Ann Inst Stat Math (2014) 66:121–140
DOI 10.1007/s10463-013-0407-z

Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition

Sanying Feng · Liugen Xue

Received: 14 October 2012 / Revised: 12 March 2013 / Published online: 27 April 2013

Abstract In this paper, we consider the statistical inference for the partially liner varying coefficient model with measurement error in the nonparametric part when some prior information about the parametric part is available. The prior information is expressed in the form of exact linear restrictions. Two types of local bias-corrected restricted profile least squares estimators of the parametric component and nonparametric component are conducted, and their asymptotic properties are also studied under some regularity conditions. Moreover, we compare the efficiency of the two kinds of parameter estimators under the criterion of Loẅner ordering. Finally, we develop a linear hypothesis test for the parametric component. Some simulation studies are conducted to examine the finite sample performance for the proposed method. A real dataset is analyzed for illustration.

Keywords Partially linear varying coefficient model · Errors-in-variables · Local bias-corrected · Restricted estimator · Profile Lagrange multiplier test · Asymptotic normality

1 Introduction

Partially linear varying coefficient model has attracted lots of attention due to its flexibility to combine traditional linear model with varying coefficient model. For example, Zhang et al. (2002, 2011), Xia et al. (2004), Fan and Huang (2005), Ahmad et al. (2005), You and Zhou (2006), Zhou and Liang (2009), Li et al. (2011, 2012), Kai et al. (2011), and among others. The partially linear varying coefficient model assumes the following structure:

S. Feng (✉) · L. Xue
College of Applied Sciences, Beijing University of Technology, Beijing 100124, China
e-mail: fsy5801@emails.bjut.edu.cn

$$Y = \boldsymbol{X}^{\tau}\boldsymbol{\beta} + \boldsymbol{Z}^{\tau}\boldsymbol{\alpha}(T) + \varepsilon, \tag{1}$$

where $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^{\tau}$ is a q-dimensional vector of unknown coefficient functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\tau}$ is a p-dimensional vector of unknown parameters and ε is the random error with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$. In this model, the dependence of $\boldsymbol{\alpha}(\cdot)$ on T implies a special kind of interaction between the covariate $\boldsymbol{Z}$ and T. The model is quite general and includes many important statistical models. For example, when $\boldsymbol{\alpha}(\cdot) = \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a constant vector, model (1) reduces to the usual linear regression model. When $q = 1$ and $\boldsymbol{Z} = 1$, model (1) becomes the partially linear regression model. When $\boldsymbol{X} = 0$, model (1) reduces to the famous varying coefficient model.

Measurement error data are often encountered in many fields, including engineering, economics, physics, biology, biomedical sciences and epidemiology. Statistical inference methods for various parametric measurement error models have been well established over the past several decades, such as Fuller (1987), and Carroll et al. (1995) studied linear errors-in-variables (EV) models and nonlinear EV models in detail, respectively. It is well known that, if the measurement errors are ignored entirely, the resulting estimators will be biased. For the partially linear varying coefficient model (1), when the covariate $\boldsymbol{X}$ is observed with additive error, You and Chen (2006) studied the estimations of parametric and nonparametric components, they showed that the proposed modified profile least squares estimator for parameter of interest is strongly consistent and asymptotically normal. Hu et al. (2009) and Wang et al. (2011) constructed the confidence regions of the unknown parameters by the empirical likelihood method, respectively. However, in this paper, we consider the nonparametric part covariate $\boldsymbol{Z}$ is measured with additive error and both $\boldsymbol{X}$ and T are measured exactly. That is, instead of the true $\boldsymbol{Z}$, the surrogate variable $\boldsymbol{W}$ is observed by

$$\boldsymbol{W} = \boldsymbol{Z} + \boldsymbol{U}, \tag{2}$$

where $\boldsymbol{U}$ is the measurement error, which is independent of $(\boldsymbol{X}^{\tau}, \boldsymbol{Z}^{\tau}, T, \varepsilon)$ and has the known covariance $\mathrm{Cov}(\boldsymbol{U}) = \Sigma_u$. If Σ_u is unknown, we also can estimate it by repeatedly measuring $\boldsymbol{W}$. The specific details can be found in Liang et al. (1999). The researches for models (1) and (2) are seldom discussed. When $\boldsymbol{X} = 0$, You et al. (2006) considered the varying coefficient EV models with corrected local polynomial method, and studied the asymptotic properties of the estimators.

In many important statistical applications, in addition to sample information we have some prior information on regression parametric vector which can be used to improve the parametric estimators. In this paper, we consider the following restricted condition

$$A\boldsymbol{\beta} = \boldsymbol{d}, \tag{3}$$

where A is a $k \times p$ known full row rank matrix, $\boldsymbol{d}$ is a $k \times 1$ known vector. For model (1), Fan and Huang (2005) proposed the GLR test statistics and Wald test statistics to test $A\boldsymbol{\beta} = 0$; Wei and Wu (2008) constructed the profile Lagrange multiplier test statistic under the restricted condition (3). In their papers, they did not consider the case of measurement errors. When $\boldsymbol{X}$ is measured with the additive error, Zhang et al. (2011) and Wei (2012) proposed the restricted modified profile least-squares estimators for parametric and nonparametric components, and constructed the modified profile Lagrange multiplier test statistics under additional restricted condition.

In this paper, we consider models (1) and (2) based on the restricted condition (3), and investigate the estimation and testing issues. Under the additional linear restricted condition, we propose a local bias-corrected restricted profile least squares approach by combining with the profile least squares, the so-called "correction for attenuation" and Lagrange multiplier method, and we obtain two estimators of the parametric component and the coefficient functions, respectively. Moreover, we compare the performance of the two parameter estimators under the criterion of Loẅner ordering. When A is taken to the different matrix, we can derive the different constrained estimators. Therefore, the proposed method is more effective in practical application. At last, we construct the local bias-corrected profile Lagrange multiplier (BCPLM) test statistic for the unknown parameter vector $\boldsymbol{\beta}$, and show that its limiting distribution is a standard Chi-squared distribution under the null hypothesis.

The paper is organized as follows. In Sect. 2, we propose the local bias-corrected restricted profile least squares method, and investigate the asymptotic properties of the estimators. In Sect. 3, we construct the BCPLM test statistic. The asymptotic distribution of the statistic is derived under regularity conditions. In Sect. 4, some simulation studies are carried out to assess the performance of the proposed method. A real data example is used for illustration in Sect. 5. Lastly, the article is concluded with a brief discussion in Sect. 6. The proofs of the main results are given in Appendix.

2 Methodology and asymptotic properties

2.1 Local bias-corrected restricted profile least squares method

Suppose that $\{(Y_i; \boldsymbol{X}_i^\tau, \boldsymbol{W}_i^\tau, T_i), 1 \le i \le n\}$ is an independent identically distributed (iid) random sample which comes from models (1) and (2). That is, they satisfy

$$\begin{cases} Y_i = \boldsymbol{X}_i^\tau \boldsymbol{\beta} + \boldsymbol{Z}_i^\tau \boldsymbol{\alpha}(T_i) + \varepsilon_i, \\ \boldsymbol{W}_i = \boldsymbol{Z}_i + \boldsymbol{U}_i, \end{cases} \tag{4}$$

where the covariate $\boldsymbol{Z}_i$ is measured with additive errors, $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{iq})^\tau$ is the surrogate variable of $\boldsymbol{Z}_i$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})^\tau$, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \ldots, \alpha_q(\cdot))^\tau$, $\{\varepsilon_i\}_{i=1}^n$ are independent and identically distributed random errors with mean zero and variance σ^2. To avoid the curse of dimensionality, we assume that T_i is a univariate variable.

If $\boldsymbol{\beta}$ is known, we can write

$$Y_i - \boldsymbol{X}_i^\tau \boldsymbol{\beta} = \boldsymbol{Z}_i^\tau \boldsymbol{\alpha}(T_i) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{5}$$

Obviously, model (5) can be treated as the usual varying coefficient model. Thus, the local linear regression approximation can be used to estimate the varying coefficient functions $\{\alpha_j(\cdot), j = 1, \ldots, q\}$. For T in a small neighborhood of t, approximate each $\alpha_j(T)$ by

$$\alpha_j(T) \approx \alpha_j(t) + \alpha_j'(t)(T - t) \equiv a_j + b_j(T - t), \quad j = 1, \ldots, q,$$

where $\alpha_j'(t) = \partial\alpha_j(t)/\partial t$. This lead to the following weighted local least-squares problem: finding $\boldsymbol{a}$ and $\boldsymbol{b}$ to minimize

$$\sum_{i=1}^{n}\{Y_i - \boldsymbol{X}_i^{\tau}\boldsymbol{\beta} - \boldsymbol{Z}_i^{\tau}[\boldsymbol{a} + \boldsymbol{b}(T_i - t)]\}^2 K_h(T_i - t), \tag{6}$$

where $\boldsymbol{a} = (a_1, \ldots, a_q)^{\tau}$, $\boldsymbol{b} = (b_1, \ldots, b_q)^{\tau}$ and $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function and h is a bandwidth.

