Journal of Mathematical Research & Exposition
Nov., 2008, Vol. 28, No. 4, pp. 877–885
DOI:10.3770/j.issn:1000-341X.2008.04.017
Http://jmre.dlut.edu.cn

Asymptotic Normality of Estimators in Partially Linear Varying Coefficient Models

WEI Chuan Hua[1], *WU Xi Zhi*[2]
(1. Department of Statistics, School of Science, Central University for Nationalities, Beijing 100081, China;
2. School of Statistics, Renmin University of China, Beijing 100872, China)
(E-mail: chweisd@yahoo.com.cn)

Abstract Partially linear varying coefficient model is a generalization of partially linear model and varying coefficient model and is frequently used in statistical modeling. In this paper, we construct estimators of the parametric and nonparametric components by Profile least-squares procedure which is based on local linear smoothing. The resulting estimators are shown to be asymptotically normal with heteroscedastic error.

Keywords asymptotic normality; Heteroscedasticity; profile least-squares approach; partially linear varying coefficient model; local linear smoothing.

Document code A
MR(2000) Subject Classification 62G05; 62G08; 62G20
Chinese Library Classification O212.1

1. Introduction

Over the last three decades, with the improvement of computing facilities, many useful semi-parametric models have been proposed to capture the underlying relationships between response variables and their associated covariates. Like parametric models, semiparametric models have various forms, say, additive models, partially linear models, single-index models, varying coefficient models and their hybrids. One popular semiparametric specification is partially linear varying coefficient model which is a generalization of the partially linear model and varying coefficient regression model. The partially linear varying coefficient model assumes the following structure:

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\alpha}(u_i) + \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1.1}$$

where $y_i^{,}s$ are responses; $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^{\mathrm{T}}$, $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iq})^{\mathrm{T}}$ and u_i are associated covariates; $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)^{\mathrm{T}}$ is a vector of q-dimensional unknown parameters and $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \alpha_1(\cdot), \ldots, \alpha_p(\cdot))^{\mathrm{T}}$ is a p-dimensional vector of unknown functions and $\varepsilon_i^{,}s$ are independent random errors with $\mathrm{E}(\varepsilon_i|\mathbf{x}_i, \mathbf{z}_i, u_i) = 0$ and $\sigma^2(\mathbf{x}_i, \mathbf{z}_i, u_i) = \mathrm{E}[\varepsilon_i^2|\mathbf{x}_i, \mathbf{z}_i, u_i]$.

Obviously, model (1.1) is an extension of the varying coefficient model which was introduced by Hastie and Tibshirani[1]. Due to its flexibility, varying coefficient model has been studied

Received date: 2006-10-12; **Accepted date**: 2007-07-13
Foundation item: the National Natural Science Foundation of China (No. 10431010).

in many different contexts and has been successfully applied to nonlinear time series analysis, longitudinal and functional data analysis, spatial data analysis, and time-varying models in finance. See, for example, the work of Cai et al.[2,3], Fan and Zhang[4], Fotheringham et al.[5], Hoover et al.[6] and Huang et al.[7] among others. When $q = 1$ and $\mathbf{z}_i = 1$, the model (1.1) becomes a partially linear model, which was proposed by Engle et al.[8] when they studied the effect of weather on electricity demand. Partially linear model has been widely studied among statisticians and econometricians. More references and techniques can be found in Hardle et al.[9,10]

Model (1.1) has been studied by many authors recently. Zhang et al.[11] developed the one-stage and two-stage procedures to estimate linear part and nonparametric part. Zhou and You[12] constructed estimators of the parametric and nonparametric components by wavelet procedure. Xia et al.[13] proposed a new estimator for parametric component by local linear method. Ahmad et al.[14] used a general series method to estimate unknown quantities in model (1.1). Fan and Huang[15] introduced a profile least-squares approach and showed the estimator of parametric component is root-n consistent. In addition, they proposed profile generalized likelihood ratio statistics for testing problems on the parametric components.

Most of related works in literatures were focused on estimating unknown constant coefficients $\boldsymbol{\beta}$ and coefficient functions $\alpha_j(\cdot)$ with homoscedastic errors. In this paper, we study the asymptotic properties of profile least-squares estimators of parametric and nonparametric components with heteroscedastic errors.

The rest of this paper is organized as follows. In Section 2, we introduce the profile least-squares procedure and construct the estimators. Large sample properties of the estimators are derived in Section 3. Proofs of the main results are relegated to Section 5.

