Econometrics and Statistics 9 (2019) 140–155

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Estimation of a semiparametric varying-coefficient mixed regressive spatial autoregressive model

Yanqing Sun[a,1,*], Yuanqing Zhang[b,1], Jianhua Z. Huang[c]

[a] Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[b] School of Finance and Business, Shanghai Normal University, Shanghai, China
[c] Department of Statistics, Texas A&M University, College Station, TX 77843, USA

ARTICLE INFO

Article history:
Received 2 March 2016
Revised 23 May 2017
Accepted 26 May 2017
Available online 9 June 2017

Keywords:
Asymptotic theory
Semiparametric varying coefficient
Series approximation
Spatial mixed regression
Teen pregnancy analysis
Two-stage least squares estimation

ABSTRACT

A semiparametric varying-coefficient mixed regressive spatial autoregressive model is used to study covariate effects on spatially dependent responses, where the effects of some covariates are allowed to vary with other variables. A semiparametric series-based least squares estimating procedure is proposed with the introduction of instrumental variables and series approximations of the conditional expectations. The estimators for both the nonparametric and parametric components of the model are shown to be consistent and their asymptotic distributions are derived. The proposed estimators perform well in simulations. The proposed method is applied to analyze a data set on teen pregnancy to investigate effects of neighborhood as well as other social and economic factors on the teen pregnancy rate.

1. Introduction

It has become increasingly clear that in many practical situations parametric modelling is not capable of capturing the relationship between the response variable and covariates of interest. Varying-coefficient models have been developed to model such associations nonparametrically, where the effects of covariates vary with other variables (e.g., see Hastie and Tibshirani (1993)). An important advantage of varying-coefficient models is that they partially ameliorate the curse of dimensionality problem by restricting the nonparametric functions to a subset of variables. The varying-coefficient partially linear models are extensions of the varying-coefficient models by allowing some covariate effects to be constant which further increases modeling flexibility; see Zhang et al. (2002), Li et al. (2002) and Fan and Huang (2005).

Spatial models have been extensively studied in the econometrics and geography literature. In recent years issues concerning spatial dependence among cross sectional units have received increased attention. Kelejian and Prucha (1999) introduced a method of moments (MOM) estimator for the autoregressive parameter in a spatial autoregressive (SAR) model. Lee (2001) developed a generalized MOM (GMM) estimator for spatial autoregressive processes to improve efficiency. Lee (2007) proposed a mixed regressive spatial autoregressive (MRSAR) model, which assumes constant covariate effects and accounts for spatially dependent responses. Approaches accounting for spatial dependence via spatial random effects have also been studied. Gelfand et al. (2003) modelled spatially correlated responses by introducing a second-order stationary

* Corresponding author.
E-mail address: yasun@uncc.edu (Y. Sun).
[1] Yanqing Sun and Yuanqing Zhang are joint-first authors.

http://dx.doi.org/10.1016/j.ecosta.2017.05.005

process as spatial random effects. Schliep et al. (2015) proposed a hierarchical autoregressive spatially varying coefficients model to predict particulate matter in the atmosphere using satellite AOT data. Schliep et al. (2015) modeled both time and space, where for each given space location, the model is autoregressive in time. In this paper, we adopt the approach of Lee (2007) by modelling covariate effects on spatially dependent responses through spatial weights that are known constants. We extend the MRSAR model to allow the effects of some covariates to vary nonparametrically with another covariate. This extension allows for more flexible modelling of covariate effects. The nonlinear nature of the effects can only be discovered with a model that allows for varying coefficients. To the best of our knowledge, varying-coefficient mixed regressive spatial autoregressive models have not been considered in the existing literature.

By accounting for spatial dependence among responses in addition to the influence of covariates, spatial models have applications in economics and the social sciences. For example, a nation's per capita GDP growth rate is affected not only by the value of its indexes, such as savings, income tax rate, and population growth rate, but also by its neighboring nations' per capita GDP growth rates and the values of these indexes. In a social science example that we analyze in Section 5, a county's teenage pregnancy rate is affected not only by its own social and economic variables, but also by the teenage pregnancy rates of its neighboring counties.

Let y_i be the response variable and (u_i, x_i^T, z_i^T) be covariates for subject i, where x_i and z_i are exogenous covariates of dimensions p and q, respectively, and u_i is a scalar covariate. The semiparametric varying-coefficient mixed regressive spatial autoregressive (SVMRSAR) model assumes that

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + z_i^T \beta + x_i^T \alpha(u_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where β is a q-dimensional coefficient vector, $\alpha(u)$ is a p-dimensional vector of functions of u with support on a one-dimensional set $\mathcal{U}$, and λ is a spacial effect parameter. The error term ε_i is a random variable satisfying $E[\varepsilon_i|x_i, z_i, u_i] = 0$ and $E[\varepsilon_i^2|x_i, z_i, u_i] = \sigma_i^2$. In addition, in model (1) w_{ij} are spatial weights of known constants with the diagonal elements $w_{ii} = 0$. The term $\sum_{j=1}^{n} w_{ij} y_j$ represents weighted spatial lag variables and the coefficient λ reflects the spatial influence on y_i by its neighbors' responses.

In many applications, it is common practice to choose the row-normalized weight matrix $W = (w_{ij})$ such that the sum of elements in each row of W is unity. For example, the ith row of W may be constructed as $w_i = (d_{i1}, d_{i2}, \ldots, d_{in})/\sum_{j=1}^{n} d_{ij}$, where $d_{ij} > 0$, for $i \neq j$, represents a function of the spatial distance of the ith and jth units such as the inverse of the distance of the i and jth units in some (characteristic) space. The weighting operation may be interpreted as an average of neighboring values. The spatial effect coefficient λ measures the average influence of neighboring observations on the responses that usually lies between $(-1, 1)$ when W is row-normalized. For a general W which is not row-normalized, λ is assumed to be in a parameter space which guarantees that the determinant of $I_n - \lambda W$ is positive, where I_n is the $n \times n$ identity matrix.

We assume that (u_i, x_i^T, z_i^T), $i = 1, \ldots, n$, are exogenous regressors. The response y_i depends on (u_i, x_i^T, z_i^T) and is correlated with other responses y_j for $j \neq i$ through the known weights w_{ij} and unknown parameter λ. Model (1) permits the interaction between the covariates u_i and x_i such that different levels of the covariate u_i are associated with different linear models. This allows examination of the extent to which the effects of covariates x_i vary over different levels of the covariate u_i. Setting $x_i = 0$, model (1) becomes the MRSAR model, which has been widely studied in the literature; see, for example, the work of Kelejian and Prucha (1999), 2001), Lee (2001), 2007) and Lin and Lee (2010), among others. When $x_i = 1$, model (1) is a partially linear MRSAR model, see the work of Su and Jin (2010) for details. When $\lambda = 0$, that is, there exists no spatial dependency impact, model (1) becomes a semiparametric varying-coefficient partially linear model. We refer to Zhang et al. (2002), Li et al. (2002) and Fan and Huang (2005) for related work.

This article investigates the semiparametric varying-coefficient mixed regressive spatial autoregressive model (1). We generalize the result of Newey (1997) from the independent case to the spatially dependent case. We propose a semiparametric series-based least squares estimation procedure that makes use of the newly introduced techniques of instrumental variables and series approximations of conditional expectations. We show that our estimators of the parametric and nonparametric components are consistent and asymptotically normal. Our simulation study demonstrates that the proposed estimators perform well. The proposed method is applied to investigate whether neighborhoods and other social economic factors affect the teen pregnancy rate based on data from the study "Health and Healthcare in the United States-County and Metro Area Data" (Thomas, 2000) and the 1990 US census (US Census Bureau, 1992).

The rest of the paper is organized as follows. The estimation procedure for model (1) is presented in Section 2. Section 3 derives the asymptotic properties of the proposed estimators and provides estimators for the asymptotic covariance matrix of the estimators. Section 4 presents a Monte Carlo simulation study evaluating the finite-sample performance of the proposed estimators. In Section 5, the proposed method is applied to analyze the factors that affect the teen pregnancy rate. Some concluding remarks are given in Section 6. All the proofs are collected in the Appendix.

2. Semiparametric series-based least squares estimation

In this section, we develop an estimation procedure for model (1). The estimation of the parametric part is obtained using the two-stage least squares estimation method with the introduction of appropriate instrumental variables. The two-stage least squares estimation involves some unknown conditional expectations which in turn are estimated by nonparametric series regression estimation. The estimation of the varying-coefficient part is also accomplished using a series approximation.

2.1. Series-based two-stage least squares estimation of parametric model components

In this section, we derive the estimators for the parametric components of model (1). Let $w_i = (w_{i1}, \dots, w_{in})^T$, $Y = (y_1, ..., y_n)^T$, $X = (x_1, \dots, x_n)^T$ and $Z = (z_1, \dots, z_n)^T$. Multiplying x_i on both sides of Eq. (1) yields

$$x_i y_i = \lambda x_i w_i^T Y + x_i z_i^T \beta + x_i x_i^T \alpha(u_i) + x_i \varepsilon_i. \tag{2}$$

Taking the conditional expectation on both sides yields

$$E(x_i y_i|u_i) = \lambda E(x_i w_i^T Y|u_i) + E(x_i z_i^T|u_i)\beta + E(x_i x_i^T|u_i)\alpha(u_i) + E(x_i\varepsilon_i|u_i).$$

Since $E(x_i\varepsilon_i|u_i) = E\{x_i E(\varepsilon_i|x_i, z_i, u_i)|u_i\} = 0$, we have

$$\alpha(u_i) = \{E(x_i x_i^T|u_i)\}^{-1}\{E(x_i y_i|u_i) - \lambda E(x_i w_i^T Y|u_i) - E(x_i z_i^T|u_i)\beta\}. \tag{3}$$

Let $\tilde{y}_i = y_i - x_i^T[E(x_i x_i^T|u_i)]^{-1}E(x_i y_i|u_i)$, $\tilde{z}_i = z_i - E(z_i x_i^T|u_i)[E(x_i x_i^T|u_i)]^{-1}x_i$, and $\tilde{Y}_{w,i} = w_i^T Y - x_i^T[E(x_i x_i^T|u_i)]^{-1}E(x_i w_i^T Y|u_i)$. Plugging in the expression (3) for $\alpha(u_i)$ into (1), we get

$$\tilde{y}_i = \tilde{b}_i^T \vartheta + \varepsilon_i, \tag{4}$$

where $\tilde{b}_i = (\tilde{Y}_{w,i}^T, \tilde{z}_i^T)^T$ and $\vartheta = (\lambda, \beta^T)^T$.

Since $\tilde{Y}_{w,i}$ is correlated with ε_i, (4) is a linear model with endogenous regressor $\tilde{b}_i$. It is well known that the ordinary least squares estimators for (4) are biased, cf. Cameron and Trivedi (2008). A general approach to estimating ϑ is to find a set of instruments that are orthogonal to ε_i and then apply a method of moments procedure. Following Kelejian and Prucha (1999), we derive a consistent estimator of ϑ by introducing appropriate instrumental variables and adopting two-stage least squares estimation.

Suppose for the moment that we have observed the conditional expectations $E(x_i x_i^T|u_i)$, $E(x_i z_i^T|u_i)$, $E(x_i y_i|u_i)$, $E(x_i w_i^T Z|u_i)$, and $E(x_i w_i^T Y|u_i)$. Let $\tilde{Z}_{w,i}^T = w_i^T Z - x_i^T[E(x_i x_i^T|\ u_i)]^{-1}E(x_i w_i^T Z|u_i)$ and $\tilde{h}_i = (\tilde{Z}_{w,i}^T, \tilde{z}_i^T)^T$. It is easy to check that $\tilde{h}_i$ is orthogonal to ε_i with $E(\tilde{h}_i\varepsilon_i) = 0$ and $\tilde{h}_i$ is orthogonal to x_i. Let $\tilde{H} = (\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n)^T$ be the matrix of instrumental variables. Let $\tilde{Y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)^T$ and $\tilde{B} = (\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_n)^T$. The two-stage least squares (2SLS) infeasible estimator of ϑ for model (4) is given by

$$\hat{\vartheta}_{inf} = [\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{B}]^{-1}\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{Y}. \tag{5}$$

The 2SLS infeasible estimator $\hat{\vartheta}_{inf}$ is obtained using the two-stage ordinary least squares (OLS) method. In the first stage, a OLS regression of $\tilde{B}$ on the instruments $\tilde{H}$ is used to obtain fitted values $\tilde{B}^* = P_{\tilde{H}}\tilde{B}$, where $P_{\tilde{H}} = \tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T$ is an idempotent projection matrix. The two-stage least squares (2SLS) infeasible estimator can be expressed as $\hat{\vartheta}_{inf} = (\tilde{B}^{*T}\tilde{B}^*)^{-1}\tilde{B}^{*T}\tilde{Y}$, which is obtained from a OLS regression of $\tilde{Y}$ on the resulting fit $\tilde{B}^*$ from the first stage OLS. The condition $E(\tilde{h}_i\varepsilon_i) = 0$ is necessary for asymptotic unbiasedness of $\hat{\vartheta}_{inf}$ for ϑ while the orthogonality $E(\tilde{h}_i x_i^T) = 0$ ensures the asymptotic normality in Theorem 2 (b) in Section 3.

