^{N}K_t^\ell(\tau)v_{itk_1}v_{isk_2}-\frac{1}{N^2T^2h}\sum_{i_1,i_2=1}^{N}\sum_{t,s=1}^{T}K_t^\ell(\tau)v_{i_1tk_1}v_{i_2sk_2}\\
&:= \widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,1)}-\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,2)}.
\end{aligned}
$$

Below we show that $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,1)}=o_{\mathrm{P}}(1)$.

Since $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,1)}=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T^2h}\sum_{t,s=1}^{T}K_t^\ell(\tau)v_{itk_1}v_{isk_2}\right)$, it is sufficient to show for each given i

$$
\frac{1}{T^2h}\sum_{t,s=1}^{T}K_t^\ell(\tau)v_{itk_1}v_{isk_2}=o_{\mathrm{P}}(1),
$$

which is equivalently to prove

$$\mathrm{E}\left(\frac{1}{T^2h}\sum_{t,s=1}^{T}K_t^{\ell}(\tau)v_{itk_1}v_{isk_2}\right)^2=o(1),$$

Here we adapt the proof of Theorem 1 in Doukhan and Louhichi (1999) to show this result.

Note that

$$\begin{aligned}\left|\mathrm{E}\left(\frac{1}{T^2h}\sum_{t,s=1}^{T}K_t^{\ell}(\tau)v_{itk_1}v_{isk_2}\right)^2\right| &\leq \left(\frac{1}{T^2h}\right)^2\sum_{t_1,t_2,t_3,t_4=1}^{T}K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{E}\left(v_{it_1k_1}v_{it_2k_2}v_{it_3k_1}v_{it_4k_2}\right)\right|\\ &\leq \sum_{m=1}^{4!}\Xi_{\widehat{\Delta}}(m)\end{aligned}$$

where 4! terms of $\Xi_{\widehat{\Delta}}(m)$ have similar forms with each of them being one possible arrangement of $t_1,\cdots,t_4$. For instance,

$$\Xi_{\widehat{\Delta}}(1)=\left(\frac{1}{T^2h}\right)^2\sum_{t_1\leq t_2\leq t_3\leq t_4}^{T}K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{E}\left(v_{it_1k_1}v_{it_2k_2}v_{it_3k_1}v_{it_4k_2}\right)\right|.$$

Then $\Xi_{\widehat{\Delta}}(2),\cdots,\Xi_{\widehat{\Delta}}(4!)$ just have different permutations at the subscript compared to $t_1\leq t_2\leq t_3\leq t_4$ in $\Xi_{\widehat{\Delta}}(1)$. Now we show $\Xi_{\widehat{\Delta}}(1)=o(1)$ and hence $\Xi_{\widehat{\Delta}}(2),\cdots,\Xi_{\widehat{\Delta}}(4!)$ can be similarly shown to be $o(1)$. In order to simplify the proof we show a more general result

$$\left(\frac{1}{T^2h}\right)^2\sum_{t_1\leq t_2\leq t_3\leq t_4}^{T}K_{t_1}^{\ell}(\tau)K_{t_2}^{\ell}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}v_{it_2k_2^*}v_{it_3k_3^*}v_{it_4k_4^*}\right)\right|=o(1),\tag{C.2}$$

for any $k_1^*,k_2^*,k_3^*,k_4^*\in\{1,\cdots,d\}$.

We then obtain

$$\begin{aligned}&\left(\frac{1}{T^2h}\right)^2\sum_{t_1\leq t_2\leq t_3\leq t_4}^{T}K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}v_{it_2k_2^*}v_{it_3k_3^*}v_{it_4k_4^*}\right)\right|\\ \leq\ &\left(\frac{1}{T^2h}\right)^2\sum K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}\right)\mathrm{E}\left(v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right|\\ &+\left(\frac{1}{T^2h}\right)^2\sum K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*},v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right|\end{aligned}$$

where the sum $\sum$ is considered over $\{t_1,\cdots,t_4\}$ fulfilling $1\leq t_1\leq\cdots\leq t_4\leq T$ with $r=t_{m+1}-t_m=$

$\max_{1\le j<4}(t_{j+1}-t_j)$. The first term on the right-hand side of the last inequality is bounded by

$$\begin{aligned}
&\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}\right)\mathrm{E}\left(v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right| \\
\le\ &\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}\right)\mathrm{E}\left(v_{it_2k_2^*}\cdots v_{it_4k_4^*}\right)\right| \\
&+\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}v_{it_2k_2^*}\right)\mathrm{E}\left(v_{it_3k_3^*}v_{it_4k_4^*}\right)\right| \\
&+\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}\cdots v_{it_3k_3}\right)\mathrm{E}\left(v_{it_4k_4^*}\right)\right| \\
\le\ &\left\{\frac{1}{T^2h}\sum_{t_1\le t_2}^{T} K^\ell_{t_1}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}v_{it_2k_2^*}\right)\right|\right\}\left\{\frac{1}{T^2h}\sum_{t_3\le t_4}^{T} K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_3k_3^*}v_{it_4k_4^*}\right)\right|\right\}
\end{aligned}$$

since $\mathrm{E}(v_{itk})=0$.

