Efficient estimation for partially linear varying-coefficient errors-in-variables models with heteroscedastic errors

Hongxia Xu[1,*], Xiaoming Duan[2]

1. Department of Mathematics, Shanghai Maritime University, Shanghai 201306, China

2. School of Economics and Management, Shanghai Maritime University, Shanghai 201306, China

hongxiaxu2008@163.com

ABSTRACT. This paper studies the varying-coefficient heteroscedastic partially linear models where some covariates are measured with additive errors. To eliminate the bias of the usual profile least squares estimation when measurement errors are ignored, a modified profile least squares estimator of the regression parameter is suggested and the local polynomial smoother is applied to constructing estimators of the varying coefficient function and error variance function. Further, for the purpose of accounting for heteroscedasticity and the estimation accuracy, re-weighted estimations of the regression parameter and varying coefficient function are proposed. Asymptotic behaviors of the above estimators are established and the re-weighted estimator is shown to be more efficient than the modified profile least-squares estimator. Both simulated and real data examples are conducted to illustrate the applications of the proposed approaches.

RÉSUMÉ. Cet article examine les Modèles partiellement linéaires à coefficient variable avec erreurs hétéroscédastiques, dans lesquels certaines covariables sont mesurées avec des erreurs additives. Pour résoudre le problème d'erreur de négliance de mesure dans l'estimation des moindres carrés du profil classique, un estimateur de moindres carrés du profil modifié du paramètre de régression a été mis au point et le lisseur polynomial local a été appliqué pour construire des estimateurs de la fonction de coefficient variable et de la fonction de variance d'erreur. En outre, des estimations re-pondérées du paramètre de régression et une fonction de coefficient variable ont été proposées, en tenant compte de l'hétéroscédasticité et de la précision de l'estimation. En comparant les exemples de données simulées aux données réelles, les comportements asymptotiques des estimateurs ci-dessus se sont avérés valables et il a été démontré que l'estimateur repondéré était plus efficace que l'estimateur des moindres carrés du profil modifié, indiquant que la stratégie proposée est réalisable et applicable.

KEYWORDS: varying-coefficient partially linear model, profile least squares, errors-in-variables, heteroscedasticity, re-weighted estimation.

MOTS-CLÉS: partiellement lineaire a coefficient variable, profil des moindres carres, erreurs dans les variables, l'heteroscedasticite, estimation reponderee.

DOI:10.3166/I2M.17.295-314

1. Introduction

In the last decades, driven by many practical applications and fueled by modern computing power, lots of useful data-analytic modeling techniques have been proposed to relax traditional parametric models and to exploit possible hidden structure. Recently, partially linear varying-coefficient model (PLCVM) which was introduced by Fan and Huang (2005) has attracted much attention from statisticians and practitioners and has the following form:

$$Y_i = X_i^T \beta + Z_i^T \alpha(U_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where Y_i are the response variables, $\beta = (\beta_1, \cdots, \beta_p)^T$ is a vector of p-dimensional unknown parameters, $\alpha(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_q(\cdot))^T Y_i = X_i^T \beta + Z_i^T \alpha(U_i) + \varepsilon_i, \ i = 1, \cdots, n,$ is a q-dimensional vector of unspecified smooth coefficient functions, $X_i = (X_{i1}, \cdots, X_{ip})^T$, $Z_i = (Z_{i1}, \cdots, Z_{iq})^T$ and U_i are the regressors, and ε_i are the random errors with $E(\varepsilon_i | X_i, Z_i, U_i) = 0$, $\mathrm{Var}(\varepsilon_i | X_i, Z_i, U_i) = \sigma^2(U_i)$.

Model (1.1) can reduce the high risk of misspecification relative to a fully parametric model and avert some critical drawbacks of purely nonparametric approaches such as the curse of dimensionality, poverty of interpretation, and lack of extrapolation ability. This model has been widely studied in the literature and the majority of the work done so far assumed that the errors ε_i are homoscedastic. For example, Fan and Huang (2005) employed a profile least squares (PLS) technique to estimate the parameter in the semi-varying coefficient model and obtained the asymptotic normality for the PLS estimator. In addition, they proposed the profile likelihood ratio test for the semi-varying coefficient model and demonstrated that it follows an asymptotically chi-squared distribution under the null hypothesis. You and Zhou (2006) applied empirical likelihood method to semi-varying coefficient model and derived a nonparametric version of the Wilk's theorem. Based on this, the confidence regions for parametric components with asymptotically correct coverage probabilities can be constructed. More references and techniques of the semi-varying coefficient models can be found in Zhang *et al*. (2002), Kai *et al*. (2011), Fan *et al*. (2013) among others.

The above referred articles assumed that the random errors are homoscedastic, that is, ε_i are independent of (X_i, Z_i, U_i). However, in applications, heteroscedasticity is often found in the model error terms and has been extensively studied by many authors. For example, Ma (2006) discussed efficient semiparametric estimator in heteroscedastic partially linear models; Zhou *et al*. (2009) proposed a new method for estimating the unknown transformation and regression parameters in non-parametric heteroscedastic transformation regression models for skewed data; Lu (2009) employed empirical likelihood method to discuss heteroscedastic partially linear models; Shen *et al*. (2014) studies the semi-varying coefficient model with heteroscedastic errors. An estimation procedure for the error variance function was suggested to obtain a consistent estimator. Then a re-weighted estimation of the unknown parameter was proposed and asymptotic normalities of the resulting estimators were established; Menictas and Wand (2015) developed fast mean field variational methodology for Bayesian heteroscedastic semiparametric regression, in which both the mean and variance are smooth, but otherwise arbitrary, functions of the predictors.

In addition, measurement error data are often encountered in many fields, including biomedical sciences, engineering, economics and biology *et al*. For instance, it has been well documented in the literature that covariates such as serum cholesterol level (Carroll *et al*., 1995), urinary sodium chloride level (Wang *et al*., 1996) and exposure to pollutants (Tosteson *et al*., 1989) are often subject to measurement errors (errors-in-variables). Simply ignoring measurement errors, known as the naive method, will result in biased estimators. So the measurement errors models are somewhat more practical than the ordinary regression model. Some work has been done in lots of regression models with measurement errors. For example, Wei (2012) investigated statistical inference for the semiparametric model when the covariates in the linear part are measured with additive error and some additional linear restrictions on the parametric component are available. They proposed a restricted modified profile least-squares estimator for the parametric component, and proved the asymptotic normality of the proposed estimator. More work of measurement error data models can be founded in Fuller (1987), You *et al*. (2006), Zhou and Liang (2009), Ma *et al*. (2013), Fan *et al*. (2013) and Feng and Xue (2014) among others. However, the above articles worked under homoscedastic assumption.

In this paper, we consider the following partially linear varying-coefficient errors-in-variables model (PLVCEVM) model with heteroscedastic errors

$$\begin{cases} Y_i = X_i^T\beta + Z_i^T\alpha(U_i) + \varepsilon_i, \\ \xi_i = X_i + \eta_i, \end{cases} \tag{1.2}$$

where $\mathrm{Var}(\varepsilon_i|X_i, Z_i, U_i) = \sigma^2(U_i)$ is an unknown function of U_i representing possible heteroscedasticity and the covariate variable X_i is measured with additive error. That is, instead of the true X_i, the surrogate variable ξ_i are observed by $\xi_i=X_i+\eta_i$, where η_i are the measurement errors, which are mean zero, independent of (X_i, Z_i, U_i, ε_i) and have the same covariance matrix $\mathrm{Cov}(\eta)=\sum_\eta$. In order to identify the model, we further assume that the covariance matrix $\sum_\eta$ is known as in Hwang (1986), Zhu and Cui (2003), You *et al*. (2006) and Feng & Xue (2014) among others. When $\sum_\eta$ is unknown, we can obtain a $\sqrt{n}$-consistent estimator as long as we have replicates of ξ_i; see Liang *et al*. (1999) for more details. For PLVCEVM with heteroscedastic errors, Fan and Huang (2013) employed empirical likelihood method to study the parameter β. Based on the result of Fan & Huang (2013), one can only obtain the confidence regions of β. However, the point estimator and its asymptotic normality of β and the statistical inference of coefficient function $\alpha(\bullet)$ are still not studied up to now.

The aim of this paper is to extend the results in Shen *et al*. (2014) from the PLVCM with heteroscedastic errors to the PLVCEVM with heteroscedastic errors (1.2) and develop a modified PLS approach and a re-weighted estimation approach to improve the accuracy of traditional estimation methods. Compared with You and Chen (2006) and Ma *et al*. (2013), the error in model (1.2) is assumed to be heteroscedastic. In order to improve the estimation accuracy by taking account for heteroscedasticity, the re-weighted estimations of the regression parameter and varying coefficient function are proposed in this paper.

The rest of this paper is organized as follows. The modified PLS estimation of the unknown parameter and the estimators of the coefficient function and the error variance function are proposed in Section 2. The re-weighted estimations of the parametric vector and coefficient function are also introduced in this section. Section 3 gives the main results of the proposed estimators. In Section 4, simulation studies and a real-data example are conducted to examine the finite-sample behavior of the proposed methods. Some remarks are put in Section 5 and all the technical proofs are relegated to Appendix.