For the sake of descriptive convenience, we denote $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^{\tau}$, $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)^{\tau}$, $\mathbf{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)^{\tau}$, $\mathbf{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)^{\tau}$, $\boldsymbol{\omega}_t = \text{diag}(K_h(T_1 - t), \ldots, K_h(T_n - t))$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\tau}$, and

$$\mathbf{D}_t^Z = \begin{pmatrix} \boldsymbol{Z}_1^{\tau} & \frac{T_1 - t}{h} & \boldsymbol{Z}_1^{\tau} \\ \vdots & \vdots & \\ \boldsymbol{Z}_n^{\tau} & \frac{T_n - t}{h} & \boldsymbol{Z}_n^{\tau} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \boldsymbol{Z}_1^{\tau}\boldsymbol{\alpha}(T_1) \\ \vdots \\ \boldsymbol{Z}_n^{\tau}\boldsymbol{\alpha}(T_n) \end{pmatrix}.$$

Then the solution to problem (6) is given by

$$[\hat{\boldsymbol{a}}^{\tau}, h\hat{\boldsymbol{b}}^{\tau}]^{\tau} = \left\{(\mathbf{D}_t^Z)^{\tau}\boldsymbol{\omega}_t\mathbf{D}_t^Z\right\}^{-1}(\mathbf{D}_t^Z)^{\tau}\boldsymbol{\omega}_t(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{7}$$

However, in our case we observe $\boldsymbol{W}_i$ instead of $\boldsymbol{Z}_i$. If we directly replace $\boldsymbol{Z}_i$ with $\boldsymbol{W}_i$ in (7), we will get an inconsistent estimator due to the measurement error. In order to overcome the effect of the measurement error, referring the method of You et al. (2006), we modify (7) to define the corrected local linear estimator by

$$[\hat{\boldsymbol{a}}^{\tau}, h\hat{\boldsymbol{b}}^{\tau}]^{\tau} = \left\{(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega\right\}^{-1}(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{D}_t^W$ has the same form as $\mathbf{D}_t^Z$ except that $\boldsymbol{Z}_i$ are replaced by $\boldsymbol{W}_i$, and

$$\Omega = \sum_{i=1}^{n}\Sigma_u \otimes \begin{pmatrix} 1 & \frac{T_i - t}{h} \\ \frac{T_i - t}{h} & \left(\frac{T_i - t}{h}\right)^2 \end{pmatrix} K_h(T_i - t),$$

where $\otimes$ is the Kronecker product. Therefore, when $\boldsymbol{\beta}$ is given, we can obtain the estimate of the vector $\boldsymbol{\alpha}(t)$ of coefficient functions by

$$\hat{\boldsymbol{\alpha}}(t, \boldsymbol{\beta}) = (\mathbf{I}_q \quad \mathbf{0}_q)\left\{(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega\right\}^{-1}(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{8}$$

where $\mathbf{I}_q$ denotes a q-dimensional identity matrix and $\mathbf{0}_q$ is the $q \times q$ matrix with all the entries being zero.

Denote $\boldsymbol{Q} = (Q_1^\tau, \ldots, Q_n^\tau)^\tau$, $\boldsymbol{S} = (Q_1^\tau \boldsymbol{W}_1, \ldots, Q_n^\tau \boldsymbol{W}_n)^\tau$, $\tilde{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{Y}$, $\tilde{\mathbf{X}} = (\boldsymbol{I} - \boldsymbol{S})\mathbf{X}$, where $Q_i = (\mathbf{I}_q \quad \mathbf{0}_q)\{(\mathbf{D}_{T_i}^W)^\tau \omega_{T_i} \mathbf{D}_{T_i}^W - \Omega\}^{-1} (\mathbf{D}_{T_i}^W)^\tau \omega_{T_i}, i = 1, \ldots, n$. Then we can obtain the local bias-corrected profile least-square estimator of $\boldsymbol{\beta}$ by minimizing

$$\sum_{i=1}^{n} \{Y_i - \boldsymbol{X}_i^\tau \boldsymbol{\beta} - \boldsymbol{W}_i^\tau \hat{\boldsymbol{\alpha}}(T_i, \boldsymbol{\beta})\}^2 - \sum_{i=1}^{n} \hat{\boldsymbol{\alpha}}^\tau (T_i, \boldsymbol{\beta}) \Sigma_u \hat{\boldsymbol{\alpha}}(T_i, \boldsymbol{\beta}), \tag{9}$$

that is

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}^\tau \tilde{\mathbf{X}} - \mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q}\mathbf{X})^{-1} (\tilde{\mathbf{X}}^\tau \tilde{\boldsymbol{Y}} - \mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q}\boldsymbol{Y}). \tag{10}$$

By the estimator (10) of $\boldsymbol{\beta}$, we can define the estimator of $\boldsymbol{\alpha}(t)$ by

$$\tilde{\boldsymbol{\alpha}}(t) = (\mathbf{I}_q \quad \mathbf{0}_q) \left\{ (\mathbf{D}_t^W)^\tau \boldsymbol{\omega}_t \mathbf{D}_t^W - \Omega \right\}^{-1} (\mathbf{D}_t^W)^\tau \boldsymbol{\omega}_t (\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{11}$$

Similar to the case in linear model, if some prior information for regression coefficients of interest can be obtained, the efficiency of the estimator can be improved by using such prior information. Up to this point, we make use of exact linear restrictions for the parameters of interest in model (1). Based on the linear restricted condition (3), a modified Lagrange function of $\boldsymbol{\beta}$ is defined as

$$\begin{aligned} F(\boldsymbol{\beta}, \boldsymbol{\lambda}) = &\sum_{i=1}^{n} \{Y_i - \boldsymbol{X}_i^\tau \boldsymbol{\beta} - \boldsymbol{W}_i^\tau \hat{\boldsymbol{\alpha}}(T_i, \boldsymbol{\beta})\}^2 - \sum_{i=1}^{n} \hat{\boldsymbol{\alpha}}^\tau (T_i, \boldsymbol{\beta}) \Sigma_u \hat{\boldsymbol{\alpha}}(T_i, \boldsymbol{\beta}) \\ &+ 2\boldsymbol{\lambda}^\tau (A\boldsymbol{\beta} - \boldsymbol{d}). \end{aligned}$$

Differentiating $F(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ and setting the result to zero, respectively, we obtain

$$\begin{cases} \frac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} = -2\tilde{\mathbf{X}}^\tau (\tilde{\boldsymbol{Y}} - \tilde{\mathbf{X}}\boldsymbol{\beta}) + 2\mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}) + \mathbf{2A}^\tau \boldsymbol{\lambda} = \mathbf{0}, \\ \frac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = A\boldsymbol{\beta} - \boldsymbol{d} = 0. \end{cases} \tag{12}$$

The solution to the problem (12) is given by

$$\begin{cases} \hat{\boldsymbol{\beta}}^{R_1} = \hat{\boldsymbol{\beta}} - P^{-1} A^\tau (A P^{-1} A^\tau)^{-1} (A\hat{\boldsymbol{\beta}} - \boldsymbol{d}), \\ \hat{\boldsymbol{\lambda}} = (A P^{-1} A^\tau)^{-1} (A\hat{\boldsymbol{\beta}} - \boldsymbol{d}), \end{cases} \tag{13}$$

where $P = \tilde{\mathbf{X}}^\tau \tilde{\mathbf{X}} - \mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q}\mathbf{X}$.

