

β -Spline Estimation in a Semiparametric Regression Model with Nonlinear Time Series Errors

¹Jinhong You, ²Gemai Chen and ³Xian Zhou

¹Department of Biostatistics, University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-7400, USA

²Department of Mathematics and Statistics, University of Calgary
Calgary, Alberta, Canada T2N 1N4

³Department of Applied Mathematics, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong, P.R. China

Abstract: We study the estimation problems for a partly linear regression model with a nonlinear time series error structure. The model consists of a parametric linear component for the regression coefficients and a nonparametric nonlinear component. The random errors are unobservable and modeled by a first-order Markov bilinear process. Based on a B-spline series approximation of the nonlinear component function, we propose a semiparametric ordinary least squares estimator and a semiparametric generalized least squares estimator of the regression coefficients, a least squares estimator of the autoregression parameter for the errors, and a B-spline series estimator of the nonparametric component function. The asymptotic properties of these estimators are investigated and their asymptotic distributions are derived. We also provide a consistent estimator for the asymptotic covariance matrix of the semiparametric generalized least squares estimator of the regression coefficients. Our results can be used to make asymptotically efficient statistical inferences. In addition, a small simulation is conducted to evaluate the performance of the proposed estimators, which shows that the semiparametric generalized least squares estimator of the regression coefficients is more efficient than the semiparametric ordinary least squares estimator.

Key words: Semiparametric regression model, first-order Markovian bilinear process, β -splines series estimation, semiparametric least squares estimators, asymptotic normality

INTRODUCTION

Partly linear regression models have attracted a lot of research interests due to their flexibility to allow both linear and nonlinear components, as well as serially correlated errors, which enables them to better describe increasingly complex data from the real world than pure parametric or nonparametric models. Consider a partly linear regression model of the following form:

$$y_i = x_i' \beta + g(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

where, y_i 's are responses, $x_i = (x_{i1}, \dots, x_{ip})'$ and t_i are called the *design points*, $\beta = (\beta_1, \dots, \beta_p)'$ is an unknown parameter vector representing the linear component, $g(\cdot)$ is an unknown function defined on $[0, 1]$ for the nonlinear component, and ε_i 's are unobservable random errors.

The model in (1.1) has been extensively studied by many researchers. A brief review of relevant literature is as follows. When ε_i are i.i.d. Random variables^[1-5]

used various estimation methods, such as the kernel method, spline method, series estimation, local linear estimation, M-estimation and two-stage estimation, to obtain estimates of the unknown quantities in (1.1). They also discussed the asymptotic properties of these estimators. However, the independence assumption for the errors is not always appropriate in applications, especially for sequentially collected economic data, which often exhibit evident serial dependence in the errors. For example, in the process of fitting the relationship between temperature and electricity usage^[6] found that the data are serially correlated. When ε_i is an autoregressive (AR) series^[7] studied an estimator of the autocorrelation coefficient^[8] considered the estimation problem for model (1.1) with linear time series errors.

It is well known that not all correlated errors can be fitted well by linear time series errors. Therefore, much attention has been shifted to nonlinear time series models in the recent literature^[9] and the references therein. There have been many papers concerned with the ordinary linear models with nonlinear time series errors. For example, under the assumption of random coefficient autoregressive errors^[10] investigated the

Corresponding Author: Xian Zhou, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, P.R. China

limit distribution of the least squares estimators of the regression and auto regression parameters. Moreover^[11] used the framework of the information matrix (IM) test to develop a test for the linear regression model when the errors are from an autoregressive conditional heteroscedastic (ARCH) process^[12]. Derived the Wald and Rao's Score test statistics for testing the effects of additional regression parameters. There has, however, been little work on the partial linear regression model with nonlinear time series errors in the literature except^[13].^[13] investigated the estimating problems of partial linear regression models with random coefficient autoregressive errors.

In this study, by approximating the nonparametric component with B-spline series we study the problems of estimating the parametric and nonparametric components of the partial linear regression model (1.1) with a nonlinear time series error structure. More specifically, we consider a first-order Markovian bilinear error process \mathcal{E}_i , which is a stationary solution of :

$$\mathcal{E}_i = (\phi + \theta e_i)\mathcal{E}_{i-1} + e_i, \quad i = 1, \dots, n, \quad (1.2)$$

where, $\{e_i\}$ is a zero mean process consisting of i.i.d. random variables with finite second moments σ_e^2 . Obviously, the model (1.2) includes the usual AR (1) structure.^[14]discussed the estimation and test problems of the ordinary linear regression model with error structure (1.2). Model (1.2) has been extensively discussed in the control theory literature^[15,16].^[17]applied model (1.2) to study the well-known Wolfersunspot numbers for the years from 1700 to 1955 and a seismic record obtained from an underground nuclear explosion that was carried out in the USA on October 29th, 1966. Recently, model (1.2) has been extended to the case of space time^[18,19]. More references about the theoretical results, applications and the extensions of the model (1.2) can be found in the monograph of^[9].