2. Profile least-squares estimation

For convenience, we first introduce the profile least-squares estimation. If $\boldsymbol{\beta}$ is known, the model (1.1) can be written as

$$y_i^* = \alpha_1(u_i)x_{i1} + \cdots + \alpha_p(u_i)x_{ip} + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{2.1}$$

where $y_i^* = y_i - \mathbf{z}_i^{\mathrm{T}}\boldsymbol{\beta}$. This transforms the partially linear varying coefficient model (1.1) into the varying coefficient model (2.1). Now, we apply a local linear regression technique to estimate the varying coefficient functions $\{\alpha_j(\cdot),\ j = 1, 2, \ldots, p\}$. For u in a small neighborhood of u_0, one can approximate $\alpha_j(\cdot)$ locally by a linear function

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha_j'(u_0)(u - u_0), \quad j = 1, 2, \ldots, p. \tag{2.2}$$

This leads to the following weighted local least-squares problems: find $\{(\alpha_j(u_0), \alpha_j'(u_0)),\ j = 1, 2, \ldots, p\}$ to minimize

$$\sum_{i=1}^{n} \Big[y_i^* - \sum_{j=1}^{p}\{\alpha_j(u_0) + \alpha_j'(u_0)(u - u_0)\}x_{ij}\Big]^2 \mathrm{K}_h(u_i - u_0), \tag{2.3}$$

where K is a kernel function, h is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$.

Let

$$\mathbf{X}=\begin{pmatrix}\mathbf{x}_1^{\mathrm{T}}\\ \mathbf{x}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{x}_n^{\mathrm{T}}\end{pmatrix}=\begin{pmatrix}x_{11}&\cdots&x_{1p}\\ x_{21}&\cdots&x_{2p}\\ \vdots&\ddots&\vdots\\ x_{n1}&\cdots&x_{np}\end{pmatrix},\mathbf{Z}=\begin{pmatrix}\mathbf{z}_1^{\mathrm{T}}\\ \mathbf{z}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{z}_n^{\mathrm{T}}\end{pmatrix}=\begin{pmatrix}z_{11}&\cdots&z_{1q}\\ z_{21}&\cdots&z_{2q}\\ \vdots&\ddots&\vdots\\ z_{n1}&\cdots&z_{nq}\end{pmatrix},$$

$$\mathbf{Y}=\begin{pmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix},\ \mathbf{M}=\begin{pmatrix}\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\alpha}(u_1)\\ \mathbf{x}_2^{\mathrm{T}}\boldsymbol{\alpha}(u_2)\\ \vdots\\ \mathbf{x}_n^{\mathrm{T}}\boldsymbol{\alpha}(u_n)\end{pmatrix},\ \mathbf{D}_{u_0}=\begin{pmatrix}\mathbf{x}_1^{\mathrm{T}}&\mathbf{x}_1^{\mathrm{T}}\frac{u_1-u_0}{h}\\ \mathbf{x}_2^{\mathrm{T}}&\mathbf{x}_2^{\mathrm{T}}\frac{u_2-u_0}{h}\\ \vdots&\vdots\\ \mathbf{x}_n^{\mathrm{T}}&\mathbf{x}_n^{\mathrm{T}}\frac{u_n-u_0}{h}\end{pmatrix}$$

and

$$\boldsymbol{\varepsilon}=(\varepsilon_1,\dots,\varepsilon_n)^{\mathrm{T}},\mathbf{W}_{u_0}=\mathrm{diag}\{K_h(u_1-u_0),K_h(u_1-u_0),\dots,K_h(u_n-u_0)\}.$$

Then model (2.1) can be written as

$$\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta}=\mathbf{M}+\boldsymbol{\varepsilon}. \tag{2.4}$$

The solution to the problem (2.3) is given by

$$[\hat{\alpha}_1(u_0),\dots,\hat{\alpha}_p(u_0),h\hat{\alpha}_1'(u_0),\cdots,h\hat{\alpha}_p'(u_0)]^{\mathrm{T}}=\{\mathbf{D}_{u_0}^{\mathrm{T}}\mathbf{W}_{u_0}\mathbf{D}_{u_0}\}^{-1}\mathbf{D}_{u_0}^{\mathrm{T}}\mathbf{W}_{u_0}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta}). \tag{2.5}$$