An alternative strategy to obtain the consistent estimator of $\boldsymbol{\beta}$ satisfying the restrictions is to minimize $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\tau (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, subject to the linear restrictions $A\boldsymbol{\beta} = \boldsymbol{d}$, similar to the argument of above, we obtain the following restricted estimator:

$$\hat{\boldsymbol{\beta}}^{R_2} = \hat{\boldsymbol{\beta}} - A^{\tau}(AA^{\tau})^{-1}(A\hat{\boldsymbol{\beta}} - \boldsymbol{d}). \tag{14}$$

We call the estimators $\hat{\boldsymbol{\beta}}^{R_k}$, $k = 1, 2$, as the local bias-corrected restricted profile least squares estimator (BCRPLSE) of $\boldsymbol{\beta}$. By (8), (13) and (14), we can define the estimators of the coefficient functions $\boldsymbol{\alpha}(\cdot)$ as follows

$$\hat{\boldsymbol{\alpha}}^{R_k}(t) = (\mathbf{I}_q \ \ \mathbf{0}_q)\left\{(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega\right\}^{-1}(\mathbf{D}_t^W)^{\tau}\boldsymbol{\omega}_t(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{R_k}), \quad k = 1, 2. \tag{15}$$

2.2 Theoretical results

We begin this subsection with the following assumptions required to derive the main results. These assumptions are quite mild and can be easily satisfied.

(A1) The random variable T has a compact support $\mathcal{T}$. The density function $f(\cdot)$ of T is Lipschitz continuous and bounded away from zero on $\mathcal{T}$.
(A2) There is a $s > 2$ such that $E||\boldsymbol{X}_1||^{2s} < \infty$, $E||\boldsymbol{Z}_1||^{2s} < \infty$, $E|\varepsilon|^{2s} < \infty$ and $E||\boldsymbol{U}_1||^{2s} < \infty$, and for some $r < 2 - s^{-1}$ there is $n^{2r-1}h \to \infty$ as $n \to \infty$.
(A3) $\{\alpha_j(\cdot), j = 1, \ldots, q\}$ have continuous second derivatives on $\mathcal{T}$.
(A4) The kernel function $K(\cdot)$ is a symmetric probability density functions with bounded support and the bandwidth h satisfies $nh^8 \to 0$ and $nh^2(\log n)^{-2} \to \infty$ as $n \to \infty$.
(A5) The $q \times q$ matrix $\Gamma(t) = E(\boldsymbol{Z}_1\boldsymbol{Z}_1^{\tau}|T = t)$ is non-singular for each $t \in \mathcal{T}$. $E(\boldsymbol{X}_1\boldsymbol{X}_1^{\tau}|T = t)$, $\Gamma^{-1}(t)$ and $\Phi(t) = E(\boldsymbol{Z}_1\boldsymbol{X}_1^{\tau}|T = t)$ are all Lipschitz continuous.

Theorem 1 *Assume that the conditions* (A1)–(A5) *hold. Then* $\hat{\boldsymbol{\beta}}^{R_1}$ *is an asymptotically normal estimator, that is*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{R_1} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(0, \Sigma), \quad n \to \infty,$$

where "$\xrightarrow{\mathcal{L}}$*" denotes the convergence in distribution, and*

$$\begin{aligned}
\Sigma &= \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} - \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1} - \Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} \\
&\quad + \Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_3\Sigma_1^{-1}, \\
\Sigma_1 &= E(\boldsymbol{X}_1\boldsymbol{X}_1^{\tau}) - E(\Phi^{\tau}(T_1)\Gamma^{-1}(T_1)\Phi(T_1)), \quad \Sigma_3 = A^{\tau}[A\Sigma_1^{-1}A^{\tau}]^{-1}A, \\
\Sigma_2 &= E(\varepsilon_1 - \boldsymbol{U}_1^{\tau}\alpha(T_1))^2\Sigma_1 + \sigma^2 E\left\{\Phi^{\tau}(T_1)\Gamma^{-1}(T_1)\Sigma_u\Gamma^{-1}(T_1)\Phi(T_1)\right\}^2 \\
&\quad + E\left\{\Phi^{\tau}(T_1)\Gamma^{-1}(T_1)(\boldsymbol{U}_1\boldsymbol{U}_1^{\tau} - \Sigma_u)\alpha(T_1)\right\}^{\otimes 2}.
\end{aligned}$$

Theorem 2 *Assume that the conditions* (A1)–(A5) *hold. Then* $\hat{\boldsymbol{\beta}}^{R_2}$ *is an asymptotically normal estimator, that is*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{R_2} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(0, \Pi), \quad n \to \infty,$$

where $\Sigma_4 = A^\tau(AA^\tau)^{-1}A$,

$$\Pi = \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} - \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_4 - \Sigma_4\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} + \Sigma_4\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\Sigma_4.$$

To make statistical inference for $\boldsymbol{\beta}$ by Theorems 1 and 2, we need to estimate the asymptotic variances Σ and Π. Note that we only need to estimate Σ_1 and Σ_2, then we can get the consistent estimators of Σ and Π using plug-in method. Σ_1 and Σ_2 can be estimated, respectively, by

$$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{X}}_i^\tau - \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}^\tau Q_i^\tau \Sigma_u Q_i \mathbf{X}$$

and

$$\hat{\Sigma}_2 = \frac{1}{n}\sum_{i=1}^{n}\{\tilde{\boldsymbol{X}}_i(\tilde{Y}_i - \tilde{\boldsymbol{X}}_i^\tau\hat{\boldsymbol{\beta}}) - \Sigma_u\tilde{\boldsymbol{\alpha}}(T_i)\}^{\otimes 2},$$

where $R^{\otimes 2} = RR^\tau$.

Corollary 1 *Assume that the conditions* (A1)–(A5) *hold.* $\hat{\boldsymbol{\beta}}^{R_1}$ *is less efficient than* $\hat{\boldsymbol{\beta}}^{R_2}$ *under Löwner ordering if and only if* $\lambda_{\max}(\Pi\Sigma^{-1}) < 1$, *where* $\lambda_{\max}(\Pi\Sigma^{-1})$ *is the maximum eigenvalue of* $\Pi\Sigma^{-1}$.

Theorem 3 *Assume that the conditions* (A1)–(A5) *hold. Then, as* $n \to \infty$, *we have*

$$\sqrt{nh}\left(\hat{\boldsymbol{\alpha}}^{R_k}(t) - \boldsymbol{\alpha}(t) - \frac{h^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}''(t)\right) \xrightarrow{\mathcal{L}} N\Big(0, \Delta\Big),$$

where $\Delta = (\kappa_0^2\nu_0 + 2\kappa_0\kappa_1\nu_1 + \kappa_1^2\nu_2)f(t)^{-1}\Sigma^*$, $\kappa_0 = \mu_2/(\mu_2 - \mu_1^2)$, $\kappa_1 = -\mu_1/(\mu_2 - \mu_1^2)$, $\Sigma^* = \Gamma^{-1}(t)[\sigma^2\Gamma(t) + \sigma^2\Sigma_u + E\{\xi_1\boldsymbol{\alpha}(t)\boldsymbol{\alpha}^\tau(t)\xi_1^\tau|T = t\}]\Gamma^{-1}(t)$, $\xi_1 = \Sigma_u - \boldsymbol{U}_1\boldsymbol{U}_1^\tau - \boldsymbol{X}_1\boldsymbol{U}_1^\tau$.

Theorem 4 *Assume that the conditions* (A1)–(A5) *hold. Then,*

$$\max_{1\le j\le q}\sup_{t\in\mathcal{T}}|\hat{\alpha}_j^{R_k}(t) - \alpha_j(t)| = O\{h^2 + (\log n/nh)^{\frac{1}{2}}\}.\quad a.s.$$

Remark 1 If h takes the optimal bandwidth, that is $h_{opt} = cn^{-1/5}$ where c is a constant, according to Theorem 4 we have

$$\max_{1\le j\le q}\sup_{t\in\mathcal{T}}|\hat{\alpha}_j^{R_k}(t) - \alpha_j(t)| = O\Big\{n^{-2/5}(\log n)^{1/2}\Big\}.\quad a.s.$$

This mean that the estimators of the nonparametric component in model (4) achieve the optimal strong uniform convergence rate of the usual nonparametric estimation in nonparametric regression.

3 Hypothesis testing

Applying the estimation method described in the previous section, we consider the following linear hypothesis

$$H_0 : A\boldsymbol{\beta} = \boldsymbol{d} \qquad \text{vs.} \qquad H_1 : A\boldsymbol{\beta} = \boldsymbol{d} + \boldsymbol{\delta}, \tag{16}$$

where $\boldsymbol{\delta}$ is a $k \times 1$ constant vector. For the partially linear varying coefficient model (1), when covariate $\boldsymbol{X}$ is measured with errors, Zhang et al. (2011) and Wei (2012) constructed the modified profile Lagrange multiplier test statistics for the testing problem (16) and studied the asymptotic distribution of the test statistics, respectively. In this paper, the nonparametric covariate $\boldsymbol{Z}$ is measured with additive errors. In order to eliminate the influence of measured error on hypothesis test, we proposed a local bias-corrected profile Lagrange multiplier test method such that the proposed test statistic can achieve the standard Chi-squared limit.