Based on the approximation of $g(\cdot)$ by a B-spline series and least squares estimation, we construct the following estimators for model (1.1) with error structure (1.2):

- i. Semiparametric ordinary least squares estimator (SOLSE) of β ,
- ii. Least squares estimator of the auto regression parameter ϕ ,
- iii. Semiparametric generalized least squares estimators (SGLSE) of β ,
- iv. Estimator of the asymptotic covariance matrix of the SGLSE of β , and
- v. B-spline series estimator of the nonparametric component function $g(\cdot)$.

We will further investigate the asymptotic properties of these estimators and derive their limiting distributions. In addition, a small simulation is conducted to evaluate the performance of the estimators.

Estimators: Throughout this study we will assume that the design points x_i and t_i are fixed, and they are related via:

$$x_{is} = h_s(t_i) + u_{is}, \quad i = 1, \dots, n; \quad s = 1, \dots, p. \quad (2.1)$$

The reasonableness of this relation can be found in^[2]. In addition, suppose that the vector $(1, \dots, 1)'$ is not in the space spanned by the column vectors of $X = (x_1, \dots, x_n)'$, which ensures the identifiability of model (1.1) according to^[1]. It is also assumed that the sequence of designs t_i forms an asymptotically regular sequence^[20] in the sense that:

$$\max_{1 \leq i \leq n} \left| \int_0^{t_i} p(t) dt - \frac{i-1}{n-1} \right| = o(n^{-1/2}),$$

where, $p(\cdot)$ denotes a positive density function on the interval $[0,1]$.

Let $\{x_i, t_i, y_i; i = 1, \dots, n\}$ be the model observed data from model (1.1), $X = (x_{ij})_{n \times p}$ and $y = (y_1, \dots, y_n)'$. Denote by $S_{k(n),j}$ the class of spline functions of degree ν and $k(n)$ knots. It is well known^[21] that $S_{k(n),j}$ has a basis consisting of $k(n) + \nu$ normalized B-splines $\{B_{k(n),j}(\cdot); j = 1, \dots, k(n) + \nu\}$ where $B_{k(n),j}(\cdot)$ has support $[(j-\nu+1)k^{-1}(n), (j+\nu+1)k^{-1}(n)]$, and $g(\cdot)$ can be approximated by a linear combination $\alpha' B_{k(n),j}(\cdot)$ of the basis, where $\alpha \in \mathbb{R}^{k(n)+\nu}$ and $B_{k(n),j}(\cdot) = (B_{k(n),1}(\cdot), \dots, B_{k(n),k(n)+\nu}(\cdot))'$.

For $\alpha \in \mathbb{R}^{k(n)+\nu}, \beta \in \mathbb{R}^p$,

$S_n(\beta, \alpha) = n^{-1} \sum_{i=1}^n (y_i - x_i' \beta - \alpha' B_{k(n)}(t_i))^2$ is minimized at

$(\hat{\beta}_n, \hat{\alpha}_n)$ given by $\hat{\beta}_n = (X' M_{B_{k(n)}} X)^{-1} X' M_{B_{k(n)}} Y$ and $\hat{\alpha}_n = (B_{k(n)}' B_{k(n)})^{-1} B_{k(n)}'(Y - X \hat{\beta}_n)$, where $B_{k(n)} = (B_{k(n),1}(t_1), \dots, B_{k(n),k(n)+\nu}(t_n))'$, $M_{B_{k(n)}} = 1 - P_{B_{k(n)}}$ and $P_{B_{k(n)}} = B_{k(n)}(B_{k(n)}' B_{k(n)})^{-1} B_{k(n)}$. This $\hat{\beta}_n$ is called the *semiparametric ordinary least squares estimator (SOLSE)* of β .

When the errors are correlated, the SOLSE $\hat{\beta}_n$ is not asymptotically efficient as it ignores the correlation. Hence we propose a *semiparametric generalized least square estimator (SGLSE)* of β . Since for a given ϕ :

$$\text{Cov}(\mathcal{E}\mathcal{E}') = \frac{(1-\phi^2)\sigma_e^2}{1-\phi^2-\theta^2\sigma_e^2} \Omega(\phi)$$

where, $\mathcal{E} = (\mathcal{E}_1, \dots, \mathcal{E}_n)'$, we define a SGLSE as follows:

$$\begin{aligned} \tilde{\beta}_G &= (X' M_{B_{k(n)}} \Omega^{-1}(\phi) B_{k(n)} M_{B_{k(n)}} X)^{-1} \\ & X' M_{B_{k(n)}} \Omega^{-1}(\phi) B_{k(n)} M_{B_{k(n)}} Y \end{aligned} \quad (2.2)$$

Where:

$$\Omega^{-1}(\phi) = \begin{pmatrix} 1 & -\phi & 0 & 0 & \dots & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & \dots & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & \dots & 0 & -\phi & 1 \end{pmatrix}_{n \times n}$$

Remark 1: By Lemma 3 in the Appendix, $X' M_{B_{k(n)}} X$ and $X' M_{B_{k(n)}} \Omega^{-1}(\phi) B_{k(n)} M_{B_{k(n)}} X$ are positive definite when n is large. Therefore, without loss of generality, we can assume that the inverses of these two matrices exist.