2. Estimation methodology

In this section, we shall construct the estimators for parametric component β and the estimators of the nonparametric component $\alpha(\bullet)$ and the error variance $\sigma(\bullet)$.

Considering the influence of the measurement errors, we cannot use PLS estimation method and should apply the so-called "correction for attenuation" (see Fuller, 1987 and Liang *et al*., 1999, for example) to overcome inconsistency. Motived by You *et al*. (2006) and Fan *et al*. (2013), we will employ the modified PLS estimation technique to estimate β and the local linear estimators of the coefficient function $\alpha(\bullet)$ will also be proposed in Section 2.1. The local linear estimator of the error variance function $\sigma(\bullet)$ will be introduced in Section 2.2. Take into account the heteroscedasticity, in Section 2.3, we further improve the estimations and propose re-weighted estimators for parametric component β and the nonparametric component $\alpha(\bullet)$.

2.1. Modified PLS estimation

In view of the relationship

$$E(Z_i^T|U_i = u_0)\alpha(u_0) = E(Y_i - X_i^T\beta|U_i = u_0) = E(Y_i - \xi_i^T\beta|U_i = u_0)$$

The varying-coefficient function $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ can be estimated by a local polynomial method. Specifically, for u in a neighborhood of u_0, we use a local linear approximation $\alpha_j(U) \approx \alpha_j(u_0) + \alpha_j^{'}(u_0)(U - u_0) \equiv a_j + b_j(U - u_0),\ j = 1, \cdots, q,$ where $\alpha_j^{'}(u_0)$ denote the first order derivative of $\alpha_j(u)$ at u_0. We obtain the estimates of $\alpha_j(\cdot)$ by finding $\{(a_j, b_j), j = 1, \cdots, q\}$ to minimize

$$\sum_{i=1}^{n}\{(Y_i - \sum_{j=1}^{p} \xi_{ij}\,\beta_j) - \sum_{j=1}^{q}[a_j + b_j(U_i - u_0)]\ Z_{ij}\}^2 K_{h_1}(U_i - u_0), \qquad (2.1)$$

Where $K(\bullet)$ is a given kernel function and $K_{h1}(\bullet)=K(\bullet/h_1)/h_1$ with a bandwidth h_1.

For notational simplicity, let $Y = (Y_1, \cdots, Y_n)^T$, $\xi = (\xi_1, \cdots, \xi_n)^T$, $M = (M_1, \cdots, M_n)^T = (Z_1^T\alpha(U_1), \cdots,\ Z_n^T\alpha(U_n))^T$, $\tilde{Y} = (I - S)Y$, $\tilde{\xi} = (I - S)\xi$, $W(u) = \text{diag}(K_{h_1}(U_1 - u), \cdots, K_{h_1}(U_n - u))$, where I is the $n\times n$ identity matrix,

$$S = \begin{pmatrix} (Z_1^T \quad 0^T)[D^T(U_1)W(U_1)D(U_1)]^{-1}D^T(U_1)W(U_1) \\ \vdots \\ (Z_n^T \quad 0^T)[D^T(U_n)W(U_n)D(U_n)]^{-1}D^T(U_n)W(U_n) \end{pmatrix},$$

$$D(u) = \begin{pmatrix} Z_1^T & \frac{U_1-u}{h_1} Z_1^T \\ \vdots & \vdots \\ Z_n^T & \frac{U_n-u}{h_1} Z_n^T \end{pmatrix}$$

and $\mathbf{0}$ is the $q\times 1$ null vector.

The solution of problem (2.1) is given by

$$(\hat{a}_1(u,\beta), \cdots, \hat{a}_q(u,\beta), h_1\hat{b}_1(u,\beta), \cdots, h_1\hat{b}_q(u,\beta))^T$$
$$= [D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)(Y - \xi\beta). \qquad (2.2)$$

Then the local polynomial estimator of M is

$\widehat{M}(\beta) = (Z_1^T\hat{\alpha}(U_1,\beta), \cdots, Z_n^T\hat{\alpha}(U_n,\beta))^T = S(Y - \xi\beta)$.

Motivated by You and Chen (2006) and Fan *et al*. (2013), we introduce the modified PLS approach to estimate β as follows, $\hat{\beta} = [\sum_{i=1}^{n}\tilde{\xi}_i\,\tilde{\xi}_i^T - n\Sigma_\eta]^{-1}\sum_{i=1}^{n}\tilde{\xi}_i\,\tilde{Y}_i$. Which is derived by minimizing $\sum_{i=1}^{n}[\ (Y_i - \xi_i^T\beta - Z_i^T\hat{\alpha}(U_i,\beta)^2 - \beta^T\Sigma_\eta\beta]$.

Based on $\hat{\beta}$, a plug-in estimator of M can be expressed as $\widehat{M} = (Z_1^T\hat{\alpha}(U_1,\hat{\beta}), \cdots, Z_n^T\hat{\alpha}(U_n,\hat{\beta}))^T = S(Y - \xi\hat{\beta})$.

2.2. Estimation of the error variance function

In this subsection, we introduce the estimator of the error variance function $\sigma^2(\bullet)$ which can be employed to improve the estimators of the parametric component β and the coefficient function $\alpha(\bullet)$.

Local linear estimator as a more attractive estimation methodology, has many appealing properties, such as no boundary effect, design adaptation, and mathematical efficiency (cf. Fan and Gibels, 1992; Ruppert and Wand, 1994; Hastie and Loader, 1993). Thus, we employ the local linear method to estimate $\sigma^2(u)$ in this paper. Note that $\sigma^2(u) = E\{(Y_i - \xi_i^T\beta - Z_i^T\alpha(U_i))^2 - \beta^T\Sigma_\eta\beta|U_i = u\}$.

Specifically, the local linear estimator of $\sigma^2(u)$ is defined by $\hat{\sigma}^2(\cdot) = \hat{\mu}$, where

$(\hat{\mu}, \hat{\nu}) = \text{argmin}_{\mu,\nu} \sum_{i=1}^{n} \{ [(Y_i - \xi_i^T \hat{\beta} - Z_i^T \hat{\alpha}(U_i, \hat{\beta}))^2 - \hat{\beta}^T \Sigma_\eta \hat{\beta}] - \mu - \nu(U_i - u)\}^2 K_{h_2}(U_i - u)$, and $K_{h2}(\bullet)$ has the same form as $K_{h1}(\bullet)$ except that h_1 is replaced by the bandwidth h_2=h_{2n}. Simple calculation yields that

$$\hat{\sigma}^2(u) = \sum_{i=1}^{n} W_{h_2 i}(u)[Y_i - \xi_i^T \hat{\beta} - Z_i^T \hat{\alpha}(U_i, \hat{\beta})]^2 - \hat{\beta}^T \Sigma_\eta \hat{\beta}, \qquad (2.3)$$

where the weight functions $W_{h21i}(\bullet)$ have the following explicit form (Fan and Gijbels, 1996):

$$W_{h_2 i}(u) = \frac{(nh_2)^{-1} K(h_2^{-1}(U_i - u))\{A_{n,2}(u) - (U_i - u)A_{n,1}(u)\}}{A_{n,0}(u)A_{n,2}(u) - A_{n,1}^2(u)}, \qquad (2.4)$$

With $A_{n,j}(u) = \frac{1}{nh_2} \sum_{i=1}^{n} K_{h_2}\left(\frac{U_i - u}{h_2}\right)(U_i - u)^j, j = 0,1,2.$

2.3. Re-weighted estimation

In this subsection, we introduce the re-weighted estimations of the parametric vector β and coefficient function $\alpha(\cdot)$based on the variance estimates $\hat{\sigma}^2(u)$ given in (2.3). It is show from Theorem 3.4 that the re-weighted estimate $\hat{\beta}_R$ has no greater asymptotic variance than the modified PLS estimate $\hat{\beta}$ that ignores heteroscedasticity. In order to give the re-weighted estimations of β and $\alpha(\cdot)$, we resort to the idea of the generalized least-squares approach in heteroscedastic linear models. By taking notice of the influence of measurement errors, the re-weighted estimate of the coefficient β in model (1.2) is given by

$$\hat{\beta}_R = (\hat{\beta}_{1R}, \cdots, \hat{\beta}_{pR})^T = [\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1} \tilde{\xi} - \sum_{i=1}^{n} \hat{\sigma}^{-2}(U_i)\Sigma_\eta]^{-1} \tilde{\xi}^T \hat{\Sigma}_\sigma^{-1} \tilde{Y}, \qquad (2.5)$$

Where $\hat{\Sigma}_\sigma = \text{diag}(\hat{\sigma}^2(U_1), \cdots, \hat{\sigma}^2(U_n))$ is an estimator of $\Sigma_\sigma = \text{diag}(\sigma^2(U_1), \cdots, \sigma^2(U_n))$. Furthermore, the re-weighted estimator of the coefficient function $\alpha(u) = (\alpha_1(u), \cdots, \alpha_q(u))^T$ is expressed as

$$\hat{\alpha}_R(u) = (\hat{\alpha}_{1R}(u), \cdots, \hat{\alpha}_{qR}(u))^T =$$

$$(I_q, 0_q)[D(u)W(u)D(u)]^{-1} D^T(u)W(u)(Y - \xi\hat{\beta}_R), \qquad (2.6)$$

Where I_q is the $q \times q$ identity matrix and 0_q is the $q \times q$ null matrix.

