DSSR

Discussion Paper No. 104

Estimation of Partially Linear Spatial Autoregressive Models with Autoregressive Disturbances

Takaki Sato

October, 2019

Data Science and Service Research Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Estimation of Partially Linear Spatial Autoregressive Models with Autoregressive Disturbances

Takaki Sato *

Abstract

This study considers semiparametric partially linear spatial autoregressive models with autoregressive disturbances that contain an unspecified nonparametric component and allow for spatial lags in both the dependent variables and disturbances. Having the nonparametric function approximated by basis functions, we propose a three-step estimation procedure for the proposed model. We also establish the consistency and asymptotic normality of the proposed estimators. Then, the finite sample performances of the proposed estimators are examined using Monte Carlo simulations. As an empirical application, we use the proposed model and estimation method to analyze Boston housing price data to evaluate the effect of air pollution on the value of owner-occupied homes.

Keywords: Partially linear models, Series estimation, Spatial econometrics, Instrumental variables.

1 Introduction

Recently, the spatial autoregressive (SAR) model proposed by Clif and Ord (1973) has received increasing attention in both theoretical and applied econometrics research. Specifically, the data in the field of regional, urban, and environmental economics usually show the spatial dependency of cross-sectional units and SAR models are used to capture this dependency. The class of SAR models is extended by considering spatial interaction effects in both the dependent variables and disturbances. We call these models SAR models with spatial autoregressive disturbances (SARAR).

Anselin (1988) and Lee (2004) propose the (quasi) maximum likelihood (ML) to estimate such parametric spatial econometric models. However, one drawback of ML estimation is the computational load when the sample size is large, because there is no closed-form expression of ML estimators; therefore, it is necessary to calculate the determinant of a large matrix, whose size depends on the sample size. Another approach

*Advanced Institute for Yotta Informatics, Tohoku University, Sendai 980-8577, Japan. satotaka@riec.tohoku.ac.jp

for the estimation of spatial econometric models consists of moment-based estimations. Kelejian and Prucha (1998, 2010) introduce generalized spatial two-stage least squares (2SLS) estimation methods, while Lee and Liu (2010) consider the generalized method of moments (GMM) estimation methods.

To avoid mis-specification of the data generating process in parametric models, the semiparametric extensions of spatial econometric models have received significant attention owing to the simple interpretation of parametric terms and the flexibility of nonparametric terms. A popular semiparametric regression model is the partially linear one, which contains explanatory variables nonlinearly rerated with dependent variables. As semiparametric extensions of the SAR models, Su and Jin (2010) and Du et al. (2018) propose partially linear SAR (PL-SAR), while Su (2012) considers partially linear SARAR (PL-SARAR) models. Zhang and Sun (2015) further study the spatial dynamic panel extension of PL-SAR models. Another semiparametric extension is the varying coefficient model, in which the impact of some explanatory variables depends on spatial units. Zhang and Shen (2015) consider semiparametric varying coefficient-specified spatial panel models and Hoshino (2018) proposes functional coefficient SAR models with endogenous regressors.

A popular method for estimating nonparametric terms in regression models is the kernel approach. Su (2012) applies kernel methods and proposes the estimation method for the PL-SARAR models in which the nonparametric terms are profiled out. However, as the sample size increases, the computational load of these estimation methods increases significantly, making them less manageable. Another estimation method for nonparametric terms is series estimation. One advantage of series methods is their computational simplicity. As such, we apply moment-based estimation methods for the estimation of nonparametric terms by approximating the nonparametric terms using basis functions such as polynomials and splines.

We consider the moment-based estiamtion method for PL-SARAR models for computational simplicity. Accordingly, we propose a three-step estimation procedure by applying the 2SLS and nonlinear least squares (NLS) methods for the parametric terms and series methods for the nonparametric term in the proposed model. The consistency and asymptotic normality of the proposed estimators are established and the small sample properties of the proposed estimators are then evaluated.

As an empirical analysis, we apply the SARAR and PL-SARAR models to Boston land price data to evaluate the causal effect of air pollution on housing prices. In the model, the dependent variable is the median value of owner-occupied homes and the explanatory variable is the nitrogen oxide (NOX) concentration. Our empirical findings are as follows. First, housing prices show spatial correlations even after we control for the potential determinants of housing prices. Second, air pollution has strong negative effects on housing prices in both the parametric and semiparametric models. Finally, the effect of air pollution of housing prices is not linear and the negative effect increases significantly when the proportion of NOX in the air is above a threshold value.

The rest of paper proceeds as follows. We introduce PL-SARAR models and propose a three-step estimation method in section 2. The asymptotic properties of the proposed estimators are established in section 3. Section 4 examines the small sample properties of the proposed estimators using Monte Carlo simulations. In section 5, we apply the proposed models to Boston land price data to investigate the empirical properties of the proposed model. Section 6 presents the concluding remarks. The proofs of Lemmas and Theorems are provided in the Appendix.

Notation: We use I_n to denote an $n \times n$ identity matrix. For matrix A_n, $||A_n||$ denotes its Frobenius norm: $||A_n|| = \{tr(A_n' A_n)\}^{1/2}$, where $tr(\cdot)$ is the trace operator. When A_n is a symmetric matrix, $\gamma_{\max}(A_n)$ and $\gamma_{\min}(A_n)$ denote the largest and smallest eigenvalues of A_n, respectively.

2 Model Specification and Estimation

Let us consider the following PL-SARAR models:

$$
\begin{aligned}
y_{n,i} &= \lambda_0 \sum_{j=1}^{n} w_{n,i,j} y_{n,j} + x_{n,i}^{'} \beta_0 + g_0(s_{n,i}) + u_{n,i}, \qquad (1)\\
u_{n,i} &= \rho_0 \sum_{j=1}^{n} m_{n,i,j} u_{n,j} + \varepsilon_{n,i},
\end{aligned}
$$

where n is the number of spatial units, $y_{n,i}$ is an observed dependent variable, $x_{n,i} = (x_{n,i}^{(1)}, \ldots, x_{n,i}^{(d_x)})'$ is a $d_x \times 1$ vector of exogenous regressors, $s_{n,i}$ is a nonparametric regressor, $g_0(\cdot)$ is an unknown function, $\varepsilon_{n,i}$ is an independently and identically distributed (i.i.d.) disturbance with mean zero and variance σ_0^2, and $w_{n,i,j}$ and $m_{n,i,j}$ are the (i,j)th elements of predetermined $n \times n$ spatial weight matrices W_n and M_n, respectively. Scalar parameters λ_0 and ρ_0 are SAR parameters and β_0 is a coefficient vector.

We apply the series approximation method to estimate the nonparametric term. Let $\{p_k(\cdot) : k = 1, 2, \ldots\}$ be a sequence of basis functions such as polynomials, splines, and Fourier series. We assume that nonparametric function $g_0(s_{n,i})$ can be approximated by $P^K(s_{n,i})'\alpha_0$, where $P^K(\cdot) = (p_1(\cdot), \ldots, p_K(\cdot))'$, K is the number of basis functions, and α_0 is a $K \times 1$ vector of parameters. Therefore, the series approximation error of the nonparametric function is given by:

$$
v_{n,i} = g_0(s_{n,i}) - P^K(s_{n,i})\alpha_0,
$$

and model (1) is expressed as

$$
\begin{aligned}
y_{n,i} &= \lambda_0 \sum_{j=1}^{n} w_{n,i,j} y_{n,j} + x_{n,i}^{'} \beta_0 + P^K(s_{n,i})\alpha_0 + v_{n,i} + u_{n,i}, \qquad (2)\\
u_{n,i} &= \rho_0 \sum_{j=1}^{n} m_{n,i,j} u_{n,j} + \varepsilon_{n,i}.
\end{aligned}
$$

For notational simplicity, we consider the following matrix notation of the proposed model. Let $Y_n = (y_{n,1}, \ldots, y_{n,n})'$, $X_n = (x_{n,1}, \ldots, x_{n,n})'$, $B_n = (W_n Y_n, X_n)$, $\delta_0 = (\rho_0, \beta_0)'$, $P_n = (P^K(s_{n,1}), \ldots, P^K(s_{n,n}))'$, $V_n = (v_{n,1}, \ldots, v_{n,n})'$, and $\varepsilon_n = (\varepsilon_{n,1}, \ldots, \varepsilon_{n,n})'$. When $I_n - \rho_0 M_n$ are nonsingular, model (2) is rewritten as:

$$
Y_n = B_n \delta_0 + P_n \alpha_0 + V_n + (I_n - \rho_0 M_n)^{-1} \varepsilon_n. \qquad (3)
$$

For the estimation of the parameters in model (3), we propose a three-step estimation procedure. In the first step, we apply 2SLS to model (3) to estimate δ_0 because the spatial lagged dependent variable, $W_n Y_n$, is correlated with the error term, $(I_n - \rho_0 M_n)^{-1}\varepsilon_n$. In the second step, we estimate the coefficient of the basis function, α_0, and the unknown function, $g_0(\cdot)$, by ordinary least squares (OLS). In the third step, the spatial autoregressive parameter and variance of disturbances, ρ_0 and σ_0^2, respectively, are estimated by applying NLS to the residuals obtained in the first and second steps.