Then the estimator for $\mathbf{M}$ is $\hat{\mathbf{M}}=\mathbf{S}(\mathbf{Y}-\mathbf{Z}\boldsymbol{\beta})$, where

$$\mathbf{S}=\begin{pmatrix}(\mathbf{x}_1^{\mathrm{T}}\ \mathbf{0})\{\mathbf{D}_{u_1}^{\mathrm{T}}\mathbf{W}_{u_1}\mathbf{D}_{u_1}\}^{-1}\mathbf{D}_{u_1}^{\mathrm{T}}\mathbf{W}_{u_1}\\ (\mathbf{x}_2^{\mathrm{T}}\ \mathbf{0})\{\mathbf{D}_{u_2}^{\mathrm{T}}\mathbf{W}_{u_2}\mathbf{D}_{u_2}\}^{-1}\mathbf{D}_{u_2}^{\mathrm{T}}\mathbf{W}_{u_2}\\ \vdots\\ (\mathbf{x}_n^{\mathrm{T}}\ \mathbf{0})\{\mathbf{D}_{u_n}^{\mathrm{T}}\mathbf{W}_{u_n}\mathbf{D}_{u_n}\}^{-1}\mathbf{D}_{u_n}^{\mathrm{T}}\mathbf{W}_{u_n}\end{pmatrix}.$$

Substituting $\hat{\mathbf{M}}$ into (2.4), we can obtain the following linear regression model

$$(\mathbf{I}-\mathbf{S})\mathbf{Y}=(\mathbf{I}-\mathbf{S})\mathbf{Z}\boldsymbol{\beta}+\boldsymbol{\varepsilon}. \tag{2.6}$$

Applying the least-squares approach to model (2.6) results in the following profile least squares estimator of parametric component $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}=[\mathbf{Z}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{Z}]^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{Y}. \tag{2.7}$$

Moreover, the estimator of $\boldsymbol{\alpha}(u)$ can be defined as

$$\hat{\boldsymbol{\alpha}}(u)=(\hat{\alpha}_1(u),\dots,\hat{\alpha}_p(u))^{\mathrm{T}}=(\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u(\mathbf{Y}-\mathbf{Z}\hat{\boldsymbol{\beta}}). \tag{2.8}$$

Let $\hat{\mathbf{Y}}=(\hat{y}_1,\hat{y}_2,\dots,\hat{y}_n)^{\mathrm{T}}$ be the vector of the fitted values of $\mathbf{Y}$ and $\hat{\boldsymbol{\varepsilon}}=(\hat{\varepsilon}_1,\hat{\varepsilon}_2,\dots,\hat{\varepsilon}_n)^{\mathrm{T}}$ be the vector of residuals. Then according to the above fitting procedure and the results in (2.7) and (2.8), we have

$$\hat{\mathbf{Y}}=\mathbf{X}\hat{\boldsymbol{\beta}}+\hat{\mathbf{M}}=\mathbf{L}\mathbf{Y}\quad\text{and}\quad\hat{\boldsymbol{\varepsilon}}=\mathbf{Y}-\hat{\mathbf{Y}}=(\mathbf{I}-\mathbf{L})\mathbf{Y}, \tag{2.9}$$

where

$$\mathbf{L}=\mathbf{S}+(\mathbf{I}-\mathbf{S})\mathbf{X}[\mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S}).$$

The above profile least-squares estimation depends on the choice of bandwidth. Here, we use the Cross-Validation technique to choose bandwidth, and the optimal value of the bandwidth is chosen to minimize the expression

$$CV(h) = \sum_{i=1}^{n} \left(\frac{\hat{\varepsilon}_i}{1 - l_{ii}} \right)^2,$$

where $\hat{\varepsilon}_i$ and l_{ii} $(i = 1, 2, \ldots, n)$ are respectively the residuals and the diagonal elements of the matrix $\mathbf{L}$ which are shown in (2.9).

3. Asymptotic properties of proposed estimators

We begin with the following assumptions required to derive the large sample property of the estimators descried in the last section, which are quite mild and can be easily satisfied. Let $\mu_i = \int t^i K(t)\mathrm{d}t$, $\nu_i = \int t^i K^2(t)\mathrm{d}t$, $\Gamma(U) = E(\mathbf{x}\mathbf{x}^{\mathrm{T}}|U)$, $\Phi(U) = E(\mathbf{x}\mathbf{z}^{\mathrm{T}}|U)$, $\bar{\mathbf{Z}} = (\bar{\mathbf{z}}_1, \bar{\mathbf{z}}_2, \ldots, \bar{\mathbf{z}}_n)^{\mathrm{T}} = (\mathbf{I} - \mathbf{S})\mathbf{Z}$ and $\bar{\mathbf{Y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n)^{\mathrm{T}} = (\mathbf{I} - \mathbf{S})\mathbf{Y}$.