3. Main results

In this section, we give the asymptotic properties for variance estimator $\hat{\sigma}^2(u)$, modified PLS estimator $\hat{\beta}$, local polynomial estimator $\hat{\alpha}(\cdot, \hat{\beta})$ and re-weighted estimators $\hat{\beta}_R$ and $\hat{\alpha}_R(\cdot)$. For convenience, let $\Gamma(u) = E[Z_1 Z_1^T | U = u]$, $\Phi(u) = E[Z_1 X_1^T | U = u]$, $\psi_i = X_i - \Phi^T(U_i)\Gamma^{-1}(U_i)Z_i$ and $A^{\otimes 2} = AA^T$.

Theorem 3.1 *Suppose that conditions (C1)-(C8) in the Appendix hold. Then* $\sup_{u\in\mathcal{D}}|\hat{\sigma}^2(u)-\sigma^2(u)| = O_p(b_n)$, *where* $\mathcal{D}$ *is domain of the variable* U *and* $b_n = (\frac{\log n}{nh_2})^{1/2} + h_2^2$.

Theorem 3.2 *Under the conditions (C1)-(C7), the modified PLS estimator* $\hat{\beta}$ *follows the following asymptotic normality,* $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1})$ as $n \to \infty$, *where* $\Sigma_1 = E(\psi_1\psi_1^T)$ *and* $\Sigma_2 = E\{\psi_1\psi_1^T(\varepsilon_1 - \eta_1^T\beta)^2]\} + \Sigma_\eta E(\sigma^2(U_1)) + E\{(\eta_1\eta_1^T - \Sigma_\eta)\beta\beta^T(\eta_1\eta_1^T - \Sigma_\eta)\}$. *Further,* $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$ *is a consistent estimator of* $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$, *where* $\hat{\Sigma}_1 = n^{-1}\tilde{\xi}^T\tilde{\xi} - \Sigma_\eta$ *and* $\hat{\Sigma}_2 = \frac{1}{n}\sum_{i=1}^n\{[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^T\hat{\beta}) + \Sigma_\eta\hat{\beta}]\}^{\otimes 2}$.

Theorem 3.3 *Suppose that conditions (C1)-(C8) in the Appendix hold, we have* $\sqrt{n}(\hat{\beta}_R - \beta) \xrightarrow{d} N(0, \Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1})$, as $n \to \infty$, *where* $\Sigma_{1R} = E(\psi_1\psi_1^T\sigma^{-2}(U_1))$ *and* $\Sigma_{2R} = E\{\psi_1\psi_1^T(\varepsilon_1 - \eta_1^T\beta)^2\sigma^{-4}(U_1)]\} + \Sigma_\eta E(\sigma^{-2}(U_1)) + E\{(\eta_1\eta_1^T - \Sigma_\eta)\beta\beta^T(\eta_1\eta_1^T - \Sigma_\eta)\sigma^{-4}(U_1)\}$. *Further,* $\hat{\Sigma}_{1R}^{-1}\hat{\Sigma}_{2R}\hat{\Sigma}_{1R}^{-1}$ *is a consistent estimator of* $\Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1}$, *where* $\hat{\Sigma}_{1R} = n^{-1}\tilde{\xi}^T\hat{\Sigma}_\sigma^{-1}\tilde{\xi} - \frac{1}{n}\sum_{i=1}^n \hat{\sigma}^{-2}(U_i)\Sigma_\eta$ *and* $\hat{\Sigma}_{2R} = \frac{1}{n}\sum_{i=1}^n \hat{\sigma}^{-4}(U_i)\{[\tilde{\xi}_i(\tilde{Y}_i - \tilde{\xi}_i^T\hat{\beta}) + \Sigma_\eta\hat{\beta}]\}^{\otimes 2}$.

Theorem 3.4 *Suppose that conditions (C1)-(C8) in the Appendix hold, then the leading term of the asymptotic variance of the re-weighted estimate* $\hat{\beta}_R$ *is not greater than that of the modified PLS estimate* $\hat{\beta}$, *saying* $\Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1}\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$. *That is,* $\Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1} - \Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}$ *is a positive semidefinite matrix.*

In the following three theorems, we give the asymptotic behaviors for local linear estimator $\hat{\alpha}(\cdot,\hat{\beta}) := \hat{\alpha}(u)$ *in (2.2) and re-weighted estimator* $\hat{\alpha}_R(u)$ *in (2.6}). Define,* $\alpha''(u) = (\alpha_1'(u),\cdots,\alpha_q'(u))^T$, $\mu_j = \int t^j K(t)dt$, $v_j = \int t^j K^2(t)dt$ *and let* $p(u)$ *be the density function of U.*

Theorem 3.5 *Suppose that conditions (C1)-(C7) hold. Then for any* $u \in \mathcal{D}$, *we have* $\sup_{u\in\mathcal{D}}\|\hat{\alpha}(u) - \alpha(u)\| = O_p(c_n)$ *and,* $\sup_{u\in\mathcal{D}}\|\hat{\alpha}_R(u) - \alpha(u)\| = O_p(c_n)$, *where* $c_n = \{\log n / nh_1\}^{1/2} + h_1^2$.

Theorem 3.6 *Suppose that conditions (C1)-(C7) hold. Then for any* $u \in \mathcal{D}$, *we have the following asymptotic normality* $\sqrt{nh_1}\{\hat{\alpha}(u) - \alpha(u) - \frac{1}{2}h_1^2\mu_2\alpha''(u)\} \xrightarrow{d} N(0, v_0\Gamma^{-1}(u)(\sigma^2(u) + \beta^T\Sigma_\eta\beta)p^{-1}(u))$.

Theorem 3.7 Suppose that conditions (C1)-(C8) hold. Then for any $u \in \mathcal{D}$, we have the following asymptotic normality

$$\sqrt{nh_1}\left\{\alpha_R(u) - \alpha(u) - \frac{1}{2}h_1^2\mu_2\alpha''(u)\right\} \xrightarrow{d} N\left(0, v_0\Gamma^{-1}(u)(\sigma^2(u) + \beta^T\Sigma_\eta\beta)p^{-1}(u)\right).$$

Remark 3.1 *Theorems 3.6-3.7 imply that the local polynomial estimator* $\hat{\alpha}(u)$ *and re-weighted estimator* $\hat{\alpha}_R(u)$ *have the same asymptotic distributions, which just embodies the characteristic of the local regression in nonparametric models.*

4. Simulation study

In this section, we carry out a simulation to investigate the finite sample behavior of the modified PLS estimator $\hat{\beta}$, local polynomial estimators $\hat{\alpha}(\cdot)$and $\hat{\sigma}^2(\cdot)$and the re-weighted estimators $\hat{\beta}_R$ and $\hat{\alpha}_R(\cdot)$which are proposed in section 2. In particular, the simulation in the first subsection aims to investigate the consistency of the estimators; the simulation in the second subsection is to show how good the asymptotic normality is by histograms and QQ-plots of the estimators; whereas the third applies the proposed methods to Boston housing data.

4.1. Consistency

Example 4.1 Consider the following VCPLEVM with heteroscedastic errors:

$$\begin{cases} Y_i = X_i^T\beta + Z_i^T\alpha(U_i) + \varepsilon_i, \\ \xi_i = X_i + \eta_i, \end{cases} i = 1, \cdots, n, \tag{4.1}$$

where $\beta = (\beta_1, \beta_2, \beta_3)^T = (1,2,3)^T$, $\alpha_1(u) = 0.5 + cos(6\pi u)$, $\alpha_2(u) = 2 + cos(2\pi u)$, Z_{i1} and Z_{i2} are independently generated from $U(-1,1)$, $U_i \sim U(0,1)$ and the error variance function is taken as $\sigma^2(u) = 0.25 + [\gamma\{1 + sin(2\pi u)\}]^2$. X_i is generated from multivariate normal distribution with mean 1 and pairwise covariance $\mathrm{Cov}(X_{ij}, X_{ik}) = 0.5^{|j-k|}$. The measurement error $\eta_i \sim N(0, \Sigma_\eta)$, with $\Sigma_\eta = 0.25 I_3$ and $0.5 I_3$ which represent different levels of measurement errors and I_3 is the 3×3 identity matrix.

The kernel is chosen to be the Gauss kernel, i.e., $K(u) = exp(-u^2/2)/\sqrt{2\pi}$ and the bandwidths are selected by cross-validation. To show the good performance of our proposed BC estimator $\hat{\sigma}^2(u)$, we compare it with the naive estimator $\breve{\sigma}^2(u)$ defined in Remark 2.1. We run 500 replications for $\gamma = 1$, 2 and 4 respectively. Then the \$500\$ estimated values of the error variance function at each grid point i/n are generated, and the averaged value of these 500 estimated values is taken as the final estimated value of error variance function at i/n. The final estimated curves of $\hat{\sigma}^2(u)$ and $\breve{\sigma}^2(u)$ are depicted in Figure 1.