The estimator is infeasible because the conditional expectations involved in the right-hand side of (5) are unknown. To obtain a feasible estimator of ϑ, we need to estimate the conditional expectations $E(x_i x_i^T|u_i)$, $E(x_i z_i^T|u_i)$, $E(x_i y_i|u_i)$, and $E(x_i w_i^T Y|u_i)$, and plug the estimates into the right-hand side of (5). Following Newey (1997), we adopt nonparametric series regression estimation of these conditional expectations using the basis functions approach. Sieve methods such as B-splines, polynomial splines, wavelets and Fourier series are often used for nonparametric estimation. In particular, Ai and Chen (2003) considered sieve minimum distance estimation under conditional moment restrictions containing unknown functions. Chen and Pouzo (2009) investigated penalized sieve minimum distance estimation for conditional moment models with unknown parametric components and unknown functions of endogenous variables. Zhang and Sun (2015) employed the sieve approach for spatial dynamic panel data regression with fixed effects. In practice, a variety of polynomial basis functions such as Hermite polynomials, polynomial splines, and B-splines can be used for the nonparametric approximation.

Let $p^K(u) = (p_{1K}(u), \dots, p_{KK}(u)))^T$ be a sequence of K known basis functions. Let $\tilde{P}(u) = I_p \otimes (p^K(u))^T$, where $\otimes$ is the Kronecker product and $\tilde{P}(u)$ is a $p \times pK$ matrix. The series approximation of $E(x_i x_i^T|u_i)$ is given by

$$E(x_i x_i^T|u_i) \approx \tilde{P}(u_i)\theta_{1K}, \tag{6}$$

where $\theta_{1K} = (\theta_{1\cdot}^T, \dots, \theta_{p\cdot}^T)^T$, and each $\theta_{l\cdot} = (\theta_{l1}, \dots, \theta_{lK})^T$ is a K-dimensional vector of parameters for $l = 1, \dots, p$. A series estimate of $E(x_i x_i^T|u_i)$ is obtained by linearly regressing $x_i x_i^T$ on $\tilde{P}(u_i)$, yielding the closed-form expression

$$\hat{E}(x_i x_i^T|u_i) = \tilde{P}(u_i)(\vec{P}^T\vec{P})^{-1}\sum_{i=1}^{n}\tilde{P}(u_i)^T x_i x_i^T, \tag{7}$$

where $\tilde{P} = (\tilde{P}^T(u_1), \ldots, \tilde{P}^T(u_n))^T$ is a $pn \times pK$ matrix. Similarly, the series estimates of $E(x_i z_i^T|u_i)$, $E(x_i y_i|u_i)$, and $E(x_i w_i^T Y|u_i)$ can be obtained as

$$\hat{E}(x_i z_i^T|u_i) = \tilde{P}(u_i)(\tilde{P}^T\tilde{P})^{-1}\sum_{i=1}^{n}\tilde{P}(u_i)^T x_i z_i^T, \tag{8}$$

$$\hat{E}(x_i y_i|u_i) = \tilde{P}(u_i)(\tilde{P}^T\tilde{P})^{-1}\sum_{i=1}^{n}\tilde{P}(u_i)^T x_i y_i, \tag{9}$$

$$\hat{E}(x_i w_i^T Y|u_i) = \tilde{P}(u_i)(\tilde{P}^T\tilde{P})^{-1}\sum_{i=1}^{n}\tilde{P}(u_i)^T x_i w_i^T Y, \tag{10}$$

respectively.

Lemma 1 in Section 3 shows that the estimates $\hat{E}(x_i x_i^T|u_i)$, $\hat{E}(x_i z_i^T|u_i)$, and $\hat{E}(x_i w_i^T Y|u_i)$ are consistent and also provides the rates of convergence. We replace the conditional expectations $E(x_i x_i^T|u_i)$, $E(x_i z_i^T|u_i)$, $E(x_i y_i|u_i)$ and $E(x_i w_i^T Y|u_i)$ with $\hat{E}(x_i x_i^T|u_i)$, $\hat{E}(x_i z_i^T|u_i)$, $\hat{E}(x_i y_i|u_i)$, $\hat{E}(x_i w_i^T Z|u_i)$ and $\hat{E}(x_i w_i^T Y|u_i)$, in the definitions of $\tilde{y}_i$, $\tilde{z}_i$, $\tilde{b}_i$ and $\tilde{h}_i$ to obtain $\hat{y}_i$, $\hat{z}_i$, $\hat{b}_i$ and $\hat{h}_i$, respectively. Let $\hat{B} = (\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_n)^T$, $\hat{H} = (\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_n)^T$ and $\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)^T$. Replacing $\tilde{B}$, $\tilde{H}$ and $\tilde{Y}$ with $\hat{B}$, $\hat{H}$ and $\hat{Y}$, respectively, in (5), we obtain the following feasible least squares estimator of ϑ:

$$\hat{\vartheta} = [\hat{B}^T\hat{H}(\hat{H}^T\hat{H})^{-1}\hat{H}^T\hat{B}]^{-1}\hat{B}^T\hat{H}(\hat{H}^T\hat{H})^{-1}\hat{H}^T\hat{Y}. \tag{11}$$

The estimators of λ and β, denoted by $\hat{\lambda}$ and $\hat{\beta}$, are the first component and the last q components of $\hat{\vartheta}$, respectively.

2.2. Series-based least squares estimation of nonparametric model components

The expression of $\alpha(u_i)$ given in (3) is only for the covariate values u_i, $i = 1, \ldots, n$, not for an arbitrary u. Therefore plugging in the estimates of the conditional expectations (7)–(10) and the first-stage estimates of β and λ into (3) does not yield an estimate of $\alpha(u)$ right away. In this section we derive the estimators for the nonparametric regression functions $\alpha(u)$.

Consider the series approximation for each component $\alpha_l(u)$, $l = 1, \ldots, p$,

$$\alpha_l(u) = \sum_{k=1}^{K}\gamma_{lk}p_{kK}(u) + v_l(u), \quad u \in \mathcal{U} \tag{12}$$

where $p_{1K}(u), \ldots, p_{KK}(u)$ are K known basis functions, $\gamma_{1K}, \ldots, \gamma_{KK}$ are unknown coefficient parameters and $v_l(u)$ is the approximation error term. Using this series approximation, model (1) can be written as

$$y_i - \lambda w_i^T Y - z_i^T\beta = \sum_{l=1}^{p}\sum_{k=1}^{K}x_{il}\gamma_{lk}p_{kK}(u_i) + \sum_{l=1}^{p}x_{il}v_l(u_i) + \varepsilon_i. \tag{13}$$

Let $\gamma_l = (\gamma_{l1}, \ldots, \gamma_{lK})^T$ and $\gamma = (\gamma_1^T, \ldots, \gamma_p^T)^T$. Consider the plug-in series-based least squares objective function

$$L(\gamma) = \sum_{i=1}^{n}\left\{y_i - \hat{\lambda}w_i^T Y - z_i^T\hat{\beta} - \sum_{l=1}^{p}\sum_{k=1}^{K}x_{il}\gamma_{lk}p_{kK}(u_i)\right\}^2. \tag{14}$$

Let $\hat{\gamma} = (\hat{\gamma}_1^T, \ldots, \hat{\gamma}_p^T)^T$ be the minimizer of $L(\gamma)$, where $\hat{\gamma}_l = (\hat{\gamma}_{l1}, \ldots, \hat{\gamma}_{lK})^T$ for $l = 1, \ldots, p$. The estimator of $\alpha_l(u)$ is given by $\hat{\alpha}_l(u) = \sum_{k=1}^{K}\hat{\gamma}_{lk}p_{kK}(u)$, which is referred to as the least squares series estimator of $\alpha_l(u)$.

Let $D_i = x_i^T\tilde{P}(u_i)$, and $D = (D_1^T, \ldots, D_n^T)^T$. The object function (14) is equivalent to

$$L(\gamma) = \sum_{i=1}^{n}\{y_i - \hat{\lambda}w_i^T Y - z_i^T\hat{\beta} - D_i\gamma\}^2. \tag{15}$$

Suppose that $\sum_{i=1}^{n}D_i^T D_i$ is invertible. Then the plug-in least squares estimator $\hat{\gamma}$ is given by

$$\hat{\gamma} = \left(\sum_{i=1}^{n}D_i^T D_i\right)^{-1}\sum_{i=1}^{n}D_i^T(y_i - \hat{\lambda}w_i^T Y - z_i^T\hat{\beta}) = (D^T D)^{-1}D^T(Y - \hat{\lambda}WY - Z\hat{\beta}). \tag{16}$$

The least squares series estimator of $\alpha(u)$, for $u \in \mathcal{U}$, has expression

$$\hat{\alpha}(u) = \tilde{P}(u)\hat{\gamma} = [I_p \otimes p^K(u)^T](D^T D)^{-1}D^T(Y - \hat{\lambda}WY - Z\hat{\beta}). \tag{17}$$

We note that since model (1) includes the weighted spatial lag variables $\sum_{j=1}^{n} w_{ij}y_j$, which are endogenous variables, estimating both the parametric and nonparametric components simultaneously through minimizing the least squares objective function (15) may not yield consistent estimators, cf. Kelejian and Prucha (1999). We show in the next section that this inconsistency problem can be avoided by first estimating λ and β using the two-stage least squares estimator facilitated with instrumental variables and then estimating $\alpha(u)$ through the profile least squares approach. The profile estimation method is often used in semiparametric estimation to improve computational efficiency, whereby the estimators for the parametric components converge at a faster rate.

3. Asymptotic results and variance estimation

In this section, we investigate the asymptotic properties of the proposed estimators for model (1). The estimators for the asymptotic covariances of these estimators are also given to enable statistical inferences.

Let $W = (w_{ij})_{n\times n}$ be the $n \times n$ spatial weights matrix of known constants, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, and $U = (u_1, \ldots, u_n)^T$. Let X be the $n \times np$ block diagonal matrix with x_i^T on the ith diagonal. The semiparametric varying-coefficient mixed regressive spatial autoregressive (SVMRSAR) model (1) can be rewritten as

$$Y = \lambda WY + X\vec{\alpha}(U) + Z\beta + \varepsilon, \tag{18}$$

where $\vec{\alpha}(U) = (\alpha^T(u_1), \ldots, \alpha^T(u_n))^T$ is an np dimensional vector of unknown coefficient functions that depends on u.

Let $\mathcal{U}$ be the support of u_i. The following regularity conditions are assumed for studying the asymptotic properties.

Condition A.

- A.1 The diagonal element of the spatial weighting matrix W is zero. The matrix $I - \lambda W$ is nonsingular for $|\lambda| < 1$. The row and column sums of the matrices W and $(I - \lambda W)^{-1}$ are bounded uniformly in n.
- A.2 The elements of X, Z are uniformly bounded by some constants uniformly in n.
- A.3 The random errors ε_i are independently identically distributed with $E[\varepsilon_i|x_i, z_i, u_i] = 0$ and $E[\varepsilon_i^2|x_i, z_i, u_i] = \sigma_i^2$ is bounded away from 0, and $E(\varepsilon_i^4|x_i, z_i, u_i)$ is bounded by a constant.
- A.4 Let $p^K(u) = (p_{1K}(u), \ldots, p_{KK}(u))^T$. The largest eigenvalue of $E[p^K(u_i)p^K(u_i)^T]$ is bounded uniformly in K. The smallest eigenvalue of $E[p^K(u_i)p^K(u_i)^T]$ is bounded away from zero uniformly in K. There exists a sequence of constants $\zeta_0(K)$ such that $sup_{u\in\mathcal{U}}\|p^K(u)\| \le \zeta_0(K)$ and $K = K(n) \to \infty$ such that $\zeta_0(K)^2K/n \to 0$ as $n \to \infty$, where $\|\cdot\|$ is the Euclidean norm of a vector.
- A.5 Let $g_0(u)$ represent one of the components $E(x_ix_i^T|u_i)$, $E(x_iz_i^T|u_i)$, $E(x_iy_i|u_i)$, $E(x_iw_i^TY\ |u_i)$ and $\alpha(u)$, and let θ_{0K} be the corresponding vector of coefficients under the series approximation similar to (6) under the basis functions $p^K(u) = (p_{1K}(u), \ldots, p_{KK}(u))^T$. There exist $\delta > 0$ and θ_K such that $\sup_{u\in\mathcal{U}}\|g_0(u) - (p^K(u))^T\theta_{0K}\| = O(K^{-\delta})$ as $K \to \infty$.
- A.6 The eigenvalues $\lambda_1, \ldots, \lambda_p$ of $E[x_ix_i^T|u_i]$ are bounded and bounded away from 0. The eigenvalues of $E[e_{XY}e_{XY}^T|U]$ are bounded and bounded away from zero, where $e_{XY} = X^TY - E(X^TY|U)$ and $E(X^TY|\ U) = (E(x_1^Ty_1|u_1), \ldots, E(x_n^Ty_n|u_n))^T$. Moreover, the eigenvalues of $E[e_{XwY}e_{XwY}^T|U]$ are bounded uniformly in n, where $e_{XwY} = X^TWY - E(X^TWY|U)$ and $E(X^TWY|U) = (E(x_1w_1^TY|u_1), \ldots, E(x_nw_n^T\ Y|u_n))^T$.
- A.7 The limits $n^{-1}\tilde{H}^T\tilde{B} \xrightarrow{P} Q_{\tilde{H}^T\tilde{B}}, n^{-1}\tilde{H}^T\tilde{H} \xrightarrow{P} Q_{\tilde{H}^T\tilde{H}}, n^{-1}\tilde{H}^T\Lambda\tilde{H} \xrightarrow{P} Q_{\tilde{H}^T\Lambda\tilde{H}}$ exist and are nonsingular, where $\Lambda = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$.