The first step is the estimation of parameter δ_0 by 2SLS because the correlation of the spatial lagged dependent variable and the error term leads to the inconsistency of the OLS estimator (see, e.g., Kelejian and Prucha (1998)). Let Z_n be an $n \times d_z$ matrix of instrumental variables. For example, we may use matrices $(X_n, W_n X_n, W_n W_n X_n)$ as instrumental variables.

Following Zhang and Sun (2015) and Du et al. (2018), we partial out the series approximation. Let $\Pi_n = P_n (P_n^{'} P_n)^{-1} P_n^{'}$ denote the projection matrix onto the space spanned by P_n. Then, we obtain:

$$
(I_n - \Pi_n) Y_n = (I_n - \Pi_n) B_n \delta_0 + (I_n - \Pi_n) V_n + (I_n - \Pi_n)(I_n - \rho M_n)^{-1} \varepsilon_n. \qquad (4)
$$

Applying 2SLS to model (4) with instrument variables Z_n, we propose the following 2SLS estimator for parameter δ_0:

$$
\hat{\delta} = (B_n^{'}(I_n - \Pi_n) H_n (I_n - \Pi_n) B_n^{'})^{-1} B_n^{'} (I_n - \Pi_n) H_n (I_n - \Pi_n) Y_n^{'},
$$

where $H_n = Z_n (Z_n^{'} Z_n)^{-1} Z_n^{'}$.

In the second step, we consider the estimation of the coefficient on the series approximation, α_0, by applying the OLS method and derive the estimator of the unknown function, $g_0(\cdot)$. Using OLS, we obtain the following estimator for α and $g_0(s_{n,i})$:

$$\begin{aligned}\hat{\alpha} &= (P_nP_n)^{-1}P_n(Y_n - B_n\hat{\delta}),\\ \hat{g}(s_{n,i}) &= P^K(s_{n,i})\hat{\alpha},\end{aligned}$$

where δ_0 is the 2SLS estimator obtained in the first step.

The third step represents the estimation of the spatial autoregressive parameter and the variance of the disturbances, ρ_0 and σ_0^2, respectively by NLS. Let $\overline{u}_n = W_nu_n, \overline{\overline{u}}_n = W_n\overline{u}_n$ and $\overline{\varepsilon}_n = W_n\varepsilon_n$. Moreover, we denote the i-th elements of $u_n, \overline{u}_n, \overline{\overline{u}}_n$, and $\overline{\varepsilon}_n$ by $u_{n,i}, \overline{u}_{n,i}, \overline{\overline{u}}_{n,i}$, and $\overline{\varepsilon}_{n,i}$, respectively.

The spatial correlation of the disturbance term indicates the following moment condition:

$$u_n - \rho\overline{u}_n = \varepsilon_n, \tag{5}$$

$$\overline{u}_n - \rho\overline{\overline{u}}_n = \overline{\varepsilon}_n. \tag{6}$$

Following Kelejian and Prucha (1999), we define the two matrices for the NLS estimation based on (5) and (6), respectively:

$$G_n = \frac{1}{n}\begin{pmatrix} 2\sum_{i=1}^n E(u_{n,i}\overline{u}_{n,i}) & -\sum_{i=1}^n E(\overline{u}_{n,i}^2) & n \\ 2\sum_{i=1}^n E(\overline{u}_{n,i}\overline{\overline{u}}_{n,i}) & -\sum_{i=1}^n E(\overline{\overline{u}}_{n,i}^2) & tr(M_n'M_n) \\ \sum_{i=1}^n E(u_{n,i}\overline{\overline{u}}_{n,i} + \overline{u}_{n,i}^2) & -\sum_{i=1}^n E(\overline{u}_{n,i}\overline{\overline{u}}_{n,i}) & 0 \end{pmatrix}, \tag{7}$$

$$g_n = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n E(u_{n,i}^2) \\ \sum_{i=1}^n E(\overline{u}_{n,i}^2) \\ \sum_{i=1}^n E(u_{n,i}\overline{u}_{n,i}) \end{pmatrix}. \tag{8}$$

We derive the objective function for the NLS estimation by replacing the disturbances in (7) and (8) with the sample moments. Let $\hat{u}_n = Y_n - B_n\hat{\delta} - P_n\hat{\alpha}$, $\hat{\overline{u}}_n = W_n\hat{u}_n$ and $\hat{\overline{\overline{u}}}_n = W_n\hat{\overline{u}}_n$. Moreover, we denote the i-th elements of $\hat{u}_n, \hat{\overline{u}}_n$, and $\hat{\overline{\overline{u}}}_n$ by $\hat{u}_{n,i}, \hat{\overline{u}}_{n,i}$, and $\hat{\overline{\overline{u}}}_{n,i}$, respectively. We also define

$$\hat{G}_n = \frac{1}{n}\begin{pmatrix} 2\sum_{i=1}^n \hat{u}_{n,i}\hat{\overline{u}}_{n,i} & -\sum_{i=1}^n \hat{\overline{u}}_{n,i}^2 & n \\ 2\sum_{i=1}^n \hat{\overline{u}}_{n,i}\hat{\overline{\overline{u}}}_{n,i} & -\sum_{i=1}^n \hat{\overline{\overline{u}}}_{n,i}^2 & tr(M_n'M_n) \\ \sum_{i=1}^n(\hat{u}_{n,i}\hat{\overline{\overline{u}}}_{n,i} + \hat{\overline{u}}_{n,i}^2) & -\sum_{i=1}^n \hat{\overline{u}}_{n,i}\hat{\overline{\overline{u}}}_{n,i} & 0 \end{pmatrix},$$

$$\hat{g}_n = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n \hat{u}_{n,i}^2 \\ \sum_{i=1}^n \hat{\bar{u}}_{n,i}^2 \\ \sum_{i-1}^n \hat{u}_{n,i}\hat{\bar{u}}_{n,i} \end{pmatrix}.$$

Let $\eta = (\rho, \rho^2, \sigma^2)'$. Then, the NLS estimators for ρ and σ^2, $\hat{\rho}$ and $\hat{\sigma}^2$, respectively, are defined as the minimizers of $(\hat{G}_n - g_n\eta)'(G_n - g_n\eta)$. Therefore, the third step estimator is defined by

$$(\hat{\rho}, \hat{\sigma}^2) = \text{argmin}\left\{(\hat{G}_n - \hat{g}_n\eta)'(\hat{G}_n - \hat{g}_n\eta)\right\}.$$

3 Asymptotic Properties

Here, we consider the asymptotic properties of the proposed estimators. We introduce the following assumptions.

Assumption 1

1. All the diagonal elements of W_n and M_n are zero.
2. Matrices $I_n - \lambda W_n$ and $I_n - \rho M_n$ are nonsingular for all $|\lambda| < 1$ and $|\rho| < 1$.
3. The row and column sums of matrices $W_n, M_n, (I_n - \lambda_0 W_n)^{-1}$ and $(I_n - \rho_0 M_n)^{-1}$ are uniformly bounded in absolute value.

Assumption 2 Disturbance $\varepsilon_{i,n}$ is i.i.d. with $E(\varepsilon_{i,n}) = 0$ and $V(\varepsilon_{i,n}) = \sigma_0^2$. Moreover, the disturbance has a finite fourth moment.

Assumption 3

1. Exogenous regressors X_n are non-stochastic and the elements of X_n are uniformly bounded in absolute value.
2. Instrumental variables Z_n are non-stochastic and the elements of Z_n are uniformly bounded in absolute value.
3. Nonparametric regressor $S_n = (s_{n,1}, \ldots, s_{n,n})'$ is non-stochastic and the set of possible values for $s_{n,i}$, S, is a compact space.

