Estimation for Semiparametric Nonlinear Regression of Irregularly Located Spatial Time-series Data

Dawlah Al-Sulami[a], Zhenyu Jiang[b], Zudi Lu[c,*], Jun Zhu[d]

[a]*School of Mathematical Sciences, University of Adelaide and Department of Statistics, King Abdulaziz University*
[b]*Statistical Sciences Research Institute, University of Southampton*
[c]*Statistical Sciences Research Institute and School of Mathematical Sciences, University of Southampton*
[d]*Department of Statistics and Department of Entomology, University of Wisconsin–Madison*

Abstract

Large spatial time-series data with complex structures collected at irregularly spaced sampling locations are prevalent in a wide range of applications. However, econometric and statistical methodology for nonlinear modeling and analysis of such data remains rare. A semiparametric nonlinear regression is thus proposed for modelling nonlinear relationship between response and covariates, which is location-based and considers both temporal-lag and spatial-neighbouring effects, allowing data-generating process nonstationary over space (but turned into stationary series along time) while the sampling spatial grids can be irregular. A semiparametric method for estimation is also developed that is computationally feasible and thus enables application in practice. Asymptotic properties of the proposed estimators are established while numerical simulations are carried for comparisons between estimates before and after spatial smoothing. Empirical application to investigation of housing prices in relation to interest rates in the United States is demonstrated, with a nonlinear threshold structure identified.

Keywords: Irregularly spaced sampling locations; Large spatial time series data; Semiparametric spatio-temporal model and estimation; Spatial smoothing.

1. Introduction

Large amounts of spatial time-series data with complex structures collected at irregularly spaced sampling locations are prevalent in a wide range of disciplines such as economics, sociology, environmental sciences. For example, it is of economic interest to study the housing price in relation to other economic index, say interest rate, based on the available quarterly, state-level data collected in the United States (Figure 4). While there is a growing body of literature on statistical tools for analyzing spatial time-series data, most methods assume linearity and stationarity on the data-generating process (see, e.g., Cressie and Wikle (2011)), which may be violated in practice. This paper therefore aims to develop more effective econometric and statistical analytical techniques for modelling nonlinear relationship between

*Corresponding author: Building 54, Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom. Tel: +44(0)23 8059 7169.

Email addresses: dalsulamid@kau.edu.sa (Dawlah Al-Sulami), zhenyujiang1@gmail.com (Zhenyu Jiang), Z.Lu@soton.ac.uk (Zudi Lu), jzhu@stat.wisc.edu (Jun Zhu)

response and covariates, needed in analysis of spatial time series or spatio-temporal data in applications.

Study of nonlinear spatio-temporal modeling is still rather rare (Cressie and Wikle (2011), pp. 437). In contrast, nonlinear analysis of time series data have been well studied in the literature (see, e.g., Tong (1990), Fan and Yao (2003), Gao (2007)). Exceptions for nonlinear spatio-temporal modelling may be found in Wikle and Hooten (2010) and Wikle and Holan (2011) who developed polynomial nonlinear spatio-temporal integro-difference equation models, and in Lu et al. (2009) who proposed semiparametric adaptively varying coefficient spatio-temporal models. Note that the models in Wikle and Hooten (2010) and Wikle and Holan (2011) are parametric which are reasonable when prior information, such as the laws in geophysics, is readily available for model specification. However, in many applications, in particular in socio-economic studies, prior knowledge is often lacking and parametric relationships among variables may suffer from model mis-specification. We are therefore, in this paper, applying semiparametric approaches that are appealing and help to uncover complex, often nonlinear, relationships (see, e.g., Li and Racine (2007), Terasvirta et al. (2010)).

Efforts to explore nonlinearity by nonparametric and semiparametric approaches for purely spatial data, particularly lattice data (i.e., with regular sampling grids), under stationarity have been well attempted in the last decade. For example, curse of dimensionality with spatial interactions when applying nonparametric approaches have been well addressed and various semiparametric approaches proposed under spatial stationarity; see, e.g., Lu and Chen (2002), Gao et al. (2006), Lu et al. (2007), Hallin et al. (2009), Robinson (2011), to list a few. For spatial data on irregular sampling grids, even under assumption of stationarity over space, there are still fewer results with nonparametric approaches; see, e.g., Sun et al. (2014) and Lu and Tjøstheim (2014) for some recent progress. However, in practice, spatial data is usually non-stationary, which may require some kind of transformations prior to application of these methods developed.

Nonparametric analysis of spatio-temporal data is still at its early stage. There are quite many challenges that we need to overcome. See Rao (2008) and Lu et al. (2009) for some recent discussions. Although there are various methods, e.g., differencing operations, to turn non-stationary time series into stationary one with unilateral time, it becomes more difficult for spatial data owing to the multi-lateral nature of space. To get across the difficulty, we will follow Rao (2008) and Lu et al. (2009) and assume that spatial time series data is non-stationary over space but stationary along time in the sequel. By this, we (Lu et al. (2009)) recently extended Fan et al. (2003) and proposed adaptively varying-coefficient spatio-temporal models which are location-dependent. These models can accommodate nonlinearity with temporal-lag and spatial-neighbouring effects (noting that Rao (2008) did not consider the spatial-neighbouring impacts). However, in Lu et al. (2009), a regular grid of spatial sampling locations is actually required for specifying appropriate neighboring variables in the models and it is also a challenge to investigate nonlinear effects of exogenous covariates due to computational complexity with these models. For example, we are facing the difficulty of irregular sampling grids with spatio-temporal modelling for the US housing price data set in Section 5 below.

The purpose of this paper is therefore to propose a family of semiparametric nonlinear regression models, with the ideas in Lu et al. (2009) and Gao et al. (2006) extended, for analysis of nonlinear relationship between response and covariates of spatial time series data. The features of these models include not only the concerned nonlinear impacts of covariates on response but also that they are location-dependent with both temporal-lag and spatial-neighbouring effects taken account of, therefore allowing data-generating process to be nonstationary over space (but stationary along time) while the sampling spatial grids can be irregular. It is worth noting that practical econometric and statistical methods for analysis of such complex spatial time-series data remain elusive, as irregular sampling grids and non-stationarity in space generally lead to the challenging large curse of dimensionality due to spatio-temporal interactions. For instance, possibly nonlinear effect of interest rate on the housing prices at the state level in the United States will be considered in Section 5, where among 49 states (excluding Alaska and Hawaii, but counting District of Columbia as a state), a convenient way to specify the neighbouring variables is by seeing all other states as the neighborhood of a state, and due to non-stationarity of the response of house price returns over 49 states, the dimension of nonlinear regression of the response at a state on its 5 temporal lags at 49 states plus one covariate of interest rate increment is as high as $(1+49\times 5)=246$. See Section 5, where by the methodology in this paper, we will be able to analyse the data by combining a popular idea of spatial weight matrix in econometrics (see, e.g., Anselin (1988)). We will develop a computationally feasible method of two-step procedure for estimation and thus enable our methodology to be readily applicable in practice. Asymptotic properties of our proposed estimates are established and numerical comparisons are made theoretically and empirically between estimates before and after spatial smoothing in the two steps.

The remainder of the paper is organized as follows. In Section 2, we present the semiparametric spatio-temporal autoregressive partially nonlinear regression model and develop a two-step procedure for estimation. We provide the asymptotic properties of the estimators in Section 3 and study the finite-sample properties via a simulation study in Section 4. In Section 5, our methodology is demonstrated to investigate housing price in relation to interest rate in the United States. We show that more insight into the effect of interest rate on housing price in different states and time periods can be gained from our method, with a threshold structure found to be helpful for prediction. Conclusions and discussion are given in Section 6. The technical details including proofs are relegated to web-based supplementary materials.

2. Methodology

2.1. Model

Let $Y_t(s)$ and $X_t(s)$ denote two spatio-temporal processes at discrete time points $t = 1,\ldots,T$ and continuous locations s in a spatial domain $S \subset \mathbb{R}^2$. The relationship of between

Y and X is of interest, with Y denoting the response variable and X the covariate vector of dimension d, respectively. Assume that both processes are observed at T time points $t = 1, \ldots, T$ and at N spatial sampling locations $s_j = (u_j, v_j)' \in S$ for $j = 1, \ldots, N$ on a possibly irregular grid. That is, the data comprise $\{(Y_t(s_j), X_t(s_j))$: $t = 1, \ldots, T$ and $j = 1, \ldots, N\}$.

As in the housing price example in Section 5, note that for a given state, not taking into account the effects of the housing prices from neighboring states like the model in Rao (2008) could result in biased estimates of the relationship between interest rates and housing price. However, the irregular grid of states makes it difficult to specify a small, same number of neighboring variables over all states as in Lu et al. (2009). By extending Rao (2008); Gao et al. (2006), we therefore propose a class of location-dependent spatio-temporal autoregressive partially (non)linear regression (STAR-PLR) models in the form of

$$Y_t(s_j) = g(X_t(s_j), s_j) + \sum_{i=1}^{p} \lambda_i(s_j) Y^{\mathrm{sl}}_{t-i}(s_j) + \sum_{l=1}^{q} \alpha_l(s_j) Y_{t-l}(s_j) + \varepsilon_t(s_j). \tag{1}$$

Here $g(X_t(s_j), s_j)$ is a fixed, nonparametric function that we are concerned with, which varies by location and characterizes the relationship between the response and exogenous covariates that are of interest in applications. To account for spatial *neighboring* effects, a spatially lagged response variable, $Y^{\mathrm{sl}}_t(s_j) = \sum_{k=1}^{N} w_{jk} Y_t(s_k)$, is defined, where w_{jk} is a spatial weight for $1 \leq j, k \leq N$ such that $w_{jj} = 0$ and the spatial weight matrix $W = [w_{jk}]_{j,k=1}^{N}$ is assumed to be specified *a priori*, the idea of which is popular in econometrics (see, e.g., Chapter 3 of Anselin (1988)). The choice of spatial weights will be discussed in the context of the data example in Section 5. To further account for temporal effects, two temporally lagged response variables, $Y^{\mathrm{sl}}_{t-i}(s_j)$ involving neighboring locations of site s_j and $Y_{t-l}(s_j)$ at location s_j, are included in the model, with temporal lags $i = 1, \ldots, p$ up to the pth lag and $l = 1, \ldots, q$ up to the qth lag, respectively. Both $Y^{\mathrm{sl}}_{t-i}(s_j)$ and $Y_{t-l}(s_j)$ are in linear relation to $Y_t(s_j)$ with spatially-varying autoregressive coefficients $\lambda_i(s_j)$ and $\alpha_l(s_j)$, respectively. The random error (or, innovation) $\varepsilon_t(s_j)$ is assumed to be independently and identically distributed (iid) over time with mean 0 and spatially-varying variance $\sigma^2(s_j)$. The processes $\{X_t(s_j)\}$, $\{Y^{\mathrm{sl}}_t(s_j)\}$, and $\{Y_t(s_j)\}$ are assumed to be stationary over time. Further, $X_t(s_j)$, $Y^{\mathrm{sl}}_{t-i}(s_j)$, and $Y_{t-l}(s_j)$ are independent of the innovation process $\varepsilon_t(s_j)$ for any t and j with $i > 0$ and $l > 0$.

Since the form of the function $g(X_t(s_j), s_j)$ is left unspecified, the model is more flexible than the traditional spatio-temporal linear regression (see, e.g., Section 6.8, Cressie (1993)). While temporal stationarity is assumed, the model allows for nonstationarity over space, because the function $g(\cdot, s_j)$ and the autoregressive coefficients $\lambda_i(s_j)$ and $\alpha_l(s_j)$ vary by location (Rao, 2008; Lu et al., 2009). Further, the innovation has a variance that varies by location and the distribution of innovation, unlike in the traditional spatio-temporal linear regression model, does not need to be Gaussian. In essence, the STAR-PLR model (1) is semiparametric and partially nonlinear, as the form of $g(X_t(s_j), s_j)$ is left unspecified and the innovation process is distribution-free. For ease of presentation, we consider $X_t(s_j) \in \mathbb{R}^1$ with $d = 1$ below. The method and theory to be developed, however, apply to a general d

dimension with a minor modification though d should not be too big in application.

2.2. Estimation

Next, we develop a two-step procedure for estimating the unknown function g and the autoregressive coefficients λ_i's and α_l's in the STAR-PLR model (1). As we will demonstrate, the computation in this two-step procedure is quite fast, making it computationally feasible for handling spatial time-series data that are becoming increasingly bigger and more complex.

Let $Z_t(s_j) = (Y^{\text{sl}}_{t-1}(s_j), \ldots, Y^{\text{sl}}_{t-p}(s_j), Y_{t-1}(s_j), \ldots, Y_{t-q}(s_j))'$ denote the vector of spatio-temporally lagged variables and let $\beta(s_j) = (\lambda_1(s_j), \ldots, \lambda_p(s_j), \alpha_1(s_j), \ldots, \alpha_q(s_j))'$ denote the corresponding vector of autoregressive coefficients. Then, the STAR-PLR model given in (1) can be rewritten as

$$Y_t(s_j) = g(X_t(s_j), s_j) + Z_t(s_j)'\beta(s_j) + \varepsilon_t(s_j), \tag{2}$$

where $t = r+1, \ldots, T$, with $r = \max\{p, q\}$, and $j = 1, \ldots, N$.

From (2), the unknown function $g(X_t(s_j), s_j)$ is given by $g(X_t(s_j), s_j) = Y_t(s_j) - Z_t(s_j)'\beta(s_j) - \varepsilon_t(s_j)$. Taking expectation conditional on the covariate leads to

$$g(X_t(s_j), s_j) = E[Y_t(s_j)|X_t(s_j)] - E[Z_t(s_j)|X_t(s_j)]'\beta(s_j),$$

which can be estimated by

$$\hat{g}(X_t(s_j), s_j) = \hat{E}[Y_t(s_j)|X_t(s_j)] - \hat{E}[Z_t(s_j)|X_t(s_j)]'\hat{\beta}(s_j),$$

provided that the unknowns involved in the two terms on the right-hand side can be well-approximated. Therefore, as in Lu et al. (2009), we propose a two-step procedure as follows and describe the details in Sections 2.3 and 2.4, to simplify the computational burden with a large set of spatial time series data.

Step 1 (Time-series based estimation): For each fixed location $s = s_j$, consider time-series based estimation.

(i) Both $E[Y_t(s)|X_t(s)]$ and $E[Z_t(s)|X_t(s)]$ are estimated by a local linear regression method.
(ii) The estimators $\hat{E}[Y_t(s)|X_t(s)]$ and $\hat{E}[Z_t(s)|X_t(s)]$ are then used to estimate the unknown vector of autoregressive coefficients, $\beta(s)$, by a least squares method.

Step 2 (Spatial smoothing): The time-series based estimators are further improved by pooling information at neighboring locations.

2.3. Time-Series Based Estimation

In Step 1, at a fixed location $s = s_j$, we estimate $g(x, s)$ with covariate $x = x(s)$ and autoregressive coefficients $\beta(s)$ by

$$\hat{g}(x, s) \quad = \quad \hat{g}_1(x, s) - \hat{g}_2(x, s)'\beta(s),$$

where $\hat{g}_1(x, s)$ and $\hat{g}_2(x, s)$ are the estimators of $g_1(x, s) = E[Y_t(s)|X_t(s) = x]$ and $g_2(x, s) = E[Z_t(s)|X_t(s) = x]$, respectively, based on local linear regression as follows.