On the other hand, in order to compare the performance of the re-weighting estimator $\hat{\beta}_R$ with the modified PLS estimator $\hat{\beta}$ and the naive estimator $\breve{\beta}_n$ (neglecting the measurement errors), we calculate the sample means, the average model errors (MEs) defined as $ME(\cdot) = (\cdot - \beta)^\top E(X^T X)(\cdot - \beta)$, and the mean squared errors (MSEs) for the three estimators with $\gamma = 1, 2, 4, 6$, $\Sigma_\eta = 0.25 I_3$ and

$0.5I_3$ respectively for $n = 800$ based on 500 replications. The naive estimator of $(\beta_1, \beta_2, \beta_3)^T$ is defined as $\breve{\beta}_n = (\breve{\beta}_1, \breve{\beta}_2, \breve{\beta}_3)^T = (\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1} \tilde{\xi})^{-1} \tilde{\xi}^T \hat{\Sigma}_\sigma^{-1} \tilde{Y}$.

Moreover, we give global mean square errors (GMSEs) of $\hat{\alpha}_R(\cdot)$and $\hat{\alpha}(\cdot)$in Table 1, where GMSE of $\hat{\alpha}_{1R}$ is defined by $\mathrm{GMSE}(\hat{\alpha}_{1R}) = (1/Mn) \sum_{l=1}^{M} \sum_{k=1}^{n} \{\hat{\alpha}_{1R}(u_k, l) - \alpha_1(u_k, l)\}^2$ with $M = 500$.

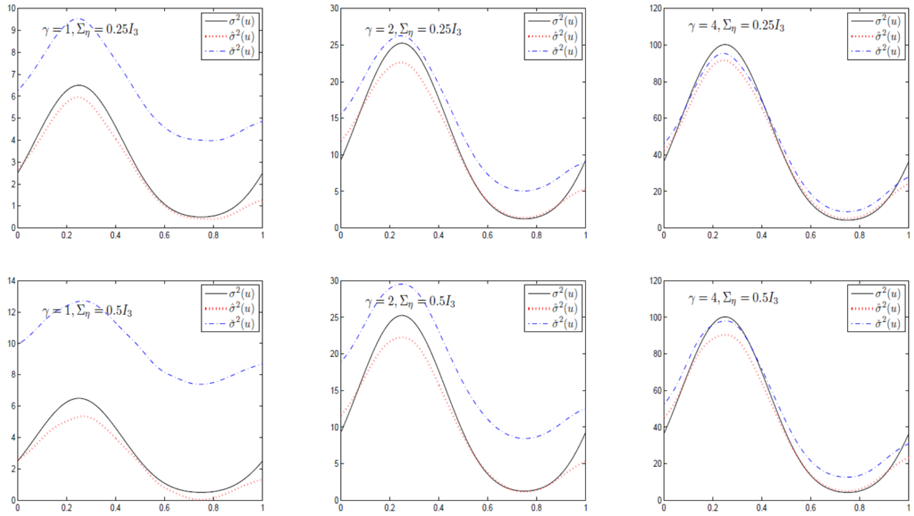

Figure 1. The estimated curves for BC estimator $\hat{\sigma}^2(u)$ and naive estimator $\breve{\sigma}^2(u)$ with n = 500

From Table 1 and Figure 1 the following observations are obtained:

(1) Shown in Figure 1, the proposed estimator of error variance performs better than the naive estimator especially for large measurement error for each fixed γ.

(2) It is shown from Table 1 that the re-weighted estimator performs best for most circum- stances among three estimators since the re-weighted estimator gives the smallest ME. It is also interesting to note that when the measurement error is negligibly small, the re-weighted estimator may perform worse than the naive estimator. Also, when the error heteroscedasticity is rather weak, the re-weighted estimator may perform worse than the modified PLS estimator.

(3) Seen from GMSEs in Table 1, the weighted estimator of coefficient function performs better than the estimator which ignores the heteroscedasticity.

Table 1. Simulated results of BC estimator $\hat{\beta}$, RWBC estimator $\hat{\beta}_R$ and the naive estimator $\breve{\beta}_n$ for true $\beta =$ (1,2,3)^T, and GMSEs for $\hat{\alpha}(\cdot)$ and $\hat{\alpha}_R(\cdot)$

		$\Sigma_\eta = 0.25I_3$				$\Sigma_\eta = 0.5I_3$			
Items		$\gamma = 1$	γ=2	γ=4	γ=6	$\gamma = 1$	γ=2	γ=4	γ=6
Mean	$\hat{\beta}_1$	1.0180	0.9796	1.0261	0.9615	0.9590	0.9524	1.0241	0.8760
	$\hat{\beta}_{1R}$	1.0079	0.9597	1.0153	0.9770	0.9484	0.9599	1.0102	0.9109
	$\breve{\beta}_1$	1.2604	1.2322	1.2611	1.2363	1.3164	1.3237	1.3575	1.3081
	$\hat{\beta}_2$	1.9680	2.0296	1.9291	2.0060	2.0237	2.0291	2.0360	1.9838
	$\hat{\beta}_{2R}$	1.9705	2.0412	1.9372	2.0161	2.0288	2.0275	2.0374	1.9905
	$\breve{\beta}_2$	1.8826	1.9260	1.8682	1.9137	1.8348	1.8292	1.8264	1.7997
	$\hat{\beta}_3$	3.0447	3.0042	3.0984	3.0561	3.0607	3.0895	2.9838	3.1716
	$\hat{\beta}_{3R}$	3.0496	3.0135	3.0914	3.0227	3.0695	3.0816	2.9995	3.1439
	$\breve{\beta}_3$	2.5751	2.5529	2.6094	2.5607	2.3110	2.314	2.2724	2.3529
ME	$\hat{\beta}$	0.0592	0.1250	0.3185	0.7592	0.1388	0.2232	0.5635	0.9877
	$\hat{\beta}_R$	0.0572	0.1179	0.2809	0.6305	0.1483	0.2151	0.5435	0.8870
	$\breve{\beta}$	0.2769	0.3122	0.3693	0.6047	0.7710	0.7492	0.9537	0.9721
MSE	$\hat{\beta}_1$	0.0218	0.0578	0.1435	0.3865	0.0717	0.1448	0.3045	0.5787
	$\hat{\beta}_{1R}$	0.0203	0.0610	0.1342	0.2957	0.0784	0.1330	0.3202	0.5576
	$\breve{\beta}_1$	0.0753	0.0772	0.1243	0.1893	0.1135	0.1338	0.2002	0.2123
	$\hat{\beta}_2$	0.0289	0.0658	0.1687	0.3339	0.0789	0.1271	0.2895	0.5756
	$\hat{\beta}_{2R}$	0.0269	0.0615	0.1369	0.2916	0.0882	0.1206	0.2815	0.5130
	$\breve{\beta}_2$	0.0243	0.0282	0.0758	0.1376	0.0425	0.0527	0.0850	0.1573
	$\hat{\beta}_3$	0.0382	0.0689	0.1777	0.4006	0.0791	0.0997	0.3295	0.5468
	$\hat{\beta}_{3R}$	0.0385	0.0613	0.1504	0.3223	0.0814	0.0983	0.2923	0.4623
	$\breve{\beta}_3$	0.1944	0.2235	0.2138	0.3269	0.4902	0.4680	0.5887	0.5145
GMSE	$\hat{\alpha}_1$	0.3225	0.6437	1.8775	3.9901	0.5202	0.8167	1.9131	4.1134
	$\hat{\alpha}_{1R}$	0.3214	0.6458	1.8756	3.9981	0.5235	0.8207	1.9039	4.0861
	$\hat{\alpha}_2$	0.2605	0.6247	1.8569	4.0876	0.4336	0.7699	2.1216	4.4723
	$\hat{\alpha}_{2R}$	0.2614	0.6248	1.8588	4.0784	0.4339	0.7662	2.1064	4.4637

4.2. Asymptotic normality

In this subsection, we shall investigate the problem of asymptotic normality of the estimator $\hat{\alpha}(\cdot)$and the re-weighted estimator $\hat{\alpha}_R(\cdot)$. Specifically, in Example 4.1, we take $\Sigma_\eta = 0.25 I_3$ and$\gamma = 2$, 4 and 6, respectively. The sample size n is chosen to be 100 and the number of simulated realizations is 500. The other variables are chosen as that in Example 4.1. The histograms for $\hat{\alpha}_1(0.5)$ and QQ-normality plots for re-weighted estimator $\hat{\alpha}_{2R}(0.5)$ with$\gamma = 2$, 4 and 6 are put in Figure 2. It can be seen from Figure 2 that the sampling distributions of the estimators fit the normal reasonably, this fit is better as decreasing the value of γ or increasing the sample size.