When ϕ is unknown, the $\tilde{\beta}_G$ in (2.2) is not directly usable, and we need a suitable estimator, say $\hat{\phi}_n$, of ϕ . Then we can replace $\Omega(\phi)$ by $\Omega(\hat{\phi}_n)$ in (2.2) and obtain an estimate of β that can be computed from the data. Noting that \mathcal{E}_i is unobservable, a reasonable estimator of ϕ is the least squares estimator :

$$\hat{\phi}_n = \sum_{i=2}^n \hat{\mathcal{E}}_i \hat{\mathcal{E}}_{i-1} \left[\sum_{i=2}^n \hat{\mathcal{E}}_{i-1}^2 \right]^{-1}$$

based on the estimated residuals:

$$\hat{\mathcal{E}}_i = y_i - x_i' \hat{\beta}_n - \hat{\alpha}_n' B_{k(n)}(t_i), \quad i = 1, \dots, n. \quad (2.3)$$

Consequently, we define our SGLSE of β , denoted by $\hat{\beta}_G$, as:

$$\begin{aligned} \hat{\beta}_G &= (X' M_{B_{k(n)}} \Omega^{-1}(\hat{\phi}_n) B_{k(n)} M_{B_{k(n)}} X)^{-1} \\ & X' M_{B_{k(n)}} \Omega^{-1}(\hat{\phi}_n) B_{k(n)} M_{B_{k(n)}} Y \end{aligned} \quad (2.4)$$

Based on this SGLSE $\hat{\beta}_G$, we can construct the following estimator of the nonparametric component function $g(\cdot)$:

$$\hat{g}_G(t) = B_{k(n)} (B_{k(n)}' B_{k(n)})^{-1} B_{k(n)}' (Y - X \hat{\beta}_G) \quad (2.5)$$

Furthermore, an estimator of the asymptotic covariance matrix of $\hat{\beta}_G$ is given by $n^{-1} \hat{\Sigma}_2$, where $\hat{\Sigma}_2 = n \hat{\sigma}_\varepsilon^2 (X' M_{B_{k(n)}} \Omega^{-1}(\hat{\phi}_n) B_{k(n)} M_{B_{k(n)}} X)^{-1}$ and:

$$\hat{\sigma}_\varepsilon^2 = n^{-1} \sum_{i=1}^n (y_i - x_i' \hat{\beta}_n - \hat{\alpha}_n' B_{k(n)}(t_i))^2 \quad (2.6)$$

Large sample properties: We begin with the following assumptions required to derive the main results, which are quite mild and can be easily satisfied (Remark 2 below).

Assumption 1: For $i = 1, \dots, n$ and $j = 1, \dots, p$, u_{ij} satisfy:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n-h} u_{ki} u_{k+h|j} = b_{hij}, \quad \text{for } h = 0, \pm 1, \pm 2, \dots; \quad i, j = 1, \dots, p, \quad (3.1)$$

where, the matrix $B = (b_{oij})$ is nonsingular, and:

$$\max_{1 \leq i \leq p} \|Au_i^*\| = O([\text{tr}(A'A)]^{1/2}) \text{ for any matrix } A \quad (3.2)$$

where, $u_i^* = (u_{1i}, \dots, u_{ni})'$ and $\|\cdot\|$ denotes the Euclidean norm.

Assumption 2: The functions $g(\cdot)$ and $h_1(\cdot), \dots, h_p(\cdot)$ are ν times continuously differentiable on the interval $[0, 1]$, where $\nu > 1$.

Assumption 3: The coefficients θ and ϕ , and the variance σ_e^2 of $\{e_i\}$ in model (1.2) satisfy $|\phi| < 1$ and $|\phi^2 + \theta^2 \sigma_e^2| < 1$.

Remark 2: The above u_{ij} behave like zero mean, uncorrelated random variables and $h_j(t_i)$ are the regression of x_{ij} on t_i . Specifically, suppose that the design points (x_i, t_i) are i.i.d. Random variables, and let $h_j(t_i) = E[x_{ij} | t_i]$ and $u_{ij} = x_{ij} - h_j(t_i)$ with $E[u_i u_i'] = B$. Then by the law of large numbers, (3.1) holds with probability 1. Moreover, according to [22] (3.2) holds when u_{ij} behave like zero mean, uncorrelated random variables. Assumption 2 is mild and holds for most commonly used functions, such as the polynomial and trigonometric functions.

The first theorem below shows the asymptotic normality of the SOLSE $\hat{\beta}_n$.

Theorem 3.1: Suppose that Assumptions 1 to 3 hold and $c_1 n^{\nu/(2\nu+1)} \leq k(n) \leq c_2 n^{\nu/(2\nu+1)}$ where c_1 and c_2 are positive constants. Then:

$$\sqrt{n}(\hat{\beta}_n - \beta) \rightarrow_D N\left(0, \frac{(1-\phi^2)\sigma_e^2}{1-\phi^2-\theta^2\sigma_e^2} B^{-1} G_1 B^{-1}\right) \text{ as } n \rightarrow \infty$$

where, $G_1 = \lim_{n \rightarrow \infty} n^{-1} U' \Omega(\phi) U$ provided it is positive definite, $U = (u_1, \dots, u_n)'$, B is defined in Assumption 1 and " \rightarrow_D " denotes convergence in distribution.