Assumption 4

1. There exist $\alpha_0 \in \mathbb{R}^K$ and $r_s > 0$ so that $\sup_{s\in S} |P^K(s)'\alpha_0 - f(s)| = O(K^{-r_s})$ for each K.

2. $\sup_{s\in S} ||P^K(s)|| = O(K^{1/2})$.

3. $\sqrt{n}K^{-r_s} \to 0$ and $\frac{K^2}{n} \to 0$ as $n \to \infty$.

4. There exist constants $\underline{c}_{P_n}$ and $\bar{c}_{P_n}$ so that $0 < \underline{c}_{P_n} < \gamma_{\min}\left(\frac{P_n' P_n}{n}\right) \leq \gamma_{\max}\left(\frac{P_n' P_n'}{n}\right) < \bar{c}_{P_n} < \infty$.

5. $g_0(s)$ is uniformly bounded in absolute value.

Assumption 5 Let $\tilde{B}_n = W_n(I_n - \lambda_0 W_n)^{-1}(B_n + g_0(S_n))$. There exist constants $\underline{c}_{\tilde{B}_n}$ and $\bar{c}_{\tilde{B}_n}$ so that $0 < \underline{c}_{\tilde{B}_n} < \gamma_{\min}\left(\frac{\tilde{B}_n \tilde{B}_n'}{n}\right) \leq \gamma_{\max}\left(\frac{\tilde{B}_n \tilde{B}_n'}{n}\right) < \bar{c}_{\tilde{B}_n} < \infty$.

We define

$$
\begin{aligned}
\Sigma_{n,1} &= \frac{1}{n}\tilde{B}_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)\tilde{B}_n, \\
\Sigma_{n,2} &= \frac{1}{n}B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)(I - \rho_0 M_n)^{-1}(I - \rho_0 M_n)'^{-1}(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n.
\end{aligned}
$$

Assumption 6 $\Sigma_1 = \lim_{n\to\infty} \Sigma_{n,1}$ and $\Sigma_2 = \lim_{n\to\infty} \Sigma_{n,2}$ exist and are bounded away from zero and infinity.

Assumption 7 There exists constant $\underline{c}_{G_n}$ so that $0 < \underline{c}_{G_n} < \gamma_{\min}(G_n' G_n)$.

Assumption 1.1 leads to the normalization of the proposed model and Assumption 1.2 to the existence condition of the model. We say that the row sums of matrix A_n are uniformly bounded in absolute value if there exists constant c_A so that

$$
\max_{1\leq i\leq n, n\geq 1} \sum_{j=1}^{n} |a_{n,i,j}| < c_A,
$$

where $a_{n,i,j}$ is the (i,j)th element of A_n. The uniform boundedness of column sums is similarly defined. Assumption 1.3 limits the spatial correlation between the elements of Y_n and ε_n. Assumption 2 provides the essential features of the disturbances. Assumption 3 is the standard set of assumptions in spatial econometrics literatures. Assumption 4.1 indicates the approximation error reduction at K^{-r_s}, assumption 4.2 imposes a restriction on the basis functions, assumption 4.3 ensures that the series approximation bias does not affect the limiting distribution of the proposed estimators, and assumptions 4.4 and 4.5 are required for the derivation of the asymptotic properties of the proposed estimators. Assumption 5 limits spatial correlation to a certain degree and is required to establish the asymptotic properties of the proposed estimator. Assumption

6 is required to derive the limiting distribution of the first-step estimator. Assumption 7 is required for the identifiability of the third-step nonlinear estimator.

First, we consider the asymptotic behaviors of the first-step estimator, $\hat{\delta}$. The limiting distribution of this estimator is centered at δ_0 and is asymptotically normal.

Theorem 1. If Assumptions 1–6 hold, then,

$$\sqrt{n}(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \sigma_0^2 \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}).$$

Second, we consider the asymptotic properties of the second-step estimators, $\hat{\alpha}$ and $\hat{g}(\cdot)$. We define:

$$\sigma^2(s) = \sigma^2(P^K(s)(P_n' P_n)^{-1} P_n' (I_n - \rho_0 M_n)^{-1} (I_n - \rho_0 M_n)'^{-1} P_n (P_n' P_n)^{-1} P'^K(s)).$$

Then, the convergence rates of $|\hat{\alpha} - \alpha_0|$ and $\sup_s |\hat{g}(s) - g_0(s)|$ are derived. Moreover, the limiting distribution of estimator $\hat{g}(\cdot)$ is centered at $g_0(\cdot)$ and asymptotically normal for a given $s \in S$.

Theorem 2. If Assumptions 1–6 hold, then,

1. $\hat{\alpha} = \alpha_0 + O_p(\sqrt{K/N} + K^{-r_s})$.

2. $\sup_s |\hat{g}(s) - g_0(s)| = O_p(K/\sqrt{n} + K^{(1-2r_s)/2})$.

3. $(\hat{g}(s) - g_0(s)) \xrightarrow{d} N(0, \sigma^2(s))$.

Finally, we show the consistency of the third-step estimator.

Theorem 3. If Assumptions 1–7 hold, then,

$$\hat{\rho} \xrightarrow{p} \rho_0,$$
$$\hat{\sigma}^2 \xrightarrow{p} \sigma_0^2.$$

4 Monte Carlo Simulation

Here, we examine the small sample performances of the proposed three-step estimators through a set of simulation experiments. We consider the following data generating process for the Monte Carlo simulations:

$$y_{n,i} = \lambda_0 \sum_{j=1}^{n} w_{n,i,j} y_{n,j} + x_{n,i}\beta_0 + g_0(s_{n,i}) + u_{n,i},$$

$$u_{n,i} \quad = \quad \rho_0 \sum_{j=1}^{n} w_{n,i,j} u_{n,j} + \varepsilon_{n,i},$$

where $x_{n,i} \sim$ i.i.d. $N(0,1)$, $s_{n,i} \sim$ i.i.d. Uniform$[0,1]$, $g_0(s_{n,i}) = \sin(3\pi s_{n,i})$ and $\varepsilon_{n,i} \sim$ i.i.d. $N(0, \sigma_0^2)$ for all $i = 1, \ldots, n$. Spatial weight matrix W_n is defined according to rook contiguity with row normalization (see, e.g., Arbia (2014)). As basis functions for the approximation of the nonparametric function, we use cubic B-splines (see, e.g., Hastie et al. (2009)). Following a simple rule-of-thumb, we set the numbers of the basis functions as $\lfloor n^{1/5} \rfloor + 2 \times 4$, where $\lfloor n^{1/5} \rfloor$ denotes the integer part of $n^{1/5}$.

We set $\beta_0 = 2$ and $\sigma_0^2 = 1$ as true values. As pairs of spatial autoregressive parameters (λ_0, ρ_0), we consider the following four cases: $(\lambda_0, \rho_0) \in \{(0.2, 0.2), (0.8, 0.8), (0.2, 0.8), (0.8, 0.2)\}$. For each parameter value, we generate a sample of size n (= 400, 900) and calculate the estimators. This step is repeated 1000 times. For the estimators of $\lambda_0, \rho_0, \beta_0$ and σ_0^2, we report the bias and root mean squared errors (RMSE). To evaluate the estimation performance of the nonparametric term, we use the average RMSE (ARMSE):

$$ARMSE = \frac{1}{1000} \sum_{l=1}^{1000} \left\{ \frac{1}{n} \sum_{i=1}^{n} [\hat{g}^l(s_{n,i}) - g_0(s_{n,i})]^2 \right\}^{1/2},$$

where $\hat{g}^l(\cdot)$ indicates the estimate from the l-th replicated dataset.

Table 1 summarizes the estimation results of $\lambda_0, \rho_0, \beta_0, \sigma_0^2$, and $g_0(\cdot)$. As the sample size of observations increases, estimations become more accurate. The results demonstrate the consistency of the proposed estimators. The ARMSE of the estimator for the nonparametric function, $\hat{g}(\cdot)$, is larger than the RMSE of the estimator for the parametric functions because the convergence rate of $\hat{g}(\cdot)$ is slower than root-N. Moreover, the bias and RMSE of the third-step estimator, $\hat{\rho}$ and σ_0^2, are larger than those of the 2SLS estimator, $\hat{\lambda}$ and $\hat{\beta}$, respectively. With regard to the magnitude of the spatial autoregressive parameters, λ_0 and ρ_0, their degree does not affect the estimation accuracy of the parametric terms. However, the bias and RMSE of the estimators for the nonparametric function tend to increase as ρ_0 increases.

