Ann Inst Stat Math (2018) 70:745–759
https://doi.org/10.1007/s10463-017-0607-z

CrossMark

Asymptotic theory for varying coefficient regression models with dependent data

Soutir Bandyopadhyay[1] · Arnab Maity[2]

Received: 1 December 2015 / Revised: 21 January 2017 / Published online: 11 March 2017

Abstract The varying coefficient models (VCMs) are extremely important tools in the statistical literature and are widely used in many subject areas for data modeling and exploration. In linear VCMs, typically the errors are assumed to be independent. However, in many situations, especially in spatial or spatiotemporal settings, this is not a viable assumption. In this article, we consider nonparametric VCMs with a general dependent error structure which allows for both spatially autoregressive and spatial moving average models as special cases. We investigate asymptotic properties of local polynomial estimators of the model components. Specifically, we show that the estimates of the unknown functions and their derivatives are consistent and asymptotically normally distributed. We show that the rate of convergence and the asymptotic covariance matrix depend on the error dependence structure and we derive the explicit formula for the convergence results.

SB's research is partially supported by NSF DMS-1406622. AM's research is partially supported by Award Number R00ES017744 from the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-017-0607-z) contains supplementary material, which is available to authorized users.

✉ Soutir Bandyopadhyay
sob210@lehigh.edu

Arnab Maity
amaity@ncsu.edu

1 Department of Mathematics, Lehigh University, 14 E. Packer Avenue, Bethlehem, PA 18015, USA

2 Department of Statistics, North Carolina State University, 2311 Stinson Drive, Campus Box 8203, Raleigh, NC 27695, USA

Keywords Kernel smoothing · Local polynomial estimation · Spatially autoregressive errors · Varying coefficient model

1 Introduction

Varying coefficient models (VCMs) have been proven to be very useful statistical tools for flexible regression modeling in many areas of study. These could be seen as a natural extension of typical parametric models so that the models are linear in the regressors, but their coefficients are allowed to vary smoothly over another covariate (Ahmad et al. 2005). VCMs are also generalizations of nonparametric regression models in the sense that one can model multiple covariates with corresponding nonparametric coefficients depending on another covariate.

There is a rich existing literature on VCMs. Fan and Zhang (2008) provide a detailed review of VCMs and their theoretical development. However, most of the existing literature depends on two assumptions. First, the errors in the regression model are independent and identically distributed (iid). Second, the response–covariate pairs are also iid across the sampling units. However, in many applications, we observe dependent data and our main objective lies in developing regression models which can be used for prediction purpose. While doing the analysis of dependent data, usually the strict stationarity assumption on the process is assumed to be true, but in reality this assumption rarely holds. While working with a Gaussian random field, we can relax the strict stationarity assumption. In such situations, we only need the mean of the random field to be constant over the whole region and the covariance function to depend only on the locations, see Cressie (1993) and Stein (1999) for more details. Though, in general we cannot relax such an assumption.

To handle the underlying nonstationarity of a time series, Tran et al. (1996) have considered linear nonparametric regression estimators for fixed design. In contrast to the usual assumption of iid residuals, they have assumed stationary dependent residuals. Moreover, no mixing condition is imposed on the dependence structure. Robinson (1989) and Cai (2007) study a time-varying coefficient time series model with a time trend function and serially correlated errors to characterize the nonlinearity, nonstationarity and trending phenomenon. In Robinson (1989), a Nadaraya–Watson-type estimator is developed to estimate the time trend and coefficient functions, whereas Cai (2007) considered a more general local polynomial approach. However, the results are proved under the assumption that the time points $t_i = i/n$, for $i = 1, \ldots, n$, and hence, the increasingly intense sampling of data points derives the consistent estimation in both Robinson (1989) and Cai (2007). Spatially, VCMs have already been considered before (see, for example, Aykroyd 1998; Aykroyd and Zimeras 1999; Dreesman and Tutz 2001; Higdon et al. 1997 and Johnson et al. 1991). However, all of these works have either considered completely parametric structures for the coefficients or the observations are equally spaced on a regular lattice.

Robinson (2011) considers a general error structure which assumes that the errors are, up to a random scalar, generated as a linear process of independent innovations that are independent of the regressors. This formulation enables us to model both lattice linear autoregressive moving average and spatially autoregressive (SAR) models. As in Robinson (2011), this framework also allows us for a form of strong dependence which is analogous to the long-range dependent time series. As a special case, one could consider the situation where the errors are generated on irregularly spaced locations over the whole region and follow a SAR model as discussed in Arbia (2006), Cliff and Ord (1981), Kelejian and Prucha (1999), Lee (2002), Robinson (2011), and the references therein. These models can be thought of as higher-dimensional extensions of the time series autoregressive models. This class of models allows us to express n spatial observations as linear transformations of n iid unobservable random variables. Also, the $n \times n$ transformation matrix is usually known apart from finitely many unknown parameters; see Sect. 2.1 for a detailed discussion on the generality of this error structure.