Assumption 3.1 The random variable u has a bounded support $\mathbf{\Omega}$. Its density function $f(\cdot)$ is Lipschitz continuous and bounded away from 0 on its support.

Assumption 3.2 The $q \times q$ matrix $E(\mathbf{x}\mathbf{x}^{\mathrm{T}}|u)$ is non-singular for each $u \in \mathbf{\Omega}$. $E(\mathbf{x}\mathbf{x}^{\mathrm{T}}|u)$, $E(\mathbf{x}\mathbf{x}^{\mathrm{T}}|u)^{-1}$ and $E(\mathbf{x}\mathbf{z}^{\mathrm{T}}|u)$ are all Lipschitz continuous.

Assumption 3.3 There is an $s > 2$ such that $E \parallel \mathbf{x} \parallel^{2s} < \infty$, $E \parallel \mathbf{z} \parallel^{2s} < \infty$ and $n^{2k-1}h \to \infty$ for some $k < 2 - s^{-1}$.

Assumption 3.4 $\{\alpha_j(\cdot), j = 1, \ldots, p\}$ have continuous second derivative in $u \in \mathbf{\Omega}$.

Assumption 3.5 The function $K(\cdot)$ is a symmetric density function with compact support and the bandwidth h satisfies $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$.

The following theorems gives the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

Theorem 3.1 *Suppose that Assumptions 3.1–3.5 hold, the profile least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is asymptotically normal, namely,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \mathbf{\Xi}^{-1}\mathbf{\Omega}\mathbf{\Xi}^{-1}$, $\mathbf{\Xi} = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}]$, $\mathbf{\Omega} = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\sigma^2(\mathbf{x}_i, \mathbf{z}_i, u_i)]$, *and* $\boldsymbol{\eta}_i = \mathbf{z}_i - \mathbf{\Phi}^{\mathrm{T}}(u_i)\mathbf{\Gamma}^{-1}(u_i)\mathbf{x}_i$.

Remark 3.1 If ε is homoscedastic, that is, $\sigma^2(\mathbf{x}_i, \mathbf{z}_i, u_i) = \sigma^2, i = 1, 2, \ldots, n$, then we have

$$\mathbf{\Omega} = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\sigma^2] = \sigma^2\mathbf{\Xi}, \qquad \mathbf{\Sigma} = \mathbf{\Xi}^{-1}\mathbf{\Omega}\mathbf{\Xi}^{-1} = \sigma^2\mathbf{\Xi}^{-1}.$$

Therefore,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\sigma^2\Xi^{-1}}).$$

This result is consistent with that of Fan and Huang[15].

To apply Theorem 3.1 for making statistical inferences, we need to estimate the asymptotic

$\mathbf{\Sigma}$. Define

$$\hat{\mathbf{\Sigma}} = (\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}})^{-1}(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}\hat{\varepsilon}_i^2)(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}})^{-1}.$$

Then we have the following results.

Theorem 3.2 *Suppose that Assumptions 3.1–3.5 hold, we have*

$$\hat{\mathbf{\Sigma}} \to^{\mathrm{P}} \mathbf{\Sigma}.$$

Combining Theorems 3.1 and 3.2, we obtain the following corollary

Corollary 3.1 *Suppose that Assumptions 3.1–3.5 hold, we have*

$$\sqrt{n}\hat{\mathbf{\Sigma}}^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_q).$$

To improve upon $\hat{\boldsymbol{\beta}}$,we construct a generalized profile least squares estimator of $\boldsymbol{\beta}$ by taking the heteroscedasticity of model (1.1) into consideration. Let $\sigma_i = \sqrt{\sigma^2(\mathbf{x}_i, \mathbf{z}_i, u_i)}$. Then our new estimator of $\boldsymbol{\beta}$ is defined as

$$\bar{\boldsymbol{\beta}} = (\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}/\sigma_i^2)^{-1}(\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{y}}_i/\sigma_i^2). \tag{3.7}$$

The following theorem gives the asymptotic normality of $\bar{\boldsymbol{\beta}}$.