Let $C_0 = [A\hat{\Sigma}_1^{-1}A^\tau]^{-1}$, $L_0 = A\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}A^\tau$, where $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$ is a consistent estimate of $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$ which is defined in Theorem 1. By the estimator of Lagrange multiplier defined in (13), we can construct the local bias-corrected profile Lagrange multiplier statistic under null hypothesis H_0 as follows

$$\hat{T}_{\text{BCPLM}} = \frac{1}{n}\hat{\boldsymbol{\lambda}}^\tau \left(C_0 L_0 C_0^\tau\right)^{-1}\hat{\boldsymbol{\lambda}}.$$

Theorem 5 *Assume that the conditions* (A1)–(A5) *hold, then*

(i) *under H_0 in* (16), $\hat{T}_{BCPLM} \xrightarrow{\mathcal{L}} \chi_k^2$, *as $n \to \infty$;*
(ii) *under H_1 in* (16), $\hat{T}_{BCPLM}$ *follows the asymptotic noncentral $\chi^2(k, \varsigma)$ distribution with k degrees of freedom, and the noncentral parameter is*

$$\varsigma = \lim_{n\to\infty} (A\boldsymbol{\beta} - \boldsymbol{d})^\tau (CLC)^{-1}(A\boldsymbol{\beta} - \boldsymbol{d}),$$

where $C = (A\Sigma_1^{-1}A^\tau)^{-1}$, $L = A\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}A^\tau$.

4 Simulation examples

To demonstrate the finite sample performance of the proposed method and the testing procedure, we conduct two simulation studies. The data are generated from the following model:

$$Y_i = X_{1i}\beta_1 + X_{2i}\beta_2 + X_{3i}\beta_3 + Z_i\alpha(T_i) + \varepsilon_i, \quad W_i = Z_i + u_i, \quad i = 1, \ldots, n, \tag{17}$$

where $\boldsymbol{\beta} = (1, 2, 1)^\tau$, $\alpha(T_i) = 2\cos(2\pi T_i)$, $T_i \sim U(0, 1)$, $\varepsilon_i \sim N(0, 0.5^2)$ and $(\boldsymbol{X}_i^\tau, Z_i)^\tau \sim N_4(\mu, \Sigma)$ with $\mu = (1, 1, 1, 1)^\tau$ and $\sigma_{kl} = 0.5^{|k-l|}$, $k, l = 1, 2, 3, 4$.

Furthermore, we assume $U_i \sim N(0, \sigma_u^2)$, and throughout the simulations, we take two variances of measurement errors for comparison: $\sigma_u^2 = 0.25$ and $\sigma_u^2 = 0.5$.

Throughout the simulations, we use the Epanechnikov kernel function $K(t) = 0.75(1 - t^2)_+$, and the bandwidth h is selected by the least squares cross-validation (CV) method. The CV statistic is defined as follows

$$\begin{aligned} \mathrm{CV}(h) = &\frac{1}{n}\sum_{i=1}^{n}\{Y_i - \boldsymbol{X}_i^\tau \hat{\boldsymbol{\beta}}^{(-i)} - \boldsymbol{W}_i^\tau \tilde{\boldsymbol{\alpha}}_{h,(-i)}(T_i)\}^2 \\ &- \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{\alpha}}_{h,(-i)}^\tau(T_i)\Sigma_u \tilde{\boldsymbol{\alpha}}_{h,(-i)}(T_i), \end{aligned} \tag{18}$$

where $\hat{\boldsymbol{\beta}}^{(-i)}$ is the local bias-corrected profile least-squares estimator defined by (9) and is computed from data with measurements of the ith observation deleted, and $\tilde{\boldsymbol{\alpha}}_{h,(-i)}(T_i)$ is the estimator defined in (11) with $\hat{\boldsymbol{\beta}}$ replaced by $\hat{\boldsymbol{\beta}}^{(-i)}$. The CV bandwidth h_{cv} is selected to minimize (18), that is $h_{cv} = \min_{h>0} \mathrm{CV}(h)$.

In the first simulation example, we compare the performance of the unrestricted estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)^\tau$ with that of the restricted estimators $\hat{\boldsymbol{\beta}}^{R_1}$, $\hat{\boldsymbol{\beta}}^{R_2}$ in terms of sample mean (Mean), sample standard deviation (SD) and sample mean squared error (MSE). Assume that the restricted condition is $A\boldsymbol{\beta} = d$ where $A = (1, 0.5, 1)$. In our simulation, the sample size is set to n =100, 200 and 400, respectively. For every case, we replicate the simulation 1000 times. The simulation results are presented in Table 1. In addition, when the sample size is 200 we compare three types of estimated curves of the nonparametric component based on different measurement errors in Fig. 1: $\hat{\alpha}^{R_1}(t)$, $\hat{\alpha}^{R_2}(t)$ and $\hat{\alpha}^{NE}(t)$, where the naive estimator $\hat{\alpha}^{NE}(t)$ is obtained by ignoring the measurement error and applying the standard profile least-squares approach.

From Table 1 we can see that all the estimators of parameters are close to the true value. As the sample size increases or the measurement error decreases, their means are generally closer to the true values, the SD and MSE of all the estimators decrease. It is noted that in all the scenarios we studied, the restricted local bias-corrected profile least-square estimators of the parametric component outperform the corresponding unrestricted estimator. In addition, by computing the average estimation errors $\|\hat{\boldsymbol{\beta}}^{R_1} - \boldsymbol{\beta}\|$ and $\|\hat{\boldsymbol{\beta}}^{R_2} - \boldsymbol{\beta}\|$ in L_2-norm we find that $\hat{\boldsymbol{\beta}}^{R_1}$ is slightly more efficient than $\hat{\boldsymbol{\beta}}^{R_2}$, the results are reported in Table 2.

Figure 1 shows that the proposed estimators of the nonparametric component $\hat{\alpha}^{R_1}(t)$ and $\hat{\alpha}^{R_2}(t)$ almost overlap, both of them are close to the true coefficient function and improves when the measurement error decreases. The naive estimator $\hat{\alpha}^{NE}(t)$ of the nonparametric component is biased, and the bias increases as the measurement variability increases.

In the second simulation, we study the performance of the proposed testing procedure. For model (17), we consider the null hypothesis $A\boldsymbol{\beta} = d$ where $A = (1, 0.5, 1)$, the corresponding alternative hypothesis is $A\boldsymbol{\beta} = d + \delta$, where $\delta = 0, 0.12, 0.24, \ldots, 1.08$. If $\delta = 0$, the alternative hypothesis becomes the null hypothesis. For sample size $n = 200$, based on 1000 simulations, Fig. 2 depicts the

Table 1 Finite sample performance of the restricted and unrestricted estimators for the parametric components

$\boldsymbol{\beta}$	σ_u^2	n	$\hat{\boldsymbol{\beta}}$			$\hat{\boldsymbol{\beta}}^{R_1}$			$\hat{\boldsymbol{\beta}}^{R_2}$		
			Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
β_1	0.25	100	1.0016	0.1264	0.0160	1.0095	0.1023	0.0105	1.0073	0.0981	0.0097
		200	0.9960	0.0739	0.0055	0.9956	0.0600	0.0036	0.9955	0.0606	0.0037
		400	0.9928	0.0511	0.0027	0.9961	0.0391	0.0015	0.9957	0.0390	0.0015
	0.5	100	1.0386	0.1180	0.0154	1.0154	0.1299	0.0171	1.0181	0.1273	0.0165
		200	0.9899	0.1138	0.0131	0.9806	0.0953	0.0095	0.9822	0.0981	0.0099
		400	1.0079	0.0508	0.0026	1.0024	0.422	0.0018	1.0031	0.0421	0.0018
β_2	0.25	100	2.0024	0.1338	0.0179	1.9944	0.1482	0.0220	2.0053	0.1475	0.0218
		200	2.0014	0.0820	0.0064	2.0020	0.0755	0.0057	2.0011	0.0894	0.0080
		400	2.0128	0.0550	0.0032	2.0096	0.0509	0.0027	2.0143	0.0626	0.0041
	0.5	100	2.0072	0.1673	0.0280	2.0242	0.1623	0.0269	1.9969	0.1798	0.0324
		200	2.0106	0.1294	0.0169	2.0177	0.1195	0.0146	2.0068	0.1445	0.0209
		400	1.9805	0.0536	0.0033	1.9828	0.0655	0.0046	1.9781	0.0539	0.0034
β_3	0.25	100	0.9844	0.1824	0.0335	0.9933	0.0872	0.0077	0.9901	0.1013	0.0104
		200	1.0045	0.0847	0.0072	1.0034	0.0541	0.0029	1.0040	0.0591	0.0035
		400	0.9943	0.0585	0.0035	0.9991	0.0359	0.0013	0.9972	0.0389	0.0015
	0.5	100	1.0039	0.2118	0.0449	0.9725	0.1258	0.0166	0.9834	0.1513	0.0232
		200	1.0220	0.1419	0.0206	1.0105	0.0807	0.0066	1.0144	0.0961	0.0095
		400	1.0126	0.0613	0.0039	1.0061	0.0307	0.0010	1.0079	0.0321	0.0011