In this article, we study a nonparametric VCM with the previously mentioned general *error* structure which is designed to include various kinds of dependent data. Our work is a further generalization of Robinson (2011). Robinson (2011) considers a standard regression setup, whereas in our work we consider a more general VCM where the response Y not only depends on L regressor variables $X_1, \ldots, X_L$ but it also depends on another covariate Z modifying the effects of X. We consider a nonparametric VCM mainly due to the fact that in dependent data analysis, the functional form is often nonlinear and cannot be described with a specific nonlinear function. In a spatial context, several works have been done on the nonparametric regression on integer lattice points (see, Hallin et al. 2004; Tran and Yakowitz 1993, and more recently, Lu et al. 2014 for spatial quantile regression with varying coefficients). Among all smoothing techniques, the Nadaraya–Watson method is probably the most standard one and it has been well documented. However, sometimes it suffers from several severe drawbacks, such as poor boundary performances, excessive bias and low efficiency. Because of these drawbacks, the local polynomial fitting methods are generally preferable (see, Fan and Gijbels 1996). In recent years, the local polynomial methods have become increasingly popular; see Fan (1992), Fan and Gijbels (1996), Loader (1999), and Ruppert and Wand (1994) for more details. In this article, we extend this approach to the context of nonparametric VCM for dependent data by defining an estimator based on local polynomial regression. Recently, Sun et al. (2014) proposed a semiparametric spatial dynamic model in order to extend the ordinary SAR models to accommodate the effects of covariates. Their model incorporates the SAR structure directly through the responses while keeping the errors iid, whereas in this work we assume the errors are generated as a linear process in independent innovations that are independent of the regressors.

Lastly, it is also worth mentioning that throughout the whole article we will only consider the *increasing domain* asymptotic framework. In contrast to the ordinary time series case where observations are usually taken at a regular interval of time and asymptotics is driven by the unidirectional flow of time, for random processes observed over space, several different types of spatial sampling designs and asymptotic structures are relevant for practical applications, for example *increasing domain* and *infill* asymptotic structure (for more detail, see Bandyopadhyay and Lahiri 2010). It has been noted that the large sample behaviors of many standard inference procedures under the infill asymptotics are noticeably different from what can be obtained under the increasing domain asymptotic frameworks; see, for example, Cressie (1993), Lahiri

(1996), Loh (2005), Stein (1999), Ying (1993) and the references therein. As a result, the case of infill asymptotics is not considered here and we concentrate only on the increasing domain asymptotic structure.

The rest of the article is organized as follows. In Sect. 2, we introduce the nonparametric VCM under a more general framework and discuss the estimation procedure for the coefficients. The main asymptotic results are presented in Sect. 3, together with simulation studies in Sect. 4. Finally, we give outlines of the proofs in Sect. 5. A complete discussion on the regularity conditions and all supporting lemmas used to prove the main results are relegated to the Supplementary material.

2 Nonparametric varying coefficient model (VCM)

2.1 Model specification

Suppose we are given observations at n locations on L regressor variables $X_1, \dots, X_L$ and the response variable Y. Based on this, let us define the following VCM,

$$\begin{aligned} Y_i &= X_{i1}\theta_1(Z_i) + \cdots + X_{iL}\theta_L(Z_i) + \xi_i, \; i = 1, \dots, n, \;\; n \geq 1 \\ &= X_i'\theta(Z_i) + \xi_i, \end{aligned}$$

where $X_i = [X_{i1}, \dots, X_{iL}]'$ is a $L \times 1$ vector, $\theta(z) = [\theta_1(z), \dots, \theta_L(z)]'$ is a vector of unknown functions, Z_i is another covariate modifying the effects of X_i and ξ_i's are the random errors. Note that one can have $X_1 \equiv 1$ if an intercept function is included in the model. As described in Sect. 1 for the random errors, we assume similar structure as in Robinson (2011). In particular, let us assume

$$\xi_i = \sigma_i(X_i, Z_i)e_i, \quad 1 \leq i \leq n,$$

where for all $n \geq 1$, $\{e_i, 1 \leq i \leq n\}$ is independent of $\{(X_i, Z_i), 1 \leq i \leq n\}$ and the first and the second moments of $\sigma_i(X_i, Z_i)$ exist. We model the dependence across i via e_i similar to Robinson (2011). Let us assume

$$e_i = \sum_{j=-\infty}^{\infty} \alpha_{ij}\epsilon_j, \tag{1}$$

where for each n, the $\epsilon_j, \; j \geq 1$ are independent random variables with zero mean and finite variance, the nonstochastic weights α_{ij}'s are at least square summable over j, where, without loss of generality, we fix $\sum_{j=-\infty}^{\infty} \alpha_{ij}^2 = 1$ for $1 \leq i \leq n, n = 1, 2, \dots$. We also assume that $var(\epsilon_i) = 1$, which implies $var(e_i) = 1$. Note that $\sigma_i^2(X_i, Z_i)$ are unknown functions and that $var(e_i) = 1$ implies $var(Y_i|X_i, Z_i) = var(\xi_i|X_i, Z_i) = \sigma_i^2(X_i, Z_i)$. Moreover,

$$E(e_i) = 0, \;\; 1 \leq i \leq n, \tag{2}$$

and an immediate consequence of (2) is $E(\xi_i|X_i, Z_i) = 0,\ 1 \le i \le n$. From the above construction, it can be observed that in this work we permit conditional heteroscedasticity. We do not assume that $\sigma_i^2(X_i, Z_i)$ are constant across i; thus, we are also allowing unconditional heteroscedasticity.