Theorem 3.3 *Suppose that Assumptions 3.1–3.5 hold, we have*

$$\sqrt{n}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Delta^{-1}}),$$

where $\mathbf{\Delta} = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}/\sigma^2(\mathbf{x}_i, \mathbf{z}_i, u_i)]$.

The next theorem provides the asymptotic normality of the nonparametric components estimators.

Theorem 3.4 *Suppose that Assumptions 3.1–3.5 hold and* $nh^5 = O(1)$, *we have*

$$\sqrt{nh}\Big(\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{1}{2}h^2\mu_2\boldsymbol{\alpha}''(u)\Big) \xrightarrow{d} N(\mathbf{0}, \nu_0\mathbf{\Psi}/f(u)),$$

where $\mathbf{\Psi} = \mathbf{\Gamma}(u)^{-1}\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\sigma^2(\mathbf{x}_1, \mathbf{z}_1, u_1)|u]\mathbf{\Gamma}(u)^{-1}$.

4. Proof of the main results

In order to prove the main results, we first introduce several lemmas. The notation $c_n = h^2 + \{\frac{\log(1/h)}{nh}\}^{1/2}$ will be used in the proofs of the lemmas and theorems.

Lemma 4.1 *Let* $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$ *be iid random sequence, where the* $\mathbf{Y}_i^{'}s$ *are scalar random variables. Further assume that* $\mathrm{E}|y|^s < \infty$ *and* $\mathrm{Sup}_x \int |y|^s f(x, y)\mathrm{d}y < \infty$, *where* f *denotes the joint density of* $(\mathbf{X}, \mathbf{Y})$. *Let* K *be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Then*

$$\sup_x |\frac{1}{n}\sum_{i=1}^{n}[K_h(\mathbf{X}_i - x)\mathbf{Y}_i - \mathrm{E}\,(K_h(\mathbf{X}_i - x)\mathbf{Y}_i)]\,| = O_p\Big(\Big\{\frac{\log(1/h)}{nh}\Big\}^{1/2}\Big),$$

provided that $n^{2\varepsilon-1}h \to \infty$ *for some* $\varepsilon < 1 - s^{-1}$.

Proof This follows immediately from the result that was obtained by Mack and Silverman[16].

Lemma 4.2 *Under the assumptions 3.1–3.5, we have*

$$\frac{1}{n}\mathbf{Z}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{Z} \to \mathbf{\Xi}.$$

Lemma 4.3 *Under the assumptions 3.1–3.5, we have*

$$n^{-1}\mathbf{Z}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{M} = O_p(c_n^2).$$

The proofs of Lemmas 4.2 and 4.3 can be found in Fan and Huang[15].

Proof of Theorem 3.1 By the definition of $\hat{\boldsymbol{\beta}}$, we have

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= [\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{Y}\\ &= [\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})(\mathbf{Z}\boldsymbol{\beta}+\mathbf{M}+\boldsymbol{\varepsilon})\\ &= \boldsymbol{\beta} + [\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{M} + [\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}.\end{aligned}$$

Hence,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \sqrt{n}[\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{M} + \sqrt{n}[\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}. \tag{4.1}$$

By Lemmas 4.2 and 4.3, it is easy to prove that

$$\sqrt{n}[\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\mathbf{M} = O_p(\sqrt{n}c_n^2) = o_p(1), \tag{4.2}$$

and

$$\sqrt{n}[\bar{\mathbf{Z}}^{\mathrm{T}}\bar{\mathbf{Z}}]^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} = n^{-1/2}\mathbf{\Xi}^{-1}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon}\{1+o_p(1)\}. \tag{4.3}$$

By Lemma 4.1, we can obtain

$$\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} = \sum_{i=1}^{n}\{\mathbf{z}_i - \Phi^{\mathrm{T}}(u_i)\Gamma^{-1}(u_i)\mathbf{x}_i\}\varepsilon_i\{1+o_p(1)\} = \sum_{i=1}^{n}\boldsymbol{\eta}_i\varepsilon_i\{1+o_p(1)\}.$$

For random variable $\boldsymbol{\eta}_i\varepsilon_i$, we have

$$\mathbf{E}(\boldsymbol{\eta}_i\varepsilon_i) = 0, \quad \mathrm{Var}(\boldsymbol{\eta}_i\varepsilon_i) = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\varepsilon_i^2] = E[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}E[\varepsilon_i^2|\mathbf{x}_i,\mathbf{z}_i,u_i]] = \mathrm{E}[\boldsymbol{\eta}_i\boldsymbol{\eta}_i^{\mathrm{T}}\sigma^2(\mathbf{x}_i,\mathbf{z}_i,u_i)] = \mathbf{\Omega}.$$