The asymptotic normality of $\hat{\phi}_n$ is established in the second theorem below.

Theorem 3.2: Under the conditions of Theorem 3.1, if, in addition, $E[\varepsilon_1^4] < \infty$, then:

$$\sqrt{n}(\hat{\phi}_n - \phi) \rightarrow_D N\left(0, \frac{\sigma_e^2 E[\varepsilon_1^2] + \theta^2 \sigma_e^2 E[\varepsilon_1^4]}{(E[\varepsilon_1^2])^2}\right) \text{ as } n \rightarrow \infty$$

The next theorem shows that the SGLSE $\tilde{\beta}_G$ defined in (2.2) and its feasible version $\hat{\beta}_G$ in (2.4) have the same limiting normal distribution.

Theorem 3.3: Under the conditions of Theorem 3.2 we have $\sqrt{n}(\hat{\beta}_G - \beta) = \sqrt{n}(\tilde{\beta}_G - \beta) + O_p(n^{-1/2})$ and the common limiting distribution of $\hat{\beta}_G$ and $\tilde{\beta}_G$ is given by:

$$\sqrt{n}(\hat{\beta}_G - \beta) \rightarrow_D N\left(0, \frac{(1-\phi^2)\sigma_e^2}{1-\phi^2-\theta^2\sigma_e^2} G_2^{-1}\right) \text{ as } n \rightarrow \infty$$

where, $G_2 = \lim_{n \rightarrow \infty} n^{-1} U' \Omega^{-1}(\phi) U$ provided it is positive definite.

Since $G_2^{-1} \leq B^{-1} G_1 B^{-1}$, according to Theorems 3.1 and 3.3, $\hat{\beta}_G$ the asymptotic covariance matrix is smaller than $\hat{\beta}_n$ in terms of asymptotic covariance matrix.

Let $B_\nu(\cdot)$ denote the ν^{th} Bernoulli polynomial, which is recursively defined by:

$$B_0(t) = 1, \quad B_i(t) = \int_0^t i B_{i-1}(z) dz + b_i \quad i = 1, 2, \dots,$$

where, $b_i = -i \int_0^1 B_{i-1}(z) dz$ is the i^{th} Bernoulli number^[23]. The following theorem establishes the asymptotic normality of the B-spline series estimator $\hat{g}_G(t)$ of $g(t)$.

Theorem 3.4: Under the conditions of Theorem 3.1, for any $t \in (i/(k(n)+\nu), (i+1)/(k(n)+\nu))$:

$$s_n^{-1} \{ \hat{g}_G(t) - (g(t) + b(t)) \} \rightarrow_D N(0,1) \text{ as } n \rightarrow \infty,$$

$$\text{where, } s_n^2 = \text{Var} \left[B_{k(n)} (B_{k(n)}' B_{k(n)})^{-1} B_{k(n)}' \varepsilon \right],$$

$$b(t) = -g^{(\nu)}(t) B_\nu(tk(n)-i) / [\nu! k(n)^\nu], \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$$

For inference about β based on the asymptotic distribution of $\hat{\beta}_G$, an estimator of its asymptotic covariance matrix is needed. Let $\hat{\Sigma}_2$ be given by (2.6)-(2.7). Then we have the following result.

Theorem 3.5: Under the conditions of Theorem 3.2 we have :

$$\hat{\Sigma}_2 - \frac{(1-\phi^2)\sigma_e^2}{1-\phi^2-\theta^2\sigma_e^2} G_2^{-1} = o_p(1)$$

From the results given in Theorems 3.3 and 3.5, it follows that:

$$n(\hat{\beta}_G - \beta)' \Sigma_2^{-1} (\hat{\beta}_G - \beta) \rightarrow_D \chi_p^2 \text{ as } n \rightarrow \infty$$

Therefore, the set $\{\beta : n(\hat{\beta}_G - \beta)' \Sigma_2^{-1} (\hat{\beta}_G - \beta) \leq \chi_{p,1-\alpha}^2\}$ constitutes a large-sample $100(1-\alpha)\%$ confidence ellipsoid for β . For small sample sizes $F_{p,n-p,1-\alpha}/p$ can be used to substitute

$$\chi_{p,1-\alpha}^2.$$

Remark 3: By applying the tensor-product B-spline technique^[24] the above results can be easily extended to the case of multivariate regressor t .