This particular structure as given in (1) can include various forms of spatial dependence and heterogeneity in the unobserved errors ξ_i, which are of interest in different economic and statistical applications. It is worth mentioning that our formulation usually does not require $\sum_{j=-\infty}^{\infty} \alpha_{ij} < \infty$ and hence covers forms of long-range dependence. This formulation also covers the case of equally spaced time series data ($\alpha_{ij} = \alpha_{|i-j|}$) as well as the lattice extension to the model. Condition (1) can also be thought of as an extension of SAR models in the sense that one can start with the parametric structure $\left(I_n - \sum_{k=1}^{\ell_1} \delta_k W_k\right)\xi = \left(I_n - \sum_{k=1}^{\ell_2} \delta_{\ell_1+k} W_{\ell_1+k}\right)\sigma\epsilon$, where the integers ℓ_1, ℓ_2 are given, I_n is the $n \times n$ identity matrix, $\xi = (\xi_1, \dots, \xi_n)'$, $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$, the δ's are known scalars, σ is an unknown scale factor, and W_k's are given $n \times n$ weight matrices satisfying further conditions in order to guarantee identifiability of the δ's. For further details about this general framework on the errors, see Robinson (2011) and Robinson and Thawornkaiwong (2011) which consider similar structure of e_i and give detailed motivation for using this structural form.

For all quantities ($Y_i, X_i, Z_i, \xi_i, \sigma_i, e_i, \alpha_{ij}$, and, ϵ_i) described above, we allow them to admit a triangular array structure throughout this work. However, we will suppress the n subscript to avoid notational complications. As noted in Robinson (2011), the triangular array framework includes the case when we need to re-label observations as n increases in lattice data or panel data.

2.2 Estimation

Let us consider the local polynomial estimator of $\theta(\cdot)$. Specifically, given z, we assume the following local expansion holds:

$$\theta_k(Z_i) \approx \theta_k(z) + \sum_{j=1}^{p} (Z_i - z)^j \theta_k^{(j)}(z)/j!,\ k = 1, \dots, L,$$

where $\theta_k^{(j)}(\cdot)$ denotes the jth derivative of function $\theta_k(\cdot)$ for $j = 1, \dots, p$.

Suppose that K is a kernel function and $h \equiv h_n$ denotes a positive scalar bandwidth sequence satisfying $h + (nh)^{-1} \to 0$ as $n \to \infty$. Define $G(t) = [I_L, (t/h)I_L, \dots, \{(t/h)^p/p!\}I_L]'$, where I_L is a $L \times L$ identity matrix. Given z, let us also define

$$\gamma_0(z) = \left[\theta_1(z), \dots, \theta_L(z), \dots, h^p\theta_1^{(p)}(z), \dots, h^p\theta_L^{(p)}(z)\right]'.$$

Thus, we can locally approximate $\theta(Z_i) \approx G(Z_i - z)'\gamma_0(z)$ and the local polynomial kernel-based estimate minimizes

$$\sum_{i=1}^{n} K\{(Z_i - z)/h\}\{Y_i - X_i' G(Z_i - z)' \gamma(z)\}^2$$

with respect to $\gamma(z)$. Equivalently, one solves for $\gamma(z)$

$$0 = \sum_{i=1}^{n} K\{(Z_i - z)/h\} G(Z_i - z) X_i \{Y_i - X_i' G(Z_i - z)' \gamma(z)\}. \tag{3}$$

The solution of (3) has the following closed form:

$$\widehat{\gamma}(z) = \left[\sum_{i=1}^{n} K\{(Z_i - z)/h\} G(Z_i - z) X_i X_i' G(Z_i - z)'\right]^{-1} \times \sum_{i=1}^{n} K\{(Z_i - z)/h\} G(Z_i - z) X_i Y_i.$$

Clearly, $\widehat{\theta}(z)$ is given by the first L elements of $\widehat{\gamma}(z)$.

3 Main results

In this section, we present the main asymptotic results regarding consistency and asymptotic normality of the estimated functions. Here, we impose the following conditions on the regression function.

(R1) For every $j = 1, \ldots, L$, $\theta_j(\cdot)$ has p derivatives.

(R2) For every $j = 1, \ldots, L$, $\theta_j^{(p)}(\cdot)$ satisfies a Lipschitz condition of degree $q \in (0, 1]$ in a neighborhood of z.

To maintain brevity, the other regularity conditions required to prove the results are given in Supplementary Materials (Section A).

Define $f_i(z)$ to be the marginal density function of Z_i, $f_{ij}(z_1, z_2)$ to be the density function of (Z_i, Z_j), $f_{ijk}(z_1, z_2, z_3)$ to be the density function of (Z_i, Z_j, Z_k) and $f_{ijk\ell}(z_1, z_2, z_3, z_4)$ to be the density function of (Z_i, Z_j, Z_k, Z_ℓ).