2.3.1. Estimation of $g_1(x,s)$

First, consider estimating the function $g_1(x,s) = E[Y_t(s)|X_t(s) = x]$ for a given covariate value x and location s. We apply a local approximation $a_0 + a_1(X_t(s) - x)$ for covariate $X_t(s)$ in the neighborhood of x, where $a_0(x,s) = g_1(x,s)$ and $a_1(x,s) = \dot{g}_1(x,s)$ is the first-order derivative of g_1 with respect to x, evaluated at (x,s). For ease of notation, we let $a_0 = a_0(x,s)$ and $a_1 = a_1(x,s)$. Let $T_0 = T - r$ denote an effective sample size with $r = \max\{p,q\}$. Let $b = b_{T_0}$ denote a temporal bandwidth. Let $K(\cdot)$ denote a bounded kernel function and $K_b(\cdot) = b^{-1}K(\cdot/b)$. We estimate a_0 and a_1 by the weighted least squares:

$$\begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \arg \min_{(a_0,a_1)' \in \mathbb{R}^2} \sum_{t=r+1}^{T} \{Y_t(s) - a_0 - a_1(X_t(s) - x)\}^2 K_b(X_t(s) - x). \tag{3}$$

Let $A(x)$ denote a $T_0 \times 2$ matrix with row $t - r$ being $(1, b^{-1}(X_t(s) - x))$ for $t = r + 1, \ldots, T$. Let $B(x) = \mathrm{diag}\{K_b(X_t(s) - x)\}_{t=r+1}^{T}$ denote a $T_0 \times T_0$ diagonal matrix. Let $Y = (Y_{r+1}(s), \ldots, Y_T(s))'$ denote a T_0-dimensional vector. The local linear estimators can be expressed as

$$(\hat{a}_0, b\hat{a}_1)' = U_{T_0}^{-1} V_{T_0},$$

where $U_{T_0} = A(x)'B(x)A(x)$ is a 2×2 matrix with entries $u_{T_0,jk}$ for $j,k = 0,1$ and $V_{T_0} = A(x)'B(x)Y = (v_{T_0,0}, v_{T_0,1})'$. In particular, with $\left(\frac{X_t(s)-x}{b}\right)^0 = 1$,

$$\begin{aligned} u_{T_0,jk} &= (T_0 b)^{-1} \sum_{t=r+1}^{T} \left(\frac{X_t(s) - x}{b}\right)^j \left(\frac{X_t(s) - x}{b}\right)^k K\left(\frac{X_t(s) - x}{b}\right), \ j,k = 0,1 \\ &\text{and} \\ v_{T_0,j} &= (T_0 b)^{-1} \sum_{t=r+1}^{T} Y_t(s) \left(\frac{X_t(s) - x}{b}\right)^j K\left(\frac{X_t(s) - x}{b}\right), \ j = 0,1. \end{aligned}$$

Thus, with $e_1 = (1,0)'$, the local linear estimator of $g_1(x,s)$ is given by

$$\hat{g}_1(x,s) = \hat{a}_0 = e_1' U_{T_0}^{-1} V_{T_0}. \tag{4}$$

2.3.2. Estimation of $g_2(x,s)$

Next, consider estimating the function $g_2(x,s) = E[Z_t(s)|X_t(s) = x]$ again by local linear regression, although the dimension of $g_2(x,s)$ is now $p + q$. Let

$$g_{21}(x,s) = \left(g_{21}^1(x,s), \ldots, g_{21}^p(x,s)\right)' = \left(E[Y_{t-1}^{\mathrm{sl}}(s)|X_t(s) = x], \ldots, E[Y_{t-p}^{\mathrm{sl}}(s)|X_t(s) = x]\right)'$$

and

$$g_{22}(x,s) = \left(g_{22}^1(x,s), \ldots, g_{22}^q(x,s)\right)' = \left(E[Y_{t-1}(s)|X_t(s) = x], \ldots, E[Y_{t-q}(s)|X_t(s) = x]\right)'.$$

Thus $g_2(x,s) = (g_{21}(x,s)', g_{22}(x,s)')'$.

Similarly to (3), we estimate the components of $g_2(x,s)$ as follows. Let

$$Z_1^i = \left(Y^{\mathrm{sl}}_{(r+1)-i}(s), \ldots, Y^{\mathrm{sl}}_{T-i}(s)\right)',$$

denote a T_0-dimensional vector for $i = 1, \ldots, p$ and let $R^i_{1T_0} = A(x)'B(x)Z_1^i = \left(r^i_{1T_0,0}, r^i_{1T_0,1}\right)'$ with

$$r^i_{1T_0,j} = (T_0 b)^{-1} \sum_{t=r+1}^{T} Y^{\mathrm{sl}}_{t-i}(s) \left(\frac{X_t(s)-x}{b}\right)^j K\left(\frac{X_t(s)-x}{b}\right), \; j = 0, 1.$$

Then for $i = 1, \ldots, p$, the local linear estimator of $g^i_{21}(x,s)$ is

$$\hat{g}^i_{21}(x,s) = e_1' U_{T_0}^{-1} R^i_{1T_0}. \tag{5}$$

Also, let $Z_2^l = \left(Y_{(r+1)-l}(s), \ldots, Y_{T-l}(s)\right)'$ denote a T_0-dimensional vector for $l = 1, \ldots, q$ and $R^l_{2T_0} = A(x)'B(x)Z_2^l = \left(r^l_{2T_0,0}, r^l_{2T_0,1}\right)'$ with

$$r^l_{2T_0,j} = (T_0 b)^{-1} \sum_{t=r+1}^{T} Y_{t-l}(s) \left(\frac{X_t(s)-x}{b}\right)^j K\left(\frac{X_t(s)-x}{b}\right), \; j = 0, 1.$$

For $l = 1, \ldots, q$, the local linear estimator of $g^l_{22}(x,s)$ is given by

$$\hat{g}^l_{22}(x,s) = e_1' U_{T_0}^{-1} R^l_{2T_0}. \tag{6}$$

Thus, the estimator of the unknown function $g_2(x,s)$ can be written as

$$\hat{g}_2(x,s) = \left(\hat{g}^1_{21}(x,s), \ldots, \hat{g}^p_{21}(x,s), \hat{g}^1_{22}(x,s), \ldots, \hat{g}^q_{22}(x,s)\right)'. \tag{7}$$

2.3.3. Estimating the unknown parameter $\beta(s)$

Since the vector of autoregressive coefficients $\beta(s)$, the unknown function $g(X_t(s), s)$ can be estimated by

$$\hat{g}(X_t(s), s; \beta) = \hat{g}_1(X_t(s), s) - \hat{g}_2(X_t(s), s)'\beta(s), \tag{8}$$

we estimate $\beta(s)$ by the least squares:

$$\begin{aligned}
\hat{\beta}(s) &= \arg\min_{\beta \in R^{p+q}} \sum_{t=r+1}^{T} \left\{Y_t(s) - Z_t'(s)\beta(s) - \hat{g}(X_t(s), s; \beta)\right\}^2 \\
&= \arg\min_{\beta \in R^{p+q}} \sum_{t=r+1}^{T} \left\{\hat{Y}_t(s) - \hat{Z}_t(s)'\beta(s)\right\}^2,
\end{aligned}$$

where $\hat{Y}_t(s) = Y_t(s) - \hat{E}[Y_t(s)|X_t(s)]$ and $\hat{Z}_t(s) = Z_t(s) - \hat{E}[Z_t(s)|X_t(s)]$. Thus,

$$\hat{\beta}(s) = \left\{\sum_{t=r+1}^{T} \hat{Z}_t(s)\hat{Z}_t(s)'\right\}^{-1} \left\{\sum_{t=r+1}^{T} \hat{Z}_t(s)\hat{Y}_t(s)\right\}. \tag{9}$$

Finally, by substituting $\hat{\beta}(s)$ into (8), $g(x,s)$ can be estimated by

$$\hat{g}(x,s) = \hat{g}_1(x,s) - \hat{g}_2(x,s)'\hat{\beta}(s). \tag{10}$$

2.4. Spatial Smoothing

To improve the estimators (9) and (10) obtained from Step 1 that is based on time-series at a given location, we consider pooling the information from neighboring locations by spatial smoothing (Lu et al., 2009). At location $s_0 \in S$ with S for the support of the spatial sampling intensity function f (c.f., Assumption **S** in Appendix), the spatial smoothing estimators of $\beta(s_0)$ and $g(x, s_0)$ can be obtained by

$$\tilde{\beta}(s_0) = \sum_{j=1}^{N} \hat{\beta}(s_j) \tilde{K}^*_{h,j}(s_0) \tag{11}$$

and

$$\tilde{g}(x, s_0) = \sum_{j=1}^{N} \hat{g}(x, s_j) \tilde{K}^*_{h,j}(s_0), \tag{12}$$

where $\hat{\beta}(s_j) = \left(\hat{\lambda}_1(s_j), \ldots, \hat{\lambda}_p(s_j), \hat{\alpha}_1(s_j), \ldots, \hat{\alpha}_q(s_j)\right)'$ and $\hat{g}(x, s_j)$ are defined in (9) and (10), respectively, and $\tilde{K}^*_{h,j}(\cdot)$ denotes a weight function on $\mathbb{R}^2$, associated with $h = h_N > 0$ a spatial bandwidth depending on the number of the spatial sampling locations N.

Here we apply local linear spatial smoothing by using the weight function $\tilde{K}^*_{h,j}(s_0) = \tilde{e}_1' \left(C'DC\right)^{-1} C'D$, which is a local linear fitting equivalent kernel, where $\tilde{e}_1 = (1, 0, 0)' \in \mathbb{R}^3$, C denotes an $N \times 3$ matrix with the jth-row $(1, (s_j - s_0)'/h)$, and $D = \text{diag}\left\{\tilde{K}_h(s_j - s_0)\right\}_{j=1}^{N}$ an $N \times N$ diagonal matrix with $\tilde{K}_h(\cdot) = h^{-2}\tilde{K}(\cdot/h)$ and $\tilde{K}(\cdot)$ a kernel function on $\mathbb{R}^2$.

3. Asymptotic Theory

For the large-sample properties stated below, the regularity conditions imposed on the time series and spatial processes are given in Appendix A and the proofs of the theorems are in a web-based Appendix B.

We first provide the asymptotic properties for the time series based estimators, $\hat{\beta}(s)$ in (9) and $\hat{g}(x, s)$ in (10), in Theorems 1–2 below.

Theorem 1. *Under Assumption* **T** *in Appendix A, together with $T_0 b^4 \to 0$ as $T_0 \to \infty$, it holds that for each $s = s_j$,*

$$T_0^{1/2} \left\{\hat{\beta}(s) - \beta(s)\right\} \xrightarrow{D} N(0, \Sigma_\beta(s))$$

as $T_0 \to \infty$, where $\xrightarrow{D}$ denotes convergence in distribution, and $\Sigma_\beta(s) = M(s)^{-1}\sigma_\varepsilon^2(s)$, with $M(s) = E\left[Z_t^*(s) Z_t^*(s)'\right]$, $Z_t^*(s) = Z_t(s) - E\left[Z_t(s)|X_t(s)\right]$ and $\sigma_\varepsilon^2(s) = Var[\varepsilon_t(s)]$.

Theorem 2. *Under Assumption* **T** *in Appendix A, (with a bandwidth b different from that in Theorem 1), for each $s = s_j$ and x in the support of $X(s)$,*

$$(T_0 b)^{1/2}\left[\{\hat{g}(x,s) - g(x,s)\} - (1/2)b^2 B_0(x,s)\right] \xrightarrow{D} N(0, \Gamma(x,s))$$

as $T_0 \to \infty$, where $B_0(x,s) = \frac{\partial^2 g(x,s)}{\partial x^2}\int u^2 K(u)du$, $\Gamma(x,s) = \sigma^2(x,s)p(x,s)^{-1}\int K^2(u)du$, $p(x,s)$ is the probability density function of $X_t(s)$, and

$$\sigma^2(x,s) = Var\left[\{Y_t(s) - Z_t(s)'\beta(s)\}|X_t(s) = x\right].$$

Next, we establish the asymptotic properties for the estimators after spatial smoothing, $\tilde{\beta}(s_0)$ in (11) and $\tilde{g}(x, s_0)$ in (12), in Theorems 3–4.

Theorem 3. *Under the conditions in Theorem 1 together with Assumption* **S** *in Appendix A,, it holds that for $s_0 \in S$, as $T_0 \to \infty$ and $N \to \infty$,*

$$\tilde{\beta}(s_0) - \beta(s_0) - (1/2)h^2 B(s_0) = T_0^{-1/2}\nu(s_0)\xi(s_0)\{1 + o_p(1)\},$$

where $\xi(s_0)$ is a $(p+q)\times 1$ Gaussian random vector with zero mean and identity covariance matrix, $B(s_0) = \mathrm{tr}\left\{\frac{\partial^2\beta(s_0)}{\partial s\partial s'}\int zz'\tilde{K}(z)dz\right\}$ and

$$\nu^2(s_0) = \sigma^2(s_0)\{Nh^2 f(s_0)\}^{-1}M(s_0)^{-1} + \sigma_1^2(s_0)M(s_0)^{-1}M_{*1}(s_0,s_0)M(s_0)^{-1},$$

with $\sigma_1^2(\cdot)$ and $M_{*1}(\cdot,\cdot)$ defined in Assumption **S**.

Theorem 4. *Under the conditions in Theorem 2 together with Assumption* **S** *in Appendix A, for $s_0 \in S$, as $T_0 \to \infty$ and $N \to \infty$,*

$$\tilde{g}(x,s_0) - g(x,s_0) - (1/2)h^2\mu_1(x,s_0) - (1/2)b^2\mu_2(x,s_0) = (T_0 b)^{-1/2}\nu_1(x,s_0)\eta(s_0)\{1 + o_p(1)\},$$

where $\eta(s_0)$ is a Gaussian random variable with zero mean and identity variance, and

$$\mu_1(x,s_0) = \mathrm{tr}\left\{\frac{\partial^2 g(x,s_0)}{\partial s\partial s'}\int zz'\tilde{K}(z)dz\right\},\ \mu_2(x,s_0) = \left\{\frac{\partial^2 g(x,s_0)}{\partial x^2}\int u^2 K(u)du\right\}$$

and

$$\nu_1^2(x,s_0) = b\sigma_1^2(s_0)p(x,s_0)^{-2}q(x,x;s_0) + \{Nh^2 p(x,s_0)f(s_0)\}^{-1}\sigma^2(s_0)\int K^2(u)du\int \tilde{K}^2(z)dz,$$

with $q(\cdot,\cdot;s_0)$ defined in Assumption **S**.

In Theorems 1–2, asymptotic normality is obtained for the time series based estimators, $\hat{\beta}(s)$ in (9) and $\hat{g}(x,s)$ in (10). For the estimators after spatial smoothing, $\tilde{\beta}(s_0)$ in (11) and $\tilde{g}(x,s_0)$ in (12), consistency results are established in Theorems 3–4. As $Nh^2 \to \infty$ and $b \to 0$, both the first term of $\nu^2(s_0)$ in Theorem 3 and $\nu_1^2(x,s_0)$ in Theorem 4 tend to 0; thus the asymptotic variances of the estimators $\tilde{\beta}(s)$ and $\tilde{g}(x,s)$ after spatial smoothing are of a smaller order than those of the time series based estimators $\hat{\beta}(s)$ and $\hat{g}(x,s)$ without spatial smoothing. Further, to minimize the mean squared error (MSE) of $\tilde{\beta}(s)$, the spatial bandwidth h should be of order $(NT)^{1/6}$. Thus, under the condition $T = o(N^2)$, the MSE $\tilde{\beta}(s)$ after spatial smoothing is smaller than that of $\hat{\beta}(s)$ without spatial smoothing. Finally, under $Nh^2 = O(b^{-1})$, the rate of the convergence for $\hat{g}(x,s)$ without spatial smoothing is $(T_0 b)^{1/2}$, whereas that of $\tilde{g}(x,s)$ is $T_0^{1/2}$ with spatial smoothing. These results hinge on the nugget effect condition in Assumption **S**, without which spatial smoothing does not appear to affect the asymptotic variance.