By central limit theorem, we have

$$n^{-1/2}\bar{\mathbf{Z}}^{\mathrm{T}}(\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} \xrightarrow{d} N(\mathbf{0},\mathbf{\Omega}). \tag{4.4}$$

Combing (4.1)–(4.4), by Slutsky theorem, there holds

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0},\mathbf{\Xi^{-1}\Omega\Xi^{-1}}).$$

Proof of Theorem 3.2 From the definition of $\hat{\varepsilon}_i$, we have

$$\hat{\varepsilon}_i = \bar{y}_i - \bar{\mathbf{z}}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} = \varepsilon_i + (\boldsymbol{\alpha}^{\mathrm{T}}(u_i)\mathbf{x}_i - \mathbf{S}_i^{\mathrm{T}}\mathbf{M}) + \bar{\mathbf{z}}_i^{\mathrm{T}}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) - \mathbf{S}_i^{\mathrm{T}}\boldsymbol{\varepsilon},$$

where $\mathbf{S}_i = (s_{i1}, s_{i2}, s_{in})^{\mathrm{T}}$, and s_{ij} is the (ij)th element of matrix $\mathbf{S}$.

By Lemma 4.1, it is easy to prove that

$$\boldsymbol{\alpha}^{\mathrm{T}}(u_i)\mathbf{x}_i - \mathbf{S}_i^{\mathrm{T}}\mathbf{M} = O_p(c_n), \quad \mathbf{S}_i^{\mathrm{T}}\boldsymbol{\varepsilon} = O_p(c_n).$$

From the result of Theorem 3.1, we have $\boldsymbol{\beta} - \hat{\boldsymbol{\beta}} = O_p(n^{-1/2})$. By the above results, we can obtain that

$$\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}\hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}\varepsilon_i^2 + o_p(1). \tag{4.5}$$

From Lemma 4.1, it is easy to prove that

$$\bar{\mathbf{z}}_i = \mathbf{z}_i - \Phi^{\mathrm{T}}(u_i)\Gamma^{-1}(u_i)\mathbf{x}_i\{1 + O_p(c_n)\}. \tag{4.6}$$

By the law of large numbers, it follows from (4.5) and (4.6),

$$\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}\varepsilon_i^2 \to^P \boldsymbol{\Omega}$$

Applying Lemma 4.2 gives

$$(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}})^{-1}(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}}\hat{\varepsilon}_i^2)(\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{z}}_i\bar{\mathbf{z}}_i^{\mathrm{T}})^{-1} \to^P \boldsymbol{\Xi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Xi}^{-1}.$$

Proof of Theorem 3.3 By the same arguments as used in the proof of Theorem 3.1, we can prove Theorem 3.3, we omit the details.

Proof of Theorem 3.4 By (2.8), we have

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}(u) = (\hat{\alpha}_1(u), \ldots, \hat{\alpha}_p(u))^{\mathrm{T}} &= (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}})\\
&= (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{M} + (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\boldsymbol{\varepsilon} +\\
&\quad (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{Z}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\\
&\doteq I_1 + I_2 + I_3.
\end{aligned}$$

Therefore, in order to complete the proof of this theorem, we just need to show

$$\begin{cases}
\text{(i)} & I_1 = \boldsymbol{\alpha}(u) + \frac{1}{2}h^2\mu_2\boldsymbol{\alpha}''(u) + o_p(h^2),\\
\text{(ii)} & \sqrt{nh}I_2 \xrightarrow{d} N(\mathbf{0}, \nu_0\boldsymbol{\Psi}/f(u)),\\
\text{(iii)} & \sqrt{nh}I_3 = o_p(1).
\end{cases}$$

Proof of (i) Because the coefficient functions $\boldsymbol{\alpha}(\cdot)$ are smooth in the neighborhood of $|u_i - u| < h$, by the Taylor's expansion, we have

$$\boldsymbol{\alpha}(u_i) = \boldsymbol{\alpha}(u) + h\boldsymbol{\alpha}'(u)(\frac{u_i - u}{h}) + \frac{h^2}{2}\boldsymbol{\alpha}''(u)(\frac{u_i - u}{h})^2 + o_p(h^2).$$