Define, for any $s, s_1, s_2 = 1, \ldots, L$,

$$\begin{aligned}
m_{i,s_1 s_2}(z) &= E\{X_{is_1} X_{is_2} | Z_i = z\} f_i(z), \\
m_{ij,s_1 s_2}(z_1, z_2) &= E\{X_{is_1} X_{js_2} | Z_i = z_1, Z_j = z_2\} f_{ij}(z_1, z_2), \\
m^*_{i,s}(z) &= E\{\sigma_i^2(X_i, Z_i) X_{is}^2 | Z_i = z\} f_i(z), \\
m^*_{i,s_1 s_2}(z) &= E\{\sigma_i^2(X_i, Z_i) X_{is_1} X_{is_2} | Z_i = z\} f_i(z), \\
m^*_{ij,s_1 s_2}(z_1, z_2) &= E\{\sigma_i(X_i, Z_i) \sigma_j(X_j, Z_j) X_{is_1} X_{js_2} | Z_i = z_1, Z_j = z_2\} f_{ij}(z_1, z_2).
\end{aligned}$$

Given z, define the $L \times L$ matrix $M_X(z)$ such that its (j, k)th element is given by $M_{X,jk}(z) = n^{-1} \sum_{i=1}^{n} m_{i,jk}(z)$ for $j, k = 1, \ldots, L$. Also, for a kernel function $K(\cdot)$, define

$$\kappa_r = \int w^r K(w) dw, \text{ and} \tag{4}$$

$$\nu_r = \int w^r K^2(w) dw. \tag{5}$$

First, we establish consistency of $\widehat{\gamma}(z)$.

Theorem 1 *Under* (R1),(R2)*, and the assumptions stated in Section A of Supplementary Materials,*

$$\widehat{\gamma}(z) - \gamma_0(z) \rightarrow^p 0$$

for any given z.

The proof is provided in Sect 5.1. Note that the result above not only ensures consistency of the estimates of the unknown functions but also the consistency of their (scaled) derivatives.

Next we show asymptotic normality of the estimators. From (D7) in Section A of Supplementary Materials note that we set $t_n = |n^{-2} \sum_{i \neq j=1}^n \beta_{ij}|$, and $s_n = (nh)^{-1}$ to define the scaling sequence

$$c_n = \begin{cases} nh & \text{if } \; t_n/s_n \rightarrow c \in [0, \infty), \\ n^2 \Big/ \left| \sum_{i \neq j=1}^n \beta_{ij} \right| & \text{if } \; t_n/s_n \rightarrow \infty, \end{cases}$$

with c being a positive constant or zero. This definition of scaling sequence is also used in Robinson (2011) in the context of nonparametric regression.

Define

$$\mathcal{B}_n = \left[\sum_{i=1}^n K\{(Z_i - z)/h\} G(Z_i - z) X_i X_i' G(Z_i - z)' \right]^{-1} \times \left[\sum_{i=1}^n K\{(Z_i - z)/h\} G(Z_i - z) X_i X_i' \{\theta(Z_i) - G(Z_i - z)' \gamma_0(z)\} \right].$$

Then, we have the following result.

Theorem 2 *Under* (R1),(R2)*, and the assumptions stated in Section A of Supplementary Materials, for any given z, entry-wise,*

$$c_n^{1/2} \left(\mathcal{I}^{-1} \Sigma \mathcal{I}^{-1} \right)^{-1/2} (\widehat{\gamma}(z) - \gamma_0(z) - \mathcal{B}_n) \rightarrow^d N(0, I),$$

where $\mathcal{B}_n = O(h^{p+q})$ *is a bias term for* p, q *as defined in* (R1), (R2), $\Sigma = \Psi$ *if* $s_n/t_n \rightarrow 0$, $\Lambda + c\Psi$ *if* $t_n/s_n \rightarrow c$*, where c is a positive constant or zero. The matrices*

Ψ and Λ are defined in Section A [see, (A.4)] and $\mathcal{I}$ is a $L(p+1) \times L(p+1)$ matrix given by

$$\mathcal{I} = \begin{bmatrix} M_X(z) & \kappa_1 M_X(z)/1! & \dots & \kappa_p M_X(z)/p! \\ \kappa_1 M_X(z)/1! & \kappa_2 M_X(z)/2! & \dots & \kappa_{p+1} M_X(z)/(p+1)! \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_p M_X(z)/p! & \kappa_{p+1} M_X(z)/(p+1)! & \dots & \kappa_{2p} M_X(z)/2p! \end{bmatrix}. \quad (6)$$

Remark 1 It is worth mentioning that if we assume the errors are generated as a linear process as discussed in Sect. 2.1, then the resulting local polynomial estimator $\widehat{\gamma}(z)$ given in Sect. 2.2 is no longer an efficient estimator of $\gamma_0(z)$. In this work to find the local polynomial estimator, we have considered the least-squared approach rather than the likelihood based approach of getting an efficient estimator. Indeed, by incorporating the dependence structure of the errors in the estimation procedure (perhaps using weighted local likelihood estimator rather than a simple local likelihood estimator) may provide more efficiency. This is an important issue from statistical perspective with our assumed model though a full discussion of an efficient estimator is beyond the scope of this article.