4. Simulation Study

We study the finite-sample performance of our proposed estimators for the unknown quantities in model (1) in a simulation study. In particular, we consider the following STAR-PLR model:

$$Y_t(s_j) = g(X_{t-1}, s_j) + \sum_{i=1}^{5} \lambda_i(s_j) Y_{t-i}^{\mathrm{sl}}(s_j) + \sum_{l=1}^{5} \alpha_l(s_j) Y_{t-l}(s_j) + \varepsilon_t(s_j), \tag{13}$$

where, as in Section 5 below, $s_j = (u_j, v_j)$ is the centroid consisting of the latitude and longitude of the jth state, $j = 1, \cdots, 49$, in the US, and for simplicity, at any location s_j, the covariate process X_{t-1} follows the same AR(1) model, $X_t = -0.450 X_{t-1} + e_t$, with iid $N(0,1)$ errors e_t's that are independent of the innovation $\varepsilon_t(s_j)$'s, and

$$g(X_{t-1}, s_j) = \log\left[1 + \left\{(b_2(s_j) + X_{t-1})^2\right\}^{b_1(s_j)}\right],$$

where $b_1(s_j) = 0.5 + 0.2\cos(u_j + v_j)$ and $b_2(s_j) = 0.6 + 0.3\sin(u_j \times v_j)$ for $s_j = (u_j, v_j)'$ is the latitude and longitude of the jth state. Further, let $\varepsilon_t(s_j)$'s follow iid normal distribution with mean 0 and standard deviation $\sigma = 0.1$ over time and space. For the other parts of model (13), we follow the set-up of the data example in Section 5 below. In particular, there are $N = 49$ spatial sampling locations and the spatially-varying autoregressive coefficients λ_i's and α_l's are set to the estimated values in the data example of Section 5.

We generate data from model (13) as follows. At each location s_j for $j = 1, \ldots, 49$, the initial values of $Y_0(s_j)$ are set to zero. Then we generate $Y_t(s_j)$ for $t = 1, 2, \ldots$. The first 50 time points are discarded and the next T time points are saved, denoted as $\{(X_t(s_j), Y_t(s_j))$ for $t = 1, \ldots, T$, and $j = 1, \ldots, N\}$. We consider two time series lengths: $T = 75$ and $T = 150$. To asses the estimate of the unknown function $g(x, s_j)$, we select 50 points of x between the 10th and 90th percentiles of the covariate X_{t-1}. The temporal bandwidth b in

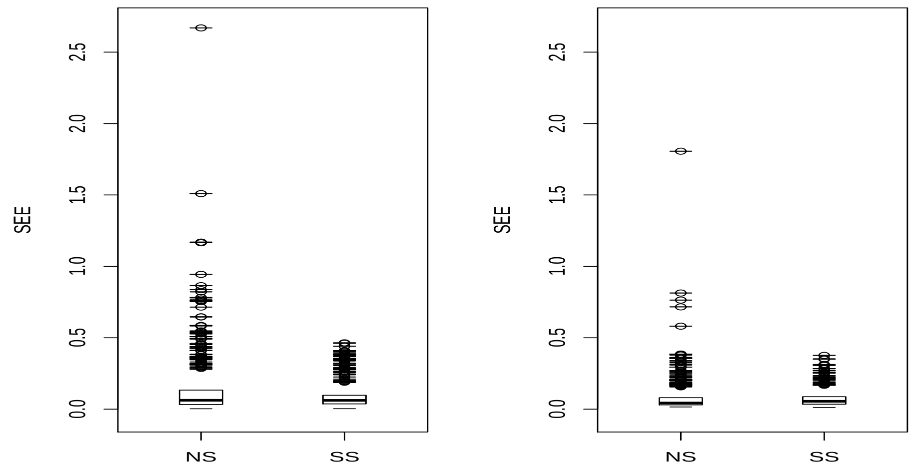

Figure 1: Boxplots of squared estimation error (SEE) for the estimation of $g(\cdot)$ without spatial smoothing (NS) or with spatial smoothing (SS), for $T = 75$ time points (left) and $T = 150$ time points (right).

Section 2.3 and the spatial bandwidth h in Section 2.4 are selected by AICc for estimation of g and coefficients (c.f.,(Hurvich et al., 2002; Lu and Zhang, 2012)).

The performance of the time-series based estimates with and without spatial smoothing will be assessed by defining a squared estimation error (SEE) as a measure of the accuracy of estimation at a location s (c.f. (Lu et al., 2009)). That is, for each location s, we define

$$\begin{aligned} \mathrm{SEE}(\hat{\lambda}_i(s)) &= \left\{\hat{\lambda}_i(s) - \lambda_i(s)\right\}^2; \qquad i = 1, \ldots, 5, \\ \mathrm{SEE}(\hat{\alpha}_l(s)) &= \left\{\hat{\alpha}_l(s) - \alpha_l(s)\right\}^2; \qquad l = 1, \ldots, 5 \text{ and} \\ \mathrm{SEE}(\hat{g}(\cdot, s)) &= \frac{1}{50}\sum_{k=1}^{50} \left\{\hat{g}(x_k, s) - g(x_k, s)\right\}^2, \end{aligned}$$

where x_k for $k = 1, \ldots, 50$ are 50 points that equally partition the interval between the 10th and the 90th percentiles of the simulated covariates X_{t-1}.

We repeat the simulation 10 times and thus have, for the 49 locations, $10 \times 49 = 490$ values in total for each type of SEE, summarized in boxplots in Figures 1–3 for the time series lengths $T = 75$ and 150. These figures clearly indicate that the estimates with spatial smoothing are more accurate than the estimates based only on individual time series data. In addition, the estimates apparently improve as the sample size increases and from the median SEE values, appear acceptable, even in the case of $N = 49$ and $T = 75$.

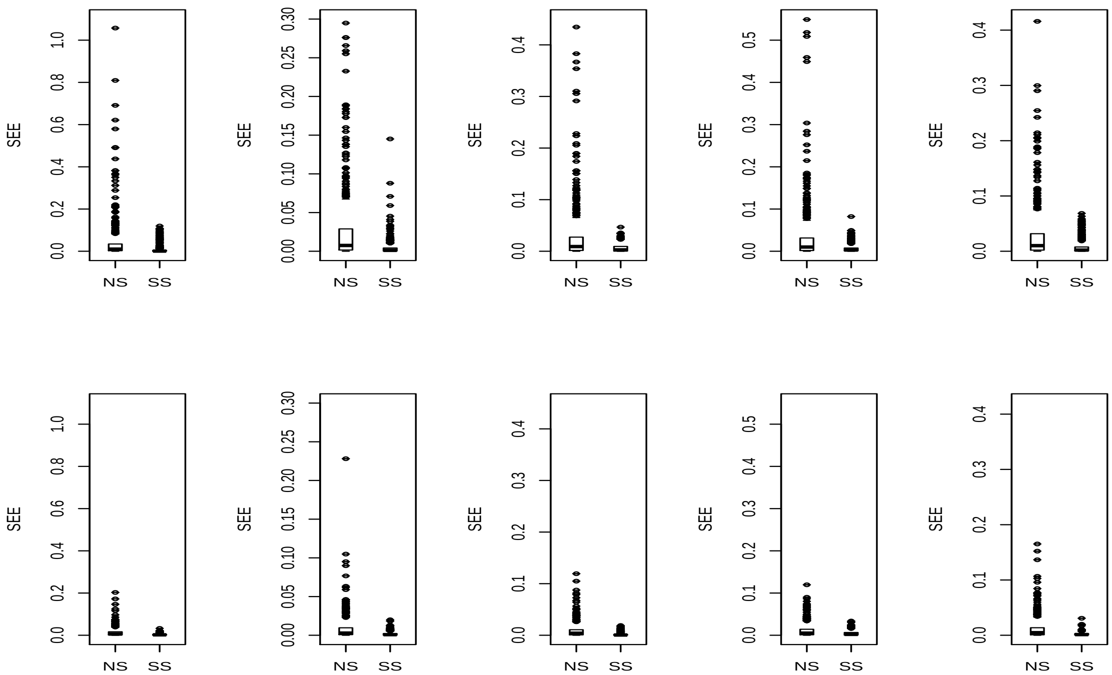

Figure 2: Boxplots of squared estimation error (SEE) for the estimation of λ_i without spatial smoothing (NS) or with spatial smoothing (SS), for $T = 75$ time points (top row) and $T = 150$ time points (bottom row), $i = 1, \ldots, 5$ (left to right per row).

5. Real Data Example

Obviously (mortgage) interest rate plays an important role in deciding housing price (c.f. Reichert (1990)). In this section, we demonstrate the methodology developed in Sections 2–3 by studying the impact of (mortgage) interest rate on housing prices in the 48 states (excluding Alaska and Hawaii) and the District of Columbia (DC) of the United States from 1991 to 2012, a time period that encompasses the US housing bubble burst, the financial crisis, and the global recession in recent years. Here we exclude Alaska and Hawaii in our consideration because they are isolated from other 49 states (counting the District of Columbia as a state). Quarterly housing price index (HPI) data from the first quarter of 1991 until the first quarter of 2012 are attained from the United States Federal Housing Financial Agency, with a time series of 85 observations for each state. It appears that the original HPI time series in all the states, shown in Figure 4, are nonstationary with increasing and then decreasing trends prior to and after the housing bubble burst in 2007. We follow the convention in economics and consider instead the geometric return of HPI, which is the change of the logarithmic HPI, for each state. Henceforth the response variable $Y_t(s_j)$ at

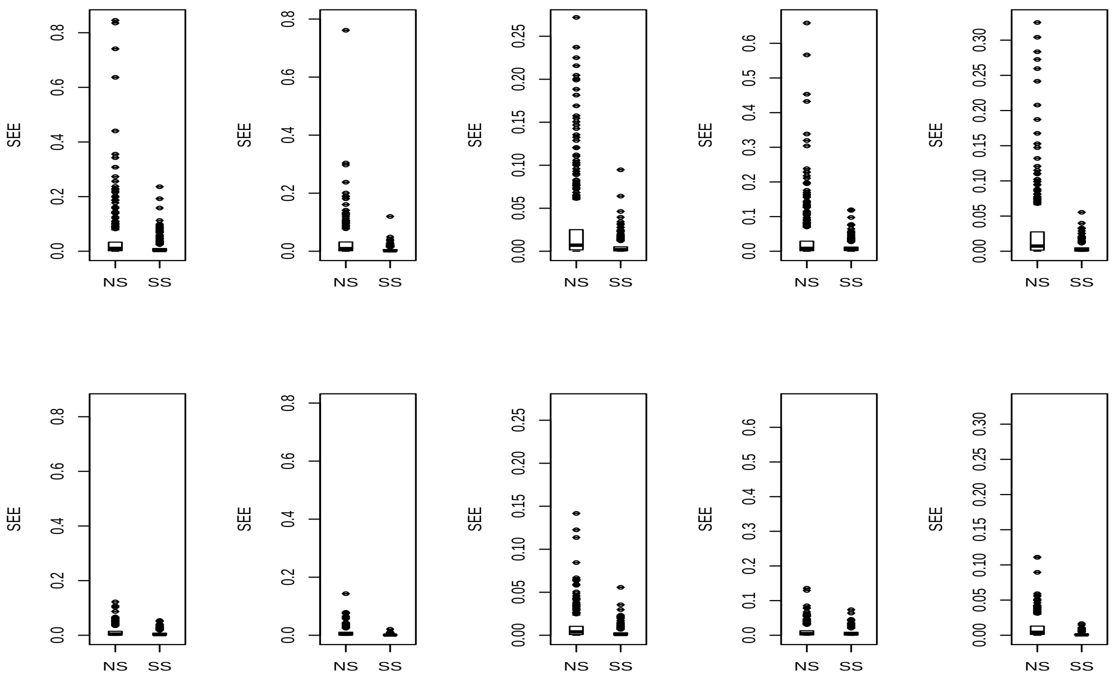

Figure 3: Boxplots of squared estimation error (SEE) for the estimation of α_l without spatial smoothing (NS) or with spatial smoothing (SS), for $T = 75$ time points (top row) and $T = 150$ time points (bottom row), $l = 1, \ldots, 5$ (left to right per row).

the tth quarter and jth state is the geometric return of housing price index (HPIGR) for $t = 2, \ldots, 85$ and $j = 1, \ldots, 49$ and the centroid $s_j = (u_j, v_j)'$ consisting of the latitude and longitude of the jth state.

The exogenous variable of interest is the quarterly change in interest rate, obtained and aggregated from monthly 30-year conventional interest rate data from the Board of Governors of the Federal Reserve System. The original quarterly interest rate data are plotted in the left panel of Figure 5 and appear to have a downward trend and thus nonstationary. However, the series of quarterly change of the interest rate, x_t, plotted in the middle panel of Figure 5, is fairly stationary, the same for all states. Further, a kernel density estimate of the quarterly change of interest rate is plotted in the right panel of Figure 5, which suggests that the distribution appears non-Gaussian.

We now assess the possibly nonlinear relationship between HPIGR, $Y_t(s_j)$, and the temporally lagged quarterly change of interest rate, $X_t(s_j) = x_{t-1}$, for $t = 2, \ldots, 85$ and $j = 1, \ldots, 49$, by specifying an STAR-PLR model (1).

First, for specifying the spatial weights w_{jk} in the spatially lagged variable $Y^{\mathrm{sl}}_{t-i}(s_j) = \sum_{k=1}^{N} w_{jk} Y_{t-i}(s_k)$, we follow a common practice in econometrics and use the inverse distance

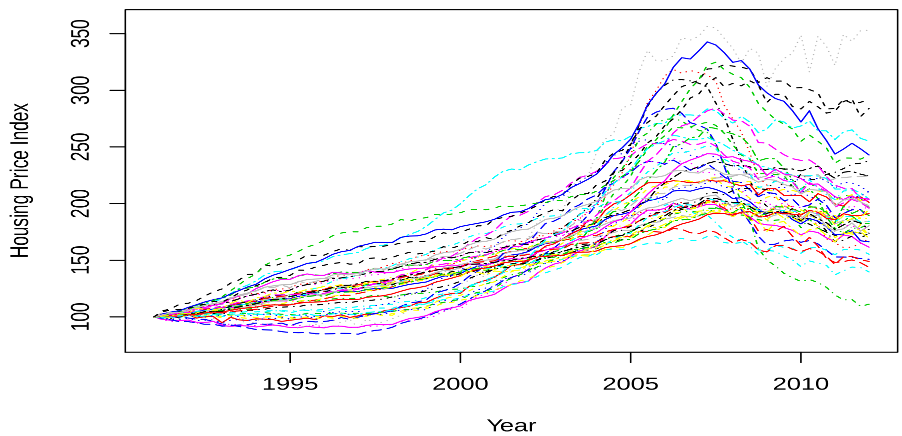

Figure 4: Time-series of the quarterly housing price index (HPI) for the 48 states (excluding Alaska and Hawaii) and District of Columbia of the United States from the first quarter of 1991 to the first quarter of 2012.

between states, such that $w_{jk} = 1/d_{jk}$ where d_{jk} is the Euclidean distance between the centroids of two states s_j and s_k, $j \neq k$, and $w_{jj} = 0$ (c.f. (Wilhelmsson, 2002)). The spatial weight matrix $W = [w_{jk}]_{j,k=1}^{N}$ is row-standardized so that $\sum_{k=1}^{N} w_{jk} = 1$. Second, to determine the orders of temporally lagged variables, p and q, we minimize the Akaike Information Criterion with correction (AICc) (c.f. (Hurvich et al., 1998))

$$\mathrm{AIC_c}(p,q) \quad = \quad \log(\hat{\sigma}^2) + \frac{1 + (T_0 N)^{-1}\mathrm{tr}(H)}{1 - (T_0 N)^{-1}\{\mathrm{tr}(H) + 2\}}, \tag{14}$$

with respect to p and q, where $\hat{\sigma}^2 = (T_0 N)^{-1} \sum_{t=r+1}^{T} \sum_{j=1}^{N} \left\{ Y_t(s_j) - \hat{Y}_t(s_j) \right\}^2$ and the hat matrix H is an $N \times N$ matrix with $N = 49$ (c.f. Appendix B.2). Finally, the bandwidth parameters b and h for time-series based estimators and those after spatial smoothing in Section 2 are determined by AICc. For the data example, $p = q = 5$ are selected and thus, the STAR-PLR model is of the form:

$$Y_t(s_j) \quad = \quad g(x_{t-1}, s_j) + \sum_{i=1}^{5} \lambda_i(s_j) Y_{t-i}^{\mathrm{sl}}(s_j) + \sum_{l=1}^{5} \alpha_l(s_j) Y_{t-l}(s_j) + \varepsilon_t(s_j), \tag{15}$$

where $t = 7, \ldots, 85$ and $j = 1, \ldots, 49$.