A simulation study: This presents a simulation study to evaluate the finite sample performance of the estimators. The observations are generated from:

$$y_i = 3.5x_i + \cos(2\pi t_i) + \varepsilon_i, \\ \varepsilon_i = (\phi + 0.1e_i)\varepsilon_{i-1} + e_i, \quad i = 1, \dots, n,$$

where, $t_i = (i-0.5)/n$, $x_i = 5t_i^2 + 0.5\eta_i$ and $\{e_i\}$ are i.i.d. $N(0,1)$. For a range of values of ϕ , we generate 10,000 samples of size 200 from the above model (the x_{ij} values are generated once for a fixed ϕ value) and estimate ϕ , β and $g(\cdot)$ for each sample. We here use the uniform knots. According to^[24] uniform knots are usually sufficient when the function $g(\cdot)$ does not exhibit dramatic changes in its derivatives. Thus, we just need to determine the number of knots to use. We use the method in^[24] to do so. Biases and sample variances (Var) of the simulated estimates are given in Tables 1 and 2, where $\hat{g}_n(t)$, $\hat{g}_G(t)$ and $\tilde{g}_G(t)$ are based on $\hat{\beta}_n$, $\hat{\beta}_G$ and $\tilde{\beta}_G$ respectively.

Table 1: Simulated biases and variances of the estimators for ϕ and β

ϕ	Bias($\hat{\phi}_n$)	Var($\hat{\phi}_n$)	Bias($\hat{\beta}_n$)	Var($\hat{\beta}_n$)	Bias($\hat{\beta}_G$)	Var($\hat{\beta}_G$)	Bias($\tilde{\beta}_G$)	Var($\tilde{\beta}_G$)
0.2	-0.0059	0.0009	0.1228	0.0178	0.1019	0.0130	0.1014	0.0127
0.3	-0.0095	0.0014	0.1154	0.0161	0.0839	0.0095	0.0829	0.0092
0.4	-0.0181	0.0019	0.1089	0.0152	0.0723	0.0082	0.0689	0.0076
0.5	-0.0204	0.0024	0.1181	0.0196	0.0688	0.0082	0.0631	0.0071
0.6	-0.0217	0.0029	0.1138	0.0189	0.0466	0.0053	0.0396	0.0046
0.7	-0.0229	0.0034	0.1198	0.0240	0.0455	0.0051	0.0336	0.0043
0.8	-0.0270	0.0039	0.1195	0.0289	0.0304	0.0048	0.0200	0.0042

Table 2: Simulated biases and variances of the estimators for $g(\cdot)$

T	$g(t)$	Bias(\hat{g}_n)	Var(\hat{g}_n)	Bias(\hat{g}_G)	Var(\hat{g}_G)	Bias(\tilde{g}_G)	Var(\tilde{g}_G)
(30-0.5)/200	-0.6064	-0.0319	0.0144	-0.0277	0.0140	-0.0269	0.0140
(60-0.5)/200	-0.2940	-0.0308	0.0095	-0.0188	0.0088	-0.0184	0.0087
(90-0.5)/200	-0.9460	-0.0856	0.0239	-0.0502	0.0182	-0.0498	0.0179
(120-0.5)/200	-0.8181	-0.0712	0.0392	-0.0661	0.0358	-0.0659	0.0356
(150-0.5)/200	-0.0157	-0.0352	0.0612	-0.0339	0.0539	-0.0338	0.0535
(170-0.5)/200	0.5750	0.0613	0.1113	-0.0009	0.0897	-0.0014	0.0897

From Table 1 we can see that in all cases the semiparametric generalized least squares estimator $\hat{\beta}_G$ has smaller bias than the semiparametric ordinary least squares estimator $\hat{\beta}_n$. The advantage of $\hat{\beta}_G$ over $\hat{\beta}_n$ is more significant when ϕ is large (high serial correlation), as one would expect since $\hat{\beta}_G$ takes the serial correlation into account whereas $\hat{\beta}_n$ does not.

Moreover, as ϕ increases, the bias and variance of $\hat{\beta}_G$ decrease, but this is not the case for $\hat{\beta}_n$. In addition, the performance of $\hat{\beta}_G$ is close to that of the ϕ -known semiparametric generalized least squares estimator $\tilde{\beta}_G$ across the values of ϕ . Table 1 also shows that the estimator $\hat{\phi}_n$ of ϕ is adequate.

From Table 2 we can see that the nonparametric estimator \hat{g}_G based on the semiparametric generalized least squares estimator $\hat{\beta}_G$ is better than the nonparametric estimator \hat{g}_n based on the semiparametric ordinary least squares estimator $\hat{\beta}_n$ in terms of bias and variance.

CONCLUSION

In this article we have studied the estimation problem of a partly linear regression model with bilinear time series errors. Using B-splines to approximate the nonparametric component, we have constructed the semiparametric ordinary and generalized least squares estimators of the parametric component, the least squares estimator of the autoregressive parameter, and the B-spline series estimator of the nonparametric component. We have also derived the asymptotic normality for these estimates. In both theory and simulation, we have

demonstrated that the semiparametric generalized least squares estimator is more efficient than the semiparametric ordinary least squares estimator.

Appendix: Proofs of Theorems: In order to prove the theorems presented earlier we first introduce several lemmas.

Lemma 1: Suppose that a function $f(\cdot)$ satisfies Assumption 2. Then we have:

$$\sup_{t \in [0,1]} |f(t) - B_{k(n)}(B_{k(n)}' B_{k(n)})^{-1} B_{k(n)}' f| = O(n^{-1}k(n)) + O(k(n)^{-\nu})$$

where $f = (f(t_1), \dots, f(t_n))'$.