5 Real Data Analysis

We apply the SARAR and PL-SARAR models to Boston housing price data collected by Harrison and Rubinfield (1978) to investigate the empirical properties of the PL-SARAR model and evaluate the effect of air pollution on house value. The data contain the median house prices in 506 Boston area census tracts, NOX concentrations per town as an index of air pollution, and other potential determinants of house values. The definitions of the variables are summarized in Table 2.

Table 1: Small sample performances of the proposed estimators by biases and root mean square errors.

		$\lambda_0 = 0.2, \rho_0 = 0.2$ $\beta_0 = 1, \sigma^2 = 1$		$\lambda_0 = 0.8, \rho_0 = 0.8$ $\beta_0 = 1, \sigma^2 = 1$		$\lambda_0 = 0.8, \rho_0 = 0.2$ $\beta_0 = 1, \sigma^2 = 1$		$\lambda_0 = 0.2, \rho_0 = 0.8$ $\beta_0 = 1, \sigma^2 = 1$	
		n = 400	n = 900	n =400	n=900	n = 400	n = 900	n=400	n=900
λ_0	Bias	-0.0001	-0.0010	-0.0128	-0.0072	-0.0008	-0.0006	0.0023	-0.0045
	RMSE	0.0511	0.0343	0.0779	0.0514	0.0309	0.0197	0.1037	0.0714
ρ_0	Bias	-0.0321	-0.0126	-0.0202	-0.0108	-0.0329	-0.0160	-0.0326	-0.0111
	RMSE	0.0893	0.0598	0.0856	0.0563	0.0899	0.0572	0.0848	0.0519
β_0	Bias	-0.0006	-0.0002	-0.0042	-0.0002	-0.0002	-0.0009	-0.0056	-0.0039
	RMSE	0.0514	0.0349	0.0512	0.0349	0.0517	0.0337	0.0708	0.0462
σ_0^2	Bias	-0.0242	-0.0124	-0.0111	-0.0023	-0.0277	-0.0102	-0.0063	-0.0047
	RMSE	0.0713	0.0478	0.0732	0.0515	0.0758	0.0489	0.0831	0.0566
$g_0(\cdot)$	ARMSE	0.1622	0.1104	0.5729	0.4171	0.1775	0.1219	0.5159	0.3716

Table 2: Variable definitions.

	Variable	Definition
Dependent variable	MEDV	Median value of owner-occupied homes.
Explanatory variables	CRIM	Per capita crime rate by town.
	RM	Average number of rooms per dwelling.
	AGE	Proportion of owner units built prior to 1940.
	TAX	Full value property tax rate per USD 10,000 per town.
	LSTAT	Proportion of lower status of the population.
	INDUS	Proportion of non-retail business acres per town.
	B	Black proportion of population.
	DIS	Weighted distances from five Boston employment centers.
	RAD	Index of accessibility to radial highways.
	PTRATIO	Pupil-teacher ratio by town school district.
	NOX	Nitrogen oxide concentration per town.

We compare the partially linear with the parametric linear models. Model 1 is defined by:

$$\begin{aligned} MEDV &= \lambda W_n MEDV + \beta_1 + \beta_2 CRIM + \beta_3 RM + \beta_4 AGE + \beta_5 TAX + \beta_6 LSTAT \\ &\quad + \beta_7 INDUS + \beta_8 B + \beta_9 DIS + \beta_1 0 RAD + \beta_{11} PTRATIO + g(NOX) + u_n, \\ u_n &= \rho W_n u_n + \varepsilon_n, \end{aligned}$$

where $g(\cdot)$ is an unknown function of NOX. We set the number of basis functions as $3 + 2 \times 4$ following a simple rule-of-thumb. In model 2, we assume explanatory variable NOX is linearly correlated with the dependent variable. Therefore, we replace $g(NOX)$ in model 1 with $\beta_{12} NOX$ in model 2. According to Pace and Gilley (1997) and Du et al. (2018), we define the (i, j)th element of the spatial weight matrix by:

$$w_{n,i,j} = \max\left(1 - \frac{d_{i,j}}{d_0}, 0\right),$$

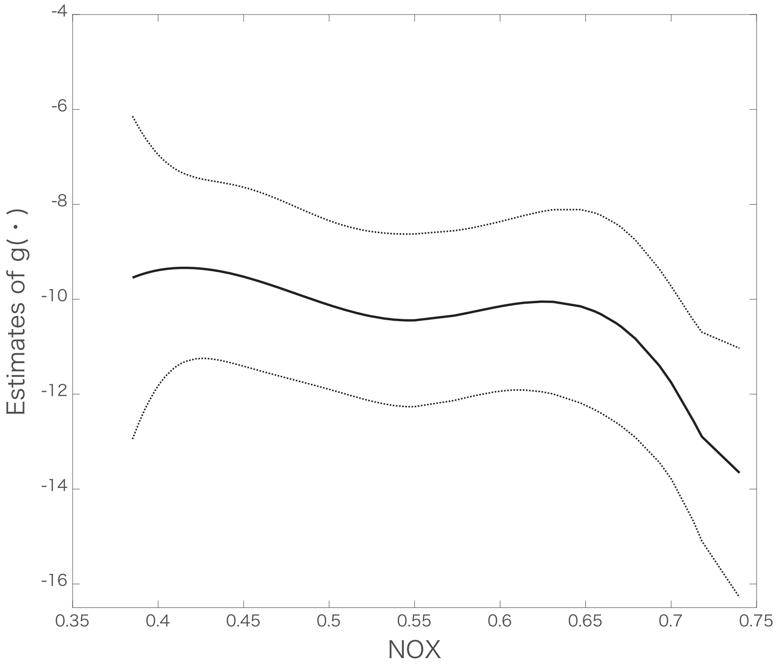

Figure 1: Estimates of nonparametric function $g(NOX)$ in model 1 and its 95% confidence interval.

where $d_{i,j}$ is the Euclidean distance calculated by the longitude and latitude coordinates of the two observations and d_0 is the threshold distance, chosen as 0.025 in this analysis. Furthermore, we normalized the weight matrix so that the sums of rows are equal to one. The parameters in model 1 are estimated by the proposed three-step estimation method and the ones in model 2 are estimated by 2SLS (see, e.g., Kelejian and Prucha (1998)).

Table 3 shows the estimation results of the regression coefficient, spatial autoregressive parameters, and variances in innovation. The estimation results of models 1 and 2 are similar and the sign and statistical significance of the regression coefficients are consistent with previous empirical research on Boston house pricing data (see, e.g. Pace and Gilley (1997) and Arbia (2014)). Figure 1 shows the estimation results of the nonparametric function in model 1. The solid line corresponds to the estimates of $g(\cdot)$ and the dotted ones to the 95% confidence interval. Our empirical findings are as follows. First, a spatial correlation between the dependent variables and disturbances exists even after we control for some of the potential determinants of housing prices. This indicates that house values in surrounding areas have a positive effect on housing prices and there may exist unobserved shocks following a spatial pattern. Second, air pollution has a strong negative effect on housing prices in both the parametric and semiparametric models because the regression coefficient on NOX in model 2 and the estimates of $g(NOX)$ in Figure 1 take negative values. Third, the

Table 3: Estimation results for the coefficients in models 1 and 2.

	Model 1		Model 2	
Variable	Coefficient	Std. error	Coefficient	Std. error
CRIM	-0.1116	0.0382	-0.1025	0.0327
RM	4.1387	0.4522	3.8561	0.4133
INDUS	-0.0449	0.0651	-0.0126	0.0616
AGE	-0.0012	0.0155	0.0020	0.0134
DIS	-0.8068	0.3227	-1.3219	0.3375
RAD	0.5106	0.1578	0.2916	0.0690
PTRARIO	-0.9877	0.1682	-0.9638	0.1351
B	0.0091	0.0037	0.0099	0.0027
LSTAT	-0.5531	0.0572	-0.5362	0.0510
TAX	-0.0215	0.0056	-0.0120	0.0038
NOX	—	—	-14.7740	4.1372
Constant	23.8648	7.7433	26.0959	7.2377
λ	0.5775	0.2847	0.4037	0.1892
ρ	0.8062	—	0.8518	—
σ^2	25.0267	—	22.1511	—

effect of air pollution of house prices is not linear and the negative effect increases when the proportion of NOX is over a threshold value. Figure 1 shows the proportion of NOX tends to negatively affect house prices and this negative effect increases rapidly for values above 0.65. These results suggest that air pollution has negative effects on house values but that people are tolerant of air pollution to a certain extent.