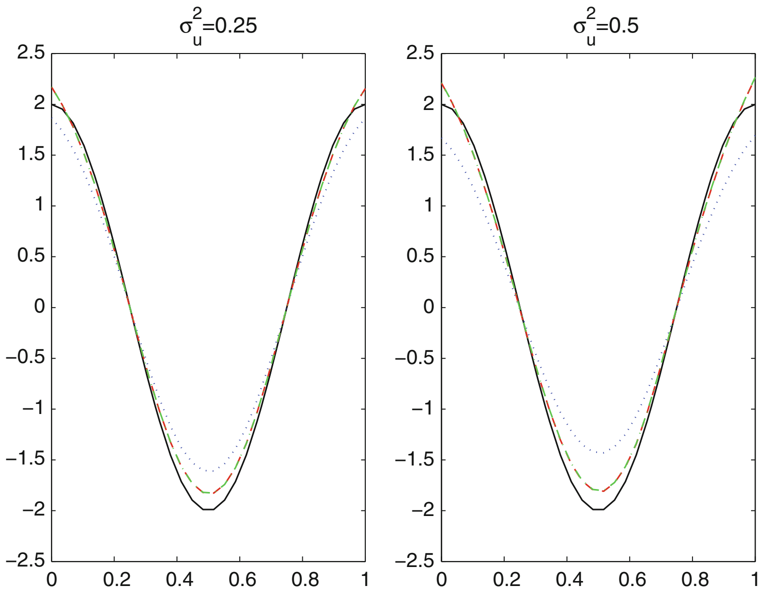

Fig. 1 Simulation results when $n = 200$. In each plot, the *solid curve* is for the true coefficient function, the *dotted curve* for $\hat{\alpha}^{NE}(t)$, the *dashed curve* for $\hat{\alpha}^{R_1}(t)$, the *dash-dotted curve* for $\hat{\alpha}^{R_2}(t)$

Table 2 The average estimation errors of the restricted and unrestricted estimators for the parametric components

Method	$\sigma_u^2 = 0.25$			$\sigma_u^2 = 0.5$		
	$n = 100$	$n = 200$	$n = 400$	$n = 100$	$n = 200$	$n = 400$
$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|$	0.0673	0.0191	0.0092	0.0840	0.0500	0.0098
$\|\hat{\boldsymbol{\beta}}^{R_1} - \boldsymbol{\beta}\|$	0.0401	0.0122	0.0055	0.0577	0.0304	0.0066
$\|\hat{\boldsymbol{\beta}}^{R_2} - \boldsymbol{\beta}\|$	0.0418	0.0152	0.0071	0.0685	0.0399	0.0075

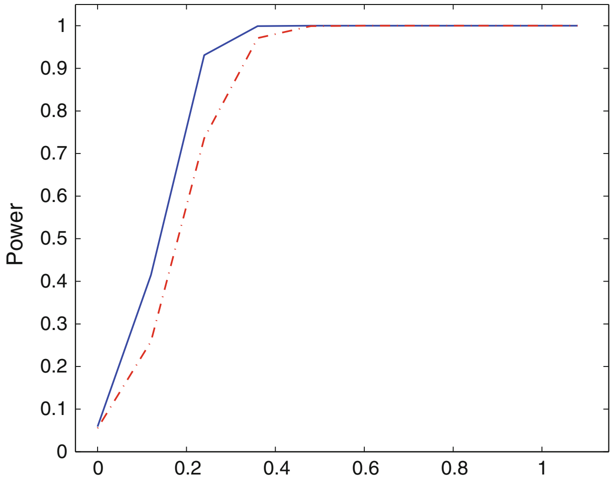

Fig. 2 The simulated power functions for sample size $n = 200$, where the *solid line* computed with $\sigma_u^2 = 0.25$, the *dash-dotted line* computed with $\sigma_u^2 = 0.5$

estimated power function curves with the significance level $\alpha = 0.05$ for $\sigma_u^2 = 0.25$ and $\sigma_u^2 = 0.5$.

We see from Fig. 2 that when the null hypothesis holds ($\delta = 0$), the size of our test is close to the nominal 5%. This demonstrates that the proposed procedure gives the right level of testing. When the alternative hypothesis is true ($\delta > 0$), the power functions increase rapidly as δ increases. These results show that the proposed test statistic performs satisfactorily. In addition, the measurement errors affect the power function, when the variance of the measurement error increases, the estimated power function decreases.

5 Application to Boston housing data

We demonstrate the effectiveness of the proposed method by an application to the Boston housing dataset, which originates from the work of Harrison and Rubinfeld (1978). The dataset consists of the median value of owner-occupied houses in 506 census tracts within the Boston metropolitan area in 1970, together with several variables which might explain the variation of housing values. Following Fan and Huang

(2005), we take per capita crime rate by town (CRIM), nitric oxide concentration parts per 10 million (NOX), average number of rooms per dwelling (RM), proportion of owner-occupied units built prior to 1940 (AGE), full value property tax per $ 10000 (TAX), pupil–teacher ratio by town school district (PTRATIO) and lower status of the population (LSTAT) as the covariates, and simply denoted as $Z_2, \ldots, Z_7$, respectively. Take $Z_1 = 1$ as the intercept term and $T = \sqrt{\text{LSTAT}}$, Fan and Huang (2005) discussed the effects of $Z_2, \ldots, Z_7$, and LSTAT on housing prices, and used the partially linear varying coefficient model

$$Y = \sum_{i=1}^{5} \alpha_i(T) Z_i + \beta_1 Z_6 + \beta_2 Z_7 + \varepsilon \tag{19}$$

to fit the given data. They employed the proposed GLR and Wald tests, and concluded that the coefficient of Z_7 is not significant at 0.01 significance level.

Now we impose a constraint on model (19) that $\beta_2 = 0$, then $A = [0, 1]$ and $d = 0$. Before applying our method, both the response and the covariates (except for Z_1) are transformed to have zero mean and unit variance. The index variable LSTAT is transformed so that its marginal distribution is $U[0, 1]$. To illustrate our method, we conducted a sensitivity analysis, as mentioned in Lin and Carroll (2000). We assume the covariate Z_5 has measurement error, that is

$$W_5 = Z_5 + U_5, \tag{20}$$

where $U_5 \sim N(0, 0.3^2)$. Thus, we apply the proposed procedure in Section 2 to models (19) and (20), and obtain the BCRPLSE of parameter $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}^{R_1} = (-0.1479, 0)^\tau$ under