Remark 2 It is interesting to note that when $t_n/s_n \to$ constant, the convergence rate in Theorem 2 is $(nh)^{1/2}$. This is the standard convergence rate of function estimates in nonparametric VCM literature when the errors are independent. This observation also matches with the convergence rate obtained in Robinson (2011) in the context of nonparametric regression.

It is possible to obtain an asymptotic expansion of $\widehat{\theta_j}(z) - \theta_j(z)$ with the corresponding expressions of bias terms. For example, let us consider the widely used local linear estimators, that is, $p = 1$. Also suppose that instead of assumptions (R1) and (R2), we impose the condition:

(R') For every $j = 1, \dots, L$, $\theta_j(\cdot)$ has the second partial derivatives and $\theta_j^{(2)}(\cdot)$ is uniformly bounded by a finite constant.

Then, it can be shown that the first L elements of A_2 (see Sect. 5.1) can be written as

$$A_2^* = h^2 b(z) + o_p(h^2)$$

for some function $b(z)$. Also, using (10) we see that there exists an $N > 0$ such that the first L elements of A_1 (see Sect. 5.1) can be written as

$$A_1^* = (nh)^{-1} \sum_{i=1}^{n} K\{(Z_i - z)/h\} \sigma_i(X_i, Z_i) G(Z_i - z) X_i \xi_i^* + o_p\left(c_n^{-1/2}\right),$$

where, $\xi_i^* = \sum_{j=1}^{N} \alpha_{ij}\epsilon_j$. Thus, we have

$$\widehat{\theta}(z) - \theta(z) = h^2 b(z) + (nh)^{-1} M_X^{-1}(z) \sum_{i=1}^{n} K\{(Z_i - z)/h\}\sigma_i(X_i, Z_i)X_i\xi_i^* \\ + o_p(h^2) + o_p\left(c_n^{-1/2}\right),$$

where we use the symmetry of the kernel function around 0 (see Assumption (K) in Section A of Supplementary Materials) to derive $\kappa_1 = 0$, thus implying $\mathcal{I}$ is a block diagonal matrix.

Based on the asymptotic expansion, it is then possible to obtain an optimal choice of bandwidth. For example, when $t_n/s_n \to$ constant, then $c_n = nh$. Thus, the mean square error (MSE) is $O_p(h^4 + (nh)^{-1})$. Minimizing this MSE over h gives us $h = O_p(n^{-1/5})$. However, we note that the optimal bandwidth still would depend on the typically unknown dependence structure. Bandwidth selection in such spatially dependent error models is still a challenging open problem and as such is out of the scope of the current article.

Remark 3 Note that according to our assumptions (H1) and (W2) (see Supplementary Materials) both s_n and t_n converge to 0 as $n \to \infty$. Also depending on the relative growths of s_n and t_n, we can relate to the short- and long-range dependence of the underlying process. We get $t_n/s_n \to c \in [0, \infty)$, i.e., t_n decays faster than s_n when ξ_i has a short-range dependence. On the other hand when ξ_i is long range dependent, then t_n decays slower than s_n implying $t_n/s_n \to \infty$.

4 Simulation study

In this section, we perform a small simulation study to validate our theoretical findings. To this end, we generate data from the following model,

$$Y_i = \theta_1(Z_i) + X_i\theta_2(Z_i) + e_i,\ i = 1, \ldots, n,$$

where Z_i and X_i's are independently generated from $Uniform(0, 1)$ and $N(0, 1)$ distributions, respectively, and $e_i = (\epsilon_{i-2} + \epsilon_{i-1} + \ldots + \epsilon_{i+2})/5$ with $\epsilon_j \sim N(0, 1), -\infty \le j \le \infty$. We further set $\theta_1(z) = cos(\pi z)$ and $\theta_2(z) = sin(\pi z)$. Three different choices of sample sizes $n = 100, 250$ and 1000 are considered, and we generate 1000 data sets for each case.

We perform local linear ($p = 1$) estimation using the Epanechnikov kernel $K(z) = (3/4)(1 - z^2)I(|z| \le 1)$ with bandwidth $h_n = sd(Z_1, \cdots, Z_n)n^{-1/5}$. Thus for each data set, we obtain $\widehat{\gamma}(z) = \left[\widehat{\theta}_1(z), \widehat{\theta}_2(z), h_n\widehat{\theta}_1^{(1)}(z), h_n\widehat{\theta}_2^{(1)}(z)\right]^T$. We perform such estimation on a grid of 11 equally spaced points between 0.1 and 0.9.