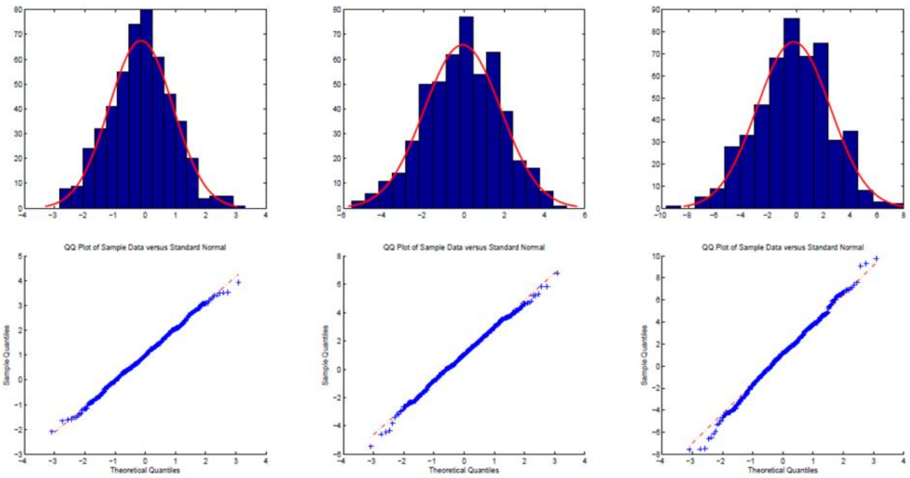

Figure 2. Histograms of re-weighted estimator $\hat{\alpha}_{1R}(0.5)$ and QQ-normality plots of re-weighted estimator $\hat{\alpha}_{2R}(0.5)$ with $\gamma = 2$ (left), γ=4 (center) and γ=6 (right)

4.3. Application to Boston housing data

We analyze the data from Boston housing data (see Harrison and Rubinfeld, 1978) to illustrate the proposed methodology developed in this paper. The data set consists of the median value of owner-occupied homes in 506 US census tracts in the Boston area in 1970, together with several variables. Following Fan and Huang (2005), several explanatory variables are per capita crime rate by town (CRIM) denoted by Z_2, nitric oxide concentration parts per 10 million (NOX) denoted by Z_3, average number of rooms per dwelling (RM) denoted by Z_4, full value property tax per \$10000 (TAX) denoted by Z_5, pupil-twacher ratio by town school district (PTRAIO) denoted by X_1,proportion of owner-occupied units built prior to 1940 (AGE) denoted by X_2 and lower status of the population (LSTAT). Take $Z_1 = 1$ as the intercept term and $U = \sqrt{\text{LSTAT}}$. Based on the observations $(Y; Z_1, Z_2, Z_3, Z_4, Z_5, X_1, X_2, U)$, Fan and Huang (2005) employed the following semi-varying coefficient model: $Y = \beta_1 X_1 +$

$\beta_2 X_2 + \sum_{i=1}^{5} \alpha_i (U) Z_i + \varepsilon$ to fit the given data. It is shown that the coefficient of X_2 is not significant at 0.01 significance level and the coefficient of X_1 is -0.7199. Due to some reasons, for example, some people did not remember the exact time, or for their own interests they gave the incorrect results whether their owner-occupied units were built prior to 1940. At the same time, for some reasons, some schools gave the wrong data of pupil-teacher ratio by town school district. Therefore, covariates X_1 and X_2 may have measurement errors and the surrogate variable $\xi = (\xi_1, \xi_2)^T$ is observed by $\xi = X + \eta$ with $X = (X_1, X_2)^T$.

To demonstrate the proposed approach, a sensitivity analysis is conducted, as mentioned in Lin and Carroll (2000). To estimate the measurement error covariance $\mathrm{Cov}(\eta)$, one either needs a validation study or replicates of AGE and PTRATIO count measures. Unfortunately, these are not available in the Boston housing data, and hence we can not estimate $\mathrm{Cov}(\eta)$ by using the Boston housing data. Similar to Lin and Carroll (2000) who assumed that the variance of the measurement errors is $1/4$ and $1/2$ and Feng and Xue (2014) who assumed the measurement error follows normal distribution $N(0, 0.3)$, we here assume that $\mathrm{Cov}(\eta)=\mathrm{diag}(1,1)$. Further more we assume the model errors are heteroscedastic which has the form $\mathrm{Var}(\varepsilon|X, Z, U) = \sigma^2(U)$.

The fitted values of the re-weighted estimators of β_1 and β_2 are $\beta_1 = -0.9985$ and $\beta_2 = -0.0072$. It can be observed that β_2 is nearly zero, so it can be conclude that the value of housing is almost not revelent to AGE (proportion of owner-occupied units built prior to 1940). Also, the estimated curves of $\alpha_i(u)$ $(i = 1, \cdots, 5)$ with our proposed method and the estimation method of Fan and Huang (2005) are reported in Figure 3. The results indicate that more crowded school in the tracts often result in lower value of housing.

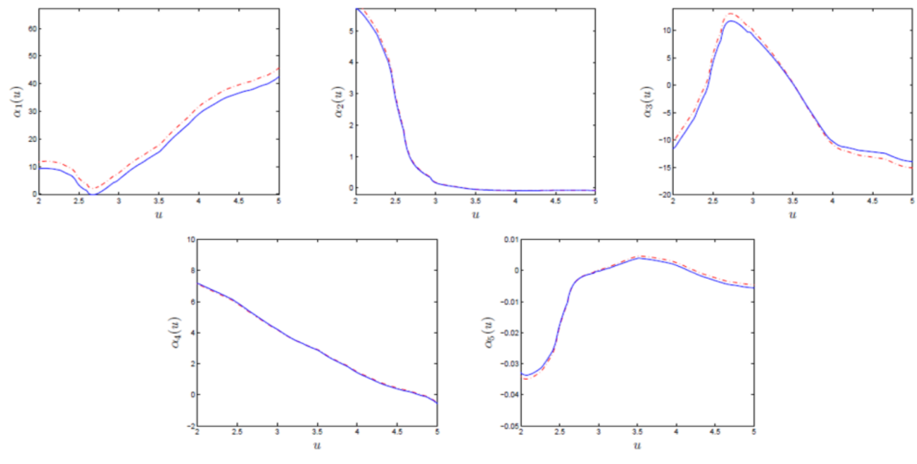

Figure 3. The estimated curve $\hat{\sigma}^2(u)$ and the estimated curves for varying-coefficient functions $\alpha_i{}^2(u)$ $(i = 1, \cdots, 5)$ with our proposed method (dashed line) and the estimation method of Fan and Huang (2005) (solid line)

References

Carroll R. J., Ruppert D., Stefanski L. A. (1995). *Measurement Error in Nonlinear Models,* Chapman & Hall, London.

Chiou J. M., Mu¨ller H. G. (1999). Nonparametric quasi-likelihood. *The Annals of Statistics,* Vol. 27, No. 1, pp. 36-64. http://doi.org/10.1214/aos/1018031100

Fan G. L., Huang Z. S. (2013). Empirical likelihood for semiparametric varying-coefficient heteroscedastic partially linear errors-in-variables models. *Communications in Statistics - Theory and Methods,* Vol. 42, pp. 2437-2450. http://doi.org/10.3770/j.issn:2095-2651.2012.01.011

Fan G. L., Liang H. Y., Wang J. F. (2013). Statistical inference for partially time-varying coefficient errors-in-variables models. *Journal of Statistical Planning and Inference,* Vol. 143, pp. 505- 519. http://doi.org/10.1214/12-EJS701

Fan J., Gijbels I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics,* Vol. 20, pp. 2008-2036. http://doi.org/10.1214/aos/1176348900

Fan J., Gijbels I. (1996). *Local Polynomial Modeling and Its Applications.* London, Chap- man and Hall.

Fan J., Huang T. (2005). Profile likelihood inferences on semiparametric varying- coefficient partially linear models. *Bernoulli,* Vol. 11, pp. 1031-1057. http://doi.org/10.3150/bj/1137421639

Feng S. Y., Xue L. G. (2014). Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. *Annals of the Institute of Statistical Mathematics,* Vol. 66, pp. 121-140. http://doi.org/10.1007/s10463-013-0407-z

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

Harrison D., Rubinfeld D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, Vol. 5, pp. 81-102. http://doi.org/10.1016/0095-0696(78)90006-2

Hastie T. J., Loader C. (1993). Local regression: automatic kernel carpentry (with Discussion). *Statistical Science,* Vol. 8, pp. 120-143. http://doi.org/10.1214/ss/1177011005

Hwang J. T. (1986). The multiplicative errors-in-variables models with applications to the recent data released by the U.S. department of energy. *Journal of The American Statistical Association,* Vol. 81, pp. 680-688. http://doi.org/10.1080/01621459.1986.10478321

Kai B., Li R., Zou H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of Statistics,* Vol. 39, pp. 305-332. http://doi.org/10.1214/10-AOS842

Liang H., H¨adle W., Carroll R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics,* Vol. 27, pp. 1519-1535. http://doi.org/10.1214/aop/1022677563

Lu X. (2009). Empirical likelihood for heteroscedastic partially linear models. *Journal of Multivariate Analysis,* Vol. 100, pp. 387-396. http://doi.org/10.1016/j.jmva.2008.05.006

Ma Y., Chiou J. M., Wang N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika,* Vol. 943, pp. 75-84. http://doi.org/10.1093/biomet/93.1.75