The estimates of $g(x, s_j)$ as a function of the quarterly change of interest rate, x, for the jth state where $j = 1, \ldots, 49$, are plotted in Figure 6, with or without spatial smoothing

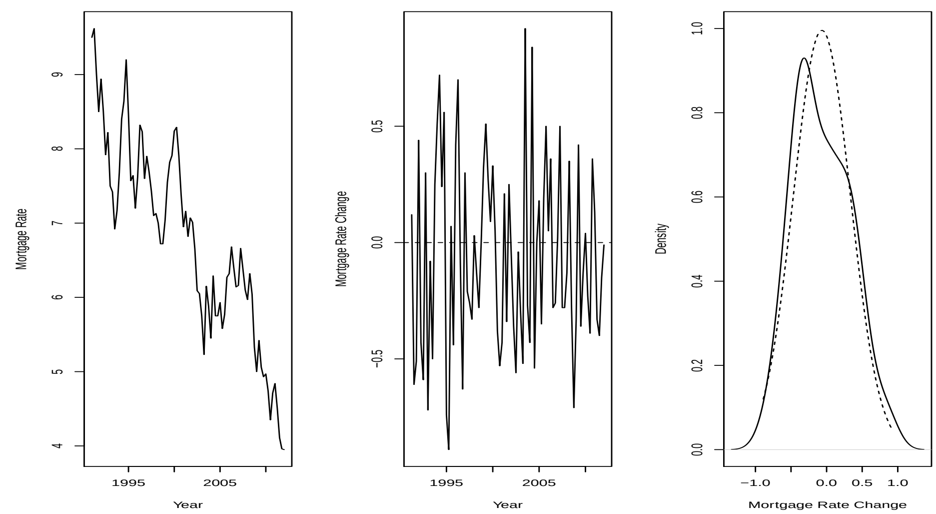

Figure 5: Time-series plot of quarterly interest rate data (left), time-series plot of quarterly change of interest rate data (middle), and kernel density estimate (solid curve) superimposed with Gaussian density estimate of same mean and variance (dashed curve) (right) in the United States from the first quarter of 1991 to the first quarter of 2012.

after the time-series based estimation. The effect of spatial smoothing is apparent. The estimated functions appear quite variable before spatial smoothing, while after spatial s-moothing, they are smoother and show clearer patterns. The relationship between HPIGR and the interest rate change among the 48 states and the DC are quite similar except for Florida in dotted line which looks slightly different from the others on the right-hand side in Figure 6(b). There is a nonlinear structure with changing points occurring approximately at $x = -0.3$, 0.1 and 0.4. In particular, for each state, the relationship is negative when the interest rate change x is smaller than -0.3 or between 0.1 and 0.4, but is positive when the interest rate change x is between -0.3 and 0.1 and appears to be constant for x greater than 0.4 (except for Florida). For Florida, the pattern seems special, which is non-constant and negative when x is larger than 0.4. According to the website of 'state of florid living' (http://www.stateoffloridaliving.com/good-time-buy-house-florida/), Florida is a highly transient state that has a real estate market that rises and falls like a yoyo. This may partly explain that a large increase of interest rate could have a large, negative impact on the return of the housing price in Florida while has a little impact in other states, as indicated for $x > 0.4$ in Figure 6(b). Furthermore, interestingly, the threshold values at $x = -0.3, 0.1$ and 0.4 appear consistent with the changing patterns of the previous kernel density estimate that exhibits a mixture pattern (Figure 5), and also with the suggestions of nonlinear relationships in McQuinn and OReilly (2007).

Like the estimates of the g function, the estimates of the autoregressive coefficients,

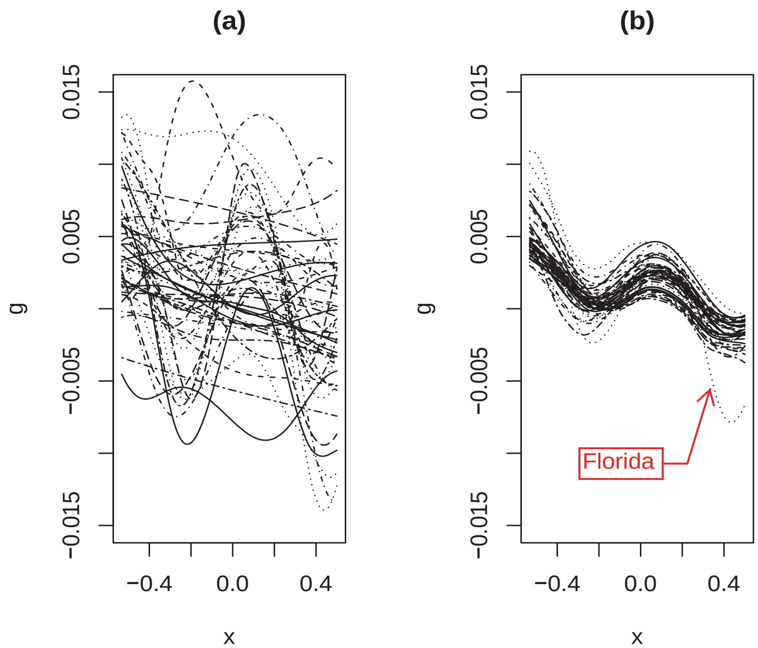

Figure 6: Estimates of g as a function of interest rate change x for the 48 states (excluding Alaska and Hawaii) and District of Columbia of the United States: (a) without spatial smoothing (left), and (b): with spatial smoothing (right).

$\lambda_i(s_j)$ and $\alpha_l(s_j)$, for $i, l = 1, \ldots, 5$ and $j = 1, \ldots, 49$, are considerably smoother after time-series based estimates are smoothed over space. To save space, we only present the maps of the estimated coefficients after spatial smoothing in Figure 7. The temporal effects among neighboring states are apparent in the maps of $\hat{\lambda}_i$ (Figure 7 left panel). While the coefficient estimates $\hat{\lambda}_1(s_j)$ and $\hat{\lambda}_4(s_j)$ at temporal lags 1 and 4 are positive in all the states with larger values in the northwest for lag 1 and in the west for lag 4, those at the other three lags 2, 3, and 5, $\hat{\lambda}_2(s_j)$, $\hat{\lambda}_3(s_j)$, and $\hat{\lambda}_5(s_j)$, are negative except for some in the east for lag 2 and some in the west for lag 3. Further, the temporal effects for a given state are also apparent in the maps of $\hat{\alpha}_l(s_j)$ in the right panel of Figure 7. It appears that $\hat{\alpha}_1(s_j)$ and $\hat{\alpha}_2(s_j)$ are mostly negative except for the southwestern states and Florida for lag 1 and California for lag 2, $\hat{\alpha}_3(s_j)$'s are positive, and $\hat{\alpha}_4(s_j)$ and $\hat{\alpha}_5(s_j)$ are positive in the northeastern states but negative in the other states. We comment that from the methodology perspective, our proposed models are location-dependent, allowing for data to be non-stationary over space with spatial site features characterised (such as varying coefficients from one region to another), so it is meaningful to use the proposed models on sub-regions of the US, e.g., the west, mid-west, south, east, although they may vary significantly from one region to another.

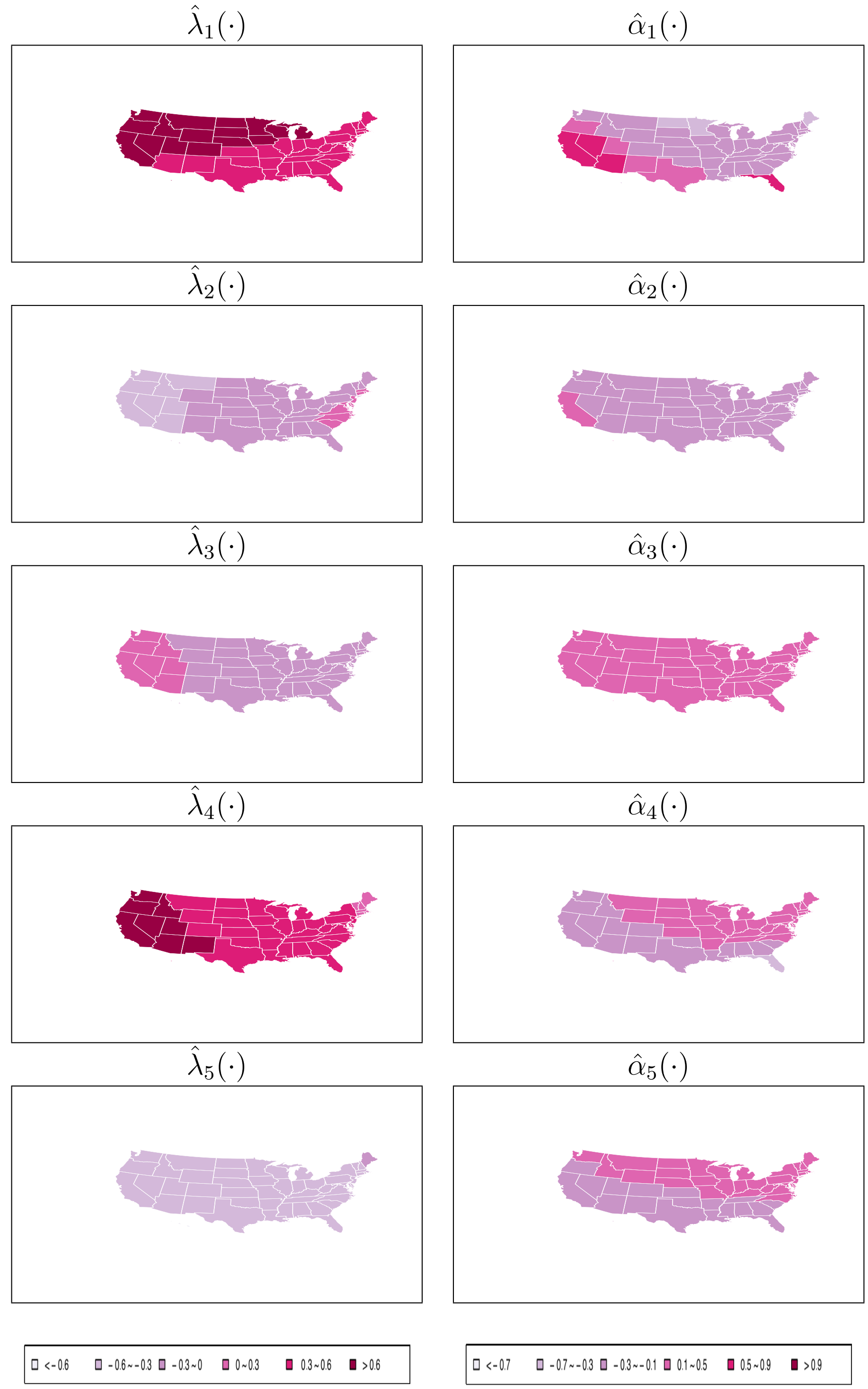

Figure 7: Maps of estimated coefficients $\hat{\lambda}_i$ for $i = 1, \ldots, 5$ (top to bottom in the first column) and $\hat{\alpha}_l$ for $l = 1, \ldots, 5$ (top to bottom in the second column).

To further evaluate our method, we consider comparison of the prediction based on different parametric forms of g function. The first is a linear function $g_{\rm L}(x,s) = a_0(s)+b_0(s)x$ where $a_0(s)$ and $b_0(s)$ are spatially-varying linear coefficients ((Thom, 1983; Reichert, 1990; Englund and Ioannides, 1997; McGibany and Nourzad, 2004)). In general, nonparametric specification can help to explore the parametrization of possibly nonlinear relationship, but itself may not give optimal prediction. Thus, a nonlinear threshold function based on Figure 6 is considered:

$$\begin{aligned} g_{\rm NL}(x,s) &= \{a_{10}(s)+a_{11}(s)x\}\,\mathcal{I}\,(x<-0.3)+\{a_{20}(s)+a_{21}(s)x\}\,\mathcal{I}\,(-0.3\le x<0.1) \\ &+ \{a_{30}(s)+a_{31}(s)x\}\,\mathcal{I}\,(0.1\le x<0.4)+\{a_{40}(s)+a_{41}(s)x\}\,\mathcal{I}\,(x\ge 0.4)\,, \end{aligned}$$

where $\mathcal{I}(\cdot)$ is an indicator function and $a_{kl}(s)$'s are the spatially-varying piecewise linear coefficients for $k=1,\ldots,4$ and $l=0,1$. With each of the parametric forms of g, we set aside the last 10 quarters for prediction and use the first $T=74$ quarters for model estimation with or without spatial smoothing after the time-series based estimation.

A mean squared prediction error (MSPE) of the one-step ahead prediction is computed for the linear and nonlinear forms of g and estimation with or without spatial smoothing. The MSPE values without spatial smoothing are 0.000782 and 0.000780 and those with spatial smoothing are 0.000624 and 0.000584 for $g_{\rm L}$ and $g_{\rm NL}$, respectively. These results demonstrate a clear advantage of using spatial smoothing in estimation for prediction, with a relative improvement approximately more than 17%. Further, compared with the linear $g_{\rm L}$, the threshold parametrization $g_{\rm NL}$ outperforms the $g_{\rm L}$ in prediction, with a relative improvement of 6.48%. These results further show that our methodology can help to uncover the relationship between the interest rate change and and geometric returns of housing prices which is more complex than linear. In addition, to assess the sensitivity of the selected model (15) with $p=q=5$, we consider the model simplified to a first order model of $p=q=1$, as suggested by a referee. Here we only report the MSPEs for the semiparametric prediction with g being a nonparametric function after spatial smoothing, which are 0.000642 and 0.000874 for $p=q=5$ and $p=q=1$, respectively. Obviously, our AIC selected model of $p=q=5$ performs much better than the model with $p=q=1$, as expected.