Therefore,

$$\mathbf{M} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\alpha}(u_1)\\ \mathbf{x}_2^{\mathrm{T}}\boldsymbol{\alpha}(u_2)\\ \vdots\\ \mathbf{x}_n^{\mathrm{T}}\boldsymbol{\alpha}(u_n) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}}\left\{\boldsymbol{\alpha}(u) + h\boldsymbol{\alpha}'(u)(\frac{u_1-u}{h}) + \frac{h^2}{2}\boldsymbol{\alpha}''(u)(\frac{u_1-u}{h})^2\right\}\\ \mathbf{x}_2^{\mathrm{T}}\left\{\boldsymbol{\alpha}(u) + h\boldsymbol{\alpha}'(u)(\frac{u_2-u}{h}) + \frac{h^2}{2}\boldsymbol{\alpha}''(u)(\frac{u_2-u}{h})^2\right\}\\ \vdots\\ \mathbf{x}_n^{\mathrm{T}}\left\{\boldsymbol{\alpha}(u) + h\boldsymbol{\alpha}'(u)(\frac{u_n-u}{h}) + \frac{h^2}{2}\boldsymbol{\alpha}''(u)(\frac{u_n-u}{h})^2\right\} \end{pmatrix} + o_p(h^2)$$

$$
= \mathbf{D}_u \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix} + \frac{h^2}{2} \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}}(\frac{u_1-u}{h})^2 \\ \mathbf{x}_2^{\mathrm{T}}(\frac{u_2-u}{h})^2 \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}}(\frac{u_n-u}{h})^2 \end{pmatrix} \boldsymbol{\alpha}''(u) + o_p(h^2). \tag{4.7}
$$

By Lemma 4.1, we can prove that

$$
\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u = \begin{pmatrix} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} K_h(u_i-u) & \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\frac{u_i-u}{h}) K_h(u_i-u) \\ \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\frac{u_i-u}{h}) K_h(u_i-u) & \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\frac{u_i-u}{h})^2 K_h(u_i-u) \end{pmatrix}
$$

$$
= nf(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \{1 + O_p(c_n)\}, \tag{4.8}
$$

and

$$
\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}}(\frac{u_1-u}{h})^2 \\ \mathbf{x}_2^{\mathrm{T}}(\frac{u_2-u}{h})^2 \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}}(\frac{u_n-u}{h})^2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\frac{u_i-u}{h})^2 K_h(u_i-u_0) \\ \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\frac{u_i-u}{h})^3 K_h(u_i-u_0) \end{pmatrix}
$$

$$
= n\mu_2(K)f(u)\boldsymbol{\Gamma}(u) \otimes (1,0)^{\mathrm{T}}\{1 + O_p(c_n)\}. \tag{4.9}
$$

From (4.7)–(4.9), we have,

$$
\begin{aligned}
I_1 = {} & (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u \begin{pmatrix} \boldsymbol{\alpha}(u) \\ h\boldsymbol{\alpha}'(u) \end{pmatrix} + \\
& \frac{1}{2}h^2\boldsymbol{\alpha}''(u)(\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}}(\frac{u_1-u}{h})^2 \\ \mathbf{x}_2^{\mathrm{T}}(\frac{u_2-u}{h})^2 \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}}(\frac{u_n-u}{h})^2 \end{pmatrix} + \\
& (\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{1}_n o_p(h^2) \\
= {} & \boldsymbol{\alpha}(u) + \frac{1}{2}h^2\mu_2\boldsymbol{\alpha}''(u) + o_p(h^2).
\end{aligned}
$$

The proof of (i) is completed.

Proof of (ii) By (4.8), we have

$$
\begin{aligned}
\sqrt{nh}I_2 &= \sqrt{nh}(\mathbf{I}_{\mathrm{p}}\ \mathbf{0}_{\mathrm{p}})\{\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{D}_u\}^{-1}\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\boldsymbol{\varepsilon} \\
&= f(u)^{-1}\boldsymbol{\Gamma}(u)^{-1}\Big(\sqrt{nh}\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i K_h(u_i-u)\varepsilon_i\Big) + o_p(1).
\end{aligned}
$$

As for $\sqrt{nh}\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i K_h(u_i-u)\varepsilon_i$, it is obviously asymptotically normal with mean 0 and variance

$$
\mathbf{V}_u = f(u)\nu_0\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\sigma^2(\mathbf{x}_1,\mathbf{z}_1,u_1)|u] + o(1).
$$

Therefore

$$\sqrt{nh}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i K_h(u_i-u)\varepsilon_i \to N\left(\mathbf{0}, \nu_0 f(u)\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\sigma^2(\mathbf{x}_1,\mathbf{z}_1,u_1)|u]\right).$$

By the Slutsky theorem, we have

$$\sqrt{nh}I_2 \to N\left(\mathbf{0}, \nu_0 f(u)^{-1}\mathbf{\Gamma}(u)^{-1}\mathrm{E}[\mathbf{x}_1\mathbf{x}_1^{\mathrm{T}}\sigma^2(\mathbf{x}_1,\mathbf{z}_1,u_1)|u]\mathbf{\Gamma}(u)^{-1}\right).$$

This completes the proof of (ii).