Results from the simulation study are displayed in Fig. 1. The first column presents the bar graphs of the mean squared error (MSE) of the four estimated components of $\widehat{\gamma}(z)$ for different sample sizes. It is evident that as sample size increases, the MSE

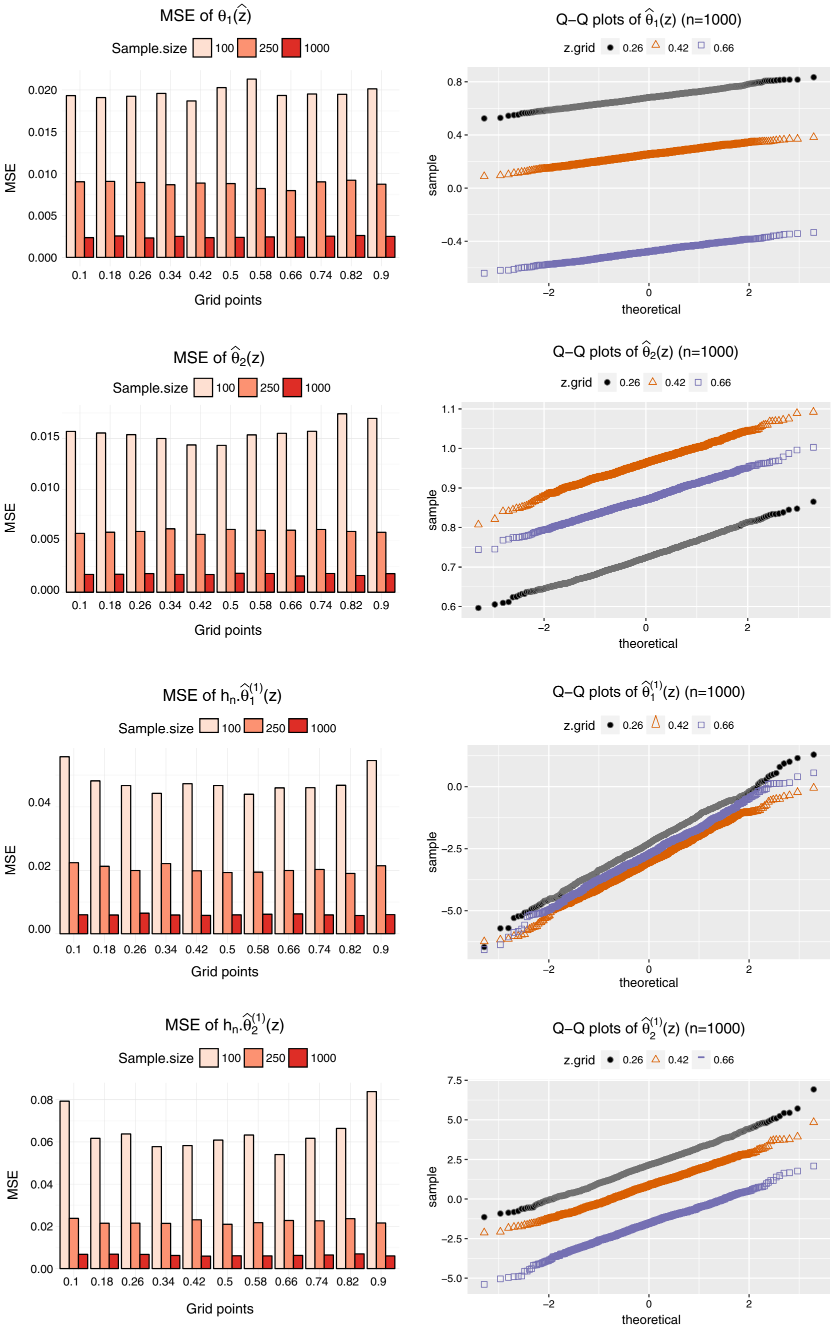

Fig. 1 Results from the simulation study based on 1000 simulated data sets. The first column displays the mean squared errors of $\widehat{\theta}_1$, $\widehat{\theta}_2$, $h_n\widehat{\theta}_1^{(1)}$ and $h_n\widehat{\theta}_2^{(1)}$ for different sample sizes. The second column displays QQ-plots for $n = 1000$ of estimates of $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\theta}_1^{(1)}$ and $\widehat{\theta}_2^{(1)}$

becomes smaller. This is expected since our theoretical result shows that the estimators converge in probability to the true value for all z.

To assess asymptotic normality of the estimators, we create QQ-plots of the estimated functions $\widehat{\theta}_1(z)$, $\widehat{\theta}_2(z)$, $\widehat{\theta}_1^{(1)}(z)$ and $\widehat{\theta}_2^{(1)}(z)$ at $z = 0.26, 0.42$ and 0.66 (values were randomly chosen) with respect to normal quantiles (second column of Fig. 1) for $n = 1000$. It is clear that the distributions of both estimated functions and their first-order derivatives are indeed close to normal.

It is worth mentioning that we considered a few other choices of the bandwidth with varying $C > 0$ where $h_n = Cn^{-1/5}$ and the results were quite similar. Therefore, to maintain brevity we refrain from reporting those results in the main article.