Finally we make some comments. (i) We have identified a threshold-like model for the impact of the interest rate change on the housing price return by our semiparametric modelling. In fact, the threshold phenomenon for interest rate has been well recognised in the literature (c.f., Pfann et al. (1997)). This seems well explain our finding, which looks reasonable and consistent with the reference of Pfann et al. (1997). (ii) In the analysis above, only the impact of interest rate is considered for simplicity of demonstration of the proposed methods. In practice, as commented by a referee, there are many other relevant variables that may impact housing price, in which case estimation of the function g in model (1) may also suffer from curse of dimensionality if the dimension of X_t is large, and further extension will hence be needed, say by allowing the function g to be of a kind of additive structure as in Gao et al. (2006). We leave this kind of partially linear additive spatio-temporal modelling of irregular sampling grids for future research. (iii) In this paper, we suppose $\{X_t\}$ is an exogenous time series variable, but its time series structure is not supposed (except α-mixing property needed in theory). As commented by a referee, it would be interesting to investigate if the lag parameters λ and α in model (1) are impacted by X_t, which need to develop a new method, quite different from what we do in this paper, for a functional-coefficient (depending on X_t) spatio-temporal model of irregular sampling grids. This is also left for future research.

6. Conclusions and Discussion

In this paper, we have developed a class of location-dependent spatio-temporally autoregressive partially (non)linear regression (STAR-PLR) models that allows for possibly nonlinear relationships between responses and covariates via a nonparametric function, possibly nonstationarity over space via spatially-varying autoregressive coefficients, and for both regular and irregular sampling spatial locations. The proposed methodology is supported by both asymptotic theory and finite sample properties via a simulation study. We have demonstrated the methodology to study housing prices and interest rate in the US, illustrating the usefulness of the proposed STAR-PLR model for uncovering complex relationships between housing prices and interest rate that are nonlinear and nonstationary over space.

Further extensions of the methodology are possible besides some discussions mentioned at the end of Section 5. For example, an important topic in housing price risk analysis is to investigate the impact of interest rates on housing price volatility. It would therefore be useful to extend the conditional mean modeling of this paper to conditional volatility modeling by developing a semiparametric spatio-temporal ARCH/GARCH type models. In addition, a data-driven approach to determine the spatial weights in the spatio-temporal models is another interesting issue in practice (Zhu et al., 2010). We leave such topics for future research.

Acknowledgement

The authors are grateful to both referees, the AE and the editor for all useful comments and suggestions, which have greatly helped to improve the presentation of this paper. This research was partially supported by an EU's Marie Curie Career Integration Grant, which is acknowledged.

References

Anselin, L., 1988. *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic Publishers.

Chow, Y., Teicher, H., 1988. *Probability Theory: Independence, Interchangibility, Martingale.* Springer.

Cressie, N., 1993. *Statistcs for Spatial Data*. New York: Wiley.

Cressie, N., Wikle, C., 2011. *Statistics for spatio-temporal data*. Wiley.

Engle, R., Granger, C., Rice, J., Weiss, A., 1986. Semi-parametric estimates of the relation between weather and electricity demand,. *Journal of the American Statistical Association 81*, 310–320.

Englund, P., Ioannides, Y., 1997. House price dynamics: an international empirical perspective. *Journal of Housing Economics 6*(2), 119–136.

Fan, J., Yao, Q., 2003. *Nonlinear time series: Nonparametric and parametric methods*. Springer Verlag.

Fan, J., Yao, Q., Cai, Z., 2003. Adaptive varying-coefficient linear models. *Journal of Royal Statistical Society B 65*(1), 57–80.

Gao, J., 1998. Semiparametric regression smoothing of non-linear time series. *Scandinavian Journal of Statistics 25*(3), 521–539.

Gao, J., 2007. *Nonlinear time series: semiparametric and nonparametric methods*. Chapman & Hall.

Gao, J., Lu, Z., Tjøstheim, D., 2006. Estimation in semiparametric spatial regression. *The Annals of Statistics 34*(3), 1395–1435.

Hallin, M., Lu, Z., Yu, K., 2009. Local linear spatial quantile regression. *Bernoulli 15*(3), 659–686.

Härdle, W., Liang, H., Gao, J., 2000. *Partially linear models*. Physica Verlag.

Hurvich, C., Simonoff, J., Tsai, C., 1998. On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society Series B 60*, 271–293.

Hurvich, C., Simonoff, J., Tsai, C., 2002. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society Series B 60*(2), 271–293.

Lahiri, S., Zhu, J., 2006. Resampling methods for spatial regression models under a class of stochastic designs. *Annals of Statistics, 34*, 1774–1813.

Li, D., Lu, Z., Linton, O., 2012. Local linear fitting under near epoch dependence: Uniform consistency with convergence rates. *Econometric Theory, 28*(05), 935–958.

Li, Q., Racine, J., 2007. *Nonparametric econometrics: Theory and Practice*. Princeton University Press, Princeton.

Lu, Z., Chen, X., 2002. Spatial nonparametric regression estimation: Non-isotropic case. *Acta Mathematicae Applicatae Sinica (English Series) 18*(4), 641–656.

Lu, Z., Linton, O., 2007. Local linear fitting under near epoch dependence. *Econometric Theory 23*(1), 37–70.

Lu, Z., Lundervold, A., Tjøstheim, D., Yao, Q., 2007. Exploring spatial nonlinearity using additive approximation. *Bernoulli 13*(2), 447–472.

Lu, Z., Steinskog, D., Tjøstheim, D., Yao, Q., 2009. Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society Series B 71*(4), 859–880.

Lu, Z., Tjøstheim, D., 2014. Nonparametric estimation of the probability density functions for irregularly observed spatial data. *Journal of the American Statistician Association 109*, 1546–1564.

Lu, Z., Tjøstheim, D., Yao, Q., 2008. Spatial smoothing, nugget effect and infill asymptotics. *Statistics & probability letters 78*(18), 3145–3151.

Lu, Z., Zhang, W., 2012. Semiparametric likelihood estimation in the survival models with informative censoring. *Journal of Multivariate Analysis 106*, 187–211.

McGibany, J., Nourzad, F., 2004. Do lower mortgage rates mean higher housing prices? *Applied Economics 36*(4), 305–313.

McQuinn, K., OReilly, G., 2007. A model of cross-country house prices. *Research Technical Paper 5*, Central Bank and Financial Services Authority of Ireland.

Pfann, G., Schotman, P., Tschernig, R., 1996. Nonlinear Interest Rate Dynamics and Implications for the Term Structure. *Journal of Ecorrometrics 74*, 149–176.

Robinson, P., 2011. Asymptotic Theory for Nonparametric Regression with Spatial Data. *Journal of Econometrics 165*, 5-19.

Rao, S., 2008. Statistical analysis of a spatio-temporal model with location-dependent parameters and a test for spatial stationarity. *Journal of Time Series Analysis 29*(4), 673–694.

Reichert, A., 1990. The impact of interest rates, income, and employment upon regional housing prices. *The Journal of Real Estate Finance and Economics 3*(4), 373–391.

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

Terasvirta, T., Tjøstheim, D., Granger, C., 2010. *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics. Oxford University Press.

Thom, R., 1983. House prices, inflation and the mortgage market. *The Economic and Social Review 15*(1), 57–68.

Tong, H., 1990. *Non-linear time series: a dynamical system approach.* Oxford University Press.

Wikle, C., Holan, S., 2011. Polynomial nonlinear spatio-temporal integro-difference equation models. *Journal of Time Series Analysis 32*(4), 339–350.

Wikle, C., Hooten, M., 2010. A general science-based framework for dynamical spatio-temporal models. *Test 19*(3), 417–451.

Wilhelmsson, M., 2002. Spatial models in real estate economics. *Housing, Theory and Society 19*(2), 92–101.

Zhang, W., Yao, Q., Tong, H., Stenseth, N., 2003. Smoothing for spatiotemporal models and its application to modeling muskrat-mink interaction. *Biometrics 59*(4), 813–821.

Zhu, J., Huang, H.-C., Reyes, P., 2010. On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society Series B 72*, 389–402.

Appendix A: Regularity Conditions

Let $S_N = \{s_1, \ldots, s_N\}$ denote the set of spatial sampling locations with a sampling density function $f(s)$ in the spatial domain $S \subset \mathbb{R}^2$. At time t, set $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_N))'$, $\mathbf{X}_t = (X_t(s_1), \ldots, X_t(s_N))'$, $G_t = (g(X_t(s_1), s_1), \ldots, g(X_t(s_N), s_N))'$, and $E_t = (\varepsilon_t(s_1), \ldots, \varepsilon_t(s_N))'$. Recall $\lambda_i(s_j) = 0$ for $q < i \le r$ or $\alpha_l(s_j) = 0$ for $p < l \le r$ for $r = \max(p, q)$. For $1 \le i \le r$, let A_i denote an $N \times N$ matrix whose (j,k)th element is $\alpha_i(s_j)$ if $j = k$ and $\lambda_i(s_j)w_{jk}$ otherwise. We can rewrite the STAR-PLR model in (1) as

$$\mathbf{Y}_t = G_t + \sum_{i=1}^{r} A_i \mathbf{Y}_{t-i} + E_t, \tag{16}$$

for $t = r+1, \ldots, T$.

For the strictly stationary time series $\{\mathbf{X}_t\}_{t=0,\pm1,\pm2,\ldots}$, we need the concept of α-mixing as follows for reference below. For $k = 1, 2, \ldots$, define

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{k}^{\infty}} |P(A)P(B) - P(AB)| \longrightarrow 0,$$

where $\mathcal{F}_i^j$ is σ-algebra generated by $\{\mathbf{X}_t\}_{i \le t \le j}$. The time series $\{\mathbf{X}_t\}$ is said to be an α-mixing process if the mixing coefficient $\alpha(k) \to 0$, as $k \to \infty$ (Fan and Yao, 2003).

We first state the regularity conditions for Theorems 1–2, given in Assumption **T** with time series data including conditions (C1)–(C7).

Assumption T:

(C1) (i) For each $s \in S_N$, the covariate process $\{X_t(s)\}$ is strictly stationary and α-mixing in time and $X_t(s)$ has a compact support $\mathbb{R}_X$ with the joint probability density function $p(x_1, x_2; s)$ of $X_{t_1}(s)$ and $X_{t_2}(s)$ being continuous and bounded from above for all $t_1 \neq t_2$ and $x_1, x_2 \in \mathbb{R}_X$. (ii) The α-mixing coefficient $\alpha(\cdot)$ satisfies $\lim_{k\to\infty} K^a \sum_{n=k}^{\infty} \{\alpha(n)\}^{\delta/(2+\delta)} = 0$ for some constant $a > \delta/(2+\delta)$.

(C2) The roots of $\det(I_N - \sum_{i=1}^{r} A_i z^i) = 0$ are outside the unit circle, where A_i is defined in (16) and I_N is an $N \times N$ identity matrix.

(C3) (i) For each $s \in S_N$, the functions $g_1(x, s) = E(Y_t(s)|X_t(s) = x)$ and $g_2(x, s) = E(Z_t(s)|X_t(s) = x)$ are continuous at all x and twice differentiable. (ii) The function $g(x, s)$ and the vector of autoregressive coefficients $\beta(s)$ are twice differentiable with respect to s.

(C4) (i) For each $s \in S$, the innovations $\{\varepsilon_t(s)\}_{t\geq r+1}$ are iid random variables independent of $\{X_t(s)\}_{t\geq r+1}$. Further, for each $t > r$, $\{\varepsilon_t(s)\}_{s\in S}$ are independent of $\left\{Y^{\text{sl}}_{t-i}(s)\right\}_{s\in S}$ for $i = 1, \ldots, p$, and $\{Y_{t-l}(s)\}_{s\in S}$ for $l = 1, \ldots, q$. (ii) For each t, the spatial covariance function $\gamma_t(s_1, s_2) \equiv Cov[\varepsilon_t(s_1), \varepsilon_t(s_2)]$ is bounded over $S \times S$. (iii) For each $s \in S$, $E\left[|\varepsilon_t(s)|^{2+\delta}\right] < \infty$ for some $\delta > 0$, $E\left[|Y_t(s)|^{2+\delta}\right] < \infty$, and $E\left[||Z_t(s)||^{2+\delta}\right] < \infty$.

(C5) The matrices $M(s)$ and $\Sigma_\beta(s)$ in Theorem 1 are positive definite for each $s \in S_N$.

(C6) (i) The kernel function $K(\cdot)$ is symmetric, uniformly bounded by some constant, and integrable. Further, $\int K(u)du = 1$ and $\int u^2 K(u)du < \infty$. (ii) $K(u)$ is Lipschitz continuous of order 1. (iii) $K(u)$ has an integrable second-order radial majorant (i.e., $Q^K(x) = \sup_{\|y\|\geq\|x\|} \left[\| y \|^2 K(y)\right]$ is integrable).

(C7) (i) The temporal bandwidth $b \to 0$ in such a way that $T_0 b \to \infty$ and $\log(T_0)/(T_0^{1/2} b) \to 0$ as $T_0 \to \infty$. (ii) There exist two sequences of positive integer vectors, $a_{T_0} \to \infty$ and $\eta_{T_0} \to \infty$, as $T_0 \to \infty$, such that $\eta_{T_0}/a_{T_0} \to 0$ and $T_0 a_{T_0}^{-1} \alpha(\eta_{T_0}) \to 0$. (iii) The temporal bandwidth $b \to 0$ in such manner that $\eta_T b = O(1)$ and $b^{-\delta/(2+\delta)} \sum_{t=\eta_{T_0}}^{\infty} \alpha(t)^{\delta/(2+\delta)} \to 0$ as $T_0 \to \infty$.

For spatial smoothing with Theorems 3–4, in additional to the above conditions, we need the following Assumption **S** including conditions (C8)–(C11).

Assumption S:

(C8) As $N \to \infty, N^{-1} \sum_{j=1}^{N} \mathcal{I}(s_j \in A) \longrightarrow \int_A f(s)ds$ for any measurable set $A \subset S \subset \mathbb{R}^2$ where the sampling density function f satisfies $f > 0$ in a neighborhood of $s_0 \in S$.

(C9) The kernel function $\tilde{K}(\cdot)$ satisfies $\int_{\mathbb{R}^2} \tilde{K}(z)dz = 1$, $\int_{\mathbb{R}^2} z\tilde{K}(z)dz = 0$ and $\int_{\mathbb{R}^2} zz'\tilde{K}(z)dz < \infty$.

(C10) (i) For each $t \geq r+1$ and $s \in S, \varepsilon_t(s) = \varepsilon_{1,t}(s)+\varepsilon_{2,t}(s)$, where $\{\varepsilon_{1,t}(s)\}$ and $\{\varepsilon_{2,t}(s)\}$ are two independent processes and both satisfy the condition C4(ii). Further, $\gamma_{1t}(s_j, s_k) \equiv Cov\,[\varepsilon_{1,t}(s_1), \varepsilon_{1,t}(s_2)]$ is continuous in (s_1, s_2) and $\gamma_{2t}(s_1, s_2) \equiv Cov\,[\varepsilon_{2,t}(s_1), \varepsilon_{2,t}(s_2)] = 0$ if $s_1 \neq s_2$ and $\gamma_{2t}(s_1, s_2) = \sigma_2^2(s_j) > 0$ is continuous in s_1. (ii) For each t, the matrix $M_*(s_1, s_2) = E\,[Z_t^*(s_1)Z_t^*(s_2)] = M_{*1}(s_1, s_2) + M_{*2}(s_1, s_2)$, where $M_{*1}(s_1, s_2)$ is continuous in (s_1, s_2), and $M_{*2}(s_1, s_2) = 0$ if $s_1 \neq s_2$ and $M_{*2}(s_1, s_2) = M_{*2}(s_1) > 0$ is continuous in s_1. (iii) For each t, the joint probability density function of $X_t(s_1)$ and $X_t(s_2)$ satisfies the following limit $\lim_{s_1,s_2\to s_0} p(x_1, x_2; s_1, s_2) = q(x_1, x_2; s_0)$ where $q(x_1, x_2; s_0)$ is continuous in both x_1 and x_2.