Ma Y., You J., Zhou Y. (2013). Generalized profile LSE in varying-coefficient partially linear models with measurement errors. *Acta Mathematicae Applicatae Sinica*, Vol. 29, pp. 477-490. http://doi.org/10.1007/s10255-013-0229-z

Mack Y. P., Silverman B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields,* Vol. 61, pp. 405-415. http://doi.org/10.1007/BF00539840

Menictas M., Wand M. P. (2015). Variational inference for heteroscedastic semiparametric regression. *Australian & New Zealand Journal of Statistics,* Vol. 57, pp. 119-138. http://doi.org/10.1111/anzs.12105

Ruppert D., Wand M. P. (1994). Multivariate weighted least squares regression. *The Annals of Statistics,* Vol. 22, pp. 1346-1370. http://doi.org/10.1214/aos/1176325632

Shen S. L., Cui J. L., Mei C. L., Wang C. W. (2014). Estimation and inference of semivarying coefficient models with heteroscedastic errors. *Journal of Multivariate Analysis,* Vol. 124, pp. 70-93. http://doi.org/10.1016/j.jmva.2013.10.010

Tosteson T., Stefanski L. A., Schafer D. W. (1989). A measurement error model for binary and ordinal regression. *Statistical Medicine,* Vol. 8, pp. 1139-1147. http://doi.org/10.1002/sim.4780080914

Wang N., Carroll R. J., Liang K. Y. (1996). Quasilikelihood and variance functions in measurement error models with replicates. *Biometrics,* Vol. 52, pp. 423-432. http://doi.org/10.2307/2532882

Wei C. H. (2012). Statistical inference for restricted partially linear varying coefficient errors-in-variables models. *Journal of Statistical Planning and Inference,* Vol. 142, pp. 2464-2472. http://doi.org/10.1016/j.jspi.2012.02.041

Xia Y. C., Li W. K. (1999). On the estimation and testing of functional-coefficient linear models. *Statistica Sinica,* Vol. 9, pp. 735-757.

You J. H., Chen G. M. (2006). Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. *Journal of Multivariate Analysis,* Vol. 97, pp. 324-341. http://doi.org/10.1016/j.jmva.2005.03.002

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,* Vol. 34, pp. 391-410. https://doi.org/10.1002/cjs.5550340303

Zhang W., Lee S. Y., Song X. (2002). Local polynomial fitting in semivarying coefficient models. *Journal of Multivariate Analysis,* Vol. 82, pp. 166-188. https://doi.org/10.1006/jmva.2001.2012

Zhou Y., Liang H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics,* Vol. 37, pp. 427-458. http://doi.org/10.1214/07-AOS561

Zhu L., Cui H. (2003). A semiparametric regression model with errors in variables. *Scandinavian Journal of Statistics,* Vol. 30, pp. 429-442. http://doi.org/10.1111/1467-9469.00340

Appendix: Assumptions and Proofs

For the convenience and simplicity, let $X = (X_1, \cdots, X_n)^T$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$, $\eta = (\eta_1, \cdots, \eta_n)^T$,

$\tilde{X} = (I - S)X$, $\tilde{M} = (I - S)M$, $\tilde{\varepsilon} = (I - S)\tilde{\varepsilon}$, $\tilde{\eta} = (I - S)\eta$ and C denote positive constant whose value may vary at each occurrence. Before proving the main theorems, we begin this section with making the following assumptions.

(C1) The kernel $K(\cdot)$is a symmetric and Lipschitz continuous function with a compact support $[-1,1]$.

(C2) The matrixes $\Gamma(u)$ and $\Phi(u)$ are non-singular, $E[\|Z_1\|^{2s}] < \infty$, $E[\|X_1\|^{2s}] < \infty$ and $E[\|\eta_1\|^{2s}] < \infty$ for some $s > 2$, where $\|\cdot\|$ is the L_2 norm.

(C3) The variable U has a bounded support $\mathcal{D}$ and its density function $p(u) > 0$ is Lipschitz continuous and has continuous second order derivative on $\mathcal{D}$.

(C4) $\{\alpha_j(\cdot), j = 1, \cdots, q\}$ have continuous second derivative.

(C5) There exist a $\delta < 1 - s^{-1}$ such that $\lim_{n\to\infty} n^{2\delta-1} h_1 = \infty$.

(C6) The variance function $\sigma^2(\cdot)$is uniformly bounded, bounded away from zero and has continuous second order derivative on its domain.

(C7) The bandwidth h_1 satisfies that $nh_1^2 \to \infty, nh_1 \to 0$ and $\lim_{n\to\infty} \frac{[log(1/h_1)]^2}{nh_1} = 0$.

(C8) The bandwidth h_2 satisfies that $nh_2^2 \to \infty$, $nh_2^8 \to 0$ and $\lim_{n\to\infty} \frac{[log(1/h_2)]^2}{nh_2} = 0$.

Lemma A.1 Assume that conditions (C1)-(C4) are satisfied. Then $\sup_{u\in\mathcal{D}} \frac{1}{nh_1} \sum_{i=1}^n K(\frac{U_i-u}{h_1})(\frac{U_i-u}{h_1})^k Z_{ij_1} Z_{ij_2} = p(u)\Gamma_{j_1j_2}(u)\mu_k + O\{h_1^2 + (\frac{log\, n}{nh_1})^{1/2}\}\ a.s.$

$\sup_{u\in\mathcal{D}} \frac{1}{nh_1} \sum_{i=1}^n K(\frac{U_i-u}{h_1})(\frac{U_i-u}{h_1})^k Z_{ij}\varepsilon_i = O\{(\frac{log\, n}{nh_1})^{1/2}\}\ a.s.$

where $j, j_1, j_2 = 1, \cdots, q$, $k = 0,1,2,4$ and $\Gamma_{j_1j_2}(U)$ is the $(j_1, j_2)th$ element of $\Gamma(U)$.

Lemma A.1 can be proved as Lemma A.2 in Xia and Li (1999).

Lemma A.2 (Mack and Silverman, 1982, Theorem B) Let $(\tilde{X}_1, \tilde{Y}_1), \cdots, (\tilde{X}_n, \tilde{Y}_n)$ be independent and identical distributed random vectors. Further assume there exists an $s > 2$ which satisfies $E|\tilde{Y}|^s < \infty$ and $\sup_x \int |y|^s \tilde{f}(x, y)dy < \infty$, where $\tilde{f}(\cdot,\cdot)$ denotes the joint probability density of $(\tilde{X}, \tilde{Y})$. $K(\cdot) > 0$ is a bounded positive function with a bounded compact support and satisfies the Lipschitz condition. Given that $\lim_{n\to\infty} n^{2\delta-1} h = \infty$ for some $\delta < 1 - s^{-1}$, then $\sup_x |\frac{\sum_{i=1}^n K_h(\tilde{X}_i - x)(\tilde{Y}_i - E\tilde{Y}_i)}{\sum_{j=1}^n K_h(\tilde{X}_i - x)}| = O_p[(\frac{log(1/h)}{nh})^{1/2}]$.

Lemma A.3 (You and Chen, 2006, Lemma A.2) Let $D_1, \cdots, D_n$ be i.i.d. random variables. If $E|D_1|^s < \infty$for $s > 1$, then $max_{1in}|D_i| = o(n^{1/s})\ a.s.$

Lemma A.4 (Chiou and Muller, 1999, Lemma 4.1) Suppose that conditions (C1)-(C8) hold. Then

$\sup_{u\in\mathcal{D}}|\sum_{i=1}^{n} W_{h_2 i}(u)\varepsilon_i^2 - \sigma^2(u)| = O_p(b_n)$, where $W_{h_2 i}(u)$ is defined in (2.4).

Lemma A.5 Suppose that conditions (C1)-(C7) hold. Then

$$n^{-1}\tilde{\xi}^T\tilde{\xi} \xrightarrow{p} \Sigma_\eta + \Sigma_1 \quad \text{(A.1)}$$

$$n^{-1}\tilde{\xi}^T\hat{\Sigma}_\sigma^{-1}\tilde{\xi} - \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}^{-2}(U_i)\Sigma_\eta = n^{-1}\tilde{X}^T\Sigma_\sigma^{-1}\tilde{X} + O_p(c_n), \quad \text{(A.2)}$$

as $n \to \infty$, where Σ_1 is defined in Theorem 3.2.

Proof. We first establish equation (A.1). Observe that $n^{-1}\tilde{\xi}^T\tilde{\xi} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i + \eta_i - S_i^T\eta)(\tilde{X}_i + \eta_i - S_i^T\eta)^T$.