By Assumption 1 and Assumption 3, we get

$$\frac{1}{T^2h}\sum_{t_1\le t_2}^{T} K^\ell_{t_1}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}v_{it_2k_2^*}\right)\right| \le \frac{C}{T}\left\{\sum_{r=0}^{T-1}\alpha_{\mathrm{mix}}(r)^{\delta/(4+\delta)}\right\}\left\{\frac{1}{Th}\sum_{t=1}^{T}K^\ell_t(\tau)\right\} = O\left(\frac{1}{T}\right).$$

So that we get

$$\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{E}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}\right)\mathrm{E}\left(v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right| = O\left(\frac{1}{T^2}\right).$$

Now it is sufficient to derive the order of

$$\begin{aligned}
&\left(\frac{1}{T^2h}\right)^2 \sum K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right| \\
=\ &\left(\frac{1}{T^2h}\right)^2 \sum_{t_1=1}^{T}\sum^{*} K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right|
\end{aligned}$$

where $\sum^*$ denotes a sum over such a collection $1\le t_1\le\cdots\le t_4\le T$ with fixed t_1, and $r=t_{m+1}-t_m=\max_{1\le j<4}(t_{j+1}-t_j)\in[0,T-1]$. Again

$$\begin{aligned}
&\left(\frac{1}{T^2h}\right)^2 \sum^{*} K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right| \\
\le\ &\left(\frac{1}{T^2h}\right)^2 \sum_{t_3=1}^{T}\sum_{r=0}^{T-1}\sum^{**} K^\ell_{t_1}(\tau)K^\ell_{t_3}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k^*_{m+1}}\cdots v_{it_4k_4^*}\right)\right|
\end{aligned}$$

where $\sum^{**}$ denotes the sum over such a collection $1\le t_1\le t_2\le t_3\le t_4\le T$ with fixed t_1, t_3 and r, and $t_{m+1}=t_m+r$ and $0\le t_{j+1}-t_j\le r$ for $j\ne m$. Hence $\sum^{**}1\le T$.

Under Assumption 1, it follows from Proposition 2.5 (i) in Fan and Yao (2008) that

$$
\begin{aligned}
&\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right| \\
\leq\;& C\alpha_{\mathrm{mix}}(r)^{\delta/(4+\delta)}\left\{\mathrm{E}\left|v_{it_1k_1^*}\cdots v_{it_mk_m^*}\right|^{(4+\delta)/m}\right\}^{m/(4+\delta)}\left\{\mathrm{E}\left|v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right|^{(4+\delta)/(4-m)}\right\}^{(4-m)/(4+\delta)}.
\end{aligned}
$$

By using Hölder's inequality successively, we have that

$$
\mathrm{E}\left|v_{it_1k_1^*}\cdots v_{it_mk_m^*}\right|^{(4+\delta)/m} \leq \left\{\mathrm{E}\left|v_{it_1k_1^*}\right|^{4+\delta}\right\}^{1/m}\cdots\left\{\mathrm{E}\left|v_{it_mk_m^*}\right|^{4+\delta}\right\}^{1/m} \quad \text{and}
$$

$$
\mathrm{E}\left|v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right|^{(4+\delta)/(4-m)} \leq \left\{\mathrm{E}\left|v_{it_{m+1}k_{m+1}^*}\right|^{4+\delta}\right\}^{1/(4-m)}\cdots\left\{\mathrm{E}\left|v_{it_4k_4^*}\right|^{4+\delta}\right\}^{1/(4-m)}.
$$

This, together with Assumption 1, leads to

$$
\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right| \leq C\alpha_{\mathrm{mix}}(r)^{\delta/(4+\delta)},
$$

and

$$
\begin{aligned}
&\left(\frac{1}{T^2h}\right)^2\sum K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right| \\
=\;&\left(\frac{1}{T^2h}\right)^2\sum_{t_1=1}^{T}\sum_{t_3=1}^{T}\sum_{r=0}^{T-1}\sum^{**}K_{t_1}^{\ell}(\tau)K_{t_3}^{\ell}(\tau)\left|\mathrm{Cov}\left(v_{it_1k_1^*}\cdots v_{it_mk_m^*}, v_{it_{m+1}k_{m+1}^*}\cdots v_{it_4k_4^*}\right)\right| \\
\leq\;&\frac{1}{T^2}\sum_{r=0}^{T-1}C\alpha_{\mathrm{mix}}(r)^{\delta/(4+\delta)}\left\{\frac{1}{Th}\sum_{t_1=1}^{T}K_{t_1}^{\ell}(\tau)\right\}\left\{\frac{1}{Th}\sum_{t_3=1}^{T}K_{t_3}^{\ell}(\tau)\right\}\sum^{**}1 = O\left(\frac{1}{T}\right).
\end{aligned}
$$

Hence, we have shown that (C.2) holds so that $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,1)} = o_{\mathrm{P}}(1)$. Similarly, we can show $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2,2)} = o_{\mathrm{P}}(1)$.

Hence, it gives $\widehat{\Delta}_{\mathbf{v}}^{(\ell,k_1,k_2)} = o_{\mathrm{P}}(1)$ which completes the whole proof.