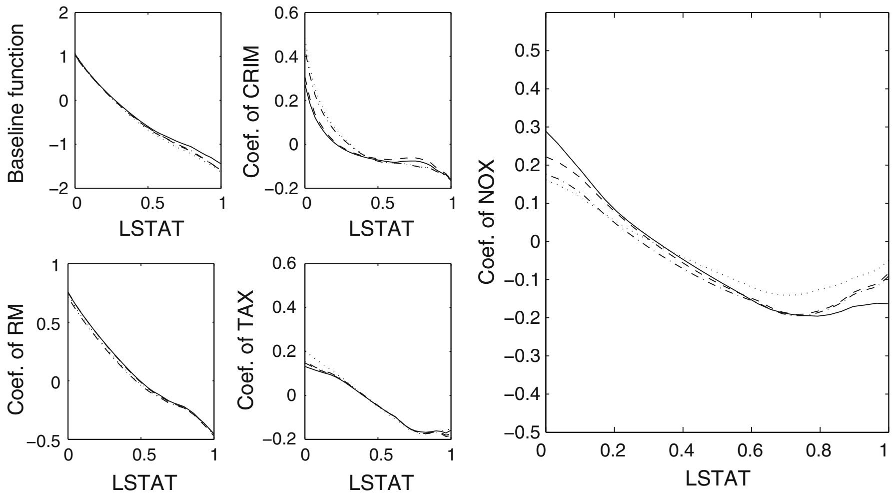

Fig. 3 Application to the Boston housing data. In each plot, the *solid curve* is for the benchmark estimated curve, the *dotted curve* for the naive estimator $\hat{\boldsymbol{\alpha}}^{NE}(t)$, the *dashed curve* for the BCRPLSE $\hat{\boldsymbol{\alpha}}^{R_1}(t)$, the *dash-dotted curve* for the profile local bias-corrected estimator $\tilde{\boldsymbol{\alpha}}(\cdot)$

the restricted condition $A\boldsymbol{\beta} = \boldsymbol{d}$, and the local bias-corrected profile least-square estimator $\hat{\boldsymbol{\beta}} = (-0.1494, 0.0543)^{\tau}$ without restriction. The estimated curves of the coefficient functions are displayed in Fig. 3. The results based on $\hat{\boldsymbol{\beta}}^{R_2}$ are similar, we omit them here.

From Fig. 3 we see that: (1) the performance of BCRPLSE $\hat{\boldsymbol{\alpha}}^{R_1}(\cdot)$ is very close to that of benchmark estimator which is estimated based on the true data. (2) The performance of BCRPLSE $\hat{\boldsymbol{\alpha}}^{R_1}(\cdot)$ is better than that of $\tilde{\boldsymbol{\alpha}}(\cdot)$, this suggests that the prior information can improve the effectiveness of the proposed local bias-correction method. (3) The local bias-corrected procedure is effective, both the BCRPLSE $\hat{\boldsymbol{\alpha}}^{R_1}(\cdot)$ and the profile local bias-corrected estimator $\tilde{\boldsymbol{\alpha}}(\cdot)$ outperform the naive estimator $\hat{\boldsymbol{\alpha}}^{NE}(\cdot)$ which ignores the measurement error.

6 Conclusion and discussion

We have studied the restricted estimation and hypothesis test of the partially liner varying coefficient model with measurement error in the nonparametric part. Based on the bias-corrected and local linear smoothing techniques, we obtained two types of restricted estimators of the parametric component and nonparametric component with the Lagrange multipliers method. Then the asymptotic normality and strong uniform convergence rates of the proposed restricted estimators were established, and the efficiency of the two kinds of parameter estimators were compared. Simulation results indicated that our proposed restricted estimators are more efficient than the unrestricted estimator. Moreover, in order to test the validity of the constraints on the parametric component, we constructed a BCPLM test statistic. By the simulation, we can find that the proposed statistic is powerful.

Statistical inferences for model (1) with measurement error in the parametric component have been studied in literature, such as You and Chen (2006); Wang et al. (2011) and Zhang et al. (2011). In this paper, our focus is on the case where the covariates in the nonparametric component are measured with errors. To our knowledge, it seems that there is no report on this issue. Furthermore, with appropriate modification, the proposed method can be easily extended to the more general case that the covariates in the parametric and nonparametric components of model (1) are both measured with additive errors. Similar results with this paper can be derived while the proofs are different, the exhaustive procedure will be presented in our future work.

Appendix: proofs of the main results

In order to prove the main results, we first introduce several lemmas. Let $\tilde{\boldsymbol{\varepsilon}} = (\boldsymbol{I} - \boldsymbol{S})\boldsymbol{\varepsilon}$, $c_n = h^2 + \left\{\frac{\log(1/h)}{nh}\right\}^{1/2}$, $\mu_j = \int t^j K(t)dt$, $\nu_j = \int t^j K^2(t)dt$, $j = 0, 1, 2, 4$.

Lemma 1 *Let $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ be i.i.d. random vectors, where the Y_i's are scale random variables. Further assume that $E|y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)dy < \infty$, where f denotes the joint density of $(\boldsymbol{X}, Y)$. Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Given that $n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - s^{-1}$, then*

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^{n} \left[K_h(\boldsymbol{X}_i - x) Y_i - E\big(K_h(\boldsymbol{X}_i - x) Y_i\big) \right] \right| = O_p\Big(\Big\{ \frac{\log(1/h)}{nh} \Big\}^{1/2} \Big).$$

This Lemma can be found in Fan and Huang (2005).

Lemma 2 *Suppose that the conditions* (A1)–(A5) *hold, then, as $n \to \infty$,*

$$(\mathbf{D}_t^W)^\tau \omega_t \mathbf{D}_t^W - \Omega = nf(t)\Gamma(t) \otimes \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \{1 + O_p(c_n)\}, \tag{21}$$

$$(\mathbf{D}_t^W)^\tau \omega_t \mathbf{X} = nf(t)\Phi(t) \otimes (1, \mu_1)\{1 + O_p(c_n)\}.$$

Proof Since the proofs of Lemma 2 are similar, we only provide the proof of (21) here. By Lemma 1, we have

$$\begin{aligned}(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W &= \sum_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^\tau \otimes \begin{pmatrix} 1 & \frac{T_i - t}{h} \\ \frac{T_i - t}{h} & (\frac{T_i - t}{h})^2 \end{pmatrix} K_h(T_i - t) \\ &\quad + \sum_{i=1}^{n} \Sigma_u \otimes \begin{pmatrix} 1 & \frac{T_i - t}{h} \\ \frac{T_i - t}{h} & (\frac{T_i - t}{h})^2 \end{pmatrix} K_h(T_i - t) + O\left\{ \left(\frac{\log n}{nh} \right)^{\frac{1}{2}} \right\}.\end{aligned}$$

Similar to the proof of (7.1) in Fan and Huang (2005), we can derive the desired result.

Lemma 3 *Suppose that the conditions* (A1)–(A5) *hold, as $n \to \infty$, then we have*

$$\frac{1}{n}[\tilde{\boldsymbol{X}}^\tau \tilde{\boldsymbol{X}} - \mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q} \mathbf{X}] \to \Sigma_1. \quad a.s.$$

The proof of Lemma 3 is similar to that of Lemma 7.2 in Fan and Huang (2005). We here omit the details.

Lemma 4 *Suppose that the conditions* (A1)–(A5) *hold, then the local bias-corrected profile least-squares estimator $\hat{\boldsymbol{\beta}}$ is asymptotically normal, namely,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(0, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}), \quad n \to \infty,$$

where Σ_1 and Σ_2 are defined in Theorem 1.

Proof By (10), we have

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{n} \left\{ \sum_{i=1}^{n} (\tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^\tau - \mathbf{X}^\tau Q_i^\tau \Sigma_u Q_i \mathbf{X}) \right\}^{-1} \\ &\quad \times \left\{ \sum_{i=1}^{n} [\tilde{\boldsymbol{X}}_i (\tilde{\boldsymbol{Z}}_i^\tau \boldsymbol{\alpha}(T_i) + \tilde{\varepsilon}_i) - \mathbf{X}^\tau Q_i^\tau \Sigma_u Q_i (\mathbf{M} + \varepsilon)] \right\}.\end{aligned}$$

By Lemmas 1 and 2, it is easy to check that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\tilde{X}_i(\tilde{\boldsymbol{Z}}_i^{\tau}\boldsymbol{\alpha}(T_i)+\tilde{\varepsilon}_i)-\mathbf{X}^{\tau}Q_i^{\tau}\Sigma_u Q_i(\mathbf{M}+\varepsilon)\right\}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i-\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\boldsymbol{W}_i(1+O_p(c_n))\right\}\\
&\quad\times\left\{\varepsilon_i+O_p(c_n)\|\boldsymbol{W}_i\|+\boldsymbol{Z}_i^{\tau}\boldsymbol{\alpha}(T_i)O_p(c_n)-\boldsymbol{U}_i^{\tau}\boldsymbol{\alpha}(T_i)(1+O_p(c_n))\right\}\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\Sigma_u\boldsymbol{\alpha}(T_i)(1+O_p(c_n))\right.\\
&\qquad\qquad\left.+\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\Sigma_u 1_q O_p(c_n)\right\}\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{[\boldsymbol{X}_i-\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\boldsymbol{Z}_i][\varepsilon_i-\boldsymbol{U}_i^{\tau}\boldsymbol{\alpha}(T_i)]-\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\boldsymbol{U}_i\varepsilon_i\right.\\
&\quad\left.+\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)(\boldsymbol{U}_i\boldsymbol{U}_i^{\tau}-\Sigma_u)\alpha(T_i)\right\}+o_p(1)\\
&\triangleq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}J_{in}+o_p(1).
\end{aligned}$$