Lemma 2: For the basis of B-splines in $S_{k(n),\nu}$ we have:

- i. $\sum_{i=1}^{k(n)+\nu} B_{k(n),i}^2 \leq 1$ for all t ; and
- ii. All eigenvalues of $n^{-1} B_{k(n)}' B_{k(n)}$ are between $c_1 k^{-1}(n)$ and $c_2 k^{-1}(n)$ for some constants $0 < c_1 < c_2 < \infty$.

The proofs of Lemmas 1 and 2 can be found^[25].

Lemma 3: Suppose that Assumptions 1 and 2 hold. Then as $n \rightarrow \infty$,

$$n^{-1} X' M_{B_{k(n)}} X \rightarrow B \quad \text{and} \\ n^{-1} X' M_{B_{k(n)}} \Omega^{-1}(\phi) B_{k(n)} M_{B_{k(n)}} X \rightarrow G_2.$$

Proof: Applying Lemmas 1 and 2, the proof of Lemma 3 is simple. We here omit the detail.

Consider the normalized weighted sum of random variables $\{ \mathcal{E}_i \}$ defined in (1.2), namely, $\left(\sum_{i=1}^n a_i^2 \right)^{-1/2} \sum_{i=1}^n a_i \mathcal{E}_i$, where $\{a_i\}$ is a sequence of real constants. It has the following asymptotic property.

Lemma 4: If $n^{-1} \sum_{k=1}^{n-h} a_i a_{i+|h|} \rightarrow \alpha_h$ for $h = \pm 1, \pm 2, \dots$, $\{a_i\}$ is bounded, and Assumption 3 holds, then:

$$\left(\sum_{i=1}^n a_i^2 \right)^{-1/2} \sum_{i=1}^n a_i \mathcal{E}_i \rightarrow_D N(0, \sigma^2) \quad \text{as } n \rightarrow \infty, \quad \text{where}$$

$$\sigma^2 = \sum_{h=-\infty}^{\infty} \alpha_h \gamma(h) \quad \text{and} \quad \gamma(h) = \sigma_e^2 \phi^{|h|} (1 - \phi^2 - \theta^2 \sigma_e^2)^{-1}$$

is the autocovariance function of $\{\mathcal{E}_i\}$.

Proof: According to the property of $\{\mathcal{E}_i\}$ and Lemmddifficult^[26], it is not difficulty to complete the proof.

We are now ready to prove the main results.

Proof of Theorem 3.1: By (2.1) it is easy to see that $u_i^* M_{B_{k(n)}} \mathcal{E} = u_i^* \mathcal{E} - u_i^* P_{B_{k(n)}} \mathcal{E} + h_i^* (I - P_{B_{k(n)}}) \mathcal{E}$, $i = 1, \dots, n$,

where $h_i = (h_1(t_i), \dots, h_p(t_i))'$. Let

$$c_0 = \sigma_e^2 (1 - \phi^2 - \theta^2 \sigma_e^2)^{-1}. \text{ Since}$$

$$\text{Var}(u_i^* P_{B_{k(n)}} \mathcal{E}) = c_0 u_i^* P_{B_{k(n)}} \Omega(\phi) P_{B_{k(n)}} u_i^* \leq c_0 \lambda_{\max}[\Omega(\phi)] u_i^* P_{B_{k(n)}} u_i^* = O(k(n) + \nu) = o(n),$$

where, $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix, we have $u_i^* P_{B_{k(n)}} \mathcal{E} = o_p(n^{1/2})$.

Moreover, by Lemma 1, there exists a real vector η such that:

$$\max_{1 \leq i \leq n} \|h_i - B_{k(n)} \eta\| = O(n^{-1}k(n)) + O(k(n)^{-\nu}).$$

This implies:

$$\text{Var}(h_i^* M_{B_{k(n)}} \mathcal{E}) = c_0 h_i^* M_{B_{k(n)}} \Omega(\phi) M_{B_{k(n)}} h_i \leq c_0 \lambda_{\max}[\Omega(\phi)] h_i^* M_{B_{k(n)}} h_i \text{ Hence}$$

$$= O(n^{-2}k(n)^2) + O(k(n)^{-2\nu}).$$

$h_i^* M_{B_{k(n)}} \mathcal{E} = o_p(n^{1/2})$. It follows that

$$X^* M_{B_{k(n)}} \mathcal{E} = U^* \mathcal{E} + o_p(n^{1/2}).$$

Next, by Lemma 1, there exists a real vector π such that:

$$\max_{1 \leq i \leq n} |g(t_i) - B_{k(n)} \pi| = O(n^{-1}k(n)) + O(k(n)^{-\nu}).$$

Therefore $x_i^* M_{B_{k(n)}} g = O(n^{-1}k(n)^2) + O(k(n)^{-2\nu}n)$ for $g = (g(t_1), \dots, g(t_n))'$, where $x_i^* = (x_{1i}, \dots, x_{ni})'$.