Condition A.1 imposes restrictions on the spatial weighting matrix. These restrictions are commonly imposed in the spatial regression literature (e.g., Lee (2003), 2007)). Condition A.2 is similar to an assumption in Kelejian and Prucha (1998). Condition A.3 is needed for establishing the asymptotic distribution of the nonparametric estimator $\widehat{\alpha}(u)$ and was used in Newey (1997).

Conditions A.4 and A.5 are imposed on the sieve approximations. Since the constant in the fixed-effects setting is not identified, we must impose some normalization on $g_0(.)$ such as $g_0(z_0) = 0$ at some point z_0 so that $g_0(.)$ can be identified. The basis functions $p^K(z)$ shall be constructed to satisfy this normalization. Condition A.4 imposes a normalization on the basis functions, bounding the second moment matrix away from singularity, and restricting the magnitude of the series terms. The bound $\zeta_0(K)$ is different for different basis functions. Newey (1997) showed that $\zeta_0(K)$ equals $C\sqrt{K}$ for splines and CK for power series where C is a constant. This implies a convergence rate at $K^2/n \to 0$ and $K^3/n \to 0$ for splines and power series, respectively. This condition is needed to ensure the convergence in probability of the sample second moment matrix of the approximating functions to their expectations.

Condition A.5 assumes that several conditional expectations can be approximated well using series and it specifies that the rate of uniform approximation error tends to zero at the rate $K^{-\delta}$. The constant δ is related to the smoothness of the conditional expectation functions and the dimensionality of u_i. If these objects have different degrees of smoothness, then the rate of convergence is determined by the least smooth component. For splines and power series, this assumption is satisfied with $\delta = s/r$, where s is the number of continuous derivatives of $g_0(u)$ that exist and r is the dimensionality of u_i; see Newey (1997). Under Condition A.5, $E(x_ix_i^T|u_i)$, $E(x_iz_i^T|u_i)$, $E(x_iy_i|u_i)$, and $\alpha(u_i)$ can be approximated well using the basis functions $p^K(u)$. Since $\lambda E(x_iw_i^TY|u_i) = E(x_iy_i|u_i) - E(x_iz_i^T|u_i)\beta - E(x_ix_i^T|u_i)\alpha(u_i)$, $E(x_iw_i^TY|u_i)$ can also be approximated well by the basis functions $p^K(u)$. Condition A.6 is needed to show that the series estimators of the conditional expectations are consistent (Lemma 1). Condition A.7 is a stability condition. The quantities are used in the expression of asymptotic variance of the estimator.

3.1. Asymptotic results

This section presents the asymptotic results of the estimators. The proofs of these results are given in the Appendix. Lemma 1 shows the consistency of the series estimators of the conditional expectations defined in (7)–(10). It also gives the rates of convergence of these estimators.

Lemma 1. *Under Conditions A.1–A.5, the following holds uniformly in u_i, $i = 1, \dots, n$ and n:*

(i) $\hat{E}(x_i x_i^T|u_i) = E(x_i x_i^T|u_i) + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$;
(ii) $\hat{E}(x_i z_i^T|u_i) = E(x_i z_i^T|u_i) + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$;
(iii) $\hat{E}(x_i y_i|u_i) = E(x_i y_i|u_i) + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$;
(iv) $\hat{E}(x_i w_i^T Y|u_i) = E(x_i w_i^T Y|u_i) + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$.
(v) $\hat{E}(x_i w_i^T Z|u_i) = E(x_i w_i^T Z|u_i) + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$.

The result in Lemma 1 is similar to the uniform convergence result given in Theorem 1 of Newey (1997). Note that δ is related to the smoothness of the conditional expectations and the dimensionality of u_i. According to Newey (1997), we can take $\delta = s/r$, where s is the number of continuous derivatives that exist of the least smooth conditional expectation and r is the dimensionality of u_i. The coefficients $\zeta_0(K)$ are different for different basis functions. The first term $O_p(\zeta_0(K)\sqrt{K}/\sqrt{n})$ essentially corresponds to the standard error of estimation and the second term $O_p(\zeta_0(K)K^{-\delta})$ corresponds to the bias of estimation. By letting the two terms $\zeta_0(K)\sqrt{K}/\sqrt{n}$ and $\zeta_0(K)K^{-\delta}$ go to zero at the same rate, we have $K = O(n^{1/(1+2\delta)})$ as $n \to \infty$, which achieves the best bias-variance trade-off in mean square error (Newey, 1997). Because $\zeta_0(K) = O_p(\sqrt{K})$ for splines and $\zeta_0(K) = O_p(K)$ for power series, the estimation errors $O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$ equal $O_p(Kn^{-1/2} + K^{1/2-\delta}) = O_p(n^{-(\delta-1/2)/(1+2\delta)})$ for splines and $O_p(K^{3/2}n^{-1/2} + K^{1-\delta}) = O_p(n^{-(\delta-1)/(1+2\delta)})$ for power series.

We next present the asymptotic results for the estimation of the parametric part of the model. The first result is about the infeasible estimator and the second result is about the feasible estimator.

Theorem 1. *Under Condition A, as $n \to \infty$,*

(a) $\hat{\vartheta}_{inf}$ *is asymptotically consistent with* $\hat{\vartheta}_{inf} \xrightarrow{P} \vartheta$;
(b) $\hat{\vartheta}_{inf}$ *is asymptotically normal with* $\sqrt{n}(\hat{\vartheta}_{inf} - \vartheta) \xrightarrow{\mathcal{D}} N\big(0, \Sigma_\vartheta\big)$, *where*

$$\Sigma_\vartheta = (Q_{\tilde{H}^T\tilde{B}}^T Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\tilde{B}})^{-1} Q_{\tilde{H}^T\tilde{B}}^T Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\Lambda\tilde{H}} Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\tilde{B}} (Q_{\tilde{H}^T\tilde{B}}^T Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\tilde{B}})^{-1}.$$

By the definition of $Q_{\tilde{H}^T\tilde{B}}$, $Q_{\tilde{H}^T\tilde{H}}$ and $Q_{\tilde{H}^T\Lambda\tilde{H}}$ given in Condition A.7, if the errors are homoscedastic, i.e., $\sigma_i^2 = \sigma^2$, then the asymptotic covariance admits a simple form $\Sigma_\vartheta = \sigma^2(Q_{\tilde{H}^T\tilde{B}}^T Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\tilde{B}})^{-1}$.

Theorem 2. *Assume that $\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}] \to 0$. Under Condition A, as $n \to \infty$,*

(a) $\hat{\vartheta}$ *is asymptotically consistent with* $\hat{\vartheta} \xrightarrow{P} \vartheta$;
(b) $\hat{\vartheta}$ *is asymptotically normal with* $\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{\mathcal{D}} N\big(0, \Sigma_\vartheta\big)$.

To study the asymptotic property of the estimator $\hat{\alpha}(u)$, we introduce a distance measure to assess its performance. Let $\|\hat{\alpha}(u) - \alpha(u)\|^2 = \big(\hat{\alpha}(u) - \alpha(u)\big)^T\big(\hat{\alpha}(u) - \alpha(u)\big)$. The asymptotic results about $\hat{\alpha}(u)$ are given in the following theorem.

Theorem 3. *Under Conditions A.1–A.7, as $n \to \infty$, we have*

(a) $\sup_{u\in\mathcal{U}} \|\hat{\alpha}(u) - \alpha(u)\| = O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$;
(b) $\Sigma_\alpha(u)^{-\frac{1}{2}}\{\hat{\alpha}(u) - \alpha(u)\} \xrightarrow{\mathcal{D}} N(0, I_p)$ *for* $u \in \mathcal{U}$,
where $\Sigma_\alpha(u) = \mathrm{Cov}\{\tilde{P}(u)(D^TD)^{-1}D^T\varepsilon|U = u\}$ equals $\tilde{P}(u)(D^TD)^{-1}D^T\Lambda D(D^TD)^{-1}\ \tilde{P}(u)^T$. If the errors are homoscedastic, then $\Sigma_\alpha(u) = \sigma^2\tilde{P}(u)(D^TD)^{-1}\tilde{P}(u)^T$.

Theorems 2 and 3 allow us to conduct statistical inference on both ϑ and $\alpha(u)$, provided that consistent estimators of the asymptotic covariance matrices are available.

3.2. Estimation of asymptotic covariance matrices

Under the assumption that the errors are homoscedastic, $E[\varepsilon_i^2|x_i, z_i, u_i] = \sigma^2$, we derive estimators of the asymptotic covariance matrices Σ_ϑ and $\Sigma_\alpha(u)$ for $\hat{\vartheta}$ and $\hat{\alpha}(u)$, and show that these estimators are consistent.

Let $b_i = (w_i^TY, z_i^T)^T$ and note that $x_i^T\hat{\alpha}(u_i) = x_i^T\tilde{P}(u_i)\hat{\gamma} = D_i\hat{\gamma}$. The error term ε_i can be estimated by $\hat{\varepsilon}_i = y_i - b_i\hat{\vartheta} - D_i\hat{\gamma}$. A consistent estimator for σ^2 is given by $\hat{\sigma}^2 = n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}$, where $\hat{\varepsilon} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^T$. The asymptotic covariance matrix Σ_ϑ for $\hat{\vartheta}$ can be estimated by

$$\hat{\Sigma}_\vartheta = \hat{\sigma}^2(\hat{Q}_{\tilde{H}^T\tilde{B}}^T \hat{Q}_{\tilde{H}^T\tilde{H}}^{-1} \hat{Q}_{\tilde{H}^T\tilde{B}})^{-1},$$

where $\hat{Q}_{\tilde{H}^T\tilde{B}} = n^{-1}\hat{H}^T\hat{B} = n^{-1}\sum_{i=1}^n \hat{h}_i\hat{b}_i^T$, $\hat{Q}_{\tilde{H}^T\tilde{H}} = n^{-1}\hat{H}^T\hat{H} = n^{-1}\sum_{i=1}^n \hat{h}_i\hat{h}_i^T$.