6 Conclusions

In this study, we consider the PL-SARAR model and series estimation methods are employed to estimate the nonparametric term of the proposed model. For model estimation, we propose a three-step estimation procedure. The first step is the estimation of the parametric regression coefficient and spatial autoregressive parameters for the dependent variables using 2SLS. The series approximation coefficient for the nonparametric function is then estimated by OLS in the second step. The third step entails the estimation of variances and spatial autoregressive parameters in disturbances using NLS. We then establish the consistency and asymptotic normality of the proposed estimators. Monte Carlo simulations indicate that the small sample performances of the proposed estimator are reasonably good. Subsequently, we apply the proposed model and estimators to Boston land price data. We find that the proportion of NOX in the air tends to negatively affect house prices, the negative effect rapidly increasing for values above 0.65.

In future studies, some extensions of this study could be considered as follows. First, GMM could be used for the estimation of spatial autoregressive parameters in the proposed model instead of 2SLS and NLS. Lee and Liu (2010) indicate that GMM estimators are more efficient for the estimation of spatial autoregressive parameters. Applying GMM estimation procedures to the proposed model improves the efficiency of estimation. Second, the extension of the proposed model to spatial dynamic panel data models could be considered. Such models can control the dynamics of economic activities and unobserved time invariant heterogeneity across spatial units. This spatial dynamic panel extension would be helpful to investigate dynamic spatial spillover and causal effects in the empirical analysis.

References

[1] Anselin, L. (1988). Spatial econometrics: methods and models. Kluwer Academic Publishers, Boston.

[2] Arbia, G. (2014). A primer for spatial econometrics: with applications in R. Palgrave Macmillan, UK.

[3] Bernstein, D. S. (2009). Matrix mathematics: theory, facts, and formulas. Princeton university press, Princeton.

[4] Cliff, A.D., & Ord, J.K. (1973). Spatial Autocorrelation. Pion, London.

[5] Du, J., Sun X., Cao, R. & Zhang, Z., (2018). Statistical inference for partially linear additive spatial autoregressive models. Spatial Statistics, 25, 52-67.

[6] Harrison Jr, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. Journal of environmental economics and management, 5(1), 81-102.

[7] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer Verlag, New York.

[8] Hoshino, T. (2018). Semiparametric spatial autoregressive models with endogenous regressors: With an application to crime data. Journal of Business & Economic Statistics, 36(1), 160-172.

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

[10] Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review, 40(2), 509-533.

[11] Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of econometrics, 157(1), 53-67.

[12] Lee, L. F. (2004). Asymptotic Distributions of QuasiMaximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica, 72(6), 1899-1925.

[13] Lee, L. F., & Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory, 26(1), 187-230.

[14] Pace, R. K., & Gilley, O. W. (1997). Using the spatial configuration of the data to improve estimation. The Journal of Real Estate Finance and Economics, 14(3), 333-340.

[15] Ptscher, B. M., & Prucha, I. R. (1997). Dynamic nonlinear econometric models, Asymptotic theory. Springer Verlag, New York.

[16] Su, L. (2012). Semiparametric GMM estimation of spatial autoregressive models. Journal of Econometrics, 167(2), 543-560.

[17] Su, L., & Jin, S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. Journal of Econometrics, 157(1), 18-33.

[18] Zhang, Y., & 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.

[19] Zhan, Y. & Sun, Y., (2015). Estimation of partially specified dynamic spatial panel data models with fixed-effects, Regional Science and Urban Economics. 51, 37-46.

Appendix

The following facts summarize some basic properties on matrix algebras.

Fact 1. If the row and column sums of $n \times n$ matrices C_1 and C_2 are uniformly bounded in absolute value, then the row and column sums of C_1C_2 and C_2C_1 are also uniformly bounded in absolute value (see, e.g., Kelejian and Prucha (1998)).

Fact 2. Let C_1 be a symmetric matrix and C_2 be a positive semidefinite matrix. Then, $\gamma_{\min}(C_1)tr(C_2) \leq tr(C_1C_2) \leq \gamma(\max)(C_1)tr(C_2)$.

Fact 3. For an $n \times n$ matrix C, its spectral radius is bounded by $\max_i \sum_{j=1}^{n} |c_{n,i,j}|$, with $c_{n,i,j}$ being the (i, j)-th element of C_n (see, the appendix of Hoshino (2018)).

The following lemmas are essential for the proofs of the main results of this paper.

Lemma 1. Let A_n be an $n \times n$ matrix whose row and column sums are uniformly bounded in absolute value, and D_n be a symmetric and idempotent matrix. Suppose that Assumptions 1-5 hold. Then,

$$B_n^{'} A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n B_n = \tilde{B}_n{}' A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n\tilde{B}_n + O_p(\sqrt{n}),$$

where $\tilde{B}_n = (W_n(I_n - \lambda_0 W_n)^{-1}(X_n\beta_0 + g_0(S_n)), X_n)$.

Proof. By the definition of the matrix B_n, we have

$$\begin{aligned} B_n &= (W_nY_n, X_n), \\ &= (W_n(I_n - \lambda_0 W_n)^{-1}(X_n\beta_0 + g_0(S_n), X_n) + (W_n(I_n - \lambda_0 W_n)^{-1}(I_n - \rho_0 M_n)^{-1}\varepsilon_n, 0_{n\times d_x}), \\ &= \tilde{B}_n + \tilde{\varepsilon}_n. \end{aligned}$$

where $\tilde{B}_n = (W_n(I_n - \lambda_0 W_n)^{-1}(X_n\beta_0 + g_0(S_n), X_n)$ and $\tilde{\varepsilon}_n = (W_n(I_n - \lambda_0 W_n)^{-1}(I_n - \rho_0 M_n)^{-1}\varepsilon_n, 0_{n\times d_x})$ and $0_{n\times d_x}$ is an $n \times d_x$ matrix whose components are zero.

Thus,

$$\begin{aligned} B_n^{'} A_n^{'}(I_n - D_n)H_n(I_n - D_n)B_n &= (\tilde{B}_n + \tilde{\varepsilon}_n)' A_n^{'}(I_n - D_n)H_n(I_n - D_n)(\tilde{B}_n + \tilde{\varepsilon}_n), \\ &= \tilde{B}_n{}' A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n\tilde{B}_n + \tilde{B}' A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n\tilde{\varepsilon}_n \\ &\quad + \tilde{\varepsilon}_n'(I_n - D_n)H_n(I_n - D_n)A_n\tilde{B}_n + \tilde{\varepsilon}_n' A_n(I_n - D_n)H_n(I_n - D_n)A_n\tilde{\varepsilon}, \\ &= R11 + R12 + R13 + R14, \end{aligned}$$

where $R11 = \tilde{B}_n{}' A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n\tilde{B}_n$, $R12 = \tilde{B}' A_n^{'}(I_n - D_n)H_n(I_n - D_n)A_n\tilde{\varepsilon}_n$, $R13 = \tilde{\varepsilon}_n'(I_n - D_n)H_n(I_n - D_n)A_n\tilde{B}_n$ and $R14 = \tilde{\varepsilon}_n' A_n(I_n - D_n)H_n(I_n - D_n)A_n\tilde{\varepsilon}$.

Firstly, we consider R14. Let $T_n = A_nW_n(I_n - \lambda_0 W_n)^{-1}(I_n - \rho_0 M_n)^{-1}$. The row and column sums of T_n is uniformly bounded in absolute value by Assumption 1 and Fact 1, and $\gamma_{\max}(T_nT_n') = O(1)$ by Fact 3. Noting that the largest eigenvalue of an idempotent matrix is at most one, by Assumption 2 and Fact 2,

$$\begin{aligned} E(\varepsilon_n' T_n^{'}(I_n - D_n)H_n(I_n - D_n)T_n\varepsilon_n) &= \sigma^2 tr((Z_n'Z_n)^{\frac{1}{2}} Z_n'(I_n - D_n)T_nT_n'(I_n - D_n)Z_n(Z_n'Z_n)^{\frac{1}{2}}, \\ &\leq \sigma^2\gamma_{\max}(T_nT_n')tr((Z_n'Z_n)^{\frac{1}{2}} Z_n'(I - D_n)Z_n(Z_n'Z_n)^{\frac{1}{2}}), \\ &\leq \sigma^2\gamma_{\max}(T_nT_n')tr((Z_n'Z_n)^{\frac{1}{2}} Z_n'Z_n(Z_n'Z_n)^{\frac{1}{2}}), \\ &= O(1). \end{aligned}$$

Then, it follows by Markov's inequality that $R14 = O_p(1)$.