Note that

$$\begin{aligned}
\mathrm{Cov}(J_{in})&=E\left\{(\varepsilon_i-\boldsymbol{U}_i^{\tau}\boldsymbol{\alpha}(T_i))(\boldsymbol{X}_i-\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\boldsymbol{Z}_i)\right\}^{\otimes 2}\\
&\quad+E\left\{\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)(\boldsymbol{U}_i\boldsymbol{U}_i^{\tau}-\Sigma_u)\boldsymbol{\alpha}(T_i)\right\}^{\otimes 2}\\
&\quad+E\left\{\Phi^{\tau}(T_i)\Gamma^{-1}(T_i)\boldsymbol{U}_i\varepsilon_i\right\}^{\otimes 2},\\
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(J_{in})&=E(\varepsilon_1-\boldsymbol{U}_1^{\tau}\boldsymbol{\alpha}(T_1))^2\Sigma_1\\
&\quad+\sigma^2E\left\{\Phi^{\tau}(T_1)\Gamma^{-1}(T_1)\Sigma_u\Gamma^{-1}(T_1)\Phi(T_1)\right\}^{2}\\
&\quad+E\left\{\Phi^{\tau}(T_1)\Gamma^{-1}(T_1)(\boldsymbol{U}_1\boldsymbol{U}_1^{\tau}-\Sigma_u)\boldsymbol{\alpha}(T_1)\right\}^{\otimes 2}.
\end{aligned}$$

Invoking the Slutsky theorem, Lemma 3 and the central limit theorem, we obtain the desired result.

Proof of Theorem 1 We first denote that

$$J_0 =: \mathbf{I} - P^{-1}A^\tau[AP^{-1}A^\tau]^{-1}A$$

where $P = (\tilde{\mathbf{X}}^\tau\tilde{\mathbf{X}} - \mathbf{X}^\tau \boldsymbol{Q}^\tau \boldsymbol{I} \otimes \Sigma_u \boldsymbol{Q}\mathbf{X})$. By Lemma 3, we obtain

$$J_0 \xrightarrow{P} \mathbf{I} - \Sigma_1^{-1}A^\tau[H\Sigma_1^{-1}A^\tau]^{-1}A =: J.$$

By (13), we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{R_1} - \boldsymbol{\beta} &= \left\{\mathbf{I} - P^{-1}A^\tau[AP^{-1}A^\tau]^{-1}A\right\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= J(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (J_0 - J)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).\end{aligned}$$

Note that $J_0 - J = o_p(1)$ and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O(n^{-1/2})$. It is easy to check that

$$(J_0 - J)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(n^{-1/2}).$$

Invoking the Slutsky theorem and Lemma 4, we obtain the desired result.

Proof of Theorem 2 By the same arguments as used in the proof of Theorem 1, we can prove Theorem 2, we omit the details.

Proof of Corollary 1 Since Σ and Π are positive definite matrices, we have

$$\begin{aligned}\Sigma > \Pi &\Leftrightarrow \Sigma^{-1/2}(\Sigma - \Pi)\Sigma^{-1/2} > 0 \\ &\Leftrightarrow \mathbf{I} - \Sigma^{-1/2}\Pi\Sigma^{-1/2} > 0 \\ &\Leftrightarrow \lambda_{\max}(\Sigma^{-1/2}\Pi\Sigma^{-1/2}) < 1 \\ &\Leftrightarrow \lambda_{\max}(\Pi\Sigma^{-1}) < 1.\end{aligned}$$

Hence Corollary 1 holds.

Proof of Theorem 3 By (15), we have

$$\begin{aligned}\hat{\boldsymbol{\alpha}}^{R_k}(t) &= (\mathbf{I}_q \quad \mathbf{0}_q)((\mathbf{D}_t^W)^\tau\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega)^{-1}(\mathbf{D}_t^W)^\tau\boldsymbol{\omega}_t(\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{R_k}) \\ &= (\mathbf{I}_q \quad \mathbf{0}_q)\Big\{((\mathbf{D}_t^W)^\tau\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega)^{-1}(\mathbf{D}_t^W)^\tau\omega_t\mathbf{M} \\ &\quad +((\mathbf{D}_t^W)^\tau\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega)^{-1}(\mathbf{D}_t^W)^\tau\omega_t\boldsymbol{\varepsilon} \\ &\quad +((\mathbf{D}_t^W)^\tau\boldsymbol{\omega}_t\mathbf{D}_t^W - \Omega)^{-1}(\mathbf{D}_t^W)^\tau\omega_t\mathbf{X}(\hat{\boldsymbol{\beta}}^{R_k} - \boldsymbol{\beta})\Big\}.\end{aligned}$$

By the Taylor expansion, we obtain

$$(\boldsymbol{D}_t^W)^\tau\omega_t\mathbf{M} = (\boldsymbol{D}_t^W)^\tau\omega_t\boldsymbol{D}_t^Z\begin{pmatrix}\boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t)\end{pmatrix} + \frac{h^2}{2}(\boldsymbol{D}_t^W)^\tau\omega_t\Psi_t\mathbf{Z}\boldsymbol{\alpha}''(t) + o_p(h^2),$$

where $\Psi_t = \operatorname{diag}\left\{\left(\frac{T_1-t}{h}\right)^2, \ldots, \left(\frac{T_n-t}{h}\right)^2\right\}$.

By Lemmas 1 and 2, it is easy to check that

$$\begin{aligned}
&\left\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\right\}^{-1} (\boldsymbol{D}_t^W)^\tau \omega_t \Psi_t \mathbf{Z}\boldsymbol{\alpha}''(t) \\
&= \left\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\right\}^{-1} \left\{(\boldsymbol{D}_t^Z)^\tau \omega_t \Psi_t \mathbf{Z}\boldsymbol{\alpha}''(t) + O\left(\left(\frac{\log n}{nh}\right)^{\frac{1}{2}}\right)\right\} \\
&= \frac{1}{\mu_2 - \mu_1^2} \begin{pmatrix} (\mu_2^2 - \mu_1\mu_3)\boldsymbol{\alpha}''(t) \\ (\mu_3 - \mu_1\mu_2)\boldsymbol{\alpha}''(t) \end{pmatrix} \{1 + o(1)\} \quad a.s. \\
&(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^Z \begin{pmatrix} \boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t) \end{pmatrix} = \left\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\right\} \begin{pmatrix} \boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t) \end{pmatrix} \\
&+ \left\{ - (\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^U + \Omega\right\} \begin{pmatrix} \boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t) \end{pmatrix},
\end{aligned}$$

Invoking Lemmas 2 and 4, we can obtain

$$(\mathbf{I}_q \quad \mathbf{0}_q)\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\}^{-1} (\boldsymbol{D}_t^W)^\tau \omega_t \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(n^{-1/2}),$$

Therefore,

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}^{R_k}(t) &= \boldsymbol{\alpha}(t) + \frac{1}{2}h^2 \frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\boldsymbol{\alpha}''(t) + (\mathbf{I}_q \quad \mathbf{0}_q)\left\{\left[\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\}^{-1}\right.\right. \\
&\quad \left.\left. \times\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{\varepsilon} - (\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^U + \Omega\} \begin{pmatrix} \boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t) \end{pmatrix}\right]\right\} + O_p(n^{-1/2}),
\end{aligned}$$

Using the same argument of (A4)–(A6) in You et al. (2006), we can obtain that

$$\begin{aligned}
&\sqrt{nh}\left[\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^W - \Omega\}^{-1}\{(\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{\varepsilon} - (\boldsymbol{D}_t^W)^\tau \omega_t \boldsymbol{D}_t^U + \Omega\}\begin{pmatrix} \boldsymbol{\alpha}(t) \\ h\boldsymbol{\alpha}'(t) \end{pmatrix}\right] \\
&\xrightarrow{\mathcal{L}} N(0, \Lambda),
\end{aligned}$$

where

$$\begin{aligned}
\Lambda &= f(t)^{-1}\Sigma^* \otimes \frac{1}{\mu_2 - \mu_1^2} \\
&\quad \times \begin{pmatrix} \mu_2^2\nu_0 - 2\mu_1\mu_2\nu_1 + \mu_1^2\nu_2 & (\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\mu_2 \\ (\mu_1^2 + \mu_2)\nu_1 - \mu_1\mu_2\nu_0 - \mu_1\mu_2 & \nu_2 - \mu_1(2\nu_1 + \mu_1\nu_0) \end{pmatrix}, \\
\Sigma^* &= \Gamma^{-1}(t)[\sigma^2\Gamma(t) + \sigma^2\Sigma_u + E\{\xi_1\boldsymbol{\alpha}(t)\boldsymbol{\alpha}^\tau(t)\xi_1^\tau | T = t\}]\Gamma^{-1}(t), \\
\xi_1 &= \Sigma_u - \boldsymbol{U}_1\boldsymbol{U}_1^\tau - \boldsymbol{X}_1\boldsymbol{U}_1^\tau.
\end{aligned}$$

From the above discussion, it implies that Theorem 3 holds.