Table 1
Summary of Bias, SEE, ESE and CP for $\lambda = -0.5, 0, 0.5$ and $\beta = 3$ and RISE of $\hat{\alpha}(u)$ with different $\sigma^2 = 9, 25$ and $K = 6$ under model (19). Each entry is based on 1000 repetitions.

n	σ^2	$\hat{\sigma}^2$	$\hat{\lambda}$				$\hat{\beta}$				$\hat{\alpha}(u)$
		Avg	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	RISE
						$\lambda = -0.5,\ \beta = 3$					
200	9	9.45	0.0031	0.229	0.140	0.931	−0.0058	0.313	0.204	0.924	0.556
300		9.08	−0.0027	0.120	0.108	0.945	0.0031	0.173	0.162	0.959	0.453
500		9.01	0.0007	0.083	0.082	0.951	−0.0033	0.125	0.123	0.942	0.337
200	25	25.26	0.0003	0.342	0.236	0.933	−0.0223	0.447	0.339	0.921	0.921
300		24.75	−0.0066	0.190	0.180	0.944	0.0005	0.282	0.267	0.959	0.754
500		24.81	−0.0005	0.140	0.137	0.949	−0.0075	0.207	0.205	0.943	0.560
						$\lambda = 0,\ \beta = 3$					
200	9	9.39	0.0008	0.156	0.098	0.930	−0.005	0.316	0.214	0.925	0.554
300		9.08	−0.0023	0.085	0.077	0.943	0.0039	0.183	0.170	0.957	0.453
500		9.01	0.0003	0.059	0.058	0.951	−0.0031	0.131	0.129	0.942	0.336
200	25	25.10	−0.0028	0.226	0.163	0.933	−0.0185	0.454	0.351	0.926	0.919
300		24.77	−0.0057	0.135	0.127	0.944	0.0023	0.297	0.281	0.957	0.754
500		24.81	−0.0009	0.099	0.097	0.948	−0.0066	0.217	0.215	0.942	0.560
						$\lambda = 0.5,\ \beta = 3$					
200	9	9.36	−0.0003	0.081	0.051	0.931	−0.0035	0.329	0.224	0.925	0.556
300		9.09	−0.0014	0.045	0.040	0.943	0.0050	0.194	0.179	0.953	0.453
500		9.01	0.0006	0.031	0.031	0.950	−0.0028	0.138	0.136	0.942	0.337
200	25	25.04	−0.0023	0.116	0.086	0.933	−0.0139	0.474	0.367	0.925	0.920
300		24.79	−0.0035	0.071	0.067	0.943	0.0048	0.313	0.295	0.955	0.755
500		24.82	−0.0008	0.053	0.051	0.946	−0.0056	0.228	0.226	0.944	0.561

The asymptotic covariance matrix $\Sigma_\alpha(u)$ for $\hat{\alpha}(u)$ can be estimated by

$$\hat{\Sigma}_\alpha(u) = \hat{\sigma}^2 \tilde{P}(u)\left\{D^T D\right\}^{-1} \tilde{P}(u)^T.$$

The following theorem establishes the consistency of these estimators.

Theorem 4. *Assume that $E[\varepsilon_i^2 | x_i, z_i, u_i] = \sigma^2$. Under Condition A, as $n \to \infty$, we have*

(a) $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$;
(b) $\hat{\Sigma}_\vartheta \xrightarrow{P} \Sigma_\vartheta$;
(c) $\hat{\Sigma}_\alpha(u) - \Sigma_\alpha(u) = o_p\left(\zeta_0(K)^2/n\right) = o_p(1)$ *for any given value* $u \in \mathcal{U}$.

4. Monte Carlo simulation

In this section, we conduct a simulation study to evaluate finite-sample performance of the proposed method. We consider the following SVMRSAR model

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + z_i^T \beta + x_i^T \alpha(u_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{19}$$

where u_i follows the uniform distribution on [0, 1], x_i follows the standard normal distribution, z_i has the exponential distribution with rate parameter 1, and the error term ε_i is normally distributed with mean zero and variance σ^2. For the sample size n that is a multiple of 10, we take the spatial weights matrix $W = I_{n/10} \otimes (1_{10} - I_{10})/9$, where 1_{10} is a 10×10 matrix with all elements equal to 1 and I_r is a $r \times r$ identity matrix for $r = 10$ and $n/10$. We consider $\alpha(u) = 6sin(2\pi u)$, $\lambda = -0.5, 0,$ or 0.5, and $\beta = -3$ or 3. In the simulations, we take $K = 6$ and $p^K(u) = (1, u, u^2, u^3, \ldots, u^{K-1})^T$. We conducted an additional simulation study using $K = 4$ and $K = 5$ to examine how the estimation accuracy is influenced by the choice of K. The simulation results using $K = 4$ and $K = 5$ presented in the Web-based Supplementary material show that the estimation accuracy is not very sensitive to the choices.

Tables 1 and 2 summarize the simulation results for estimating λ, β and $\alpha(u)$ under different settings of the true parameters for $n = 200, 300,$ and 500 using $K = 6$. Table 1 is for $\beta = 3$ and Table 2 is for $\beta = -3$. Each entry is based on 1000 repetitions.

For each estimator, Bias is the average of estimation biases from 1000 repetitions, SSE is the sample standard error of the estimates, ESE is the average of the estimated standard errors, CP is the coverage probability of a 95 percent confidence interval, and RISE is the average of the square root integrated square error of $\hat{\alpha}(u)$, where for each repetition

Table 2
Summary of Bias, SEE, ESE and CP for $\lambda = -0.5, 0, 0.5$ and $\beta = -3$ and RISE of $\hat{\alpha}(u)$ with $\sigma^2 = 9, 25,$ and $K = 6$ under model (19). Each entry is based on 1000 repetitions.

n	σ^2	$\hat{\sigma}^2$	$\hat{\lambda}$				$\hat{\beta}$				$\hat{\alpha}(u)$
		Avg	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	RISE
						$\lambda = -0.5,\ \beta = -3$					
200	9	9.24	−0.0102	0.193	0.135	0.932	0.0028	0.265	0.201	0.925	0.556
300		9.10	−0.0029	0.125	0.109	0.942	0.0090	0.178	0.162	0.956	0.452
500		9.00	−0.0045	0.083	0.083	0.949	0.0001	0.125	0.123	0.941	0.337
200	25	24.79	−0.0225	0.284	0.224	0.932	0.0048	0.390	0.330	0.928	0.920
300		24.78	−0.0089	0.199	0.181	0.945	0.0160	0.284	0.267	0.958	0.752
500		24.78	−0.0093	0.138	0.139	0.947	0.0017	0.208	0.204	0.940	0.561
						$\lambda = 0,\ \beta = -3$					
200	9	9.27	−0.0082	0.137	0.096	0.929	0.0007	0.277	0.212	0.923	0.554
300		9.10	−0.0024	0.088	0.077	0.943	0.0080	0.186	0.170	0.956	0.452
500		9.00	−0.0034	0.058	0.058	0.950	−0.0007	0.131	0.129	0.943	0.337
200	25	24.86	−0.0182	0.203	0.159	0.932	−0.0001	0.408	0.347	0.927	0.919
300		24.79	−0.0074	0.140	0.128	0.945	0.0130	0.297	0.281	0.955	0.752
500		24.79	−0.0071	0.097	0.097	0.948	−0.0001	0.218	0.215	0.941	0.560
						$\lambda = 0.5,\ \beta = -3$					
200	9	9.30	−0.0049	0.074	0.051	0.930	−0.0015	0.291	0.223	0.922	0.555
300		9.10	−0.0015	0.046	0.041	0.944	0.0069	0.195	0.179	0.954	0.452
500		9.01	−0.0019	0.031	0.031	0.952	−0.0016	0.138	0.136	0.946	0.337
200	25	24.97	−0.0110	0.110	0.085	0.933	−0.0060	0.430	0.365	0.921	0.921
300		24.80	−0.0045	0.074	0.068	0.945	0.0103	0.311	0.296	0.954	0.753
500		24.81	−0.0040	0.051	0.051	0.951	−0.0022	0.229	0.226	0.946	0.561

$$\mathrm{RISE}(\hat{\alpha}(u)) = \left[\sum_{l=1}^{p}\int_{\mathcal{U}}\{\hat{\alpha}_l(u) - \alpha_l(u)\}^2 dF(u)\right]^{1/2},$$

and $\mathcal{U}$ is the support of u, with $F(u)$ the distribution function of u.

The simulation results show that the biases for estimating λ and β are small and decrease when n increases. The variance estimation works well, with $\hat{\sigma}^2$ close to σ^2 and the difference becoming small when n becomes larger. The SEE and ESE for $\hat{\lambda}$ and $\hat{\beta}$ are close and the differences decrease with the sample size. The RISE of $\alpha(u)$ becomes very small as n becomes larger. Fig. 1 compares the true function $\alpha(u)$ with the estimated function $\hat{\alpha}(u)$.

5. Analysis of teenage pregnancy rates

Teenage pregnancy is one of the subject areas where social interaction effects are believed to be most important. Jencks and Mayer (1990) conclude that neighborhoods have a stronger effect on sexual behavior than on cognitive skills, school enrollment decisions, or even criminal activity. Many studies including Hogan and Kitagawa (1985), Crane (1991), Case and Katz (1991) and Evans et al. (1992) analyzed neighborhood effects in teenage pregnancy.

Based on the data from the study "Health and Healthcare in the United States-County and Metro Area Data" (Thomas, 2000), and the 1990 US Census (US Census Bureau, 1992), Lin and Lee (2010) studied the spatial effects at more aggregated levels using the MRSAR model and examined how county teenage pregnancy rates are affected by each other. The proposed model (1) is a semiparametric model that includes a nonparametric component for more flexible modeling. Applying the proposed method, we relate a county's teenage pregnancy rate, which is defined as the percentage of pregnancies occurring for females of 12–17 years old, to those of its neighbors and its own characteristics. Following Kelejian and Robinson (1993), we focus on counties in the 10 Upper Great Plains States, including Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, Nebraska, North Dakota, South Dakota, and Wyoming, which consist of 761 counties. A county's neighbors are referred to as its geographically neighboring counties.

For each county i, $i = 1, \ldots, 761,$ we define $Teen_i$ as the teenage pregnancy rate, Edu_i as the education service expenditure (divided by 100), Inco_i as median household income (divided by 1000), FHH_i as the percentage of female-headed households, Black_i as the proportion of the population that is black, and Phy_i as the number of physicians per 1000 population, all in county i. Lin and Lee (2010) found that Phy_i, Black_i, FHH_i, Inco_i, and Edu_i are five county level covariates that influence the teen pregnancy rates by using the MASAR model where the covariate effects are assumed constant. In the following, we fit the data using the SVMRSAR model with varying intercept $\alpha(u_i)$. We demonstrate how this model can be used to discover nonlinear effects, facilitate transformations and improve model fitting.

We consider the following SVMRSAR model:

$$Teen_i = \lambda \sum_{j=1}^{761} w_{ij} Teen_j + z_i^T\beta + \alpha(u_i) + \varepsilon_i, \tag{20}$$

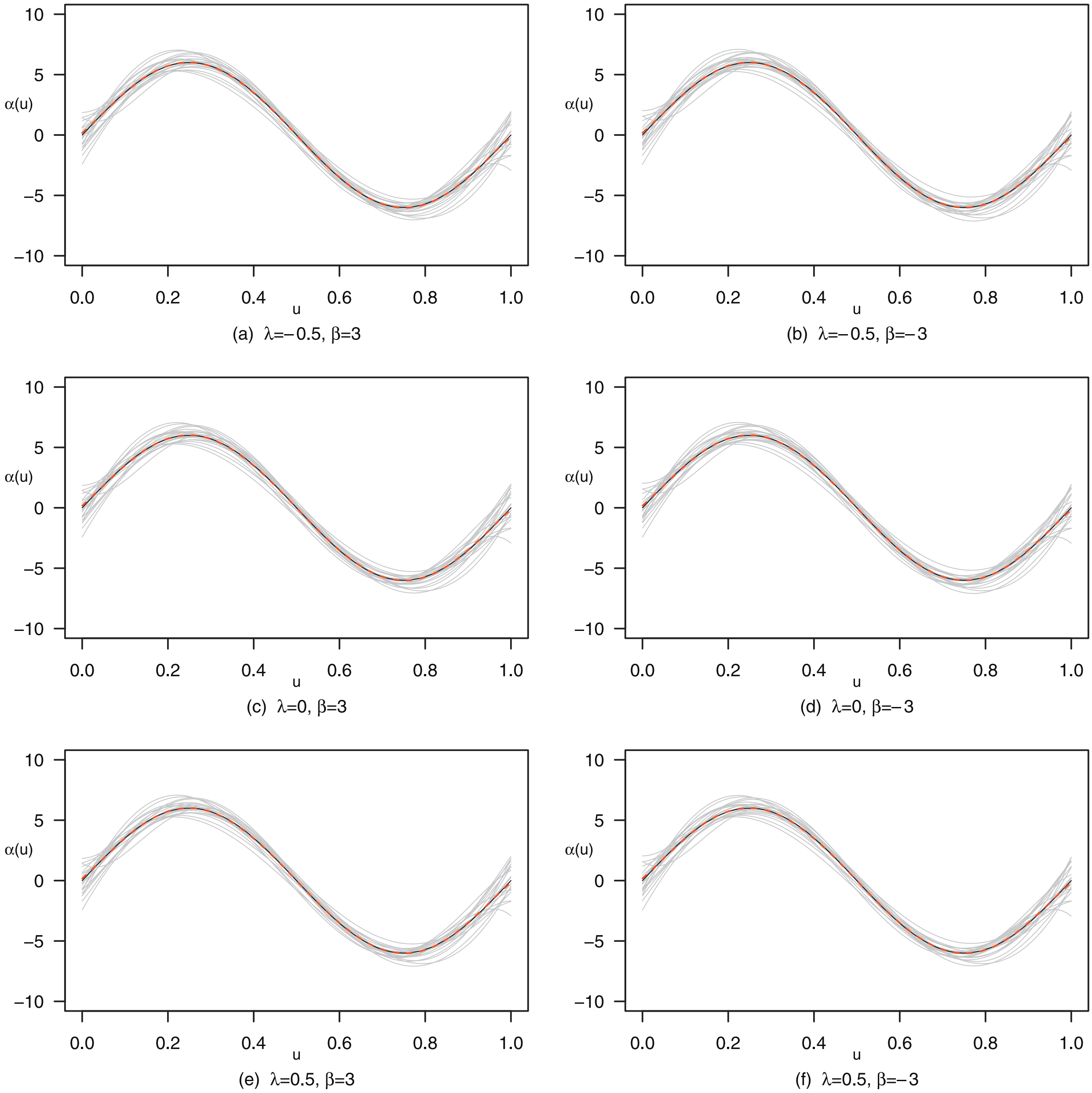

Fig. 1. Plots of the estimates of $\alpha(u) = 6\sin(2\pi u)$ under model (19) with $K = 6$ for $n = 200$ and $\sigma = 5$ and for different values of λ and β. The black dashed line is the true function. The dotted line is the average estimate over 1000 simulations. The grey lines are the estimates of $\alpha(u)$ in 20 simulations.

for $i = 1, \ldots, 761$, where z_i and u_i are the county-level covariates and w_{ij} are the entries in the spatial weights matrix. The spatial weights w_{ij} are set to zero if two counties are not neighboring counties, and all neighbors of the same county are assigned equal weight in the row-normalized spatial weights matrix. The term $\sum_{j=1}^{761} w_{ij} Teen_j$ is simply the average of the teenage pregnancy rates of county i's neighbors.