Next, we consider R12. By assumption 5,

$$
\begin{aligned}
E||\tilde{B}_n' A_n (I_n - D_n) H_n (I_n - D_n) T_n \varepsilon_n||^2 &= Etr(\varepsilon_n' T_n' (I_n - D_n) H_n (I_n - D_n) A_n' \tilde{B}_n \tilde{B}_n' A_n (I_n - D_n) H_n (I_n - D_n) T_n \varepsilon_n), \\
&\leq n\sigma^2 \overline{c}_{\tilde{B}_n} \gamma_{\max}(A_n' A_n) tr(T_n' (I_n - D_n) H_n (I_n - D_n) H_n (I_n - D_n) T_n), \\
&\leq n\sigma^2 \overline{c}_{\tilde{B}_n} \gamma_{\max}(A_n' A_n) \gamma_{\max}(T_n' T_n) tr(H_n), \\
&= O(n).
\end{aligned}
$$

Thus, $R12 = O_p(\sqrt{n})$ by Jensen's inequality and Markov's inequality. Similarly, we have $R13 = O_p(\sqrt{n})$.

By combining the convergence rate of $R12, R13$ and $R14$, we have

$$
B_n^{'} A_n^{'} (I_n - D_n) H_n (I_n - D_n) A_n B_n = R11 + O_p(\sqrt{n}).
$$

□

Lemma 2 Let A_n be an $n \times n$ matrix whose row and column sums are uniformly bounded in absolute value, D_n be a symmetric and idempotent matrix. Suppose that Assumptions 1-5 hold. Then,

$$
B_n' A_n (I_n - D_n) H_n (I_n - D_n) A_n V_n = O(nK^{-r_s}),
$$

Proof. By the definition of B_n, we have

$$
\begin{aligned}
B_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n &= \tilde{B}_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n + \tilde{\varepsilon}_n' A_n (I_n - D_n) H_n (I_n - D_n) A_n V_n, \\
&= R21 + R22,
\end{aligned}
$$

where $R21 = \tilde{B}_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n$ and $R22 = \tilde{\varepsilon}_n' A_n (I_n - D_n) H_n (I_n - D_n) A_n V_n$.

Firstly, we consider $R21$. By Assumption 4 and 5,

$$
\begin{aligned}
||\tilde{B}_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n||^2 &= tr(V_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n \tilde{B} \tilde{B}_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n), \\
&\leq n\overline{c}_{\tilde{B}_n} \gamma_{\max}(A_n A_n') tr(V_n' A_n' (I_n - D_n) H_n (I_n - D_n) A_n V_n), \\
&\leq n\overline{c}_{\tilde{B}_n} \gamma_{\max}(A_n A_n') \gamma_{\max}(A_n' A_n) ||V_n||^2, \\
&\leq n^2 \overline{c}_{\tilde{B}_n} \gamma_{\max}(A_n A_n') \gamma_{\max}(A_n' A_n) \sup_{s \in S} |p(s)' B_0 - f(s)|^2, \\
&= O(n^2 K^{-2r_s}).
\end{aligned}
$$

Thus, $R21 = O(nK^{-r_s})$ by Jensen's inequality.

Next, we consider $R22$. Similarly, by assumption 4 and 5,

$$\begin{aligned} E\|\varepsilon_n' T_n'(I_n - D_n)H_n(I_n - D_n)A_n V_n\|^2 &= Etr(V_n' A_n(I_n - D_n)H_n(I_n - D_n)T_n\varepsilon_n\varepsilon_n' T_n'(I_n - D_n)H_n(I_n - D_n)A_n V_n), \\ &\leq \sigma^2 \gamma_{\max}(T_n T_n')\gamma_{\max}(A_n' A_n)\|V_n\|^2, \\ &\leq \sigma^2 \gamma_{\max}(T_n T_n')\gamma_{\max}(A_n' A_n) n \sup_{s\in S} |p(s)' B_0 - f(s)|^2, \\ &= O(nK^{-2r_s}). \end{aligned}$$

Thus, $R22 = O_p(\sqrt{n}K^{-r_s})$ by Jensen's inequality and Markov's inequality.

By combining the convergence rate of $R21$ and $R22$, we have

$$B_n' A_n(I_n - D_n)H_n(I_n - D_n)A_n V_n = O(nK^{-r_s}).$$

□

Proof of Theorem 1 By the definition of $\hat{\delta}$,

$$\begin{aligned} \hat{\delta} &= (B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n)^{-1} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)Y_n, \\ &= \delta_0 + (B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n)^{-1} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)V \\ &\quad + (B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n)^{-1} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)(I - \rho_0 M_n)^{-1}\varepsilon_n. \end{aligned}$$

Thus,

$$\begin{aligned} \sqrt{n}(\hat{\delta} - \delta_0) &= \left(\frac{1}{n} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n\right)^{-1} \frac{1}{\sqrt{n}} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)V \\ &\quad + \left(\frac{1}{n} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n\right)^{-1} \frac{1}{\sqrt{n}} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)(I_n - \rho_0 M_n)^{-1}\varepsilon_n. \end{aligned}$$

By Lemma 1 and 2,

$$\begin{aligned} \frac{1}{n} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)B_n &\xrightarrow{p} \Sigma_2, \\ \frac{1}{\sqrt{n}} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)V &\xrightarrow{p} 0. \end{aligned}$$

By Slutsky's theorem and a central limit theorem, we have

$$\begin{aligned} \sqrt{n}(\hat{\delta} - \delta) &= \left(\frac{R11}{n} + O(n^{-1})\right)^{-1} \left(\frac{1}{\sqrt{n}} B_n'(I_n - \Pi_n)H_n(I_n - \Pi_n)(I_n - \rho_0 M_n)^{-1}\varepsilon_n + O(K^{-r_s})\right), \\ &\xrightarrow{d} N(0, \sigma^2 \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}). \end{aligned}$$

□

Proof of Theorem 2 Firstly, we consider the convergence rate of $\hat{\alpha}$. By the definition of $\hat{\alpha}$,

$$\begin{aligned}
\hat{\alpha} &= (P_n'P_n)^{-1}P_n'(Y_n - B_n\hat{\delta}), \\
&= \alpha_0 + (P_n'P_n)^{-1}P_n'B_n(\delta_0 - \hat{\delta}) + (P_n'P_n)^{-1}P_n'V_n + (P_n'P_n)^{-1}P_n'(I_n - \rho_0 M_n)^{-1}\varepsilon_n, \\
&= \alpha_0 + R31 + R32 + R33,
\end{aligned}$$

where $R31 = (P_n'P_n)^{-1}P_n'B_n(\delta_0 - \hat{\delta})$, $R32 = (P_n'P_n)^{-1}P_n'V_n$ and $R33 = (P_n'P_n)^{-1}P_n'(I_n - \rho_0 M_n)^{-1}\varepsilon_n$.

By the definition of B_n, we have

$$\begin{aligned}
R31 &= (P_n'P_n)^{-1}P_n'\tilde{B}(\delta_0 - \hat{\delta}) + (P_n'P_n)^{-1}P_n'\tilde{\varepsilon}(\delta_0 - \hat{\delta}), \\
&= R41 + R42,
\end{aligned}$$

where $R41 = (P_n'\Pi)^{-1}\Pi'\tilde{B}(\delta_0 - \hat{\delta})$ and $R42 = (P_n'P_n)^{-1}P_n'\tilde{\varepsilon}(\delta_0 - \hat{\delta})$.

Firstly we consider $R41$.

$$\begin{aligned}
||(P_n'P_n)^{-1}P_n'\tilde{B}_n(\delta_0 - \hat{\delta})||^2 &= tr((\delta_0 - \hat{\delta})'\tilde{B}_n'P_n(P_n'P_n)^{-2}P_n'\tilde{B}_n(\delta_0 - \hat{\delta})), \\
&\leq \frac{1}{n}\underline{c}_{P_n}^{-1} tr((\delta_0 - \hat{\delta})'\tilde{B}_n'P(P'P)^{-1}P'\tilde{B}_n(\delta_0 - \hat{\delta})), \\
&\leq \underline{c}_{P_n}^{-1}\overline{c}_{\tilde{B}_n} tr((\delta_0 - \hat{\delta})'(\delta_0 - \hat{\delta})), \\
&\leq c_\Pi c_{\tilde{B}} tr((\delta_0 - \hat{\delta})'(\delta_0 - \hat{\delta})), \\
&= O(n^{-1}).
\end{aligned}$$

Thus, $R41 = O(n^{-1/2})$ by Jensen's inequality.