Consequently:

$$\sqrt{n}(\hat{\beta}_n - \beta) = \sqrt{n}(X^* M_{B_{k(n)}} X)^{-1} X^* M_{B_{k(n)}} \mathcal{E} + \sqrt{n}(X^* M_{B_{k(n)}} X)^{-1} X^* M_{B_{k(n)}} g$$

$$= \sqrt{n}(U^* U)^{-1} U^* \mathcal{E} + o_p(1).$$

The conclusion of Theorem 3.1 then follows from Lemma 3 and Assumption 1.

Proof of Theorem 3.2: It is easy to see that the following equation holds:

$$\sum_{i=2}^n \hat{\mathcal{E}}_i \hat{\mathcal{E}}_{i-1} - \sum_{i=2}^n \mathcal{E}_i \mathcal{E}_{i-1} = \sum_{i=2}^n (\hat{\mathcal{E}}_i - \mathcal{E}_i)(\hat{\mathcal{E}}_{i-1} - \mathcal{E}_{i-1}) + \sum_{i=2}^n (\hat{\mathcal{E}}_i - \mathcal{E}_i)\mathcal{E}_{i-1} + \sum_{i=2}^n (\hat{\mathcal{E}}_{i-1} - \mathcal{E}_{i-1})\mathcal{E}_i. \tag{A.1}$$

According to the definition of $\hat{\mathcal{E}}_i$ in (2.3), we have:

$$\sum_{i=2}^n (\hat{\mathcal{E}}_i - \mathcal{E}_i)\mathcal{E}_{i-1} = \sum_{i=2}^n (\beta - \hat{\beta}_n)' x_i \mathcal{E}_{i-1}$$

$$- \sum_{i=2}^n (\beta - \hat{\beta}_n)' X^* B_{k(n)} (B_{k(n)}^* B_{k(n)})^{-1} B_{k(n)}(t_i) \mathcal{E}_{i-1}$$

$$+ \sum_{i=2}^n [g(t_i) - B_{k(n)} \pi(t_i) (B_{k(n)}^* B_{k(n)})^{-1} B_{k(n)}(t_i) g] \mathcal{E}_{i-1}$$

$$- \sum_{i=2}^n \mathcal{E}^* B_{k(n)} (B_{k(n)}^* B_{k(n)})^{-1} B_{k(n)}(t_i) \mathcal{E}_{i-1} \tag{A.2}$$

$$= I_1 - I_2 + I_3 - I_4, \quad \text{say.}$$

By the root- n consistency of $\hat{\beta}_n$ and the proof of Theorem 3.1, we have $I_1 = O_p(1)$. Next, let $B_{k(n)}^* = (B_{k(n)}(t_2), \dots, B_{k(n)}(t_n))'$, $\mathcal{E}^* = (\mathcal{E}_1, \dots, \mathcal{E}_{n-1})'$, and $\Omega_{n-1}(\phi) = \text{Var}(\mathcal{E}^*)$. Then the Cauchy-Schwarz inequality gives

$$|I_2| \leq \mathcal{E}^* B_{k(n)} (B_{k(n)}^* B_{k(n)})^{-1} B_{k(n)}^* \mathcal{E}^* + (\beta - \hat{\beta}_n)' X^* P_{B_{k(n)}} X (\beta - \hat{\beta}_n) = I_{21} + I_{22}, \text{ say.}$$

Since $E|I_{21}| = O(k(n))$, we have $I_{21} = O_p(k(n))$. By Theorem 3.1, it is easy to see $I_{22} = O_p(1)$. So $I_2 = o_p(n^{1/2})$. Then Lemma 1 yields $E[I_3^2] = O(nk(n)^{-2\nu}) + O(n^{-1}k(n)^2)$ so that $I_3 = O(n^{1/2}k(n)^{-\nu}) + O(n^{-1/2}k(n))$. The proof of

$I_4 = O_p(k(n))$ is similar to that for I_{21} . Therefore, by (A.2), the middle term in the right hand side of (A.1) is $o_p(n^{1/2})$, and similar arguments show that the other two terms are $o_p(n^{1/2})$ as well. Consequently,

$$\sum_{i=2}^n \hat{\mathcal{E}}_i \hat{\mathcal{E}}_{i-1} = \sum_{i=2}^n \mathcal{E}_i \mathcal{E}_{i-1} + o_p(n^{1/2}).$$

By the same reason, we

$$\text{have } \sum_{i=2}^n \hat{\mathcal{E}}_i^2 = \sum_{i=2}^n \mathcal{E}_i^2 + o_p(n^{1/2}). \text{ It follows that:}$$

$$\sqrt{n}(\hat{\phi}_n - \phi) = \sqrt{n} \left\{ \left(\sum_{i=2}^n \mathcal{E}_i^2 \right)^{-1} \sum_{i=2}^n \mathcal{E}_i \mathcal{E}_{i-1} - \phi \right\} + o_p(1).$$

Combining this with the result of^[27], we complete the proof.