The histograms of the these variables show that the data distributions for Phy, Black, FHH, and Edu are very skewed with very sparse observations at the right tails. The data distribution of Inco is also slightly skewed. The sparsity of the observations at the right tails makes the nonparametric estimation unstable. We use log transformations for the variables Phy, Black, Inco, Edu and the square root transformation for FHH. The value one is added before the log transformation to avoid $-\infty$. The transformations reduce sparseness in the right tails.

The preliminary analysis indicates that $\log(1+\text{Phy}_i)$, $\log(1+\text{Black}_i)$, $\text{sqrt}(\text{FHH}_i)$ and $\log(1+ \text{Inco}_i)$ have constant spatial effects, while the effect of $\log(1+\text{Edu}_i)$ demonstrates a nonlinear pattern. As a result, we take the final fitted model to be (20), where u_i is $\log(1+\text{Edu}_i)$, and z_i is a vector consisting of $\log(1+\text{Phy}_i)$, $\log(1+\text{Black}_i)$, $\text{sqrt}(\text{FHH}_i)$ and $\log(1+ \text{Inco}_i)$. The estimate of $\alpha(u)$ is given by $\hat{\alpha}(u) = 16.08 + 1.22u - 0.36u^2$. Fig. 2 shows $\hat{\alpha}(u)$ with 95% pointwise confidence bands.

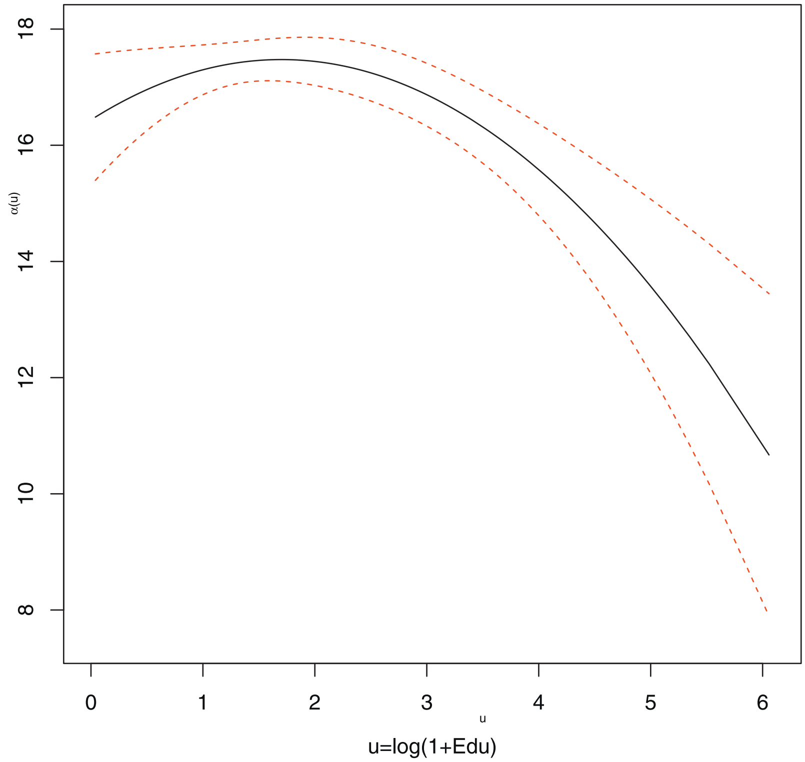

Fig. 2. Plots of the estimates $\hat{\alpha}(u) = 16.44 + 1.22u - 0.36u^2$ and pointwise 95% confidence intervals under model (20), where u_i = log(1+Edu$_i$), and z_i is a vector of log(1+Phy$_i$), log(1+Black$_i$), sqrt(FHH$_i$) and log(1+ Inco$_i$).

The estimation for λ and the spatial effects β of log(1+Phy$_i$), log(1+Black$_i$), sqrt(FHH$_i$) and log(1+ Inco$_i$) are given in the first row of the following table. The second row shows the corresponding Wald statistics, which are the estimated effects divided by their respective estimated standard errors. The values of the test statistics show that all five county level covariates Phy$_i$, Black$_i$, FHH$_i$, Inco$_i$, and Edu$_i$ have significant effects on teen pregnancy rate.

	λ	$\log(1+Phy)$	$\log(1+Black)$	$\sqrt{FHH}$	$\log(1+Inco)$
$\hat{\lambda}$ or $\hat{\beta}$	0.273	−1.14	1.19	4.48	−6.15
Wald-statistic	3.163	−2.07	3.43	10.76	−5.30

Our analysis shows that a higher rate of teen pregnancy is associated with a higher proportion of black population and a higher percentage of female-headed households in the county. On the other hand, the counties with increased accessibility to physicians, higher median household income and higher education service expenditure have lower teen pregnancy rates. Finally, conditional on the above covariates, a county's teen pregnancy rate is statistically significantly geographically affected by the teen pregnancy rates of its neighbors. A county's teen pregnancy rate increases by an estimated 0.273 percent for each percent increase in the average of the teenage pregnancy rates of its neighbors.

6. Concluding remarks

We investigated a semiparametric varying-coefficient mixed regressive spatial autoregressive model that extends the MR-SAR model. Our model can flexibly model covariate effects while allowing spatial dependence. The regression coefficients under our model can be constant and/or vary nonparametrically with another covariate. We proposed a semiparametric series-based least squares estimation procedure that utilizes instrumental variables and series approximations of the conditional expectations. We showed that the proposed estimators of both the parametric and nonparametric components are consistent and asymptotically normal. The estimators for the asymptotic covariances of these estimators are also derived to enable statistical inferences via the model. We conducted a simulation study to investigate the finite-sample performance of the proposed estimators. The simulation showed that the proposed estimators perform well with satisfactory finite-sample performance.

We applied our method to analyze teen pregnancy rates based on data of 761 counties in the 10 Upper Great Plains States from the study "Health and Healthcare in the United States-County and Metro Area Data" (Thomas, 2000), and the

1990 US Census (US Census Bureau, 1992). The model considered counties' geographic spatial dependence and possible nonlinear covariate effects of social and economic factors including a counties education service expenditure, median household income, the percentage of female-headed households, the population proportion that is black, and the number of physicians per 1000 population. Our analysis showed that a higher rate of teen pregnancy is associated with a higher proportion of black population and a higher percentage of female-headed households in the county. On the other hand, counties with increased accessibility to physicians, higher median household income and higher education service expenditure have lower teen pregnancy rates. We also found that a county's teen pregnancy rate is geographically affected by the teen pregnancy rates of its neighbors.

The proposed model can be used to discover the nonlinear nature of covariate effects, via plotting of point and confidence interval estimates for $\alpha(u)$. Formal analysis can be carried out through hypothesis testing of $\alpha(u) \equiv$ constant versus $\alpha(u) \neq$ constant. This hypothesis testing analysis can be implemented by constructing the test statistics as functionals of the integrated estimator $\hat{A}(u) = \int_a^u \hat{\alpha}(v)\,dv$ for $[a,u] \subset \mathcal{U}$; see Gilbert and Sun (2014). The rate of convergence of $\hat{A}(u)$ is expected to be $O_p(n^{-1/2})$. This development requires establishing weak convergence results for the integrated estimator $\hat{A}(u)$ for $[a,u] \in \mathcal{U}$. This is an interesting problem that merits further investigation.

The leave-one-subject-out cross-validation approach proposed in Rice and Silverman (1991) and its modification of the leave-subjects-out approach have been very popular in choosing smoothing parameters. However, selecting K and basis functions is particularly difficult for the SVMRSAR model under study. The cross-validation approach is not readily adaptable to the SVMRSAR model because the response variables y_i, $i = 1, \ldots, n$, are dependent. The problem is that leaving out an observation, say y_k, is equivalent to setting the spacial weight w_{ik} to zero, thus changing the structure of associations among observations. The leave-one-subject-out cross-validation method can be adapted under some structural assumptions on the spatial weights where by leaving-one-subject one also leaves out the neighbors of this subject. This problem needs further investigation.

Acknowledgments

The authors thank the reviewers for their constructive comments that have improved the paper. The authors also thank Professor Xu Lin at Wayne State University for kindly sending us the data set on teen pregnancy and Dr. Peter Gilbert for proofreading the paper. The research of Yanqing Sun was partially supported by the National Institutes of Health NIAID (grant number R37 AI054165), the National Science Foundation (grant number DMS-1513072) and the Reassignment of Duties fund provided by The University of North Carolina at Charlotte. The work of Yuanqing Zhang was partially supported by a grant from National Natural Science Foundation of China (grant number 71371118), Humanity and Social Science Youth foundation of Ministry of Education (grant number 15YJC790150), National Statistical Key Research project of National Bureau of Statistics (grant number 415 2016LZ16). The research of Jianhua Huang was partially supported by National Science Foundation (grant number DMS-1208952).

Appendix A. Proofs

We denote $tr(A) = trace(A)$ for a square matrix A.

Proof of Lemma 1. Proofs of the assertions (i) and (ii) in Lemma 1 are similar to Newey (1997) and thus are omitted. *Proof of assertion (iii).* Denote $E\{p_{lk}(u_i) \times p_{mj}(u_i)\} = \tilde{E}_{lkmj}$. By Condition A.4, we can get $sup_{u \in \mathcal{U}} \|p^K(u)\| \le \zeta_0(K)$ and $\zeta_0(K)^2 K/n \to 0$ as $n \to \infty$. Let $\|\cdot\|$ denote the Euclidean norm. Let $\tilde{E} = E[\tilde{P}(u_i)^T \tilde{P}(u_i)]$ and $\tilde{Q}_1 = \vec{P}^T \vec{P}/n$. We have

$$
\begin{aligned}
E\big[\|\tilde{Q}_1 - \tilde{E}\|^2\big] &= \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K} E\left\{\sum_{i=1}^{n} p_{lk}(u_i) \times p_{mj}(u_i)/n - \tilde{E}_{lkmj}\right\}^2 \\
&= \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K} \frac{1}{n^2}\sum_{i=1}^{n} E\{p_{lk}(u_i) \times p_{mj}(u_i) - \tilde{E}_{lkmj}\}^2.
\end{aligned}
$$

Employing the same line of proof as in Newey (1997, pp. 161–162) and applying Condition A.3 and 4, we obtain

$$
\begin{aligned}
E\big[\|\tilde{Q}_1 - \tilde{E}\|^2\big] &\le \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K} \frac{1}{n^2}\sum_{i=1}^{n} E[\{p_{lk}(u_i) \times p_{mj}(u_i)\}^2] \\
&\le \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K} \frac{1}{n} E[p_{lk}(u_i) \times p_{mj}(u_i)]^2 \\
&\le E\left[\frac{2C^2\zeta_0(K)^2}{n} E\left\{\sum_{m=1}^{p}\sum_{j=1}^{K} \big(p_{mj}(u_i)\big)^2\right\}\right] \\
&\le \frac{2C^2\zeta_0(K)^2 K}{n} \to 0.
\end{aligned}
$$

Hence

$$\|\tilde{Q}_1 - \tilde{E}\|^2 = O_p(\zeta_0(K)\sqrt{K}/\sqrt{n}) = o_p(1). \tag{21}$$

It follows that the smallest eigenvalue of $\tilde{Q}_1$ converges in probability to a positive value τ_1. Let 1_n be the indicator function for the smallest eigenvalue of $\tilde{Q}_1$ being greater than $0.5\tau_1$. Then $\lim_{N\to\infty} P(1_n = 1) = 1$.