(C11) The spatial smoothing bandwidth $h \to 0$ and $Nh^2 \to \infty$, as $N \to \infty$.

The above regularity conditions are fairly mild. Condition (C1) assumes that the covariate process $X_t(s)$ has smooth, bounded probability density functions and is α-mixing over time (c.f. Fan and Yao (2003), pp. 68), which are quite standard in nonparametric time series analysis. The boundedness of $X_t(s)$ is for simplicity of proof; otherwise, we may use truncation argument for $X_t(s)$ as usually done in the literature (c.f., Gao et al. (2006)), with more tedious proof needed. (C2) is a stationarity condition assumed about the autoregressive coefficient matrices A_i's in (16), whereas (C3) assumes smoothness conditions about the functions g, g_1, g_2 and the vector of autoregressive coefficients $\beta(s)$ given in Section 2. Conditions (C4) and (C5) impose conditions on the model regarding the innovation processes as well as $Y_t(s)$ and $Z_t(s)$, which are mild. (C6) is a standard regularity condition imposed on the kernel function $K(\cdot)$ for the time-series based estimation while (C9) on $\tilde{K}(\cdot)$ is for spatial smoothing. Conditions (C7) and (C11) are the requirements about the temporal bandwidth $b = b_{T_0}$ and the spatial bandwidth $h = h_N$, respectively. Furthermore, over space, we impose (C8) on the spatial sampling intensity (density) (c.f. Lahiri and Zhu (2006)) and (C10) on the nugget effects for $\{\varepsilon_t(s)\}$, $Z_t^*(s)$ and $X_t(s)$, which are needed for spatial smoothing.

The conditions imposed on the time series in (C1)–(C7) are fairly mild and used in the literature (c.f. Fan and Yao (2003) and Gao (2007)). Similarly, conditions (C8)–(C11) have been used for spatial smoothing; c.f., Zhang et al. (2003), Lu et al. (2008) and Lu et al. (2009).

Web-based Supplementary Materials: Appendix B

"Estimation for Semiparametric Nonlinear Regression of Irregularly Located Spatial Time-series Data"

B.1. Proof of Theorems

Let $\xrightarrow{P}$ denote convergence in probability, $a^*_{T_0} = \{\log(T_0)/(T_0 b)\}^{1/2} + b^2$, and

$$U = p(x,s)\begin{pmatrix} 1 & 0 \\ 0 & \int u^2 K(u)du \end{pmatrix}.$$

Also, note that owing to (16), Y_t may not be α-mixing in general (c.f., Lu and Linton (2007) and Li et al. (2012)), and hence the Theorems may not follow from the α-mixing asymptotic results in the literature.

B.1.1. Proof of Theorem 1

Notation. Let $\hat{Y}_t(s) = Y_t(s) - \hat{E}\left[Y_t(s)|X_t(s)\right]$ and $\hat{Z}_t(s) = Z_t(s) - \hat{E}\left[Z_t(s)|X_t(s)\right]$, where $\hat{E}\left[Y_t(s)|X_t(s)\right] = \hat{g}_1(X_t(s), s)$ and $\hat{E}\left[Z_t(s)|X_t(s)\right] = \hat{g}_2(X_t(s), s)$. Recall also that $Z^*_t(s) = Z_t(s) - E[Z_t(s)|X_t(s)]$ in Theorem 1. Let $\Delta^Y_t(s) = E\left[Y_t(s)|X_t(s)\right] - \hat{E}\left[Y_t(s)|X_t(s)\right]$ and $\Delta^Z_t(s) = E\left[Z_t(s)|X_t(s)\right] - \hat{E}\left[Z_t(s)|X_t(s)\right]$.

Proof. Since by (9),

$$\begin{aligned} \hat{\beta}(s) - \beta(s) &= \left\{T_0^{-1}\sum_{t=r+1}^{T}\hat{Z}_t(s)\hat{Z}_t(s)'\right\}^{-1} T_0^{-1}\sum_{t=r+1}^{T}\hat{Z}_t(s)\left\{\hat{Y}_t(s) - \hat{Z}_t(s)'\beta(s)\right\} \\ &= A_{ZZ}^{-1}A_{ZY}, \end{aligned}$$

where

$$\begin{aligned} A_{ZZ} &= T_0^{-1}\sum_{t=r+1}^{T} Z^*_t(s)Z^*_t(s)' + T_0^{-1}\sum_{t=r+1}^{T} Z^*_t(s)\Delta^Z_t(s)' \\ &\quad + T_0^{-1}\sum_{t=r+1}^{T}\Delta^Z_t(s)Z^*_t(s)' + T_0^{-1}\sum_{t=r+1}^{T}\Delta^Z_t(s)\Delta^Z_t(s)' = \sum_{l=1}^{4} A_{ZZ,l} \end{aligned}$$

and

$$\begin{aligned} A_{ZY} &= T_0^{-1}\sum_{t=r+1}^{T}\varepsilon_t(s)Z^*_t(s) + T_0^{-1}\sum_{t=r+1}^{T} Z^*_t(s)\left\{\Delta^Y_t(s) - \Delta^Z_t(s)'\beta(s)\right\} \\ &\quad + T_0^{-1}\sum_{t=r+1}^{T}\varepsilon_t(s)\Delta^Z_t(s) + T_0^{-1}\sum_{t=r+1}^{T}\Delta^Z_t(s)\left\{\Delta^Y_t(s) - \Delta^Z_t(s)'\beta(s)\right\} = \sum_{l=1}^{4} A_{ZY,l}, \end{aligned}$$

it suffices to show that

$$A_{ZZ} \xrightarrow{P} M(s) \text{ and } T_0^{1/2} A_{ZY} \longrightarrow^{D} N(0, M(s)\sigma_\varepsilon^2(s)), \tag{B.1}$$

where $M(s)$ is defined in Theorem 1.

We first show that

$$\sum_{t=r+1}^{T} \left\{\hat{g}^{(m)}(X_t(s), s) - g^{(m)}(X_t(s), s)\right\}^2 = o_p(T_0^{1/2}),$$

for $m = 0, 1, \ldots, (p+q)$, where $g^{(0)}(x, s) = g_1(x, s)$ and $g^{(m)}(x, s)$ is the m-th component of $g_2(x, s)$, defined in Section 2.3, for $m = 1, \ldots, (p+q)$. By the uniform convergence theorem (Li et al. (2012), page 942),

$$\sup_{x \in \mathbb{R}_X} |\hat{g}^{(m)}(x, s) - g^{(m)}(x, s)| = O_p\left(a_{T_0}^*\right).$$

Since $\log(T_0)/(T_0^{1/2} b) \to 0$ and $T_0 b^4 \to 0$,

$$T_0^{1/2} \left[\{\log(T_0)/(T_0 b)\}^{1/2} + b^2\right]^2 \;=\; O(1)\left[\left\{\log(T_0)/(T_0^{1/2} b)\right\} + T_0^{1/2} b^4\right] \to 0.$$

Thus, for $\Delta_t^{(m)}(s) = \hat{g}^{(m)}(X_t(s), s) - g^{(m)}(X_t(s), s)$,

$$\sum_{t=r+1}^{T} \left\{\Delta_t^{(m)}(s)\right\}^2 = \sum_{t=r+1}^{T} \left\{\hat{g}^{(m)}(X_t(s), s) - g^{(m)}(X_t(s), s)\right\}^2 = o_p(T_0^{1/2}). \tag{B.2}$$

By the Cauchy-Schwarz inequality, as $T_0 \to \infty$, the (m, n)th element of $A_{ZZ,4}$ satisfies

$$\begin{aligned} |A_{ZZ,4}(m, n)| &= T_0^{-1} \Big| \sum_{t=r+1}^{T} \Delta_t^{(m)}(s) \Delta_t^{(n)}(s) \Big| \\ &\le T_0^{-1} \left[\sum_{t=r+1}^{T} \left\{\Delta_t^{(m)}(s)\right\}^2\right]^{1/2} \left[\sum_{t=r+1}^{T} \left\{\Delta_t^{(n)}(s)\right\}^2\right]^{1/2} = o_p(1). \end{aligned}$$

Similarly, we have $A_{ZZ,2} = o_p(1)$ and $A_{ZZ,3} = o_p(1)$. Thus, as $T_0 \to \infty$,

$$A_{ZZ} \xrightarrow{P} M(s) = E\left[Z_t^*(s) Z_t^*(s)'\right]. \tag{B.3}$$

Moreover, by condition (C4) and $T_0 b^4 \to 0$ together with the Cauchy-Schwarz inequality and (B.2), we have

$$T_0^{1/2} \sum_{l=2}^{4} A_{ZY,l} = o_p(1). \tag{B.4}$$

Finally, by the martingale central limit theorem (c.f., Chow and Teicher (1988), page 318), we have,

$$T_0^{1/2} A_{ZY,1} = T_0^{-1/2} \sum_{t=r+1}^{T} \varepsilon_t(s) Z_t^*(s) \longrightarrow^{D} N(0, M(s)\sigma_\varepsilon^2(s)). \tag{B.5}$$

With (B.3), (B.4) and (B.5), the proof is completed. □

B.1.2. Proof of Theorem 2

Notation. For $g_1(x,s) = E[Y_t(s)|X_t(s) = x]$, define

$$\begin{aligned} H_{T_0}^v = \begin{pmatrix} \hat{g}_1(x,s) - g_1(x,s) \\ \{\hat{\dot{g}}_1(x,s) - \dot{g}_1(x,s)\}b \end{pmatrix} &= U_{T_0}^{-1} V_{T_0} - \begin{pmatrix} g_1(x,s) \\ \dot{g}_1(x,s)b \end{pmatrix} \\ &= U_{T_0}^{-1} \left\{ V_{T_0} - U_{T_0} \begin{pmatrix} g_1(x,s) \\ \dot{g}_1(x,s)b \end{pmatrix} \right\} = U_{T_0}^{-1} W_{T_0}^v, \end{aligned} \tag{B.6}$$

where $\dot{g}_1(x,s)$ is the first order derivative of $g_1(x,s)$ with respect to x, and $Y_t^*(s) = Y_t(s) - a_0 - a_1\{X_t(s) - x\}$, $W_{T_0}^v = (W_{T_0 0}^v, W_{T_0 1}^v)'$ is given by, for $\left(\frac{X_t(s)-x}{b}\right)^0 = 1$,

$$\left(W_{T_0}^v\right)^j = (T_0 b)^{-1} \sum_{t=r+1}^{T} Y_t^*(s) \left(\frac{X_t(s) - x}{b}\right)^j K\left(\frac{X_t(s) - x}{b}\right), \tag{B.7}$$

for $j = 0, 1$.

Proof. Now for $g_2(x,s) = E[Z_t(s)|X_t(s) = x]$, recall that $g_2(x,s) = (g_{21}(x,s)', g_{22}(x,s)')'$, where $g_{21}(x,s) = (g_{21}^i(x,s))'$ with $g_{21}^i(x,s) = E[Y_{t-i}^{\mathrm{sl}}(s)|X_t(s) = x]$ for $i = 1, \ldots, p$ and $g_{22}(x,s) = \left(g_{22}^l(x,s)\right)'$ with $g_{22}^l(x,s) = E[Y_{t-l}(s)|X_t(s) = x]$ for $l = 1, \ldots, q$. Then, for $i = 1, \ldots, p$, let

$$\begin{aligned} H_{1T_0}^{ri} = \begin{pmatrix} \hat{g}_{21}^i(x,s) - g_{21}^i(x,s) \\ \{\hat{g}_{21}^{(1)i}(x,s) - g_{21}^{(1)i}(x,s)\}b \end{pmatrix} &= U_{T_0}^{-1} R_{1T_0}^i - \begin{pmatrix} g_{21}^i(x,s) \\ g_{21}^{(1)i}(x,s)b \end{pmatrix} \\ &= U_{T_0}^{-1} \left\{ R_{1T_0}^i - U_{T_0} \begin{pmatrix} g_{21}^i(x,s) \\ g_{21}^{(1)i}(x,s)b \end{pmatrix} \right\} = U_{T_0}^{-1} W_{1T_0}^{ri}, \end{aligned} \tag{B.8}$$

where $W_{1T_0}^{ri}$ comprises, for $j = 0, 1$,

$$\left(W_{1T_0}^{ri}\right)_j = (T_0 b)^{-1} \sum_{t=r+1}^{T} Y_{t-i}^{\mathrm{sl}*}(s) \left(\frac{X_t(s) - x}{b}\right)^j K\left(\frac{X_t(s) - x}{b}\right), \tag{B.9}$$

with $Y^{\mathrm{sl}*}_{t-i}(s) = Y^{\mathrm{sl}}_{t-i}(s) - c_0 - c_1\{X_t(s) - x\}$. Similarly, for $l = 1, \ldots, q$, let

$$H^{rl}_{2T_0} = \begin{pmatrix} \hat{g}^l_{22}(x,s) - g^l_{22}(x,s) \\ \{\hat{g}^{(1)l}_{22}(x,s) - g^{(1)l}_{22}(x,s)\}b \end{pmatrix} = U^{-1}_{T_0} R^l_{2T_0} - \begin{pmatrix} g^l_{22}(x,s) \\ g^{(1)l}_{22}(x,s)b \end{pmatrix}$$
$$= U^{-1}_{T_0}\left\{ R^l_{2T_0} - U_{T_0} \begin{pmatrix} g^l_{22}(x,s) \\ g^{(1)l}_{22}(x,s)b \end{pmatrix} \right\} = U^{-1}_{T_0} W^{rl}_{2T_0}, \tag{B.10}$$

where $W^{rl}_{2T_0}$ comprises, for $j = 0, 1$,

$$\left(W^{rl}_{2T_0}\right)_j = (T_0 b)^{-1} \sum_{t=r+1}^{T} Y^*_{t-l}(s) \left(\frac{X_t(s) - x}{b}\right)^j K\left(\frac{X_t(s) - x}{b}\right),$$

with $Y^*_{t-l}(s) = Y_{t-l}(s) - c_0 - c_1\{X_t(s) - x\}$.

Since $\hat{g}(x,s) = \hat{g}_1(x,s) - \hat{g}_2(x,s)'\hat{\beta}(s)$ estimates $g(x,s) = g_1(x,s) - g_2(x,s)'\beta(s)$,

$$\hat{g}(x,s) - g(x,s) = \{\hat{g}_1(x,s) - g_1(x,s)\} - \{\hat{g}_2(x,s) - g_2(x,s)\}' \beta(s) - \hat{g}_2(x,s)' \left\{\hat{\beta}(s) - \beta(s)\right\}.$$

From Theorem 1, $T_0^{1/2}\left\{\hat{\beta}(s) - \beta(s)\right\} = O_p(1)$ and $(T_0 b)^{1/2}\hat{g}_2(x,s)'\left\{\hat{\beta}(s) - \beta(s)\right\} = O_p(b^{1/2}) = o_p(1)$. Thus, to establish the asymptotic normality of $\hat{g}(x,s)$, it suffices to establish the asymptotic normality of $\hat{g}_1(x,s) - g_1(x,s)$ and $\hat{g}_2(x,s) - g_2(x,s)$.