□

Lemma C.5. *Under Assumptions 1, 3 - 7, for any $NT\times NT$ UB matrix A,*

$$
\frac{\mathrm{tr}(P_{N,T}A)}{NT} = \frac{\mathrm{tr}(A)}{NT} + o_{\mathrm{P}}(1)
$$

for sufficiently large N and T. For instance, A can take matrix I_{NT}, $G_{N,T}$, $G_{N,T}^2$ and $G_{N,T}G_{N,T}^{\top}$ since these matrices are UB under Assumptions 4 and 5.

Proof. Due to the symmetry of $P_{N,T}$, we have

$$
\mathrm{tr}(P_{N,T}A) = \mathrm{tr}(P_{N,T}A^{\top}) = \mathrm{tr}\left(P_{N,T}\frac{A+A^{\top}}{2}\right)
$$

where $(A+A^{\top})/2$ is a symmetric matrix. Then, we only need to show the lemma for the case of symmetric matrix A. From now on, we assume A is symmetric.

By calculation,

$$\begin{aligned}
\frac{\mathrm{tr}(P_{N,T}A)}{NT} - \frac{\mathrm{tr}(A)}{NT} &= \frac{\mathrm{tr}\{S^\top S - S - S^\top - (I_{NT}-S)^\top \tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)\}A}{NT} \\
&= \frac{1}{NT}\mathrm{tr}\{(S^\top S - S - S^\top)A\} - \frac{1}{NT}\mathrm{tr}\{\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)A(I_{NT}-S)^\top\}
\end{aligned} \tag{C.3}$$

It can be show the elements of S are $O_{\mathrm{P}}(1/NTh)$ uniformly. As matrix A is UB, we have

$$\frac{1}{NT}\mathrm{tr}(SA) = \frac{1}{NT}\mathrm{tr}(S^\top A) = \frac{1}{NT}O_{\mathrm{P}}\left(\frac{NT}{NTh}\right) = O_{\mathrm{P}}\left(\frac{1}{NTh}\right) = o_{\mathrm{P}}(1)$$

by Lemma A.8 in Lee (2004). Similarly we can approve that $S^\top S$ is also with elements being $O_{\mathrm{P}}(1/NTh)$ uniformly so that $(NT)^{-1}\mathrm{tr}(S^\top SA) = o_{\mathrm{P}}(1)$.

Thus,

$$\frac{1}{NT}\mathrm{tr}\left\{(S^\top S - S - S^\top)A\right\} = o_{\mathrm{P}}(1). \tag{C.4}$$

As $I_{NT} - S$ is UB in probability for sufficient large N and T, the product $(I_{NT}-S)A(I_{NT}-S)^\top$ is also UB in probability for sufficient large N and T by Lemma C.6 so that

$$\begin{aligned}
&\sup_{N,T\geq 1}\|(I_{NT}-S)A(I_{NT}-S)^\top\|_2 \\
\leq &\sqrt{\sup_{N,T\geq 1}\|(I_{NT}-S\Phi)A(I_{NT}-S)^\top\|_1 \sup_{N,T\geq 1}\|(I_{NT}-S)A(I_{NT}-S)^\top\|_\infty} = O_{\mathrm{P}}(1).
\end{aligned}$$

Since $\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top$ is an idempotent matrix and $(I_{NT}-S)A(I_{NT}-S)^\top$ is symmetric, by 6.77 on page 120 in Seber (2007) it can be shown

$$\begin{aligned}
&\mathrm{tr}\{\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)A(I_{NT}-S)^\top\} \\
= &\mathrm{tr}\{(\tilde{D}^\top\tilde{D})^{-1/2}\tilde{D}^\top(I_{NT}-S)A(I_{NT}-S)^\top\tilde{D}(\tilde{D}^\top\tilde{D})^{-1/2}\} \\
\leq &(N-1)\sup_{N,T\geq 1}\|(I_{NT}-S)A(I_{NT}-S)^\top\|_2 = O_{\mathrm{P}}(N-1).
\end{aligned}$$

Thus,

$$\frac{1}{NT}\mathrm{tr}\left\{\tilde{D}(\tilde{D}^\top\tilde{D})^{-1}\tilde{D}^\top(I_{NT}-S)A(I_{NT}-S)^\top\right\} = o_{\mathrm{P}}(1). \tag{C.5}$$

Combining (C.3)-(C.5), we accomplish the proof.

□

Lemma C.6. *Let $A_n = (a_{ij})$ and $B_n = (b_{ij})$ be $n \times n$ matrices. If A_n and B_n are UB, the product A_nB_n is also UB.*

Proof. See the proof of Lemma 1 in Kelejian and Prucha (1999). $\square$

Lemma C.7. *(1) Suppose Assumptions 1, 3- 7 hold. Then, $S_{N,T}^{-1}(\rho)$, $G_N(\rho)$ and $G_{N,T}(\rho)$ and are UB matrices uniformly in $\rho \in \Delta$, especially when $\rho = \rho_0$, we have $S_{N,T}^{-1}$, G_N and $G_{N,T}$ are UB matrices. (2) Suppose Assumptions 1-7 hold, $R_{N,T}$, S and $P_{N,T}$ are UB in probability.*