For the series-based approximation $E(x_iy_i|u_i) \approx \tilde{P}(u_i)\theta_{1K}$, the estimator of θ_{1K} is given by $\hat{\theta}_{1K} = (\vec{P}^T\vec{P})^{-1}\vec{P}^T\eta_1$, where $\eta_1 = X^TY$. We have $\hat{E}(x_iy_i|u_i) = \tilde{P}(u_i)\hat{\theta}_{1K}$. Denote $g_1 = E(X^TY|U) = (E(x_1^Ty_1|u_1), \dots, E(x_n^Ty_n|u_n))^T$ and $e_{XY} = \eta_1 - g_1$. By Condition A.5, there exist θ_{1K} and $\delta > 0$ such that $n^{-1}(g_1 - \vec{P}\theta_{1K})^T(g_1 - \vec{P}\theta_{1K}) = n^{-1}\sum_{i=1}^n \|E(x_i^Ty_i|u_i) - \tilde{P}(u_i)\theta_{1K}\|^2 = O_p(K^{-2\delta})$.

Let $\vec{Q} = \vec{P}^T\vec{P}$. Consider the decomposition

$$\hat{\theta}_{1K} - \theta_{1K} = \vec{Q}^{-1}\vec{P}^T(\eta_1 - g_1)/n + \vec{Q}^{-1}\vec{P}^T(g_1 - \vec{P}\theta_{1K})/n. \tag{22}$$

The largest eigenvalue of $\vec{Q}^{-1}$ is bounded. Thus, by the triangle inequality,

$$\begin{aligned} 1_n\|\hat{\theta}_{1K} - \theta_{1K}\| &\le 1_n\|\vec{Q}^{-1}\vec{P}^Te_{XY}/n\| + 1_n\|\vec{Q}^{-1}\vec{P}^T(g_1 - \vec{P}\theta_{1K})/n\| \\ &\le 1_n\|\vec{Q}^{-1/2}\vec{P}^Te_{XY}/n\| + 1_n\|\vec{Q}^{-1/2}\vec{P}^T(g_1 - \vec{P}\theta_{1K})/n\|. \end{aligned} \tag{23}$$

By Condition A.6, the largest eigenvalue of $E[e_{XY}e_{XY}^T|U]$ is bounded by a constant C uniformly in n. Note that $1_n\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^T$ is idempotent. We have

$$\begin{aligned} E[1_n\|\vec{Q}^{-1/2}\vec{P}^Te_{XY}/n\|^2|U] &= 1_nE[e_{XY}^T\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^Te_{XY}|U]/n = 1_nE[tr\{\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^Te_{XY}e_{XY}^T\}|U]/n \\ &= 1_ntr\{\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^TE[e_{XY}e_{XY}^T|U]\}/n \le C1_ntr\{\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^T\}]/n \\ &\le pCK/n. \end{aligned}$$

Hence, $1_n\|\vec{Q}^{-1/2}\vec{P}^Te_{XY}/n\| = O_p(\sqrt{K}/\sqrt{n})$ by the Markov inequality.

Next, since $\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^T$ is a projection matrix, we have that

$$\begin{aligned} 1_n\|\vec{Q}^{-1/2}\vec{P}^T(g_1 - \vec{P}\theta_{1K})/n\| &= 1_n[(g_1 - \vec{P}\theta_{1K})^T]\vec{P}(\vec{P}^T\vec{P})^{-1}\vec{P}^T(g_1 - \vec{P}\theta_K)/n]^{1/2} \\ &\le O_p(1)[(g_1 - \vec{P}\theta_{1K})^T(g_1 - \vec{P}\theta_{1K})/n]^{1/2} = O_p(K^{-\delta}). \end{aligned}$$

Combining these results, we obtain that $1_n\|\hat{\theta}_{1K} - \theta_{1K}\| = O_p(\sqrt{K}/\sqrt{n} + K^{-\delta})$.

Now, consider the decomposition

$$\hat{E}(x_iy_i|u_i) - E(x_iy_i|u_i) = \tilde{P}(u_i)(\hat{\theta}_{1K} - \theta_{1K}) - E(x_iy_i|u_i) + \tilde{P}(u_i)\theta_{1K}.$$

By Conditions A.4 and A.5 and the triangle inequality,

$$\begin{aligned} 1_n\|\hat{E}(x_iy_i|u_i) - E(x_iy_i|u_i)\| &\le \|\tilde{P}(u_i)\|\|\hat{\theta}_{1K} - \theta_{1K}\| + \|E(x_iy_i|u_i) - \tilde{P}(u_i)\theta_{1K}\| \\ &\le \zeta_0(K)1_n\|\hat{\theta}_{1K} - \theta_{1K}\| + O(K^{-\delta}) \\ &= O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}]), \end{aligned}$$

uniformly in u_i and n. Since $1_n \xrightarrow{P} 1$ as $n \to \infty$, the assertion (iii) is proved.

Proofs of the assertions (iv) and (v) are similar to the proof of (iii) and thus are omitted.

Proof of Theorem 1. *Proof of Part (a).* By (4) and (5),

$$\begin{aligned} \hat{\vartheta}_{inf} - \vartheta &= \{[\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{B}]^{-1}\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{Y} - \vartheta\} \\ &= \{[\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{B}]^{-1}\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T(\tilde{B}\vartheta + \varepsilon) - \vartheta\} \\ &= [\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\tilde{B}]^{-1}\tilde{B}^T\tilde{H}(\tilde{H}^T\tilde{H})^{-1}\tilde{H}^T\varepsilon \\ &= \left[\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\tilde{B}}{n}\right]^{-1}\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\varepsilon}{n}. \end{aligned} \tag{24}$$

By Condition A.7, $n^{-1}\tilde{H}^T\tilde{B} \xrightarrow{P} Q_{\tilde{H}^T\tilde{B}}$, $n^{-1}\tilde{H}^T\tilde{H} \xrightarrow{P} Q_{\tilde{H}^T\tilde{H}}$. It follows by the nonsingularity conditions given in Condition A.7 that

$$\left[\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\tilde{B}}{n}\right]^{-1}\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1} \xrightarrow{P} [Q_{\tilde{H}^T\tilde{B}}^TQ_{\tilde{H}^T\tilde{H}}^{-1}Q_{\tilde{H}^T\tilde{B}}]^{-1}Q_{\tilde{H}^T\tilde{B}}^TQ_{\tilde{H}^T\tilde{H}}^{-1}.$$

Recall that $\tilde{h}_i$ is an instrumental variable for $\tilde{b}_i$ introduced for model (4) with $E(\tilde{h}_i\varepsilon_i) = 0$. Since $\{\tilde{h}_i\varepsilon_i,\ i = 1, \dots, n\}$ is a sequence of uncorrelated random variables, $\mathrm{Cov}(n^{-1}\sum_{i=1}^n \tilde{h}_i\varepsilon_i) = n^{-2}\sum_{i=1}^n \mathrm{Cov}(\tilde{h}_i\varepsilon_i) = n^{-1}\ \mathrm{Cov}(\tilde{h}_1\varepsilon_1) \to 0$. We have $n^{-1}\tilde{H}^T\varepsilon = n^{-1}\sum_{i=1}^n \tilde{h}_i\varepsilon_i \xrightarrow{P} 0$. Thus, $\hat{\vartheta}_{inf} \xrightarrow{P} \vartheta$. *Proof of Part (b).* By (24),

$$\sqrt{n}(\hat{\vartheta}_{inf} - \vartheta) = \left[\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\tilde{B}}{n}\right]^{-1}\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\varepsilon}{\sqrt{n}}. \tag{25}$$

Moreover,

$$n^{-1}\sum_{i=1}^{n}\mathrm{Cov}(\tilde{h}_i\varepsilon_i|u_i,x_i,Z)=n^{-1}\sum_{i=1}^{n}E(\tilde{h}_i\varepsilon_i{}^2\tilde{h}_i^T|u_i,x_i,Z)=n^{-1}\tilde{H}^T\Lambda\tilde{H}\xrightarrow{P}Q_{\tilde{H}^T\Lambda\tilde{H}}.$$

Applying the central limit theorem (cf. Wooldridge (2010, pp. 95–96) or Kelejian and Prucha (1998, Theorem A.1)), we have $n^{-1/2}\tilde{H}^T\varepsilon\xrightarrow{\mathcal{D}}N(0,Q_{\tilde{H}^T\Lambda\tilde{H}})$. Hence, $\sqrt{n}(\hat{\vartheta}_{inf}-\vartheta)\xrightarrow{\mathcal{D}}N(0,\Sigma_\vartheta)$.

Proof of Theorem 2. *Proof of Part (a).* Note from (4) that $\varepsilon=\tilde{Y}-\tilde{B}\vartheta$. Let $\breve{\varepsilon}=\hat{Y}-\hat{B}\vartheta$. We have

$$\begin{aligned}\hat{\vartheta}-\vartheta&=[\hat{B}^T\hat{H}(\hat{H}^T\hat{H})^{-1}\hat{H}^T\hat{B}]^{-1}\hat{B}^T\hat{H}(\hat{H}^T\hat{H})^{-1}\hat{H}^T\hat{Y}-\vartheta\\&=\left[\frac{\hat{B}^T\hat{H}}{n}\left(\frac{\hat{H}^T\hat{H}}{n}\right)^{-1}\frac{\hat{H}^T\hat{B}}{n}\right]^{-1}\frac{\hat{B}^T\hat{H}}{n}\left(\frac{\hat{H}^T\hat{H}}{n}\right)^{-1}\frac{\hat{H}^T\breve{\varepsilon}}{n}\end{aligned}\tag{26}$$

We shall show that this term is close to $\hat{\vartheta}_{inf}-\vartheta$ given in (24).

By Lemma 1, we have that

$$\hat{E}(x_ix_i^T|u_i)=E(x_ix_i^T|u_i)+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}]),$$
$$\hat{E}(x_iz_i^T|u_i)=E(x_iz_i^T|u_i)+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}]).$$

Note that x_i and z_i are uniformly bounded by Condition A.2. Under Condition A.6 it follows that

$$[\hat{E}(x_ix_i^T|u_i)]^{-1}\hat{E}(x_iz_i^T|u_i)=[E(x_ix_i^T|u_i)]^{-1}E(x_iz_i^T|u_i)+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}]).$$

Therefore $\hat{z}_i=\tilde{z}_i+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])$, uniformly in i. Using the same argument we can show that $\hat{y}_i=\tilde{y}_i+O_p(\zeta_0((K)[\sqrt{K}/\sqrt{n}+K^{-\delta}]),\hat{b}_i=\tilde{b}_i+O_p(\zeta_0((K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])$, and $\hat{h}_i=\tilde{h}_i+O_p(\zeta_0((K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])$, uniformly in i. Moreover,

$$\breve{\varepsilon}_i=\hat{y}_i-\hat{b}_i^T\vartheta=\tilde{y}_i-\tilde{b}_i^T\vartheta+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])=\varepsilon_i+O_p(\zeta_0((K)[\sqrt{K}/\sqrt{n}+K^{-\delta}]).$$

The results in the previous paragraph together with Condition A.7 yields

$$\begin{aligned}\frac{\hat{B}^T\hat{H}}{n}&=\frac{1}{n}\sum_{i=1}^{n}\hat{b}_i\hat{h}_i^T=\frac{1}{n}\sum_{i=1}^{n}\tilde{b}_i\tilde{h}_i^T+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])\xrightarrow{P}Q_{\tilde{H}^T\tilde{B}}^T,\\\frac{\hat{H}^T\hat{H}}{n}&=\frac{1}{n}\sum_{i=1}^{n}\hat{h}_i\hat{h}_i^T=\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_i\tilde{h}_i^T+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])\xrightarrow{P}Q_{\tilde{H}^T\tilde{H}},\\\frac{\hat{H}^T\breve{\varepsilon}}{n}&=\frac{1}{n}\sum_{i=1}^{n}\hat{h}_i\breve{\varepsilon}_i=\frac{1}{n}\sum_{i=1}^{n}\tilde{h}_i\varepsilon_i+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])\xrightarrow{P}0.\end{aligned}\tag{27}$$