Similarly, we consider $R42$.

$$\begin{aligned}
E||(P_n'P_n)^{-1}P_n'T_n\varepsilon_n(\lambda_0 - \hat{\lambda})||^2 &= (\lambda_0 - \hat{\lambda})^2\sigma^2 tr(T_n'P_n(P_n'P_n)^{-2}P'T_n), \\
&= (\lambda_0 - \hat{\lambda})^2\sigma^2\frac{1}{n}\underline{c}_{P_n}^{-1} tr(T_n'T_n), \\
&= O(n^{-1})
\end{aligned}$$

Thus, $R42 = O_p(n^{-1/2})$ by Jensen's inequality and Markov's inequality.

Therefore, we have $R31 = O_p(n^{-1/2})$ by combining the convergence rate of $R41$ and $R42$.

Next, we consider $R32$.

$$
\begin{aligned}
||(P_n'P_n)^{-1}P_n'V_n||^2 &= tr(V_n'P_n(P_n'P_n)^{-2}P_nV_n), \\
&\leq \frac{1}{n}\underline{c}_{P_n}^{-1}tr(V'V), \\
&\leq \underline{c}_{P_n}^{-1}\sup_{s\in S}|p(s)'B_0 - f(s)|^2, \\
&= O(K^{-2r_s}).
\end{aligned}
$$

Thus, $R32 = O(K^{-r_s})$ by Jensen's inequality.

Finally, we consider $R33$.

$$
\begin{aligned}
E||(P_n'P_n)^{-1}P_n'(I_n-\rho_0 M_n)^{-1}\varepsilon_n||^2 &= Etr(\varepsilon_n(I_n-\rho_0 M_n)^{'-1}P_n(P_n'P_n)^{-2}P_n'(I_n-\rho_0 M_n)^{-1}\varepsilon_n, \\
&\leq \frac{1}{n}\sigma^2\underline{c}_{P_n}^{-1}\gamma_{\max}((I_n-\rho_0 M_n)^{'-1}(I_n-\rho_0 M_n)^{-1})tr(P_n(P_n'P_n)^{-1}P_n'), \\
&= O\left(\frac{K}{n}\right).
\end{aligned}
$$

Thus, $R33 = O_p(\sqrt{K}/\sqrt{n})$by Jensen's inequality and Markov's inequality.

Therefore, we obtain $\hat{\alpha} = \alpha_0 + O_p\left(\frac{\sqrt{K}}{\sqrt{n}} + K^{-r_s}\right)$.

Next, we consider the uniform convergence rate of $\hat{g}(\cdot)$. By the triangle inequity and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
\sup_s||\hat{g}(s) - g_0(s)|| &\leq \sup_s||P^K(s)(\hat{\alpha}-\alpha_0)|| + \sup_s||P^K(s)\alpha_0 - g_0(s)||, \\
&\leq ||\hat{\alpha}-\alpha_0||\sup_s||P^K(s)|| + O(K^{-r_s}), \\
&= O_p\left(\frac{K}{\sqrt{n}} + K^{(1-2r_s)/2}\right).
\end{aligned}
$$

Finally, we consider the limiting distribution of $\hat{g}(\cdot)$. By the defintion of $\hat{g}(s)$,

$$
\begin{aligned}
\hat{g}(s) - g_0(s) &= P^K(s)\hat{\alpha} - (p^K\alpha_0 + O(K^{-r_s}), \\
&= p^K(R31 + R32 + R33) + O(K^{-r_s}).
\end{aligned}
$$

It follows by the above discussion that

$$
\begin{aligned}
||p^K R31|| &= ||P^K(s)(P_n'P_n)^{-1}P_n'B(\delta - \hat{\delta})||, \\
&\leq ||P^K(s)||\ ||(P_n'P_n)^{-1}P_n'B(\delta - \hat{\delta})||,
\end{aligned}
$$

$$= O\left(\frac{\sqrt{K}}{\sqrt{n}}\right),$$

and

$$\begin{aligned} ||p^K R32|| &= ||P^K(s)(P_n'P_n)^{-1}P_n'V||, \\ &\leq ||P^K(s)||\ ||(P_n'P_n)^{-1}P_n'V_n||, \\ &= O(K^{(1-2r_s)/2}). \end{aligned}$$

Thus,

$$\hat{g}(s) - g_0(s) = P^K(s)(P_n'P_n)^{-1}P_n'(I_n - \rho_0 M_n)^{-1}\varepsilon_n + O\left(\frac{\sqrt{K}}{\sqrt{n}} + K^{(1-2r_s)/2}\right).$$

Let us consider the variance of the first term of the above equation.

$$\begin{aligned} \sigma^2(s) &= E(P^K(s)(P_n'P_n)^{-1}P_n'(I_n - \rho_0 M_n)^{-1}\varepsilon_n\varepsilon_n'(I_n - \rho_0 M_n)^{'-1}P_n(P_n'P_n)^{-1}P^{'K}(s)), \\ &= \sigma^2(P^K(s)(P_n'P_n)^{-1}P_n'(I_n - \rho_0 M_n)^{-1}(I_n - \rho_0 M_n)^{'-1}P_n(P_n'P_n)^{-1}P^{'K}(s)), \\ &\leq \sigma^2\gamma_{\max}((I_n - \rho_0 M_n)^{-1}(I_n - \rho_0 M_n)^{'-1})(P^K(s)(P_n'P_n)^{-1}P^{'K}(s)), \\ &\leq \sigma^2\gamma_{\max}((I_n - \rho_0 M_n)^{-1}(I_n - \rho_0 M_n)^{'-1})\underline{c}_{P_n}\frac{1}{n}(P^K(s)P^{'K}(s)), \\ &= O\left(\frac{K}{n}\right). \end{aligned}$$

Similarly, $\sigma^2(s) \geq O(K/n)$, Thus $\sigma^2(s) = O(K/n)$.

By Slutsky's theorem and a central limit theorem, we obtaine

$$\hat{g}(s) - g_0(s) \xrightarrow{d} N(0, \sigma^2(s)).$$

□

Proof of Theorem 3 Let $\hat{u}_n = Y_n - B_n\hat{\delta} - P_n\hat{\alpha}$. As the first step, we show that

$$\frac{1}{n}\hat{u}_n'A_n\hat{u}_n - E\frac{1}{n}u_n'A_nu_n = o_p(1),$$

where A_n is a matrix whose row and column sums are uniformly bounded in absolute values.

Note that

$$\begin{aligned}\frac{1}{n}\hat{u}_n'A_n\hat{u}_n - E\frac{1}{n}\hat{u}_n'A_n\hat{u}_n &= \frac{1}{n}\hat{u}_n'A_n\hat{u}_n - \frac{1}{n}u_n'A_nu_n, \\ &\quad +\frac{1}{n}u_n'A_nu_n - E\frac{1}{n}u_n'A_nu_n.\end{aligned}$$

Firstly, we show that $\frac{1}{n}u_n'A_nu_n - E\frac{1}{n}u_n'A_nu_n = o_p(1)$. By the definition of u_n

$$\begin{aligned}\frac{1}{n}u_n'A_nu_n &= \frac{1}{n}\varepsilon_n'(I_n-\rho_0M_n)'A_n(I_n-\rho_0M_n)\varepsilon_n, \\ &= \frac{1}{n}\varepsilon_n'A_n^*\varepsilon_n,\end{aligned}$$

where $A_n^* = (I_n-\rho_0M_n)'A_n(I_n-\rho_0M_n)$ and the row and column sums of A_n^* are uniformly bounded in absolute value by Fact 1. Thus it follows that $\frac{1}{n}u_n'A_nu_n - E\frac{1}{n}u_n'A_nu_n = o_p(1)$ immediately from the basic property of laws of large numbers in Lee (2004).