Proof. (1) According to Assumptions 4 and 5, we have W and $S_N^{-1}(\rho)$ are UB in $\rho \in \Delta$. Then, it holds that $\sup_{NT\geq 1}\|S_{N,T}^{-1}(\rho)\|_1 = \sup_{NT\geq 1}\|I_T \otimes S_N^{-1}(\rho)\|_1 = \sup_{N>1}\|S_N^{-1}(\rho)\|_1 < \infty$ and $\sup_{NT\geq 1}\|S_{N,T}^{-1}(\rho)\|_\infty = \sup_{NT\geq 1}\|I_T \otimes S_N^{-1}(\rho)\|_\infty = \sup_{N\geq 1}\|S_N^{-1}(\rho)\|_\infty < \infty$ uniformly in $\rho \in \Delta$ so that $S_{N,T}^{-1}(\rho)$ is UB uniformly in $\rho \in \Delta$. Since $G_N(\rho) = WS_N^{-1}(\rho)$, we have $G_N(\rho)$ is UB matrices uniformly in $\rho \in \Delta$ by Lemma C.6. As $G_{N,T}(\rho) = I_T \otimes G_N(\rho)$, similarly to $S_{N,T}^{-1}(\rho)$ we can show $G_{N,T}(\rho)$ is UB uniformly in $\rho \in \Delta$. (2) Based on assumption 1, 3- 7, the proof can be similarly shown as Lemma A.5 in Su and Jin (2010). $\square$

Lemma C.8. *Under Assumption 2, for any $NT \times NT$ matrix $A_{N,T}$ with elements being a function of $\{X_t, t = 1, \cdots, T\}$, we have*

(1) $\mathrm{E}(\mathbf{e}^\top A_{N,T}\mathbf{e}|\mathcal{F}_V) = \sigma_0^2 \mathrm{tr}(A_{N,T})$,

(2) $\mathrm{Var}(\mathbf{e}^\top A_{N,T}\mathbf{e}|\mathcal{F}_V) = (m_4 - 3\sigma_0^4)\sum_{i=1}^{NT}\left(a_{ii}^2\right) + \sigma_0^4 \mathrm{tr}(A_{N,T}A_{N,T}^\top + A_{N,T}^2)$,

where a_{ii} is the i-th diagonal element of A_{NT}.

Proof. (1) Assumption 2 implies that $\mathrm{E}(\mathbf{e}_t\mathbf{e}_t^\top|\mathcal{F}_V) = \mathrm{E}(\mathrm{E}(\mathbf{e}_t\mathbf{e}_t^\top|\mathcal{E}_{t-1})|\mathcal{F}_V) = \mathrm{E}(\sigma_0^2 I_N|\mathcal{F}_V) = \sigma_0^2 I_N$ and $\mathrm{E}(\mathbf{e}_t\mathbf{e}_s^\top|\mathcal{F}_V) = \mathrm{E}(\mathrm{E}(\mathbf{e}_t\mathbf{e}_s^\top|\mathcal{E}_{\min\{t,s\}-1})|\mathcal{F}_V) = \mathrm{E}(0_{N\times N}|\mathcal{F}_V) = 0_{N\times N}$. Hence, $\mathrm{E}(\mathbf{e}\mathbf{e}^\top|\mathcal{F}_V) = \sigma_0^2 I_{NT}$ and

$$\begin{aligned}\mathrm{E}(\mathbf{e}^\top A_{N,T}\mathbf{e}|\mathcal{F}_V) &= \mathrm{tr}(\mathrm{E}(A_{N,T}\mathbf{e}\mathbf{e}^\top|\mathcal{F}_V)) \\ &= \mathrm{tr}(A_{N,T}\mathrm{E}(\mathbf{e}\mathbf{e}^\top|\mathcal{F}_V)) \\ &= \sigma_0^2\mathrm{tr}(A_{N,T}).\end{aligned}$$

(2) The proof is similar to Lemma A.11 in Lee (2004). The only difference here is that $A_{N,T}$ is stochastic. Given $\mathcal{F}_V$, we can treat $A_{N,T}$ non-stochastic and adopt the proof of Lemma A.11 in Lee (2004) to show

$$\mathrm{Var}(\mathbf{e}^\top A_{N,T}\mathbf{e}|\mathcal{F}_V) = (m_4 - 3\sigma_0^4)\sum_{i=1}^{NT}\left(a_{ii}^2\right) + \sigma_0^4\mathrm{tr}(A_{N,T}A_{N,T}^\top + A_{N,T}^2).$$

$\square$

Lemma C.9. *Under Assumption 8, $\Sigma_{\boldsymbol{\theta}_0}$ is positive definite.*

Proof. It suffices to show

$$\frac{1}{2\sigma_0^4}\Psi_{R,R} + \frac{c_1}{2\sigma_0^4} - \frac{c_2^2}{\sigma_0^4} > 0.$$

As it can be show $c_1 - 2c_2^2 = \lim_{N,T\to\infty}\mathrm{tr}(C_{N,T}^s C_{N,T}^{s\top})/(NT) \geq 0$ where $C_{N,T}^s = G_{N,T} - (NT)^{-1}\mathrm{tr}(G_{N,T})I_{NT}$, we only need to show that

$$\frac{1}{2\sigma_0^4}\Psi_{R,R} > 0,$$

which is satisfied by Assumption 8. Thus, we conclude the proof.