It follows from (25)–(27) that

$$\begin{aligned}\hat{\vartheta}-\vartheta&=\left[\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\tilde{B}}{n}\right]^{-1}\frac{\tilde{B}^T\tilde{H}}{n}\left(\frac{\tilde{H}^T\tilde{H}}{n}\right)^{-1}\frac{\tilde{H}^T\varepsilon}{n}+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])\\&=\hat{\vartheta}_{inf}-\vartheta+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])\xrightarrow{P}0.\end{aligned}$$

Proof of Part (b). As shown in Part (a), $\hat{h}_i=\tilde{h}_i+O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])$ uniformly in i. We have

$$\begin{aligned}n^{-1/2}\hat{H}^T\breve{\varepsilon}&=n^{-1/2}\hat{H}^T(\hat{Y}-\hat{B}\vartheta)\\&=n^{-1/2}\hat{H}^T\{\hat{Y}-Y-(\hat{B}-B)\vartheta\}+n^{-1/2}\hat{H}^T(Y-B\vartheta)\\&=n^{-1/2}\hat{H}^T\{\hat{Y}-Y-(\hat{B}-B)\vartheta\}+n^{-1/2}\tilde{H}^T(Y-B\vartheta)+o_p(1).\end{aligned}\tag{28}$$

Let

$$\begin{aligned}\hat{\xi}_i(u_i)&=[\hat{E}(x_ix_i^T|u_i)]^{-1}\left[\hat{E}(x_iy_i|u_i)-\lambda\hat{E}(x_iw_i^TY|u_i)-\hat{E}(x_iz_i^T|u_i)\beta\right]\\&\quad-[E(x_ix_i^T|u_i)]^{-1}\left[E(x_iy_i|u_i)-\lambda E(x_iw_i^TY|u_i)-E(x_iz_i^T|u_i)\beta\right].\end{aligned}\tag{29}$$

Then $\hat{y}_i-\tilde{y}_i-(\hat{b}_i-\tilde{b}_i)\vartheta=x_i^T\hat{\xi}_i(u_i)$. Following the arguments used in proving part (b) of Theorem 1, $\hat{\xi}_i(u_i)=O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n}+K^{-\delta}])=o_p(1)$ uniformly in i. Because the instrumental variable $\tilde{h}_i$ is orthogonal to x_i, $E(\tilde{h}_ix_i^T|u_i)=0$. We have that $n^{-1/2}\sum_{i=1}^{n}\tilde{h}_ix_i^T$ converges in distribution to a normal random vector. It then follows that

$$n^{-1/2}\hat{H}^T\{\hat{Y}-Y-(\hat{B}-B)\vartheta\}=n^{-1/2}\sum_{i=1}^{n}\tilde{h}_ix_i^T\hat{\xi}_i(u_i)=o_p(1).\tag{30}$$

By the proof of Part (b) of Theorem 1,

$$n^{-1/2}\tilde{H}^T(Y - B\vartheta) = n^{-1/2}\tilde{H}^T\varepsilon \xrightarrow{\mathcal{D}} N(0, Q_{\tilde{H}^T\Lambda\tilde{H}}). \tag{31}$$

By (27),

$$\left[\frac{\hat{B}^T\hat{H}}{n}\left(\frac{\hat{H}^T\hat{H}}{n}\right)^{-1}\frac{\hat{H}^T\hat{B}}{n}\right]^{-1}\frac{\hat{B}^T\hat{H}}{n}\left(\frac{\hat{H}^T\hat{H}}{n}\right)^{-1} \xrightarrow{P} \{Q^T_{\tilde{H}^T\tilde{B}}Q^{-1}_{\tilde{H}^T\tilde{H}}Q_{\tilde{H}^T\tilde{B}}\}^{-1}Q^T_{\tilde{H}^T\tilde{B}}Q^{-1}_{\tilde{H}^T\tilde{H}}. \tag{32}$$

It follows from (26), (28), (30), (31), (32) and the Slutsky Theorem that $\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{\mathcal{D}} N\big(0, \Sigma_\vartheta\big)$. This completes the proof of Theorem 2.

Proof of Theorem 3. *Proof of Part (a).* Let $b_i = (w_i^TY, z_i^T)^T$ and $B = (b_1, \ldots, b_n) = (WY, Z)$. By (16),

$$\widehat{\gamma} = (D^TD)^{-1}D^T(Y - B\vartheta) + (D^TD)^{-1}D^TB(\hat{\vartheta} - \vartheta). \tag{33}$$

Note that $D = (D_1, \ldots, D_n)^T$ is an $n \times pK$ matrix. We denote $E\big\{x_{il}p_{lk}(u_i) \times x_{im}p_{mj}(u_i)\big\} = \bar{E}_{lkmj}$. By Condition A.4, $sup_{u\in\mathcal{U}}\|p^K(u)\| \le \zeta_0(K)$ and $\zeta_0(K)^2K/n \to 0$ as $n \to \infty$. Let $\|\cdot\|$ denote the Euclidean norm. Let $\bar{E} = E[D(u_i)^TD(u_i)]$ and $\tilde{Q}_2 = D^TD/n$. We have

$$\begin{aligned}E[\|\tilde{Q}_2 - \bar{E}\|^2] &= \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K}E\{\sum_{i=1}^{n}x_{il}p_{lk}(u_i) \times x_{im}p_{mj}(u_i)/n - \bar{E}_{lkmj}\}^2\\ &= \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}E\{x_{il}p_{lk}(u_i) \times x_{im}p_{mj}(u_i) - \bar{E}_{lkmj}\}^2.\end{aligned}$$

Employing the same line of proof as in Newey (1997, pp. 161–162) and applying Condition A.3 and A.4, we obtain

$$\begin{aligned}E[\|\tilde{Q}_2 - \bar{E}\|^2] &\le \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K}\frac{1}{n^2}\sum_{i=1}^{n}E\{x_{il}p_{lk}(u_i) \times x_{im}p_{mj}(u_i)\}^2\\ &\le \sum_{l=1}^{p}\sum_{k=1}^{K}\sum_{m=1}^{p}\sum_{j=1}^{K}\frac{1}{n}E\{x_{il}p_{lk}(u_i) \times x_{im}p_{mj}(u_i)\}^2\\ &\le E\left[\frac{2C^2\zeta_0(K)^2}{n}E\left\{\sum_{m=1}^{p}\sum_{j=1}^{K}\big(x_{im}p_{mj}(u_i)\big)^2\right\}\right]\\ &\le E\left[\frac{2C^4\zeta_0(K)^2}{n}E\left\{\sum_{m=1}^{p}\sum_{j=1}^{K}\big(p_{mj}(u_i)\big)^2\right\}\right]\\ &\le \frac{2C^4\zeta_0(K)^2K}{n} \to 0.\end{aligned}$$

Hence

$$\|\tilde{Q}_2 - \bar{E}\| = O_p(\zeta_0(K)\sqrt{K}/\sqrt{n}) = o_p(1). \tag{34}$$

It follows from (34) that the largest eigenvalue of $(D^TD/n)^{-1}$ is bounded with probability going to one. We have

$$\|(D^TD)^{-1}D^TB(\hat{\vartheta} - \vartheta)\|^2 = O_p(n^{-1})(\hat{\vartheta} - \vartheta)^TB^TD(D^TD)^{-1}D^TB(\hat{\vartheta} - \vartheta) = O_p(n^{-1}). \tag{35}$$

Note that $v_l(u) = \alpha_l(u) - \sum_{k=1}^{K}\gamma_{lk}p_{kK}(u)$, for $l = 1, \ldots, p$, is the error term in (12). Let $v(u) = (v_1(u), \ldots, v_p(u))^T$. Then $\alpha(u) = \tilde{P}(u)\gamma + v(u)$. Let $V_i = x_i^Tv(u_i)$ and $V = (V_1, \ldots, V_n)^T$. By (13), $y_i - \vartheta^Tb_i = D_i\gamma + V_i + \varepsilon_i$ and thus $Y - B\vartheta = D^T\gamma + V + \varepsilon$. By (33) and (35), we have

$$\hat{\gamma} = \gamma + (D^TD)^{-1}D^T\varepsilon + (D^TD)^{-1}D^TV + O_p(n^{-1/2}). \tag{36}$$

By (34), the smallest eigenvalue of $\tilde{Q}_2$ converges in probability to a positive value τ_2. Let 1_n be the indicator function for the smallest eigenvalue of $\tilde{Q}_2$ being greater than $0.5\tau_2$. Then $\lim_{n\to\infty}P(1_n = 1) = 1$. Since $E[1_n\|(n^{-1}D^TD)^{-\frac{1}{2}}D^T\varepsilon/n\|^2] = E[1_n\varepsilon^TD(D^TD)^{-1}D^T\varepsilon/n] = E[1_ntr\{(D^TD)^{-1/2}\ D^TE(\varepsilon\varepsilon^T|U)D(D^T\ D)^{-1/2}\}/n]$, which is bounded by $O(1)E[1_ntr\{I_{pK}\}/n] = O(K/n)$ by Condition A.3, we have $(n^{-1}D^TD)^{-1/2}\ D^T\varepsilon/n = O_p(\sqrt{K}/\sqrt{n})$. Hence,

$$(D^TD)^{-1}D^T\varepsilon = [(n^{-1}D^TD)^{-1/2}][(n^{-1}D^TD)^{-1/2}D^T\varepsilon/n] = O_p(\sqrt{K}/\sqrt{n}). \tag{37}$$

Under conditions A.2 and A.5, we get $v_l(u_i) = O(K^{-\delta})$ and $V_i = O_p(K^{-\delta})$. Thus, $V = O_p(K^{-\delta})$. Since $1_nD(D^TD)^{-1}D^T$ is idempotent, we have

$$1_n||(D^TD)^{-1}D^TV|| \le O_p(1)1_n[V^TD(D^TD)^{-1}D^TV/n]^{1/2} \le O_p(1)[V^TV/n]^{1/2} = O_p(K^{-\delta}).$$

Hence,

$$(D^TD)^{-1}D^TV = O_p(K^{-\delta}). \tag{38}$$

It follows from (36)–(38) that $\hat{\gamma} = \gamma + O_p(\sqrt{K/n} + K^{-\delta})$. Since

$$\begin{aligned} 1_n||\hat{\alpha}(u) - \alpha(u)|| &\le ||\tilde{P}(u)(\hat{\gamma} - \gamma)|| + ||\tilde{P}(u)\gamma - \alpha(u)|| \\ &\le O_p(1)\zeta_0(K)1_n||\hat{\gamma} - \gamma|| + O(K^{-\delta}) \\ &= O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}]), \end{aligned}$$

we obtain $sup_{u\in U}||\hat{\alpha}(u) - \alpha(u)|| = O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}])$.

Proof of Part (b). Note that at a fixed u, $\hat{\alpha}(u) - \alpha(u) = \tilde{P}(u)(\hat{\gamma} - \gamma) + \tilde{P}(u)\gamma - \alpha(u)$. By Conditions A.4 and A.5, (36) and (38),

$$\begin{aligned} \hat{\alpha}(u) - \alpha(u) &= \tilde{P}(u)(\hat{\gamma} - \gamma) + \tilde{P}(u)\gamma - \alpha(u) \\ &= \tilde{P}(u)(D^TD)^{-1}D^T\varepsilon + O_p\big(\zeta_0(K)[n^{-1/2} + K^{-\delta}]\big). \end{aligned}$$

Recall that $\Sigma_\alpha(u) = \mathrm{Cov}\big[\tilde{P}(u)(D^TD)^{-1}D^T\varepsilon|U = u\big]$. Let c be a p-dimensional column vector whose components are not all zero, and let $\xi_{in} = c^T\tilde{P}(u)(D^TD)^{-1}D_i^T\varepsilon_i/\{c^T\Sigma_\alpha(u)c\}^{1/2}$. We have $E[\xi_{in}] = 0$, $s_n^2 = \sum_{i=1}^n E[\xi_{in}^2] = 1$. In the following we check Lindeberg's condition.