Similar to the proof of (A.6) and (A.9) in Shen *et al.* (2014), we can derive $\tilde{X}_i = [X_i - \Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\{1 + O_p(c_n)\} = \psi_i\{1 + O_p(c_n)\}$,

$$\tilde{\varepsilon}_i = \varepsilon_i + O_p(c_n) \ \text{ and } \tilde{\eta}_i = \eta_i + O_p(c_n). \quad \text{(A.3)}$$

Note that $\{\psi_i, i = 1, \cdots, n\}$ is independent and identical distributed, then together with (A.3), we have

$n^{-1}\tilde{\xi}^T\tilde{\xi} = \frac{1}{n}\sum_{i=1}^{n}[\psi_i\{1 + O_p(c_n)\} + \eta_i + O_p(c_n)][\psi_i\{1 + O_p(c_n)\} + \eta_i + O_p(c_n)]^T = \frac{1}{n}\sum_{i=1}^{n}(\psi_i\psi_i^T + \eta_i + \eta_i^T) + o_p(1) = E(\psi_1\psi_1^T) + \Sigma_\eta$, which proves (A.1). On the other hand, in order to prove (A.2), we denote $\psi = (\psi_1, \cdots, \psi_n)^T$. Then by equation (A.3) and Theorem3.1, it follows that $n^{-1}\tilde{\xi}^T\hat{\Sigma}_\sigma^{-1}\tilde{\xi} - \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}^{-2}(U_i)\Sigma_\eta = n^{-1}(\tilde{X} + \tilde{\eta})^T\hat{\Sigma}_\sigma^{-1}(\tilde{X} + \tilde{\eta}) - \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}^{-2}(U_i)\Sigma_\eta = n^{-1}\tilde{X}^T\Sigma_\sigma^{-1}\tilde{X} + n^{-1}\tilde{X}^T(\hat{\Sigma}_\sigma^{-1} - \Sigma_\sigma^{-1})\tilde{X} + n^{-1}\eta^T\Sigma_\sigma^{-1}\eta - \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}^{-2}(U_i)\Sigma_\eta + O_p(c_n) = n^{-1}\tilde{X}^T\Sigma_\sigma^{-1}\tilde{X} + \Sigma_\eta \cdot \frac{1}{n}\sum_{i=1}^{n}(\sigma^{-2}(U_i) - \hat{\sigma}^{-2}(U_i)) + O_p(c_n) = n^{-1}\tilde{X}^T\Sigma_\sigma^{-1}\tilde{X} + O_p(c_n)$, which gives (A.2). Thus, the proof of Lemma A.5 is completed.

Proof of Theorem 3.1. Denote $\widetilde{W}(u) = \text{diag}(W_{h_2 1}(u), \cdots, W_{h_2 n}(u))$, where $W_{h_2 i}(u)$ is defined in (2.4). Then by (2.3) and note $\xi = X + \eta$, we obtain $\hat{\sigma}^2(u) = \hat{\varepsilon}^T\widetilde{W}(u)\hat{\varepsilon} - \hat{\beta}^T\Sigma_\eta\hat{\beta} = (Y - \xi\hat{\beta} - \widehat{M})^T\widetilde{W}(u)(Y - \xi\hat{\beta} - \widehat{M}) - \hat{\beta}^T\Sigma_\eta\hat{\beta} = [\varepsilon + X(\beta - \hat{\beta}) - \eta\hat{\beta} + (M - \widehat{M})]^T\widetilde{W}(u)[\varepsilon + X(\beta - \hat{\beta}) - \eta\hat{\beta} + (M - \widehat{M})] - \hat{\beta}^T\Sigma_\eta\hat{\beta} = \varepsilon^T\widetilde{W}(u)\varepsilon - 2(\widehat{M} - M)^T\widetilde{W}(u)\varepsilon - 2[X(\hat{\beta} - \beta)]^T\widetilde{W}(u)\varepsilon + [X(\hat{\beta} - \beta)]^T\widetilde{W}(u)[X(\hat{\beta} - \beta)] + 2(\widehat{M} - M)^T\widetilde{W}(u)X(\hat{\beta} - \beta) + (\widehat{M} - M)^T\widetilde{W}(u)(\widehat{M} - M) - 2(\eta\hat{\beta})^T\widetilde{W}(u)\varepsilon + 2(\eta\hat{\beta})^T\widetilde{W}(u)X(\hat{\beta} - \beta) + 2(\eta\hat{\beta})^T\widetilde{W}(u)(\widehat{M} - M) + (\eta\hat{\beta})^T\widetilde{W}(u)\eta\hat{\beta} - \hat{\beta}^T\Sigma_\eta\hat{\beta} := \sum_{i=1}^{10} D_i$.

Lemma A.4 implies that $\sup_{u\in\mathcal{D}}|D_1-\sigma^2(u)| = O_p(b_n)$. From Theorem 3.5 in You and Chen (2006), we have $\sup_{u\in\mathcal{D}}\|\hat{\alpha}(u)-\alpha(u)\| = O(c_n)\ a.s.$, where $\mathbf{1}_{q\times 1}$ denotes a column vector of order q whose entities are all ones. By noting $M_i = Z_i^T\alpha(U_i)$ and$\widehat{M}_i = Z_i^T\hat{\alpha}(U_i)$, it follows that

$$\widehat{M}_i - M_i = O_p(c_n),\ i=1,\cdots,n. \tag{A.4}$$

Then by Lemma A.2, we obtain that $\sup_{u\in\mathcal{D}}|D_9+D_{10}| = \sup_{u\in\mathcal{D}}|\frac{\sum_{j=1}^n K_{h_2}(X_j-x)(\eta_j\eta_j^T-\Sigma_\eta)}{\sum_{j=1}^n K_{h_2}(X_j-x)}| = O_p(b_n)$.

Similarly, we can easily get that $\sup_{u\in\mathcal{D}}|D_i| = O_p(b_n),\ i=2,\cdots,8,$

which completes the proof of Theorem 3.1.

Proof of Theorem 3.2. Let $\nabla = \sum_{i=1}^n \tilde{\xi}_i\,\tilde{\xi}_i^T - n\Sigma_\eta$. Similar to the proof of Theorem 3.1 in Fan *et al*. (2013), we write

$$\hat{\beta}-\beta = \nabla^{-1}n\Sigma_\eta\beta + \nabla^{-1}\sum_{i=1}^n\tilde{\xi}_i\,(\tilde{Y}_i-\tilde{\xi}_i^T\beta) = \nabla^{-1}n\Sigma_\eta\beta + \nabla^{-1}\sum_{i=1}^n(\tilde{X}_i+\tilde{\eta}_i)(\widetilde{M}_i+\tilde{\varepsilon}_i-\tilde{\eta}_i^T\beta). \tag{A.5}$$

From equation (A.3) and condition (C7), one can derive that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{X}_i\tilde{\varepsilon}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n[X_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\{1+O_p(c_n)\}\cdot\{\varepsilon_i+O_p(c_n)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n[X_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\varepsilon_i + o_p(1) \tag{A.6}$$

Similar to the proof of (A.6), from (A.3)-(A.4) and conditions (C7)-(C8), it is easily to prove that

$\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{X}_i\tilde{\eta}_i^T = \frac{1}{\sqrt{n}}\sum_{i=1}^n[X_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\eta_i^T + o_p(1)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\eta}_i\,\tilde{\varepsilon}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n\eta_i\,\varepsilon_i + o_p(1)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\eta}_i\,\tilde{\eta}_i^T = \frac{1}{\sqrt{n}}\sum_{i=1}^n\eta_i\,\eta_i^T + o_p(1)$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{X}_i\,\widetilde{M}_i = o_p(1)$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{\eta}_i\,\widetilde{M}_i = o_p(1)$, which together with equation (A.5) yield that

$$\sqrt{n}(\hat{\beta}-\beta) = (\frac{\nabla}{n})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\{[X_i+\eta_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i](\varepsilon_i-\eta_i^T\beta)+\Sigma_\eta\beta\}+o_p(1). \tag{A.7}$$

Note that $\{[X_i+\eta_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i](\varepsilon_i-\eta_i^T\beta)+\Sigma_\eta\beta,\ i=1,\cdots,n\}$ is independent and identical distributed with mean zero. Thus

$\mathrm{Var}\{\frac{1}{\sqrt{n}}\sum_{i=1}^n[(X_i+\eta_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i)(\varepsilon_i-\eta_i^T\beta)+\Sigma_\eta\beta]\} = E[(X_i+\eta_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i)(\varepsilon_i-\eta_i^T\beta)+\Sigma_\eta\beta]^{\otimes 2} = \Sigma_2,$

Then, applying central limit theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\{[X_i+\eta_i-\Phi^T(U_i)\Gamma^{-1}(U_i)Z_i](\varepsilon_i-\eta_i^T\beta)+\Sigma_\eta\beta\}\xrightarrow{d}N(0,\Sigma_2),$$

which combining with Lemma A.4, equation (A.7) and the Slutsky theorem lead to the asymptotic normality for $\hat{\beta}$.

By Lemma A.4 and the proof of the asymptotic normality of β , one can easy to obtain the consistency of $\hat{\Sigma}_1^{-1}\hat{\Sigma}_2\hat{\Sigma}_1^{-1}$.