5 Proof of Theorems

5.1 Proof of Theorem 1

We write

$$\widehat{\gamma}(z) - \gamma_0(z) = A_3^{-1}(A_1 + A_2),$$

where

$$A_1 = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}G(Z_i - z)X_i\xi_i,$$

$$A_2 = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}G(Z_i - z)X_iX_i'\{\theta(Z_i) - G(Z_i - z)'\gamma_0(z)\},$$

$$A_3 = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}G(Z_i - z)X_iX_i'G(Z_i - z)'.$$

To prove consistency, we will show that entry-wise, $A_3 - \mathcal{I} \to^p 0$, $A_1 \to^p 0$ and $A_2 \to^p 0$, where $\mathcal{I}$ is a $L(p+1) \times L(p+1)$ matrix defined in (6).

Starting with A_3, define for $r, j, k \geq 0$,

$$B_{rjk} = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}(Z_i - z)^r X_{ij}X_{ik}/h^r r!.$$

Note that the structure of a typical term in A_3 is given by B_{rjk} for appropriate $r, j, k \geq 0$. Therefore, entry-wise, the result

$$A_3 - \mathcal{I} = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}G(Z_i - z)X_iX_i'G(Z_i - z)' - \mathcal{I} \to^p 0 \quad (7)$$

follows by a direct application of Lemma 1 in Section B of Supplementary Materials.

Next for $r,\ j \geq 1$, let us define a typical term of A_1 as

$$C_r = (nh)^{-1}(r!)^{-1} \sum_{i=1}^{n} K\{(Z_i - z)/h\}\{(Z_i - z)/h\}^r X_{ij}\xi_i$$

for appropriate values of r. Now we observe that

$$E(C_r^2) = (nh)^{-2}(r!)^{-2} \sum_{i=1}^{n} E[h^{-2r} E(X_{ij}^2|Z_i) K^2\{(Z_i - z)/h\}(Z_i - z)^{2r}\xi_i^2]$$
$$+ (nh)^{-2}(r!)^{-2} \sum_{i \neq k=1}^{n} \times E\big[h^{-2r} E(X_{ij}X_{kj}|Z_i, Z_k) K\{(Z_i - z)/h\} K\{(Z_k - z)/h\}$$
$$\times (Z_i - z)^r (Z_k - z)^r \xi_i \xi_k\big].$$

The first term of the sum is

$$(nh)^{-1}(r!)^{-2} \int K^2(w) w^{2r} n^{-1} \sum_{i=1}^{n} m_{i,j}^{*}(wh + z) dw \to 0 \tag{8}$$

by (5) and assumptions (H1) and (D5) (see Supplementary Materials). Similarly, the second term is

$$n^{-2}(r!)^{-2} \int \sum_{i \neq k=1}^{n} \beta_{ik} m_{ik,jj}^{*}(w_1 h + z, w_2 h + z) K(w_1) K(w_2) w_1^r w_2^r dw_1 dw_2 \to 0, \tag{9}$$

where the last limit follows from the assumption that $n^{-2} \sum_{i \neq j} \beta_{ij} \to 0$ as in (8) of Supplementary Materials. Hence combining (8) and (9), we get $E(C_r^2) \to 0$. Since C_r represents a typical element in the vector A_1, it is implied that $E(||A_1||^2) \to 0$. Therefore, $A_1 \to^p 0$, entry-wise.

For the remaining term A_2, note that for a typical entry D_{rj} of A_2 for appropriate values of $r,\ j$, we have for some $Z_i^* \in\ [Z_i, z]$,

$$E|D_{rj}|$$
$$= h^p (r!)^{-1} E \left| (nh)^{-1} \sum_{i=1}^{n} K\{(Z_i - z)/h\}\{(Z_i - z)/h\}^{r+p} X_{ij} \right.$$
$$\left. \times \sum_{k=1}^{L} X_{ik}\{\theta_k^{(p)}(Z_i^*) - \theta_k^{(p)}(z)\} \right|$$
$$\leq h^p (r!)^{-1} (nh)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{L} E \left| K\{(Z_i - z)/h\}\{(Z_i - z)/h\}^{r+p} E(|X_{ij}X_{ik}||Z_i) \right|$$

$$\times \sup_{|u|\le h} |\theta_k^{(p)}(z+u) - \theta_k^{(p)}(z)|$$
$$\le h^{p+q}(r!)^{-1}n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{L}\int |K\{w\}\{w\}^r|E(|X_{ij}X_{ik}||Z_i = wz+h)f_i(wh+z)dw,$$
$$= O(h^{p+q}) \to 0,$$

where the last line follows from Assumption (R2). Therefore, $A_2 \to^p 0$ entry-wise, and hence, the proof is completed. □

5.2 Proof of Theorem 2

The main steps to prove Theorem 2 is as follows. We note that $c_n^{1/2}(\widehat{\gamma}(z) - \gamma_0(z) - \mathcal{B}_n) = A_3^{-1}\, c_n^{1/2} A_1$, where A_1 and A_3 are defined in Sect. 5.1 and $\mathcal{B}_n = A_3^{-1} A_2$. Thus, the result follows if we prove $\mathcal{B}_n = O(h^{p+q})$ and asymptotic normality of $c_n^{1/2} A_1$. To prove the later result, we use a central limit theorem for martingale differences. In fact, the error structure used in this article is needed to apply such a limit theorem. It is important to see that generally the martingale difference assumptions of time series models are hard to extend as there is no natural ordering to our data. However, we follow the same triangular array setting for observed data as in Robinson (2011) (Section 6, items 8 and 9) which ensures the martingale differences in higher dimensions to be well defined.