$\square$

Lemma C.10. *Under Assumptions 1-8, the sample estimates* $\widehat{\Sigma}_{\boldsymbol{\theta}_0}$, $\widehat{\Omega}_{\boldsymbol{\theta}_0}$ *and* $\widehat{\Sigma}_X$ *mentioned in the end of Section 4 are consistent estimators of* $\Sigma_{\boldsymbol{\theta}_0}$, $\Omega_{\boldsymbol{\theta}_0}$ *and* Σ_X, *respectively.*

Proof. By definition, we know $\Omega_{\boldsymbol{\theta}_0} = \lim_{N,T\to\infty} \widehat{\Omega}_{\boldsymbol{\theta}_0}$ so that the consistency holds.

According to the consistency of $\widehat{\boldsymbol{\theta}}$ in Theorem 1, we have $\widehat{\sigma}^2 \xrightarrow{\mathrm{P}} \sigma_0^2$. As G_{NT} is UB, the limits of $\widehat{c}_1$ and $\widehat{c}_2$ exist which are defined by c_1 and c_2. Moreover, in Lemma B.2 we have shown $\widehat{\Psi}_{R,R} = (NT)^{-1}\mathrm{E}(R_{N,T}^\top P_{N,T} R_{N,T}) \xrightarrow{\mathrm{P}} \Psi_{R,R}$. By Slutsky theorem, we get $\widehat{\Sigma}_{\boldsymbol{\theta}_0} = \begin{pmatrix} \frac{1}{\widehat{\sigma}^2}\widehat{\Psi}_{R,R} + \widehat{c}_1 & \frac{\widehat{c}_2}{\widehat{\sigma}^2} \\ \frac{\widehat{c}_2}{\widehat{\sigma}^2} & \frac{1}{2\widehat{\sigma}^4} \end{pmatrix} \xrightarrow{\mathrm{P}} \Sigma_{\boldsymbol{\theta}_0}$.

Since $X_{it} = g(\tau_t) + \mathbf{v}_{it}$, we have $\widehat{g}(\tau) = \frac{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right) X_{it}}{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right)} = \frac{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right) g(\tau_t)}{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right)} + \frac{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right) \mathbf{v}_{it}}{\sum_{i,t} K\left(\frac{\tau_t-\tau}{h}\right)} = g(\tau) + o_{\mathrm{P}}(1)$ due to the weak consistency of the nonparametric estimator. Then, $\widehat{\Sigma}_{\mathbf{v}} = (NT)^{-1}\sum_{i,t} \widehat{\mathbf{v}}_{it}\widehat{\mathbf{v}}_{it}^\top = (NT)^{-1}\sum_{i,t}(\mathbf{v}_{it} + g(\tau_t) - \widehat{g}(\tau_t))(\mathbf{v}_{it} + g(\tau_t) - \widehat{g}(\tau_t))^\top = (NT)^{-1}\sum_{i,t} \mathbf{v}_{it}\mathbf{v}_{it}^\top + o_{\mathrm{P}}(1) \xrightarrow{\mathrm{P}} \Sigma_{\mathbf{v}}$. Therefore, $\widehat{\Sigma}_X = \widehat{g}(\tau_t)\widehat{g}(\tau_t)^\top + \widehat{\Sigma}_{\mathbf{v}} \xrightarrow{\mathrm{P}} \Sigma_X$.

$\square$

Appendix D: Additional Tables

Table D.8: Means and standard deviations of bias of $\widehat{\rho}$ ($\rho_0 = 0.7, \sigma_0^2 = 1$).

(a) Our model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0816	-0.0472	-0.0177	-0.0217	-0.0129	-0.0043	-0.0175	-0.0112	-0.0037
		(0.0836)	(0.0538)	(0.0319)	(0.0354)	(0.0249)	(0.0146)	(0.0330)	(0.0240)	(0.0142)
	T=15	-0.0471	-0.0292	-0.0120	-0.0115	-0.0070	-0.0029	-0.0096	-0.0056	-0.0024
		(0.0583)	(0.0415)	(0.0257)	(0.0264)	(0.0190)	(0.0123)	(0.0255)	(0.0184)	(0.0121)
	T=30	-0.0220	-0.0141	-0.0061	-0.0050	-0.0028	-0.0010	-0.0043	-0.0023	-0.0009
		(0.0342)	(0.0265)	(0.0173)	(0.0151)	(0.0121)	(0.0081)	(0.0147)	(0.0120)	(0.0080)
(I-2)	T=10	-0.0816	-0.0472	-0.0177	-0.0141	-0.0096	-0.0029	-0.0110	-0.0076	-0.0022
		(0.0836)	(0.0538)	(0.0319)	(0.0319)	(0.0237)	(0.0143)	(0.0302)	(0.0232)	(0.0141)
	T=15	-0.0471	-0.0292	-0.0120	-0.0068	-0.0043	-0.0014	-0.0058	-0.0031	-0.0009
		(0.0583)	(0.0415)	(0.0257)	(0.0251)	(0.0187)	(0.0123)	(0.0246)	(0.0184)	(0.0123)
	T=30	-0.0220	-0.0141	-0.0061	-0.0031	-0.0013	-0.0001	-0.0023	-0.0007	0.0002
		(0.0342)	(0.0265)	(0.0173)	(0.0148)	(0.0120)	(0.0081)	(0.0147)	(0.0119)	(0.0081)
(I-3)	T=10	-0.0871	-0.0502	-0.0197	-0.0203	-0.0141	-0.0071	-0.0161	-0.0126	-0.0075
		(0.0843)	(0.0538)	(0.0324)	(0.0302)	(0.0222)	(0.0129)	(0.0260)	(0.0193)	(0.0125)
	T=15	-0.0485	-0.0302	-0.0127	-0.0123	-0.0099	-0.0081	-0.0108	-0.0092	-0.0072
		(0.0572)	(0.0410)	(0.0255)	(0.0212)	(0.0158)	(0.0107)	(0.0197)	(0.0153)	(0.0108)
	T=30	-0.0217	-0.0143	-0.0064	-0.0089	-0.0085	-0.0083	-0.0082	-0.0075	-0.0061
		(0.0344)	(0.0263)	(0.0171)	(0.0124)	(0.0104)	(0.0077)	(0.0126)	(0.0105)	(0.0079)