Next, we consider that $\frac{1}{n}\hat{u}_n'A_n\hat{u}_n - \frac{1}{n}u_n'A_nu_n = o_p(1)$. By the definition of $\hat{u}_n$,

$$\begin{aligned}\hat{u}_n &= Y_n - B_n\hat{\delta} - P_n\hat{\alpha}, \\ &= Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - P_n\hat{\alpha}, \\ &= u_n + (\lambda_0-\hat{\lambda})W_nY_n + X_n(\beta_0-\hat{\beta}) + (g_0(S_n)-P_n\hat{\alpha}), \\ &= u_n + (\lambda_0-\hat{\lambda})W_n(I_n-\lambda_0W_n)^{-1}(I_n-\rho_0M_n)\varepsilon_n \\ &\quad +(\lambda_0-\hat{\lambda})W_n(I_n-\lambda_0W_n)^{-1}(X_n\beta_0+g_0(S_n)) + X_n(\beta_0-\hat{\beta}) + (g_0(S_n)-P_n\hat{\alpha}), \\ &= u_n + \psi_1 + \psi_2 + \psi_3 + \psi_4,\end{aligned}$$

where $\psi_1 = (\lambda_0-\hat{\lambda})W_n(I_n-\lambda_0W_n)^{-1}(I_n-\rho_0M_n)\varepsilon_n, \psi_2 = (\lambda_0-\hat{\lambda})W_n(I_n-\lambda_0W_n)^{-1}(X_n\beta_0+g_0(S_n)), \psi_3 = X_n(\beta_0-\hat{\beta})$ and $\psi_4 = (g_0(S_n)-P_n\hat{\alpha})$.

Thus,

$$\begin{aligned}\frac{1}{n}\hat{u}_n'A_n\hat{u}_n - \frac{1}{n}u_n'A_nu_n &= \phi_1+\phi_2+\phi_3+\phi_4+2\phi_5+2\phi_6+2\phi_7+2\phi_8+2\phi_9+2\phi_{10} \\ &\quad +2\phi_{11}+2\phi_{12}+2\phi_{13}+2\phi_{14},\end{aligned}$$

where $\phi_1 = \frac{1}{n}\psi_1'A_n\psi_1, \phi_2 = \frac{1}{n}\psi_2'A_n\psi_2, \phi_3 = \frac{1}{n}\psi_3'A_n\psi_3, \phi_4 = \frac{1}{n}\psi_4'A_n\psi_4, \phi_5 = \frac{1}{n}u_n'A_n\psi_1, \phi_6 = \frac{1}{n}u_n'A_n\psi_2, \phi_7 = \frac{1}{n}u_n'A_n\psi_3, \phi_8 = \frac{1}{n}u_n'A_n\psi_4, \phi_9 = \frac{1}{n}\psi_1'A_n\psi_2, \phi_{10} = \frac{1}{n}\psi_1'A_n\psi_3, \phi_{11} = \frac{1}{n}\psi_1'A_n\psi_4, \phi_{12} = \frac{1}{n}\psi_2'A_n\psi_3, \phi_{13} =$

$\frac{1}{n}\psi_2' A_n \psi_4$ and $\phi_{14} = \frac{1}{n}\psi_3' A_n \psi_4$. We show that $\phi_i, i = 1, \dots, 14$, are or order $o_p(1)$. Here, note that

$$
\begin{aligned}
\hat{\rho} - \rho_0 &= O_p\left(\frac{1}{\sqrt{n}}\right), \\
\hat{\beta} - \beta_0 &= O_p\left(\frac{1}{\sqrt{n}}\right), \\
g_0(s) - P_n\hat{\alpha} &= O_p\left(\frac{K}{\sqrt{n}} + K^{(1-2r_s)/2}\right), \\
&= O_p\left(\frac{K}{\sqrt{n}} + \frac{\sqrt{K}}{\sqrt{n}}\sqrt{n}K^{-r_s}\right), \\
&= o_p(1).
\end{aligned}
$$

For example,

$$
\begin{aligned}
E\phi_1 &= E\frac{1}{n}(\lambda_0 - \hat{\lambda})^2 \varepsilon_n'(I_n - \rho_0 M_n)'(I_n - \lambda_0 W_n)'^{-1} W_n' A_n (\lambda_0 - \hat{\lambda}) W_n (I_n - \lambda_0 W_n)^{-1}(I_n - \rho_0 M_n)\varepsilon_n, \\
&= E(\lambda_0 - \hat{\lambda})^2 \frac{1}{n}\varepsilon_n' \tilde{A}_n \varepsilon_n, \\
&= \sigma_0^2 (\lambda_0 - \hat{\lambda})^2 \frac{1}{n}\sum_{i=1}^{n} \tilde{a}_{n,i,j}, \\
&= o_p(1),
\end{aligned}
$$

where $\tilde{A}_n = (I_n - \rho_0 M_n)'(I_n - \lambda_0 W_n)'^{-1} W_n' A_n (\lambda_0 - \hat{\lambda}) W_n (I_n - \lambda_0 W_n)^{-1}(I_n - \rho_0 M_n)$ and $\tilde{a}_{n,i,j}$ is the (i,j)th element of the matrix $\tilde{A}_n$. The remaining terms can be shown to be $o_p(1)$ in the same way. Therefore, $\frac{1}{n}\hat{u}_n' A_n \hat{u}_n - E\frac{1}{n}u_n' A_n u_n = o_p(1)$

We prove the consistency of the third step estimator following Kelejian and Prucha (1999). The objective function of the nonlinear least squares estimator and its corresponding counterpart are given by

$$
\begin{aligned}
R_n(\theta) &= [G_n - g_n]'[G_n - g_n], \\
\hat{R}_n(\theta) &= [\hat{G}_n - \hat{g}_n]'[\hat{G}_n - \hat{g}_n],
\end{aligned}
$$

where $\theta = (\rho, \sigma^2)'$.

Let $\theta_0 = (\rho_0, \sigma_0^2)'$. By Assumption 7,

$$
\begin{aligned}
R_n(\theta) - R_n(\theta_0) &= [\rho - \rho_0, \rho^2 - \rho_0^2, \sigma^2 - \sigma_0^2] G_N' G_n [\rho - \rho_0, \rho^2 - \rho_0^2, \sigma^2 - \sigma_0^2]', \\
&\geq \underline{c}_{G_n} [\rho - \rho_0, \sigma^2 - \sigma_0^2][\rho - \rho_0, \sigma^2 - \sigma_0^2]', \\
&= \underline{c}_{G_n} ||\theta - \theta_0||^2.
\end{aligned}
$$

It follow that for every $\varepsilon > 0$ and any N,

$$\begin{aligned} \inf_{\theta:||\theta-\theta_0||\geq\varepsilon} [R_n(\theta) - R_n(\theta_0)] &\geq \underline{c}_{G_n}\varepsilon^2, \\ &> 0. \end{aligned}$$

Thus, the identifiability of θ is proved.

Let $F_n = [G_n, -g_n], \hat{F}_n = [\hat{G}_n, -\hat{g}_n], \rho \in [-a, a]$ and $\sigma^2 \in [0, b]$.

$$\begin{aligned} |R_n(\theta) - \hat{R}_n(\theta)| &= \left| [\rho, \rho^2, \sigma^2, 1][F_n' F_n - \hat{F}_n' \hat{F}_n][\rho, \rho^2, \sigma^2, 1]' \right|, \\ &\leq ||F_n' F_n - \hat{F}_n' \hat{F}_n||\ [1 + a^2 + a^4 + b^2]. \end{aligned}$$

The elements of F_n and $\hat{F}_n$ are all of the form $\frac{1}{n}\hat{u}_n' A_n \hat{u}_n$ and $E\frac{1}{n}\hat{u}_n' A_n \hat{u}_n$ where the row and column sums of A_n are uniformly bounded in absolute value. We have shown that $\frac{1}{n}\hat{u}_n' A_n \hat{u}_n - E\frac{1}{n}\hat{u}_n' A_n \hat{u}_n = o_p(1)$. Thus, $F_n - \hat{F}_n = o_p(1)$. It follow that

$$\begin{aligned} \sup_{\rho,\sigma} |R_n(\theta) - \hat{R}_n| &\leq ||F_n' F_n - \hat{F}_n' \hat{F}_n||[1 + a^2 + a^4 + b^2], \\ &\xrightarrow{p} 0. \end{aligned}$$

The consistency of $\hat{\rho}_n$ and $\hat{\sigma}_n^2$ follows form Lemma 3.1 in Potscher and Prucha (1997).

□