For $W^r_{T_0} = \left(W^{r1}_{1T_0}, \ldots, W^{rp}_{1T_0}, W^{r1}_{2T_0}, \ldots, W^{rq}_{2T_0}\right)'$, by the arguments of Lemmas 3.2, 3.3, and 3.4 of Lu and Linton (2007), we have for $d = 1$

$$(T_0 b)^{1/2} \begin{pmatrix} U^{-1}W^v_{T_0} - U^{-1}E[W^v_{T_0}] \\ U^{-1}W^r_{T_0} - U^{-1}E[W^r_{T_0}] \end{pmatrix} \longrightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} U^{-1}\Sigma^{vv}(U^{-1})' & U^{-1}\Sigma^{vr}(U^{-1})' \\ U^{-1}\Sigma^{rv}(U^{-1})' & U^{-1}\Sigma^{rr}(U^{-1})' \end{pmatrix}\right), \tag{B.11}$$

where

$$\begin{aligned}
E[W^v_{T_0}] &= (b^2/2)\frac{\partial^2 g_1(x,s)}{\partial x^2} p(x,s) \begin{pmatrix} \int u^2 K(u)du \\ 0 \end{pmatrix} + o(b^2), \\
E[W^r_{T_0}] &= (b^2/2)\frac{\partial^2 g_2(x,s)}{\partial x^2} p(x,s) \begin{pmatrix} \int u^2 K(u)du \\ 0 \end{pmatrix} + o(b^2), \\
\Sigma^{vv} &= Var\left[Y_t(s)|X_t(s) = x\right] p(x,s) \begin{pmatrix} \int K^2(u)du & 0 \\ 0 & \int u^2 K^2(u)du \end{pmatrix}, \\
\Sigma^{vr} &= (\Sigma^{rv})' = Cov\left[Y_t(s), Z_t(s)|X_t(s) = x\right] p(x,s) \otimes \begin{pmatrix} \int K^2(u)du & 0 \\ 0 & \int u^2 K^2(u)du \end{pmatrix} \text{ and} \\
\Sigma^{rr} &= Var\left[Z_t(s)|X_t(s) = x\right] p(x,s) \otimes \begin{pmatrix} \int K^2(u)du & 0 \\ 0 & \int u^2 K^2(u)du \end{pmatrix},
\end{aligned}$$

where $\otimes$ stands for the Kroneck product. Thus,

$$(T_0 b)^{1/2}\left\{\left(\begin{array}{c}\hat{g}_1(x,s)-g_1(x,s)\\ \hat{g}_2(x,s)-g_2(x,s)\end{array}\right)-\left(\begin{array}{c}B_0^v(x,s)\\ B_0^r(x,s)\end{array}\right)\right\}\longrightarrow N\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}\Gamma^{vv}(s) & \Gamma^{vr}(s)\\ \Gamma^{rv}(s) & \Gamma^{rr}(s)\end{array}\right)\right), \tag{B.12}$$

where

$$\begin{aligned}
\Gamma^{vv}(s) &= Var\left[Y_t(s)|X_t(s)=x\right]p(x,s)^{-1}\int K^2(u)du,\\
\Gamma^{vr}(s) &= (\Gamma^{rv}(s))' = Cov\left[Y_t(s),Z_t(s)|X_t(s)=x\right]p(x,s)^{-1}\int K^2(u)du,\\
\Gamma^{rr}(s) &= Var\left[Z_t(s)|X_t(s)=x\right]p(x,s)^{-1}\int K^2(u)du,\\
B_0^v(x,s) &= (b^2/2)\frac{\partial^2 g_1(x,s)}{\partial x^2}\int u^2K(u)du+o_p(b^2) \text{ and}\\
B_0^r(x,s) &= (b^2/2)\frac{\partial^2 g_2(x,s)}{\partial x^2}\int u^2K(u)du+o_p(b^2).
\end{aligned} \tag{B.13}$$

Now by Slutsky's theorem, and noticing $g(x,s)=g_1(x,s)-g_2(x,s)'\beta(s)$,

$$(T_0 b)^{1/2}\{\hat{g}(x,s)-g(x,s)\} = (T_0 b)^{1/2}\left[\{\hat{g}_1(x,s)-g_1(x,s)\}-\{\hat{g}_2(x,s)-g_2(x,s)\}'\beta(s)\right]+o_P(1)$$

is asymptotically normal. Thus, Theorem 2 follows. □

B.1.3. Proof of Theorem 3

Proof. For kernel function $\tilde{K}^*_{h,j}(s_0)$ in Section 2.4, it is straightforward to verify that, under condition (C9),

$$N^{-1}\left(C'DC\right)\longrightarrow f(s_0)\left(\begin{array}{cc}1 & 0\\ 0 & \int zz'\tilde{K}(z)dz\end{array}\right),$$

where $C, D, \tilde{K}^*(\cdot), \tilde{K}(\cdot)$ are defined in Section 2.4. Further,

$$\sum_{j=1}^{N}\left(\frac{s_j-s_0}{h}\right)\tilde{K}^*_{h,j}(s_0)=0 \tag{B.14}$$

and

$$\sum_{j=1}^{N}\left(\frac{s_j-s_0}{h}\right)\left(\frac{s_j-s_0}{h}\right)'\tilde{K}^*_{h,j}(s_0)\longrightarrow\int zz'\tilde{K}(z)dz. \tag{B.15}$$

Now by Theorem 1, we have, for $s\in S$,

$$\hat{\beta}(s)-\beta(s)=T_0^{-1}M(s)^{-1}\sum_{t=r+1}^{T}\varepsilon_t(s)Z_t^*(s)+o_P(T^{-1/2}).$$

Then, for $s_0 \in S$, we have

$$\begin{aligned}
\tilde{\beta}(s_0) - \beta(s_0) &= \sum_{j=1}^{N} \left\{ \hat{\beta}(s_j) - \beta(s_j) \right\} \tilde{K}^*_{h,j}(s_0) + \sum_{j=1}^{N} \{\beta(s_j) - \beta(s_0)\} \tilde{K}^*_{h,j}(s_0) \\
&= \sum_{j=1}^{N} T_0^{-1} M(s_j)^{-1} \sum_{t=r+1}^{T} \varepsilon_t(s_j) Z_t^*(s_j) \tilde{K}^*_{h,j}(s_0) + \sum_{j=1}^{N} \{\beta(s_j) - \beta(s_0)\} \\
&\quad \times \tilde{K}^*_{h,j}(s_0) = A_1 + A_2, \qquad \text{(B.16)}
\end{aligned}$$

where A_1 and A_2 are associated with the variance and bias of $\tilde{\beta}(s_0)$, respectively.

For A_2, by Taylor's expansion and from (B.14) and (B.15), we have

$$\begin{aligned}
A_2 &= \sum_{j=1}^{N} \left\{ \frac{\partial \beta(s_0)}{\partial s'} (s_j - s_0) + (1/2)(s_j - s_0)' \frac{\partial^2 \beta(s_0)}{\partial s \partial s'} (s_j - s_0) \right\} \tilde{K}^*_{h,j}(s_0) \\
&= h \frac{\partial \beta(s_0)}{\partial s} \sum_{j=1}^{N} \left(\frac{s_j - s_0}{h} \right) \tilde{K}^*_{h,j}(s_0) \\
&\quad + (h^2/2) \frac{\partial^2 \beta(s_0)}{\partial s \partial s'} \sum_{j=1}^{N} \left(\frac{s_j - s_0}{h} \right) \left(\frac{s_j - s_0}{h} \right)' \tilde{K}^*_{h,j}(s_0) \\
&= (h^2/2) \mathrm{tr} \left\{ \frac{\partial^2 \beta(s_0)}{\partial s \partial s'} \int z z' \tilde{K}(z) dz \right\} \{1 + o(1)\}. \qquad \text{(B.17)}
\end{aligned}$$

For A_1, it is clear that $E[A_1] = 0$. Thus, we have

$$\begin{aligned}
& Var\left[\tilde{\beta}(s_0) - \beta(s_0) \right] \\
=& \sum_{j=1}^{N} T_0^{-2} M(s_j)^{-1} \sum_{t=r+1}^{T} Var\left[\varepsilon_t(s_j) Z_t^*(s_j) \right] \{M(s_j)^{-1}\}' \left\{ \tilde{K}^*_{h,j}(s_0) \right\}^2 \\
+& \sum_{j \neq k=1}^{N} T_0^{-2} M(s_j)^{-1} \sum_{t=r+1}^{T} Cov\left[\varepsilon_t(s_j) Z_t^*(s_j), \varepsilon_t(s_k) Z_t^*(s_k) \right] M(s_k)^{-1} \tilde{K}^*_{h,j}(s_0) \tilde{K}^*_{h,k}(s_0) \\
=& V_1 + V_2,
\end{aligned}$$

where

$$
\begin{aligned}
V_1 &= \sum_{j=1}^{N} T_0^{-2} M(s_j)^{-1} \sum_{t=r+1}^{T} E\left\{Z_t^*(s_j) Z_t^*(s_j)'\right\} Var\left[\varepsilon_t(s_j)\right] \left\{M(s_j)^{-1}\right\}' \left\{\tilde{K}_{h,j}^*(s_0)\right\}^2 \\
&= \sum_{j=1}^{N} T_0^{-1} M(s_j)^{-1} E\left[Z_t^*(s_j) Z_t^*(s_j)'\right] \Gamma(s_j, s_j) \{M(s_j)^{-1}\}' \{N^2 h^4 f^2(s_0)\}^{-1} \\
&\qquad\qquad \times \tilde{K}^2\left(\frac{s_j - s_0}{h}\right) \{1 + o(1)\} \\
&= \sigma^2(s_0)(T_0 N)^{-1} M(s_0)^{-1} \int \{h^4 f^2(s_0)\}^{-1} \tilde{K}^2\left(\frac{s - s_0}{h}\right) f(s) ds \{1 + o(1)\} \\
&= \sigma^2(s_0)(T_0 N h^2)^{-1} M(s_0)^{-1} \{f^2(s_0)\}^{-1} \int \tilde{K}^2(z) f(s_0 + hz) dz \{1 + o(1)\} \\
&= \sigma^2(s_0)\{T_0 N h^2 f(s_0)\}^{-1} M(s_0)^{-1} \int \tilde{K}^2(z) dz \{1 + o(1)\}, \qquad \text{(B.18)}
\end{aligned}
$$

and under condition C9(ii),

$$
\begin{aligned}
V_2 &= \sum_{j \neq k=1}^{N} T_0^{-2} M(s_j)^{-1} \sum_{t=r+1}^{T} E\left\{Z_t^*(s_j) Z_t^*(s_k)'\right\} Cov\left[\varepsilon_t(s_j), \varepsilon_t(s_k)\right] M(s_k)^{-1} \\
&\quad \times \tilde{K}_{h,j}^*(s_0)\, \tilde{K}_{h,k}^*(s_0) \\
&= \sum_{j \neq k=1}^{N} M(s_j)^{-1} T_0^{-1} M_*(s_j, s_k) \Gamma_1(s_j, s_k) M(s_k)^{-1} \{N^2 h^4 f^2(s_0)\}^{-1} \\
&\quad \times \tilde{K}\left(\frac{s_j - s_0}{h}\right) \tilde{K}\left(\frac{s_k - s_0}{h}\right) \{1 + o(1)\} \\
&= \sigma_1^2(s_0) T_0^{-1} M(s_0)^{-1} M_{*1}(s_0) M(s_0)^{-1} \{h^4 f^2(s_0)\}^{-1} \int \tilde{K}\left(\frac{s^* - s_0}{h}\right) f(s^*) ds^* \\
&\quad \times \int \tilde{K}\left(\frac{s^* - s_0}{h}\right) f(s^*) ds^* \{1 + o(1)\} \\
&= \sigma_1^2(s_0)\{T_0 f^2(s_0)\}^{-1} M(s_0)^{-1} M_{*1}(s_0) M(s_0)^{-1} \int \tilde{K}(z) f(s_0 + hz) dz \\
&\quad \times \int \tilde{K}(y) f(s_0 + hy) dy \{1 + o(1)\} \\
&= \sigma_1^2(s_0) T_0^{-1} M(s_0)^{-1} M_{*1}(s_0) M(s_0)^{-1} \{1 + o(1)\}. \qquad \text{(B.19)}
\end{aligned}
$$

It follows from (B.18) and (B.19) that the asymptotic variance is

$$
T_0^{-1}\left[\sigma^2(s_0)\{N h^2 f(s_0)\}^{-1} M(s_0)^{-1} \int \tilde{K}^2(z) dz + \sigma_1^2(s_0) M(s_0)^{-1} M_{*1}(s_0) M(s_0)^{-1}\right]
$$

which is $T_0^{-1}\nu^2(z, s_0)$, and together with (B.17), thus the proof for asymptotic variance and bias is completed.

Finally, as done in the proof of Theorem 1, the asymptotic normality follows from (B.16) by letting $T \to \infty$ first and then $N \to \infty$, and hence $\xi(s_0)$ is of Gaussian distribution. The proof is completed. □

B.1.4. Proof of Theorem 4

Proof. By Theorem 2, we have

$$\begin{aligned} & \hat{g}(x,s) - g(x,s) \\ = \; & \{\hat{g}_1(x,s) - g_1(x,s)\} - \{\hat{g}_2(x,s) - g_2(x,s)\}' \beta(s) + o_P((T_0 b)^{-1/2}) \\ = \; & \{\hat{g}_1(x,s) - \hat{g}_2(x,s)'\beta(s)\} - \{g_1(x,s) - g_2(x,s)'\beta(s)\} + o_P((T_0 b)^{-1/2}). \end{aligned}$$

Thus,

$$\begin{aligned} \hat{g}(x,s) - g(x,s) \; = \; & e_1' \left\{A(x)'B(x)A(x)\right\}^{-1} A(x)'B(x) \left(Y^* - G\right) \\ & + e_1' \left\{A(x)'B(x)A(x)\right\}^{-1} A(x)'B(x)G - g(x,s) + o_P((T_0 b)^{-1/2}), \end{aligned}$$

where $Y^* = (Y_{r+1}(s) - Z_{r+1}(s)'\beta(s), \ldots, Y_T(s) - Z_T(s)'\beta(s))'$, $G = (g(X_{r+1}(s), s), \ldots, g(X_T(s), s))'$, and both $A(x)$ and $B(x)$ are defined in Section 2.