Proof of (iii) By Lemma 4.1, we can obtain that

$$\mathbf{D}_u^{\mathrm{T}}\mathbf{W}_u\mathbf{Z} = nf(u)\mathbf{\Phi}(u)\otimes(1,0)^{\mathrm{T}}\{1+O_p(c_n)\}. \tag{4.10}$$

From Theorem 3.1, we have

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} = O_p(n^{-1/2}). \tag{4.11}$$

Combining (4.10), (4.11) and (4.7) yields

$$\sqrt{nh}I_3 = \sqrt{nh}\mathbf{\Gamma}(u)^{-1}\mathbf{\Phi}(u)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = o_p(1).$$

This completes the proof of (iii). □

References

[1] HASTIE T, TIBSHIRANI R. *Varying-coefficient models* [J]. J. Roy. Statist. Soc. Ser. B, 1993, **55**(4): 757–796.

[2] CAI Zongwu, FAN Jianqing, LI Runze. *Efficient estimation and inferences for varying-coefficient models* [J]. J. Amer. Statist. Assoc., 2000, **95**(451): 888–902.

[3] CAI Zongwu, FAN Jianqing, YAO Qiwei. *Functional-coefficient regression models for nonlinear time series* [J]. J. Amer. Statist. Assoc., 2000, **95**(451): 941–956.

[4] FAN Jianqing, ZHANG Wenyang. *Statistical estimation in varying coefficient models* [J]. Ann. Statist., 1999, **27**(5): 1491–1518.

[5] FOTHERINGHAM A S, CHARLTON M, BRUNSDON C. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* [M]. New York: Wiley, 2002.

[6] HOOVER D R, RICE J A, WU C O. et al. *Nonparametric smoothing estimation of time-varying coefficient models with longitudinal data* [J]. Biometrika, 1998, **85**: 809–822.

[7] HUANG Jianhua, WU C O, ZHOU Lan. *Polynominal spline estimation and inference for Varying Coefficient Models with longitudinal data* [J]. Statist. Sinica, 2004, **14**: 763–788.

[8] ENGLE R F, GRANGER W J, RICE J. et al. *Semiparametric estimates of the relation between weather and electricity sales* [J]. J. Amer. Statist. Assoc., 1986, **80**: 310–319.

[9] HARDLE W, LIANG Hua, GAO Jiti. *Partially Linear Models* [M]. Springer, Heidelberg: Physica-Verlag, 2000.

[10] HARDLE W, MULLER M, SPERLICH S. et al. *Nonparametric and Semiparametric Models* [M]. Berlin: Springer-Verlag, 2004.

[11] ZHANG Wenyang, LEE S Y, SONG Xinyuan. *Local polynomial fitting in semivarying coefficient model* [J]. J. Multivariate Anal., 2002, **82**(1): 166–188.

[12] ZHOU Xian, YOU Jinhong. *Wavelet estimation in varying-coefficient partially linear regression models* [J]. Statist. Probab. Lett., 2004, **68**(1): 91–104.

[13] XIA Yingcun, ZHANG Wenyang, TONG H. *Efficient estimation for semivarying-coefficient models* [J]. Biometrika, 2004, **91**(3): 661–681.

[14] AHMAD I, LEELAHANON S, LI Qi. *Efficient estimation of semiparametric partially linear varying coefficient model* [J]. Ann. Statist, 2005, **33**: 258–283.

[15] FAN Jianqing, HUANG Tao. *Profile likelihood inferences on semiparametric varying-coefficient partially linear models* [J]. Bernoulli, 2005, **11**(6): 1031–1057.

[16] MACK Y P, SILVERMAN B W. *Weak and strong uniform consistency of kernel regression estimates* [J]. Z. Wahrsch. Verw. Gebiete, 1982, **61**(3): 405–415.