(b) Lee–Yu model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0121	-0.0104	-0.0051	0.0294	0.0327	0.0380	0.0905	0.0915	0.0969
		(0.0538)	(0.0410)	(0.0286)	(0.0458)	(0.0394)	(0.0267)	(0.0460)	(0.0388)	(0.0269)
	T=15	-0.0077	-0.0051	-0.0032	0.0409	0.0428	0.0463	0.1040	0.1056	0.1080
		(0.0453)	(0.0357)	(0.0241)	(0.0408)	(0.0331)	(0.0233)	(0.0414)	(0.0320)	(0.0227)
	T=30	-0.0054	-0.0045	-0.0031	0.0469	0.0493	0.0493	0.1096	0.1104	0.1110
		(0.0305)	(0.0253)	(0.0169)	(0.0292)	(0.0228)	(0.0176)	(0.0275)	(0.0222)	(0.0157)
(I-2)	T=10	-0.0121	-0.0104	-0.0051	0.1256	0.1248	0.1272	0.1907	0.1897	0.1896
		(0.0538)	(0.0410)	(0.0286)	(0.0437)	(0.0356)	(0.0248)	(0.0245)	(0.0203)	(0.0139)
	T=15	-0.0077	-0.0051	-0.0032	0.1396	0.1394	0.1398	0.2028	0.2020	0.2011
		(0.0453)	(0.0357)	(0.0241)	(0.0372)	(0.0285)	(0.0207)	(0.0197)	(0.0163)	(0.0125)
	T=30	-0.0054	-0.0045	-0.0031	0.1448	0.1437	0.1428	0.2059	0.2050	0.2041
		(0.0305)	(0.0253)	(0.0169)	(0.0242)	(0.0197)	(0.0139)	(0.0142)	(0.0116)	(0.0088)
(I-3)	T=10	-0.0093	-0.0071	-0.0038	0.0781	0.0793	0.0852	0.1463	0.1470	0.1508
		(0.0440)	(0.0347)	(0.0248)	(0.0371)	(0.0325)	(0.0220)	(0.0311)	(0.0267)	(0.0184)
	T=15	-0.0054	-0.0053	-0.0040	0.0879	0.0884	0.0921	0.1571	0.1576	0.1595
		(0.0388)	(0.0309)	(0.0203)	(0.0319)	(0.0265)	(0.0192)	(0.0274)	(0.0225)	(0.0160)
	T=30	-0.0036	-0.0044	-0.0037	0.0905	0.0928	0.0947	0.1591	0.1608	0.1625
		(0.0262)	(0.0210)	(0.0150)	(0.0235)	(0.0190)	(0.0139)	(0.0207)	(0.0159)	(0.0114)

Table D.9: Means and standard deviations of bias of $\widehat{\sigma}^2$ ($\rho_0 = 0.7, \sigma_0^2 = 1$).