We start by investigating $\mathcal{B}_n$. Using similar argument as in Sect. 5.1, we can easily prove $A_2 = O_p(h^{p+q})$. Combined with the result $A_3 - \mathcal{I} \to^p 0$ (see (7)), it readily follows that $\mathcal{B}_n = O(h^{p+q})$.

To prove asymptotic normality of $c_n^{1/2} A_1$, we adopt techniques similar to that of Robinson (2011). As in Lemma 9 of Robinson (2011), we note that there exists a sequence $N = N_n$, increasing with n such that

$$A_1 = \sum_{j=1}^{N} W_j \epsilon_j + o_p(c_n^{-1/2}), \tag{10}$$

where

$$W_j = (nh)^{-1}\sum_{i=1}^{n} K\{(Z_i - z)/h\}\sigma_i(X_i, Z_i)G(Z_i - z)X_i\alpha_{ij}. \tag{11}$$

Asymptotic normality of $c_n^{1/2} A_1 \to^d N(0, \Sigma)$ follows from Lemma 5, together with Lemmas 2 and 3 (hence, $c_n(T - \Sigma) \to^p 0$), and consequently, Theorem 2 follows from noting that $A_3 - \mathcal{I} \xrightarrow{p} 0$ from (7). □

Acknowledgements The authors are extremely grateful to the Associate Editor and two anonymous referees whose suggestions greatly improved the results and presentation of the article.

References

Ahmad, I., Leelahanon, S., Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, *33*, 256–283.

Arbia, G. (2006). *Spatial econometrics*: *Statistical foundations and applications to regional analysis*. Berlin: Springer.

Aykroyd, R. G. (1998). Bayesian estimation for homogeneous and inhomogeneous gaussian random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 533–539.

Aykroyd, R. G., Zimeras, S. (1999). Inhomogeneous priors for image reconstruction. *Journal of the American Statistical Association*, *94*, 934–946.

Bandyopadhyay, S., Lahiri, S. N. (2010). Asymptotic properties of discrete Fourier transforms for spatial data. *Sankhya, Series A*, *71*, 221–259.

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

Cliff, A., Ord, J. K. (1981). *Spatial processes*: *Models & applications*. London: Pion.

Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.

Dreesman, J. M., Tutz, G. (2001). Non-stationary conditional models for spatial data based on varying coefficients. *Journal of the Royal Statistical Society, Series D*: *The Statistician*, *50*, 1–15.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of American Statistical Association*, *87*, 998–1004.

Fan, J., Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman & Hall.

Fan, J., Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, *1*, 179–195.

Hallin, M., Lu, Z., Tran, L. T. (2004). Local linear spatial regression. *Annals of Statistics*, *32*, 2469–2500.

Higdon, D. M., Bowsher, J. E., Johnson, V. E., Turkington, T. G., Gilland, D. R., Jaszczak, R. J. (1997). Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transaction on Medical Imaging*, *16*, 516–526.

Johnson, V. E., Wong, W. H., Hu, X. P., Chen, C. T. (1991). Image restoration using Gibbs priors-boundary modeling, treatment of blurring and selection of hyperparameter. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *13*, 413–425.

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

Lahiri, S. N. (1996). On inconsistency of estimators under infill asymptotics for spatial data. *Sankhya, Series A*, *58*, 403–417.

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory*, *18*, 252–277.

Loader, C. (1999). *Local regression and likelihood*. New York: Springer.

Loh, W. L. (2005). Fixed-domain asymptotics for a subclass of Matern-type Gaussian random fields. *Annals of Statistics*, *33*, 2344–2394.

Lu, Z., Tang, Q., Cheng, L. (2014). Estimating spatial quantile regression with functional coefficients: A robust semiparametric framework. *Bernoulli*, *20*(1), 164–189.

Robinson, P. M. (1989). *Nonparametric estimation of time-varying parameters*. Berlin, Heidelberg: Statistical Analysis and Forecasting of Economic Structural Change. Springer.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics*, *165*, 5–19.

Robinson, P. M., Thawornkaiwong, S. (2011). Statistical inference on regression with spatial dependence. *Journal of Econometrics*, *167*, 521–542.

Ruppert, D., Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, *22*, 1346–1370.

Stein, M. L. (1999). *Interpolation of spatial data*: *Some theory for kriging*. New York: Springer.

Sun, Y., Yan, H., Zhang, W., Lu, Z. (2014). A semiparametric spatial dynamic model. *Annals of Statistics*, *42*(2), 700–727.

Tran, L. T., Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *Journal of Multivariate Analysis*, *44*, 23–46.

Tran, L. T., Roussas, G., Yakowitz, S., Truong Van, B. (1996). Fixed-design regression for linear time series. *The Annals of Statistics*, *24*, 975–991.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Annals of Statistics*, *21*, 1567–1590.