Then, for $\varepsilon = (\varepsilon_{r+1}(s), \ldots, \varepsilon_T(s))'$, we have

$$\begin{aligned} & e_1' \left\{A(x)'B(x)A(x)\right\}^{-1} A(x)'B(x) \left(Y^* - G\right) = e_1' \left\{A(x)'B(x)A(x)\right\}^{-1} A(x)'B(x)\varepsilon \\ & = \{p(x,s)T_0 b\}^{-1} \sum_{t=r+1}^{T} \varepsilon_t(s) K\left(\frac{X_t(s) - x}{b}\right) \{1 + o_p(1)\}, \text{ and} \end{aligned}$$

$$e_1' \left\{A(x)'B(x)A(x)\right\}^{-1} A(x)'B(x)G = \{p(x,s)T_0 b\}^{-1} \sum_{t=r+1}^{T} K\left(\frac{X_t(s) - x}{b}\right) g(X_t(s), s)\{1 + o_p(1)\}.$$

By Taylor's expansion of $g(X_t(s), s)$, we have

$$\begin{aligned} & \{p(x,s)T_0 b\}^{-1} \sum_{t=r+1}^{T} K\left(\frac{X_t(s) - x}{b}\right) g(X_t(s), s) \\ & = b\{2T_0 p(x,s)\}^{-1} \frac{\partial^2 g(x,s)}{\partial x^2} \sum_{t=r+1}^{T} \left\{\frac{X_t(s) - x}{b}\right\}^2 K\left(\frac{X_t(s) - x}{b}\right) \\ & = b^2 \{2p(x,s)\}^{-1} \frac{\partial^2 g(x,s)}{\partial x^2} \int u^2 K(u) p(x + bu, s) du \{1 + o_p(1)\} \\ & = (b^2/2) \frac{\partial^2 g(x,s)}{\partial x^2} \int u^2 K(u) du \{1 + o_p(1)\}. \end{aligned}$$

Thus,

$$\begin{aligned}\hat{g}(x,s)-g(x,s) &= \{p(x,s)T_0 b\}^{-1}\sum_{t=r+1}^{T}\varepsilon_t(s)K\left(\frac{X_t(s)-x}{b}\right)\{1+o_p(1)\} \\ &\quad +(b^2/2)\frac{\partial^2 g(x,s)}{\partial x^2}\int u^2K(u)du\{1+o_p(1)\}. \end{aligned} \tag{B.20}$$

It follows from (B.20) that, as $T_0\to\infty$,

$$\begin{aligned}\tilde{g}(x,s_0)-g(x,s_0) &= \sum_{j=1}^{N}\{\hat{g}(x,s_j)-g(x,s_j)\}\tilde{K}^*_{h,j}(s_0)+\sum_{j=1}^{N}\{g(x,s_j)-g(x,s_0)\}\tilde{K}^*_{h,j}(s_0) \\ &= \sum_{j=1}^{N}\{p(x,s_j)T_0 b\}^{-1}\sum_{t=r+1}^{T}\varepsilon_t(s_j)K\left(\frac{X_t(s_j)-x}{b}\right)\tilde{K}^*_{h,j}(s_0)\{1+o_p(1)\} \\ &\quad +\sum_{j=1}^{N}(b^2/2)\frac{\partial^2 g(x,s_j)}{\partial x^2}\int u^2K(u)du\tilde{K}^*_{h,j}(s_0)\{1+o_p(1)\} \\ &\quad +\sum_{j=1}^{N}\{g(x,s_j)-g(x,s_0)\}\tilde{K}^*_{h,j}(s_0)+o_p(1) \\ &= I_1\{1+o_p(1)\}+I_2\{1+o_p(1)\}+I_3+o_p(1), \end{aligned} \tag{B.21}$$

where

$$\begin{aligned} I_2 &= \sum_{j=1}^{N}(b^2/2)\frac{\partial^2 g(x,s_j)}{\partial x^2}\int u^2K(u)du\tilde{K}^*_{h,j}(s_0) \\ &= (b^2/2)\frac{\partial^2 g(x,s_0)}{\partial x^2}\int\int u^2K(u)du\{h^2f(s_0)\}^{-1}\tilde{K}\left(\frac{s-s_0}{h}\right)f(s)ds\{1+o(1)\} \\ &= b^2/\{2h^2f(s_0)\}\frac{\partial^2 g(x,s_0)}{\partial x^2}\int\int u^2K(u)du\tilde{K}\left(\frac{s-s_0}{h}\right)f(s)ds\{1+o(1)\} \\ &= b^2/\{2h^2f(s_0)\}\frac{\partial^2 g(x,s_0)}{\partial x^2}\int u^2K(u)du\int\tilde{K}(s^*)f(s_0+hs^*)ds^*\{1+o(1)\} \\ &= (b^2/2)\frac{\partial^2 g(x,s_0)}{\partial x^2}\int u^2K(u)du\{1+o(1)\}. \end{aligned} \tag{B.22}$$

Also, from (B.14), (B.15), and by Taylor's expansion, we have

$$
\begin{aligned}
I_3 &= \sum_{j=1}^{N} \{g(x,s_j) - g(x,s_0)\} \tilde{K}^*_{h,j}(s_0) \\
&= \sum_{j=1}^{N} \frac{\partial g(x,s_0)}{\partial s}(s_j - s_0)\tilde{K}^*_{h,j}(s_0) + (1/2)\sum_{j=1}^{N}(s_j - s_0)'\frac{\partial^2 g(x,s_0)}{\partial s \partial s'}(s_j - s_0)\tilde{K}^*_{h,j}(s_0) \\
&= h\frac{\partial g(x,s_0)}{\partial s}\sum_{j=1}^{N}\left(\frac{s_j - s_0}{h}\right)\tilde{K}^*_{h,j}(s_0) \\
&\quad +(h^2/2)\frac{\partial^2 g(x,s_0)}{\partial s \partial s'}\sum_{j=1}^{N}\left(\frac{s_j - s_0}{h}\right)\left(\frac{s_j - s_0}{h}\right)'\tilde{K}^*_{h,j}(s_0) \\
&= (h^2/2)\int z'\frac{\partial^2 g(x,s_0)}{\partial s \partial s'} z\tilde{K}(z)dz\{1 + o_p(1)\}. \qquad \text{(B.23)}
\end{aligned}
$$

Further, $E[I_1] = 0$ and

$$
\begin{aligned}
E[I_1^2] &= (T_0 b)^{-2}\sum_{j=1}^{N}\{p^2(x,s_j)\}^{-1}\sum_{t=r+1}^{T} E[\varepsilon_t^2(s_j)]K^2\left(\frac{X_t(s_j) - x}{b}\right)\left\{\tilde{K}^*_{h,j}(s_0)\right\}^2 \\
&\quad +(T_0 b)^{-2}\sum_{j\neq k=1}^{N}\{p(x,s_j)p(x,s_k)\}^{-1} \\
&\quad \times \sum_{t=r+1}^{T} E[\varepsilon_t(s_j)\varepsilon_t(s_k)]K\left(\frac{X_t(s_j) - x}{b}\right)K\left(\frac{X_t(s_k) - x}{b}\right)\tilde{K}^*_{h,j}(s_0)\tilde{K}^*_{h,k}(s_0) \\
&= I_{11} + I_{12}. \qquad \text{(B.24)}
\end{aligned}
$$

In particular,

$$
\begin{aligned}
I_{11} &= (T_0 b)^{-2}\sum_{j=1}^{N}\{p(x,s_j)\}^{-2}\sum_{t=r+1}^{T}\sigma^2(s_j)EK^2\left(\frac{X_t(s_j) - x}{b}\right)\left\{\tilde{K}^*_{h,j}(s_0)\right\}^2 \\
&= (T_0 b)^{-1}\sum_{j=1}^{N}\sigma^2(s_j)\{p(x,s_j)\}^{-2}\left\{\tilde{K}^*_{h,j}(s_0)\right\}^2\int K^2(u)p(x+bu,s_j)du \\
&= (T_0 b)^{-1}\sum_{j=1}^{N}\sigma^2(s_j)\{p(x,s_j)\}^{-1}\int K^2(u)du\left[\left\{Nh^2 f(s_0)\right\}^{-1}\tilde{K}\left(\frac{s_j - s_0}{h}\right)\right]^2\{1 + o(1)\} \\
&= \sigma^2(s_0)\{T_0 bNh^2 f^2(s_0)\}^{-1}\int p(x,s_0)^{-1}K^2(u)du\int \tilde{K}^2(z)f(s_0 + hz)dz\{1 + o(1)\} \\
&= \sigma^2(s_0)\{T_0 bNh^2 p(x,s_0)f(s_0)\}^{-1}\int K^2(u)du\int \tilde{K}^2(z)dz\{1 + o(1)\} \qquad \text{(B.25)}
\end{aligned}
$$

and, under condition C9(iii),

$$
\begin{aligned}
I_{12} &= (T_0 b)^{-2} \sum_{j\neq k=1}^{N} \{p(x,s_j)p(x,s_k)\}^{-1} \sum_{t=r+1}^{T} \Gamma(s_j,s_k) E K\left(\frac{X_t(s_j)-x}{b}\right) K\left(\frac{X_t(s_k)-x}{b}\right) \\
&\quad \times \tilde{K}^{*}_{h,j}(s_0)\,\tilde{K}^{*}_{h,k}(s_0) \\
&= (T_0 b)^{-2} \sum_{j\neq k=1}^{N} \{p(x,s_j)p(x,s_k)\}^{-1} \sum_{t=r+1}^{T} \sigma_1^2(s_j) K\left(\frac{X_t(s_j)-x}{b}\right) K\left(\frac{X_t(s_k)-x}{b}\right) \\
&\quad \times \left[\{Nh^2 f(s_0)\}^{-1}\tilde{K}\left(\frac{s_j-s_0}{h}\right)\right]\left[\{Nh^2 f(s_0)\}^{-1}\tilde{K}\left(\frac{s_k-s_0}{h}\right)\right]\{1+o(1)\} \\
&= T_0^{-1} \sum_{j\neq k=1}^{N} \{p(x,s_j)p(x,s_k)\}^{-1}\sigma_1^2(s_j) E\left[K_b\left(X_t(s_j)-x\right)K_b\left(X_t(s_k)-x\right)\right] \\
&\quad \times \left[\{Nh^2 f(s_0)\}^{-1}\tilde{K}\left(\frac{s_j-s_0}{h}\right)\right]\left[\{Nh^2 f(s_0)\}^{-1}\tilde{K}\left(\frac{s_k-s_0}{h}\right)\right]\{1+o(1)\} \\
&= T_0^{-1}\sigma_1^2(s_0)\{p(x,s_0)\}^{-2} q(x,x;s_0) f(s_0)^{-2} \int \tilde{K}(z) f(s_0+hz)dz \int \tilde{K}(y) f(s_0+hy)dy\{1+o(1)\} \\
&= T_0^{-1}\sigma_1^2(s_0)\{p(x,s_0)\}^{-2} q(x,x;s_0)\{1+o(1)\}. \qquad \text{(B.26)}
\end{aligned}
$$

Thus, from (B.25) and (B.26), we have

$$
\begin{aligned}
&(T_0 b)^{-1}\left[b\sigma_1^2(s_0)\{p(x,s_0)\}^{-2}q(x,x;s_0) + \sigma^2(s_0)\{Nh^2 p(x,s_0) f(s_0)\}^{-1}\int K^2(u)du \int \tilde{K}^2(z)dz\right] \\
&\quad \times\{1+o_p(1)\} = (T_0 b)^{-1}\nu_1^2(x,s_0).
\end{aligned}
$$

Together with (B.22) and (B.23), the proof for asymptotic variance and bias is completed.

Finally, as done in the proof of Theorem 2, the asymptotic normality follows from (B.21) by letting $T\to\infty$ first and then $N\to\infty$, and hence $\eta(s_0)$ is of Gaussian distribution. The proof is completed. □

B.2. Hat Matrix *H* in (14)

We specify the hat matrix H in (14) with respect to model (1). Denote the vector of fitted values by $\hat{Y}$ such that $\hat{Y}=HY$ with

$$
\overbrace{\begin{pmatrix} \hat{Y}(s_1) \\ \hat{Y}(s_2) \\ \vdots \\ \hat{Y}(s_N) \end{pmatrix}}^{\hat{Y}} = \overbrace{\begin{pmatrix} H_{11} & H_{12} & \cdots & H_{1N} \\ H_{21} & H_{22} & \cdots & H_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ H_{N1} & H_{N2} & \cdots & H_{NN} \end{pmatrix}}^{H} \overbrace{\begin{pmatrix} Y(s_1) \\ Y(s_2) \\ \vdots \\ Y(s_N) \end{pmatrix}}^{Y}.
$$

Here, the hat matrix H is an $NT_0 \times NT_0$ matrix with $T_0 \times T_0$ sub-matrix $H_{jk} = 0$ for $j \neq k$, $j, k = 1, \ldots, N$ and $\hat{Y}(s_j) = H_{jj} Y(s_j)$, with $\hat{Y}(s_j) = (\hat{Y}_{r+1}(s_j), \ldots, \hat{Y}_T(s_j))'$ and $Y(s_j) = (Y_{r+1}(s_j), \ldots, Y_T(s_j))'$.

To define H_{jj}, note that, by model (1),

$$\hat{Y}_t(s_j) = \hat{g}(X_t(s_j), s_j) + Z_t(s_j)'\hat{\beta}(s_j),$$

where $\hat{\beta}(s_j)$ and $\hat{g}(X_t, s_j)$ are given in (9) and (10). Hence, denoting

$$\hat{g}(X, s_j) = (\hat{g}(X_{r+1}(s_j), s_j), \ldots, \hat{g}(X_T(s_j), s_j))' \text{ and } Z(s_j) = (Z_{r+1}(s_j), \ldots, Z_T(s_j))',$$

we have

$$\begin{aligned} \hat{Y}(s_j) &= \hat{g}(X, s_j) + Z(s_j)\hat{\beta}(s_j) \\ &= \hat{g}_1(X, s_j) + \{Z(s_j) - \hat{g}_2(X, s_j)\}\, \hat{\beta}(s_j) \\ &= \hat{g}_1(X, s_j) + \{Z(s_j) - \hat{g}_2(X, s_j)\} \left\{\hat{Z}(s_j)'\hat{Z}(s_j)\right\}^{-1} \hat{Z}(s_j)'\hat{Y}(s_j), \end{aligned}$$

where $\hat{Z}(s_j)$ is defined similarly to $Z(s_j)$, with $\hat{Z}_t(s_j) = Z_t(s_j) - \hat{E}[Z_t(s_j)|X_t(s_j)]$, while $\hat{g}_1(X, s_j)$ and $\hat{g}_2(X, s_j)$ are defined similarly to $\hat{g}(X, s_j)$, with $\hat{g}_1(X_t(s_j), s_j) = \hat{E}[Y_t(s_j)|X_t(s_j)]$ and $\hat{g}_2(X_t(s_j), s_j) = \hat{E}[Z_t(s_j)|X_t(s_j)]$, respectively, used in (9). It follows from Theorem 1 that in calculating $\hat{\beta}(s_j)$, a bandwidth b^* smaller than the optimal bandwidth by AICc, b, is needed, which, according to empirical experience, is set as $b^* = 0.75b$ (c.f., (Lu and Zhang, 2012)) in numerical examples. Then, we have

$$\begin{aligned} \hat{Y}(s_j) &= \hat{g}_1(X, s_j) + \{Z(s_j) - \hat{g}_2(X, s_j)\} \left\{\hat{Z}(s_j)'\hat{Z}(s_j)\right\}^{-1} \hat{Z}(s_j)' \left\{Y(s_j) - \hat{E}[Y(s_j)|X]\right\} \\ &= \hat{g}_1(X, s_j) + \{Z(s_j) - \hat{g}_2(X, s_j)\} \left\{\hat{Z}(s_j)'\hat{Z}(s_j)\right\}^{-1} \hat{Z}(s_j)'Y(s_j) \\ &\quad - \{Z(s_j) - \hat{g}_2(X, s_j)\} \left\{\hat{Z}(s_j)'\hat{Z}(s_j)\right\}^{-1} \hat{Z}(s_j)'\hat{E}[Y(s_j)|X] \\ &= H_{1,j}Y(s_j) + H_{2,j}Y(s_j) - H_{2j}H_{1,j}Y(s_j) = \{H_{1,j} + H_{2,j}(I - H_{1,j})\}\, Y(s_j) = H_{jj}Y(s_j), \end{aligned}$$

where $H_{2,j} = \{Z(s_j) - \hat{g}_2(X, s_j)\} \left\{\hat{Z}(s_j)'\hat{Z}(s_j)\right\}^{-1} \hat{Z}(s_j)'$ and $H_{1,j}$ is a $T_0 \times T_0$ matrix whose $(t-r)$th row is of the form $e_1' \{A(X_t(s_j))'B(X_t(s_j))A(X_t(s_j))\}^{-1} A(X_t(s_j))'B(X_t(s_j))$, for $t = r+1, \ldots, T$, with both $A(x)$ and $B(x)$ defined in Section 2.