(a) Our model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.1720	-0.1157	-0.0579	-0.1825	-0.1243	-0.0585	-0.1724	-0.1174	-0.0539
		(0.1355)	(0.1090)	(0.0806)	(0.1342)	(0.1070)	(0.0798)	(0.1331)	(0.1073)	(0.0800)
	T=15	-0.1409	-0.0964	-0.0478	-0.1499	-0.1019	-0.0470	-0.1431	-0.0955	-0.0443
		(0.1072)	(0.0944)	(0.0676)	(0.1060)	(0.0933)	(0.0660)	(0.1055)	(0.0928)	(0.0661)
	T=30	-0.1239	-0.0813	-0.0412	-0.1284	-0.0824	-0.0381	-0.1240	-0.0795	-0.0374
		(0.0783)	(0.0659)	(0.0466)	(0.0760)	(0.0651)	(0.0458)	(0.0756)	(0.0649)	(0.0459)
(I-2)	T=10	-0.1720	-0.1157	-0.0579	-0.1650	-0.1127	-0.0512	-0.1581	-0.1073	-0.0486
		(0.1355)	(0.1090)	(0.0806)	(0.1329)	(0.1072)	(0.0800)	(0.1320)	(0.1073)	(0.0804)
	T=15	-0.1409	-0.0964	-0.0478	-0.1368	-0.0914	-0.0410	-0.1330	-0.0876	-0.0390
		(0.1072)	(0.0944)	(0.0676)	(0.1058)	(0.0929)	(0.0668)	(0.1054)	(0.0933)	(0.0667)
	T=30	-0.1239	-0.0813	-0.0412	-0.1203	-0.0764	-0.0348	-0.1162	-0.0734	-0.0332
		(0.0783)	(0.0659)	(0.0466)	(0.0761)	(0.0654)	(0.0462)	(0.0765)	(0.0659)	(0.0466)
(I-3)	T=10	-0.1727	-0.1166	-0.0592	-0.1637	-0.1087	-0.0442	-0.1482	-0.0972	-0.0375
		(0.1342)	(0.1078)	(0.0809)	(0.1350)	(0.1083)	(0.0814)	(0.1339)	(0.1088)	(0.0819)
	T=15	-0.1396	-0.0955	-0.0475	-0.1233	-0.0767	-0.0253	-0.1165	-0.0734	-0.0272
		(0.1072)	(0.0945)	(0.0670)	(0.1070)	(0.0945)	(0.0663)	(0.1066)	(0.0948)	(0.0667)
	T=30	-0.1227	-0.0802	-0.0406	-0.1013	-0.0563	-0.0138	-0.1019	-0.0597	-0.0225
		(0.0784)	(0.0660)	(0.0465)	(0.0762)	(0.0669)	(0.0474)	(0.0779)	(0.0670)	(0.0480)

(b) Lee–Yu model

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	-0.0151	-0.0119	-0.0046	1.2822	1.2926	1.2834	1.5749	1.5826	1.5716
		(0.1508)	(0.1183)	(0.0841)	(0.4191)	(0.3452)	(0.2363)	(0.5005)	(0.4110)	(0.2766)
	T=15	-0.0041	-0.0046	-0.0011	1.2952	1.2797	1.2758	1.5676	1.5468	1.5316
		(0.1207)	(0.1023)	(0.0699)	(0.3493)	(0.2704)	(0.1964)	(0.4263)	(0.3320)	(0.2302)
	T=30	-0.0059	-0.0016	-0.0006	1.2912	1.2935	1.2762	1.5579	1.5527	1.5311
		(0.0873)	(0.0711)	(0.0484)	(0.2468)	(0.1859)	(0.1355)	(0.2888)	(0.2240)	(0.1620)
(I-2)	T=10	-0.0151	-0.0119	-0.0046	1.7496	1.7505	1.7245	2.0543	2.0588	2.0245
		(0.1508)	(0.1182)	(0.0841)	(0.5376)	(0.4394)	(0.2930)	(0.6373)	(0.5199)	(0.3455)
	T=15	-0.0041	-0.0046	-0.0011	1.7152	1.6858	1.6570	1.9542	1.9222	1.8878
		(0.1206)	(0.1023)	(0.0698)	(0.4597)	(0.3565)	(0.2447)	(0.5364)	(0.4166)	(0.2855)
	T=30	-0.0059	-0.0016	-0.0006	1.7036	1.6869	1.6531	1.9281	1.9077	1.8700
		(0.0873)	(0.0711)	(0.0484)	(0.3099)	(0.2395)	(0.1721)	(0.3569)	(0.2753)	(0.1976)
(I-3)	T=10	-0.0158	-0.0137	-0.0054	2.0293	1.9866	1.9310	2.7299	2.6319	2.4962
		(0.1503)	(0.1181)	(0.0842)	(0.5568)	(0.4361)	(0.2963)	(0.6438)	(0.5133)	(0.3403)
	T=15	-0.0051	-0.0040	-0.0003	2.0278	1.9609	1.9046	2.6653	2.5423	2.4153
		(0.1192)	(0.1017)	(0.0695)	(0.4411)	(0.3423)	(0.2460)	(0.5192)	(0.4065)	(0.2779)
	T=30	-0.0071	-0.0012	-0.0001	2.0255	1.9701	1.9029	2.6568	2.5304	2.3974
		(0.0861)	(0.0704)	(0.0484)	(0.3190)	(0.2387)	(0.1684)	(0.3728)	(0.2698)	(0.1942)

Table D.10: Means and standard deviations of MSE of $\widehat{\boldsymbol{\beta}}(\tau) = (\widehat{\beta}_1(\tau), \widehat{\beta}_2(\tau))^\top$ ($\rho_0 = 0.7, \sigma_0^2 = 1$).