Proof of Theorem 3.3. Denote $\hat{\beta}_T = (\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X})^{-1}\tilde{\xi}^T \Sigma_\sigma^{-1}\tilde{Y}$. We first establish

$$\sqrt{n}(\hat{\beta}_R - \beta) = \sqrt{n}(\hat{\beta}_T - \beta) + o_p(1). \tag{A.8}$$

Obviously, in order to prove (A.8), we need only to show that

$$\sqrt{n}(\hat{\beta}_R - \hat{\beta}_T) = o_p(1). \tag{A.9}$$

By the definition of $\hat{\beta}_R$, we have $\sqrt{n}(\hat{\beta}_R - \hat{\beta}_T) = \sqrt{n}\{[\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1}\tilde{\xi} - \sum_{i=1}^n \hat{\sigma}^{-2}(U_i)\Sigma_\eta]^{-1}\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1}\tilde{Y} - (\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X})^{-1}\tilde{\xi}^T \Sigma_\sigma^{-1}\tilde{Y}\} = \sqrt{n}\{[\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1}\tilde{\xi} - \sum_{i=1}^n \hat{\sigma}^{-2}(U_i)\Sigma_\eta]^{-1} - (\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X})^{-1}\}\tilde{\xi}^T \hat{\Sigma}_\sigma^{-1}(\tilde{X}\beta + \widetilde{M} + \tilde{\varepsilon}) + \sqrt{n}(\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X})^{-1}(\tilde{\xi}^T \widehat{\Sigma_\sigma^{-1}} - \tilde{\xi}^T \Sigma_\sigma^{-1})(\tilde{X}\beta + \widetilde{M} + \tilde{\varepsilon})$

Then according to Theorem 3.1 and Lemma A.5, we can easily get (A.9). Analogously to the proof

of Theorem 3.2, one can obtain that

$$\sqrt{n}(\hat{\beta}_T - \beta) = \left(\frac{\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X}}{n}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\{[X_i + \eta_i - \Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\sigma^{-2}(U_i)(\varepsilon_i - \eta_i^T\beta) + \Sigma_\eta\beta\sigma^{-2}(U_i)\} + o_p(1)$$

By (A.3)-(A.4), $n^{-1}\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X} = E[\psi_1\psi_1^T\sigma^{-2}(U_1)] + o_p(1)$ and $E(\varepsilon_i|X_i, Z_i, U_i) = 0$, we can derive that

$$\left(\frac{\tilde{X}^T \Sigma_\sigma^{-1}\tilde{X}}{n}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\{[X_i + \eta_i - \Phi^T(U_i)\Gamma^{-1}(U_i)Z_i]\sigma^{-2}(U_i)(\varepsilon_i - \eta_i^T\beta) + \Sigma_\eta\beta\sigma^{-2}(U_i)\} \xrightarrow{d} N(0, \Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1}) \text{ as } n \to \infty.$$

Thus, $\sqrt{n}(\hat{\beta}_R - \beta) \xrightarrow{d} N(0, \Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1})$. By Lemma A.4 and the proof of the asymptotic normality of $\hat{\beta}_R$, one can easy to obtain the consistency of $\hat{\Sigma}_{1R}^{-1}\hat{\Sigma}_{2R}\hat{\Sigma}_{1R}^{-1}$.

Proof of Theorem 3.4. Denote $A_i = [X_i + \eta_i - \Phi^T(U_i)\Gamma^{-1}(U_i)Z_i](\varepsilon_i - \eta_i^T\beta) + \Sigma_\eta\beta$, $A = (A_1, \cdots, A_n)^T$ and $\psi = (\psi_1, \cdots, \psi_n)^T$. Then from the proof of Theorem 3.2 and Theorem3.3, we have

$$\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} = \{E(\psi_1\psi_1^T)\}^{-1}E(A_iA_i^T)\{E(\psi_1\psi_1^T)\}^{-1} = \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi_i\,\psi_i^T\right)^{-1}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n A_i\,A_i^T\right)\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi_i\,\psi_i^T\right)^{-1} = n(\psi^T\psi)^{-1}A^TA(\psi^T\psi)^{-1}$$

And

$\Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1} =$
$(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi_i\,\psi_i^T\sigma^{-2}(U_i))^{-1}(\lim_{n\to\infty}\sum_{i=1}^{n}A_i\,A_i^T\sigma^{-4}(U_i))(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\psi_i\,\psi_i^T\sigma^{-2}(U_i))^{-1} =$
$(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}A^T\Sigma_\sigma^{-2}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}$.

For any given p-dimensional vector $\mathbf{a}$, we find

$\boldsymbol{a}^T(\psi^T\psi)^{-1}A^TA(\psi^T\psi)^{-1}\boldsymbol{a} = \|A(\psi^T\psi)^{-1}\boldsymbol{a}\|^2 = \|A(\psi^T\psi)^{-1}\boldsymbol{a} - \Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a} + \Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 = \|A(\psi^T\psi)^{-1}\boldsymbol{a} - \Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 + \|\Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 + 2\boldsymbol{a}^T(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}A^T\Sigma_\sigma^{-1}[A(\psi^T\psi)^{-1}\boldsymbol{a} - \Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a} = \|A(\psi^T\psi)^{-1}\boldsymbol{a} - \Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 + \|\Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 \geq \|\Sigma_\sigma^{-1}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}\|^2 = \boldsymbol{a}^T(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}A^T\Sigma_\sigma^{-2}A(\psi^T\Sigma_\sigma^{-1}\psi)^{-1}\boldsymbol{a}$, which implies that $\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1} \geq \Sigma_{1R}^{-1}\Sigma_{2R}\Sigma_{1R}^{-1}$ and the proof of Theorem 3.4 is finished.

Proof of Theorem 3.5. By the definition of $\hat{\alpha}(u)$, similar to the proof of Theorem 3.4 in

Fan *et al*. (2013), we have $\hat{\alpha}(u) - \alpha(u) = (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)(Y - \xi\hat{\beta}) - \alpha(u) = (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)[M - Z\alpha(u)] + (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\xi(\beta - \hat{\beta}) - (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\eta\beta + (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\varepsilon.$

Note that $\|\beta - \hat{\beta}\| = O_p(n^{-1/2})$ by Theorem 3.2. This, in conjunction with (A.2) and (A.4) in Shen *et al*. (2014) yields that

$(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\xi(\beta - \hat{\beta}) = (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)X(\beta - \hat{\beta}) + (I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\eta(\beta - \hat{\beta}) = \mathbf{1}_{q\times 1}O_p(n^{-1/2}).$

Furthermore, by standard argument for local linear estimator, we have

$$\sup_{u\in\mathcal{D}}(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)[M - Z\alpha(u)] = \mathbf{1}_{q\times 1}O_p(h_1^2).$$

(A.9) in Shen *et al*. (2014) implies that $(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\varepsilon = \mathbf{1}_{q\times 1}O_p(c_n).$

Similarly, we can derive $(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)\eta = \mathbf{1}_{q\times 1}O_p(c_n).$

Therefore, we have $\sup_{u\in\mathcal{D}}\|\hat{\alpha}(t) - \alpha(t)\| = O_p(c_n).$

By using the same method and together with Theorem 3.3, yields that $\sup_{u\in\mathcal{D}}\|\hat{\alpha}_R(u) - \alpha(u)\| = O_p(c_n).$

Thus the proof of Theorem 3.5 is completed.

Proof of Theorem 3.6. By the Taylor expansion and a direct simplification, one can derive

$$M = \begin{pmatrix} Z_1^T\alpha(u) + (U_1 - u)Z_1^T\alpha'(u) + \frac{1}{2}(U_1 - u)^2 Z_1^T\alpha''(u) \\ \vdots \\ Z_n^T\alpha(u) + (U_n - u)Z_n^T\alpha'(u) + \frac{1}{2}(U_n - u)^2 Z_n^T\alpha''(u) \end{pmatrix} + o(h^2).$$

Then according to the proof of Theorem 3.4, conditions (C1)-(C4) and Theorem 3.2, we have

$$\sqrt{nh_1}\{\hat{\alpha}(u) - \alpha(u) - [\tfrac{1}{2}h_1^2\mu_2\alpha''(u) + o_p(h_1^2)]\} =$$
$$\sqrt{nh_1}(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)(\varepsilon - \eta\beta) + o_p(1). \quad \text{(A.10)}$$

Denote $\frac{1}{\sqrt{nh}}\sum_{i=1}^n K\,(\frac{U_i-u}{h})Z_i(\varepsilon_i - \eta_i^T\beta) = \mathcal{H}$.It is easy to obtain the conditional expectation and variance on $\mathcal{D}$ that

$$E(\mathcal{H}|\mathcal{D}) = 0 \text{ and } \mathrm{Var}(\mathcal{H}|\mathcal{D}) = v_0\Gamma^{-1}(u)(\sigma^2(u) + \beta^T\Sigma_\eta\beta)p(u). \quad \text{(A.11)}$$

(A.2) in Shen *et al*. (2014) implies that $n^{-1}D^T(u)W(u)D(u) = p(u)\Gamma(u) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}\{1 + O_p(c_n)\}$.

This, in conjunction with the (A.11) the Slutsky and central limit theorems yields that $\sqrt{nh_1}(I_q, 0_q)[D^T(u)W(u)D(u)]^{-1}D^T(u)W(u)(\varepsilon - \eta\beta) \xrightarrow{d} N(0, v_0\Gamma^{-1}(u)(\sigma^2(u) + \beta^T\Sigma_\eta\beta)p^{-1}(u))$, which together with (A.10) and the Slutsky theorem gives the result of Theorem 3.6.

Since the proof of Theorem 3.7 is analogous to that of Theorem 3.6, we omit the details here.