For every $\epsilon > 0$,

$$\frac{1}{s_n^2}\sum_{i=1}^n E[1(|\xi_{in}| > \epsilon s_n)\xi_{in}^2] = \sum_{i=1}^n \epsilon^2 E[1(|\xi_{in}/\epsilon| > 1)(\xi_{in}/\epsilon)^2] \le \sum_{i=1}^n \epsilon^{-2}E(\xi_{in}^4). \tag{39}$$

Note that

$$\sum_{i=1}^n E(\xi_{in}^4|X, Z, U) \le n^{-2}\sum_{i=1}^n \frac{\|c^T\tilde{P}(u)(D^TD/n)^{-1}\|^4}{\{nc^T\Sigma_\alpha(u)c\}^2}E\{\|D_i^T\varepsilon_i\|^4|x_i, z_i, u_i\}. \tag{40}$$

By Conditions A.2–A.4, the largest eigenvalue of $n^{-1}D^TD$ is bounded and the smallest eigenvalue of $n^{-1}D^TD$ is bounded away from zero with probability one. Note that $\Sigma_\alpha(u) = \tilde{P}(u)(D^TD)^{-1}D^T\Lambda D(D^TD)^{-1}\ \tilde{P}(u)^T$. We have $\|c^T\tilde{P}(u)\ (D^TD/n)^{-1}\|^4 = O_p(\{nc^T\Sigma_\alpha(u)c\}^2)$. Hence,

$$\sum_{i=1}^n E(\xi_{in}^4|X, Z, U) \le O_p(1)n^{-2}\sum_{i=1}^n E\{\|D_i^T\varepsilon_i\|^4|x_i, z_i, u_i\}. \tag{41}$$

Under Condition A.2 and A.4, $\|D_i\|^2 \le O(1)(\zeta_0(K))^2$. By Condition A.3, $E[\varepsilon_i^4|x_i, z_i, u_i] \le O(1)$. We have $E\{\|D_i^T\varepsilon_i\|^4\} \le O(1)\zeta_0(K)^2E\{\|D_i\|^2E[\varepsilon_i^4|x_i,\ z_i,\ u_i]\} \le O(1)\zeta_0(K)^2K$. Hence, $\sum_{i=1}^n E(\xi_{in}^4) \le O(1)\zeta_0(K)^2\ K/n \to 0$. By the Lindeberg–Feller central limit theorem, $\sum_{i=1}^n \xi_{in} \to N(0, 1)$ in distribution. The proof of the theorem is completed following an application of the Crémer–Wald device.

Proof of Theorem 4. *Proof of part (a).* Observe that $\hat{\varepsilon} = Y - B\hat{\vartheta} - D\hat{\gamma} = \varepsilon - B(\hat{\vartheta} - \vartheta) - D(\hat{\gamma} - \gamma) - V$, where $B = (WY, Z)$ and V is defined just before (36). We have the following decomposition for $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = n^{-1}\hat{\varepsilon}^T\hat{\varepsilon} = n^{-1}\varepsilon^T\varepsilon + \Delta_n^1 + \Delta_n^2 + \Delta_n^3 + \Delta_n^4 + \Delta_n^5 + \Delta_n^6 + \Delta_n^7 + \Delta_n^8 + \Delta_n^9, \tag{42}$$

where $\Delta_n^1 = (\hat{\vartheta} - \vartheta)^T[n^{-1}B^TB](\hat{\vartheta} - \vartheta)$, $\Delta_n^2 = -2(\hat{\vartheta} - \vartheta)^T[n^{-1}B^T\varepsilon]$, $\Delta_n^3 = (\hat{\gamma} - \gamma)^T[n^{-1}D^TD]\ \ (\hat{\gamma} - \gamma)$, $\Delta_n^4 = -2(\hat{\gamma} - \gamma)^T[n^{-1}D^T\varepsilon]$, $\Delta_n^5 = (\hat{\vartheta} - \vartheta)^T[n^{-1}B^TD](\hat{\gamma} - \gamma)$, $\Delta_n^6 = -2(\hat{\vartheta} - \vartheta)^T[n^{-1}B^TV]$, $\Delta_n^7 = -2(\hat{\gamma} - \gamma)^T[n^{-1}D^TV]$, $\Delta_n^8 = n^{-1}\varepsilon^TV$, and $\Delta_n^9 = n^{-1}V^TV$.

Condition A.3 and Chebyshev's inequality imply $n^{-1}\varepsilon^T\varepsilon \xrightarrow{P} \sigma^2$. The consistency $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ follows by showing that $\Delta_n^j \xrightarrow{P} 0$ for $j = 1, \ldots, 9$.

By Theorem 2 and from the proof of Theorem 3, $\hat{\vartheta} = \vartheta + O_p(n^{-1/2})$, $\hat{\gamma} = \gamma + O_p(\sqrt{K/n} + K^{-\delta})$ and $V = O_p(K^{-\delta})$. Since $Y = \lambda WY + X\vec{\alpha}(U) + Z\beta + \varepsilon$, we have

$$Y = (I - \lambda W)^{-1}X\vec{\alpha}(U) + (I - \lambda W)^{-1}Z\beta + (I - \lambda W)^{-1}\varepsilon. \tag{43}$$

Note that

$$n^{-1}B^TB = \begin{pmatrix} n^{-1}Y^TW^TWY & n^{-1}Y^TW^TZ \\ n^{-1}Z^TWY & n^{-1}Z^TZ \end{pmatrix}$$

Plugging expression (43) for Y into $n^{-1}B^TB$ and by Condition A.1 and A.2, we get $n^{-1}B^TB = O_p(1)$. Thus $\Delta_n^1 = O_p(n^{-1}) = o_p(1)$. Because $n^{-1}B^T\varepsilon \le \sqrt{n^{-1}B^TB}\sqrt{n^{-1}\varepsilon^T\varepsilon}$, we have $n^{-1}B^T\varepsilon = O_p(1)$. Hence $\Delta_n^2 = O_p(n^{-1/2}) = o_p(1)$.

From the proof of Theorem 3, we have $n^{-1}D^TD = O_p(1)$. Hence $\Delta_n^3 = O_p(K/n + K^{-2\delta}) = o_p(1)$. Since $n^{-1}D^T\varepsilon \le \sqrt{n^{-1}D^TD}\sqrt{n^{-1}\varepsilon^T\varepsilon}$, $n^{-1}D^T\varepsilon = O_p(1)$, and $\Delta_n^4 = O_p(\sqrt{K/n} + K^{-\delta}) = o_p(1)$.

Similarly, $\Delta_n^5 = O_p[(\sqrt{K/n} + K^{-\delta})n^{-1/2}] = o_p(1)$ is followed by $\hat{\gamma} = \gamma + O_p(\sqrt{K/n} + K^{-\delta})$, $n^{-1}B^TD \le \sqrt{n^{-1}D^TD}\sqrt{n^{-1}B^TB}$, $n^{-1}D^TD = O_p(1)$ and $n^{-1}B^TD = O_p(1)$. $\Delta_n^6 = O_p(n^{-1/2}) = o_p(1)$ follows from $\hat{\vartheta} = \vartheta + O_p(n^{-1/2})$, $n^{-1}B^TV \le \sqrt{n^{-1}V^TV}\sqrt{n^{-1}B^TB}$, and $n^{-1}B^TV = O_p(1)$. By $n^{-1}D^TV \le \sqrt{n^{-1}V^TV}\sqrt{n^{-1}D^TD}$, which gives $n^{-1}D^TV = O_p(1)$, we have $\Delta_n^7 = O_p(\sqrt{K/n} + K^{-\delta}) = o_p(1)$. It is easy to see that $\Delta_n^8 = n^{-1}\varepsilon^TV = O_p(K^{-\delta}) = o_p(1)$ and $\Delta_n^9 = n^{-1}V^TV = O_p(K^{-2\delta}) = o_p(1)$. This completes the proof of part (a).

Proof of part (b). From the proof of Theorem 2, we have $\hat{Q}_{\tilde{H}^T\Lambda\tilde{H}} \xrightarrow{P} Q_{\tilde{H}^T\Lambda\tilde{H}}$,

$$\hat{Q}_{\tilde{H}^T\tilde{B}}^T = n^{-1}\sum_{i=1}^n \hat{b}_i^T\hat{h}_i = n^{-1}\sum_{i=1}^n \tilde{b}_i^T\tilde{h}_i + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}]) \xrightarrow{P} Q_{\tilde{H}^T\tilde{B}}^T,$$

$$\hat{Q}_{\tilde{H}^T\tilde{H}}^T = n^{-1}\sum_{i=1}^n \hat{h}_i^T\hat{h}_i = n^{-1}\sum_{i=1}^n \tilde{h}_i^T\tilde{h}_i + O_p(\zeta_0(K)[\sqrt{K}/\sqrt{n} + K^{-\delta}]) \xrightarrow{P} Q_{\tilde{H}^T\tilde{H}}.$$

Hence,

$$(\hat{Q}_{\tilde{H}^T\tilde{B}}^T\hat{Q}_{\tilde{H}^T\tilde{H}}^{-1}\hat{Q}_{\tilde{H}^T\tilde{B}})^{-1} \xrightarrow{P} (Q_{\tilde{H}^T\tilde{B}}^T Q_{\tilde{H}^T\tilde{H}}^{-1} Q_{\tilde{H}^T\tilde{B}})^{-1}.$$

It follows that $\hat{\Sigma}_\vartheta \xrightarrow{P} \Sigma_\vartheta$.

Proof of part (c). By Theorem 4 (a), $\hat{\sigma}^2 - \sigma^2 = o_p(1)$. We have

$$\begin{aligned} n(\hat{\Sigma}_\alpha(u) - \Sigma_\alpha(u)) &= (\hat{\sigma}^2 - \sigma^2)\tilde{P}(u)\left\{D^TD/n\right\}^{-1}\tilde{P}(u)^T \\ &= o_p(1)\tilde{P}(u)O_p(1)\tilde{P}(u)^T = \zeta_0(K)^2 o_p(1). \end{aligned}$$

Thus, $n(\hat{\Sigma}_\alpha(u) - \Sigma_\alpha(u))/\zeta_0(K)^2 = o_p(1)$.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ecosta.2017.05.005.

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71 (6), 1795–1843.
Cameron, A.C., Trivedi, P.K., 2008. Microeconometrics: Methods and Applications. Cambridge University Press, New York.
Case, A.C., Katz, L.F., 1991. The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths. Technical Report NBER working paper no. w3705. National Bureau of Economic Research, Cambridge, MA.
Chen, X., Pouzo, D., 2009. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. J. Econom. 152 (1), 46–60.
Crane, J., 1991. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. Am. J. Sociol. 96 (5), 1226–1259.
Evans, W.N., Oates, W.E., Schwab, R.M., 1992. Measuring peer group effects: a study of teenage behavior. J. Polit. Econ. 100 (5), 966–991.
Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11 (6), 1031–1057.
Gelfand, A.E., Kim, H.-J., Sirmans, C., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. J. Am. Stat. Assoc. 98 (462), 387–396.
Gilbert, P., Sun, Y., 2014. Testing for vaccine efficacy against a spectrum of pathogen sequences in stratified mark-specific proportional hazards models with missing marks, with application to the RV144 HIV vaccine efficacy trial. J. R. Stat. Soc. Ser. C 64, 49–73.
Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B (Methodol.) 55, 757–796.
Hogan, D.P., Kitagawa, E.M., 1985. The impact of social status, family structure, and neighborhood on the fertility of black adolescents. Am. J. Sociol. 90 (4), 825–855.
Jencks, C., Mayer, S.E., 1990. Thesocial consequences of growing up in a poor neighborhood. Inner-City Poverty in the United States, 111, p. 186.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Finance Econ. 17 (1), 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Int. Econ. Rev. 40 (2), 509–533.
Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. J. Econom. 104 (2), 219–257.
Kelejian, H.H., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Pap. Reg. Sci. 72 (3), 297–312.
Lee, L.-F., 2001. Generalized Methodof Moments Estimation of Spatial Autoregressive Processes. Manuscript. Department of Economics, Ohio State University.
Lee, L.-F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econom. Rev. 22 (4), 307–335.
Lee, L.-F., 2007. Gmm and 2SLS estimation of mixed regressive, spatial autoregressive models. J. Econom. 137 (2), 489–514.
Li, Q., Huang, C.J., Li, D., Fu, T.-T., 2002. Semiparametric smooth coefficient models. J. Bus. Econ. Stat. 20 (3), 412–422.
Lin, X., Lee, L.-f., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. J. Econom. 157 (1), 34–52.
Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. J. Econom. 79 (1), 147–168.
Rice, J.A., Silverman, B.W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. J. R. Stat. Soc. Ser B (Methodol.) 10, 233–243.
Schliep, E., Gelfand, A., Holland, D., 2015. Autoregressive spatially varying coefficients model for predicting daily PM 2.5 using VIIRS satellite AOT. Adv. Stat. Climatol. Meteorol. Oceanogr. 1 (1), 59.
Su, L., Jin, S., 2010. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. J. Econom. 157 (1), 18–33.
Thomas, R.K., 2000. Health and healthcare in the united states: county and metro area data.. J. Healthc. Qual. 22 (4), 47.
US Census Bureau, 1992. Census of Population and Housing 1990, Summary Tape File 3 on CD-ROM. The Bureau Producer and Distributor, Washington, DC.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts.
Zhang, W., Lee, S.-Y., Song, X., 2002. Local polynomial fitting in semivarying coefficient model. J. Multivar. Anal. 82 (1), 166–188.
Zhang, Y., Sun, Y., 2015. Estimation of partially specified dynamic spatial panel data models with fixed-effects. Reg. Sci. Urban Econ. 51, 37–46.