(a) $\widehat{\beta}_1(\tau)$

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	0.2435	0.1048	0.0315	0.0915	0.0473	0.0169	0.1709	0.0897	0.0294
		(0.3177)	(0.1290)	(0.0359)	(0.1036)	(0.0474)	(0.0147)	(0.2526)	(0.1203)	(0.0348)
	T=15	0.1130	0.0552	0.0201	0.0481	0.0263	0.0109	0.0932	0.0476	0.0206
		(0.1509)	(0.0621)	(0.0202)	(0.0515)	(0.0226)	(0.0098)	(0.1171)	(0.0600)	(0.0243)
	T=30	0.0361	0.0205	0.0085	0.0185	0.0108	0.0044	0.0333	0.0202	0.0088
		(0.0403)	(0.0210)	(0.0087)	(0.0147)	(0.0088)	(0.0039)	(0.0370)	(0.0235)	(0.0101)
(I-2)	T=10	0.6854	0.2757	0.0764	0.2556	0.1386	0.0462	0.3989	0.2245	0.0749
		(0.9448)	(0.3787)	(0.0986)	(0.3509)	(0.1926)	(0.0524)	(0.5803)	(0.3484)	(0.0968)
	T=15	0.2968	0.1397	0.0481	0.1379	0.0744	0.0328	0.2374	0.1222	0.0555
		(0.4199)	(0.1735)	(0.0549)	(0.1651)	(0.0926)	(0.0354)	(0.3274)	(0.1730)	(0.0690)
	T=30	0.0911	0.0501	0.0199	0.0530	0.0309	0.0133	0.0867	0.0521	0.0229
		(0.1117)	(0.0564)	(0.0221)	(0.0559)	(0.0331)	(0.0133)	(0.1086)	(0.0678)	(0.0265)
(I-3)	T=10	0.8859	0.3395	0.0973	0.3023	0.1511	0.0438	0.3506	0.1887	0.0681
		(1.2591)	(0.4863)	(0.1435)	(0.5140)	(0.2330)	(0.0542)	(0.6831)	(0.2734)	(0.0942)
	T=15	0.3524	0.1611	0.0552	0.1228	0.0641	0.0282	0.1696	0.0970	0.0495
		(0.5253)	(0.2135)	(0.0685)	(0.1931)	(0.1009)	(0.0339)	(0.2480)	(0.1292)	(0.0645)
	T=30	0.1036	0.0575	0.0214	0.0401	0.0259	0.0138	0.0690	0.0473	0.0251
		(0.1456)	(0.0746)	(0.0250)	(0.0520)	(0.0317)	(0.0179)	(0.0972)	(0.0635)	(0.0340)

(b) $\widehat{\beta}_2(\tau)$

		(II-1)			(II-2)			(II-3)		
		N=10	N=15	N=30	N=10	N=15	N=30	N=10	N=15	N=30
(I-1)	T=10	0.0682	0.0381	0.0166	0.0697	0.0398	0.0179	0.0612	0.0370	0.0175
		(0.0575)	(0.0315)	(0.0131)	(0.0601)	(0.0311)	(0.0131)	(0.0497)	(0.0292)	(0.0129)
	T=15	0.0445	0.0256	0.0104	0.0459	0.0274	0.0124	0.0419	0.0257	0.0126
		(0.0365)	(0.0207)	(0.0077)	(0.0366)	(0.0207)	(0.0091)	(0.0344)	(0.0199)	(0.0089)
	T=30	0.0191	0.0113	0.0052	0.0205	0.0128	0.0081	0.0196	0.0130	0.0084
		(0.0149)	(0.0084)	(0.0038)	(0.0151)	(0.0092)	(0.0054)	(0.0146)	(0.0094)	(0.0056)
(I-2)	T=10	0.0682	0.0381	0.0166	0.0578	0.0357	0.0173	0.0525	0.0344	0.0175
		(0.0575)	(0.0315)	(0.0131)	(0.0475)	(0.0294)	(0.0126)	(0.0418)	(0.0279)	(0.0123)
	T=15	0.0445	0.0256	0.0104	0.0399	0.0250	0.0126	0.0380	0.0246	0.0129
		(0.0365)	(0.0207)	(0.0077)	(0.0324)	(0.0197)	(0.0088)	(0.0303)	(0.0194)	(0.0087)
	T=30	0.0191	0.0113	0.0052	0.0192	0.0129	0.0081	0.0186	0.0132	0.0087
		(0.0149)	(0.0084)	(0.0038)	(0.0144)	(0.0092)	(0.0055)	(0.0140)	(0.0091)	(0.0056)
(I-3)	T=10	0.0658	0.0375	0.0166	0.0556	0.0336	0.0159	0.0469	0.0319	0.0173
		(0.0535)	(0.0311)	(0.0123)	(0.0511)	(0.0277)	(0.0117)	(0.0382)	(0.0254)	(0.0115)
	T=15	0.0408	0.0240	0.0099	0.0355	0.0232	0.0142	0.0336	0.0234	0.0137
		(0.0335)	(0.0189)	(0.0074)	(0.0290)	(0.0171)	(0.0090)	(0.0267)	(0.0165)	(0.0089)
	T=30	0.0173	0.0103	0.0047	0.0195	0.0167	0.0144	0.0191	0.0153	0.0105
		(0.0143)	(0.0081)	(0.0036)	(0.0130)	(0.0092)	(0.0057)	(0.0128)	(0.0092)	(0.0065)

Table D.11: Variance inflation factor (VIF) of the detrended regressors $X_{N,T}$ and $R_{N,T}$.

	$R_{N,T}$	Intercept	$log(FDI)$	$log(Asset)$	GDP_m	GDP_s	Emp_{ms}
VIF	1.8421	1.8410	1.0503	1.1277	2.4759	2.5219	1.1055