Proof of Theorem 3.3: By the definition of $\hat{\beta}_G$, Lemmas 1 to 3 and Theorem 3.1 we have:

$$\sqrt{n}(\hat{\beta}_G - \beta) = \sqrt{n}(U' \Omega^{-1}(\phi)U)^{-1}U' \Omega^{-1}(\phi)\varepsilon + o_p(1).$$

Moreover, similar to the proof of Theorem 4.3 in^[10], we have:

$$\sqrt{n}(U' \Omega^{-1}(\phi)U)^{-1}U' \Omega^{-1}(\phi)\varepsilon \rightarrow {}_D N\left(0, \frac{(1-\phi^2)\sigma_\varepsilon^2}{1-\phi^2-\theta^2\sigma_\varepsilon^2} B^{-1}G_2 B^{-1}\right) \text{ as } n \rightarrow \infty.$$

Thus the proof is complete.

Proof of Theorem 3.4: Combining the root- n consistency of $\hat{\beta}_G$ with Theorem 2.1 in^[28] we can easily complete the proof.

Proof of Theorem 3.5: This theorem follows from Lemma 2 and the proofs of Theorems 3.2–3.3.

REFERENCES

- Chen, H., 1988. Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, 16: 136-146.
- Speckman, P., 1988. Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B*, 50: 413-436.
- Chen, H. and J. Shiau, 1994. Data-driven efficient estimation for a partial linear model linear model. *Ann. Statist.*, 22: 211-237.
- Donald, G. and K. Newey, 1994. Series estimation of semilinear models. *J. Multivariate Anal.*, 50: 30-40.
- Hamilton, A. and K. Truong, 1997. Local linear estimation in partly linear models. *J. Multivariate Anal.*, 60:1-19.
- Engle, R.F., W. J. Granger, J. Rice and A. Weiss, 1986. Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.*, 80: 310-319.
- Schick, A., 1994. Estimation of the autocorrelation coefficient in the presence of a regression trend. *Statist. Probab. Lett.*, 21: 371-380.
- Gao, J.T., 1995. Asymptotic theory for partly linear models. *Comm. Statist. Theory Methods*, A24 8: 1985-2009.
- Tong, H., 1990. *Nonlinear Time Series*. Oxford University Press, Oxford.
- Hwang, S.Y. and I. V. Basawa, 1993. Parameter estimation in a regression model with random coefficient autoregressive errors. *J. Statist. Plann. Infer.*, 36: 57-67.
- Bera, A.K. and X. L. Zuo, 1996. Specification test for a linear regression model with ARCH process. *J. Statist. Plann. Infer.*, 50: 283-308.
- Dutta, H., 1999. Large sample tests for a regression model with autoregressive conditional heteroscedastic errors. *Comm. Statist. Theory Methods*, A28: 105-117.
- You, J.H. and G. Chen, 2002. Parameter estimation in partially linear regression models with random coefficient autoregressive errors. *Comm. Statist. Theory Methods*, 31: 1137-1158.
- Hwang, S.Y., 1995. A study on a regression model with nonlinear time series errors. *Korean J. Applied Stat.*, 9: 187-200.
- Rubberti, A., A. Isidori and P. d'Allesandro, 1972. *Theory of Bilinear Dynamical Systems*. Berlin: Springer-Verlag.
- Mohler, R.R., 1973. *Bilinear Control Processes*. New York: Academic Press.
- Rao, T.S., 1981. On the theory of bilinear time series models. *J. Roy. Statist. Soc. Ser. B*, 43: 244-255.
- Dai, Y. and L. Billard, 1998. A space time bilinear model and its identification. *J. Time Ser. Anal.*, 19: 657-679.
- Dai, Y. and L. Billard, 2003. Maximum likelihood estimation in space time bilinear models. *J. Time Ser. Anal.*, 24: 24-45.
- Sacks, J. and D. Ylvisacker, 1970. Design for regression problems with correlated errors III. *Ann. Math. Statist.*, 41: 2057-2074.
- de Boor, C., 1978. *A practical Guide to Splines*. Springer-Verlag, New York.
- Moyeed, R. A. and P. J. Diggle, 1994. Rates of convergence in semi-parametric modeling of longitudinal data. *Austra. J. Statist.*, 36: 21-35.
- Barrow, D.L. and P.W. Smith, 1979. Efficient L_2 approximation by splines. *Numer. Math.*, 33: 101-114.
- He, X. and P. D. Shi, 1996. Bivariate tensor-product B-spline in a partly linear model. *J. Multivariate Anal.*, 58: 162-181.
- Burman, P., 1991. Regression function estimation from dependent observations. *J. Multivariate Anal.*, 36: 263-279.
- Fuller, W.A., 1996. *Introduction to statistical time series*. Wiley, New York.
- Feignin, P.D. and R.L. Tweedie, 1985. Random coefficient autoregressive processes: A Markov chain analysis of stationarity and fitness of moments. *J. Time Ser. Anal.*, 6: 23-32.
- Zhou, S., X. Shen and D. A. Wolfe, 1998. Local asymptotic for regression splines and confidence regions. *Anal. Statist.*, 26: 1760-1782.