Proof of Theorem 4 By the same argument used in the proof of Theorem 3.1 in Xia and Li (1999), we can complete the proof. Hence, we here omit the details.

Proof of Theorem 5 Under the null hypothesis of testing problem (16), and applying Lemma 4, we can prove that

$$\sqrt{n}A(\hat{\boldsymbol{\beta}} - \boldsymbol{d}) \xrightarrow{\mathscr{L}} N(0, A\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}A^\tau), \quad n \to \infty. \tag{22}$$

By Lemmas 1–3, we have

$$L_0 = A\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}A^\tau \xrightarrow{P} A\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}A^\tau =: L, \tag{23}$$

and

$$C_0 = \left[A\left(\frac{1}{n}\tilde{\boldsymbol{X}}^\tau\tilde{\boldsymbol{X}} - \frac{1}{n}\mathbf{X}^\tau\boldsymbol{Q}^\tau\boldsymbol{I}\otimes\Sigma_u\boldsymbol{Q}\mathbf{X}\right)^{-1}A^\tau\right]^{-1} \xrightarrow{P} (A\Sigma_1^{-1}A^\tau)^{-1} =: C. \tag{24}$$

By (13) and (22), and again invoking the Slutsky theorem and Lemma 3, we can derive that

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{\boldsymbol{\lambda}} &= \left[A\left(\frac{1}{n}\tilde{\boldsymbol{X}}^\tau\tilde{\boldsymbol{X}} - \frac{1}{n}\mathbf{X}^\tau\boldsymbol{Q}^\tau\boldsymbol{I}\otimes\Sigma_u\boldsymbol{Q}\mathbf{X}\right)^{-1}A^\tau\right]^{-1}\sqrt{n}(A\hat{\beta} - \boldsymbol{d})\\ &\xrightarrow{\mathcal{L}} N(0, CLC^\tau).\end{aligned} \tag{25}$$

By (23)–(25), we have that

$$\frac{1}{n}\hat{\boldsymbol{\lambda}}^\tau(C_0L_0C_0^\tau)^{-1}\hat{\boldsymbol{\lambda}} \xrightarrow{\mathcal{L}} \chi_k^2.$$

On the other hand, under the alternative hypothesis, and again applying Lemma 4, we have

$$\sqrt{n}A(\hat{\boldsymbol{\beta}} - \boldsymbol{d}) \xrightarrow{\mathcal{L}} N(A\boldsymbol{\beta} - \boldsymbol{d}, A\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}A^\tau), \quad n \to \infty.$$

By using the same argument as (23) and (24), it can be shown that

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{\boldsymbol{\lambda}} &= \left[A\left(\frac{1}{n}\tilde{\boldsymbol{X}}^\tau\tilde{\boldsymbol{X}} - \frac{1}{n}\mathbf{X}^\tau\boldsymbol{Q}^\tau\boldsymbol{I}\otimes\Sigma_u\boldsymbol{Q}\mathbf{X}\right)^{-1}A^\tau\right]^{-1}\sqrt{n}(A\hat{\beta} - \boldsymbol{d})\\ &\xrightarrow{\mathcal{L}} N(A\boldsymbol{\beta} - \boldsymbol{d}, CLC^\tau).\end{aligned}$$

Then we have

$$\frac{1}{n}\hat{\boldsymbol{\lambda}}^\tau(C_0L_0C_0^\tau)^{-1}\hat{\boldsymbol{\lambda}} \xrightarrow{\mathcal{L}} \chi^2(k, \varsigma),$$

where $\chi^2(k, \varsigma)$ denotes the asymptotic noncentral Chi-squared distribution with k degrees of freedom, and the noncentral parameter

$$\varsigma = \lim_{n\to\infty} (A\boldsymbol{\beta} - \boldsymbol{d})^{\tau}(CLC)^{-1}(A\boldsymbol{\beta} - \boldsymbol{d}).$$

□

Acknowledgments Sanying Feng's research was supported by the National Nature Science Foundation of China (11101014, 11001118), the National Social Science Foundation of China (11CTJ004) and the Fundamental Research Foundation of Beijing University of Technology (X4006013201101). Liugen Xue's research was supported by the National Nature Science Foundation of China (11171012), the Science and Technology Project of the Faculty Adviser of Excellent PhD Degree Thesis of Beijing (20111000503) and the Beijing Municipal Education Commission Foundation (KM201110005029).

References

Ahmad, I., Leelahanon, S., Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *The Annals of Statistics*, *33*, 258–283.

Carroll, R. J., Ruppert, D., Stefanski, L. A. (1995). *Measurement Error in Non- linear Models*. London: Chapman & Hall/CRC.

Fan, J. Q., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, *11*, 1031–1057.

Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.

Harrison, D., Rubinfeld, D. L. (1978). Hedonic housing pries and the demand for clean air. *Journal of Environmental Economics and Management*, *5*, 81–102.

Hu, X. M., Wang, Z. Z., Zhao, Z. Z. (2009). Empirical likelihood for semiparametric varying-coefficient partially linear errors-in-variables models. *Statistics and Probability Letters*, *79*, 1044–1052.

Kai, B., Li, R. Z., Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics*, *39*, 305–332.

Li, G. R., Feng, S. Y., Peng, H. (2011). Profile-type smoothed score function for a varying coefficient partially linear model. *Journal of Multivariate Analysis*, *102*, 372–385.

Li, G. R., Lin, L., Zhu, L. X. (2012). Empirical likelihood for varying coefficient partially linear model with diverging number of parameters. *Journal of Multivariate Analysis*, *105*, 85–111.

Liang, H., Härdle, W., Carroll, R. J. (1999). Estimation in a semi-parametric partially linear errors-in-variables models. *The Annals of Statistics*, *27*, 1519–1533.

Lin, X. H., Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, *95*, 520–534.

Wang, X. L., Li, G. R., Lin, L. (2011). Empirical likelihood inference for semiparametric varying-coefficient partially linear EV models. *Metrika*, *73*, 171–185.

Wei, C. H. (2012). Statistical inference for restricted partially linear varying coefficient errors-in-variables models. *Journal of Statistical Planning and Inference*, *142*, 2464–2472.

Wei, C. H., Wu, X. Z. (2008). Profile Lagrange multiplier test for partially linear varying-coefficient regression models. *Journal of System Science and Mathematical Science*, *28*, 416–424.

Xia, Y. C., Li, W. K. (1999). On the estimation and testing of functional coefficient linear models. *Statistica Sinica*, *9*, 737–757.

Xia, Y. C., Zhang, W. Y., Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika*, *91*, 661–681.

You, J. H., Chen, G. M. (2006). Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis*, *97*, 324–341.

You, J. H., Zhou, Y. (2006). Empirical likelihood for semi-parametric varying-coefficient partially linear model. *Statistics and Probability Letters*, *76*, 412–422.

You, J. H., Zhou, Y., Chen, G. M. (2006). Corrected local polynomial estimation in varying coefficient models with measurement errors. *The Canadian Journal of Statistics*, *34*, 391–410.

Zhang, W. W., Li, G. R., Xue, L. G. (2011). Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition. *Computational Statistics and Data Analysis*, *55*, 3027–3040.
Zhang, W. Y., Lee, S., Song, X. Y. (2002). Local polynomial fitting in semivarying coefficient models. *Journal of Multivariate Analysis*, *82*, 166–188.
Zhou, Y., Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with generated regressors. *The Annals of Statistics*, *37*, 427–458.