Statistical Inference for Heteroscedastic Semi-Varying Coefficient EV Models

Jianhong Shi

School of Mathematics and Computer Science, Shanxi Normal University, Linfen, China, 041000

Weixing Song

Department of Statistics, Kansas State University, Manhattan, KS, 66503

Fanrong Zhao*

School of Mathematics and Computer Science, Shanxi Normal University, Linfen, China, 041000

Abstract

This paper proposes an estimation procedure for a class of semi-varying coefficient regression models when the covariates of the linear part are subject to measurement errors. Initial estimates for the regression and varying coefficients are first constructed by the profile least squares procedure without input from heteroscedasticity, a bias-corrected kernel estimate for the variance function then is proposed, which in turn is used to define re-weighted bias-corrected estimates of the regression and varying coefficients. Large sample properties of the proposed estimates are thoroughly investigated. The finite sample performance of the proposed estimates are assessed by an extensive simulation study. The simulation results show that the re-weighted bias-corrected estimates outperform the initial estimates and the naive estimates.

Keywords: Semi-varying coefficient EV model, Heteroscedasticity, Profile least-squares, Consistency, Asymptotic normality

2000 MSC: primary 62F35, secondary 62F10

1. Introduction

As an extension of the partially linear and the varying coefficient regression models, the semi-varying coefficient models have received much attention from researchers and practitioners in the past decades. To be specific, the semi-varying coefficient models can be formulated as

$$Y = X'g(U) + Z'\beta + \varepsilon, \tag{1.1}$$

where Y is a scalar response variable, and $(X', Z', U)'$ are the associated explanatory variables of dimensions p, q and 1, respectively. Here and after, the apostrophe denotes the transpose of a vector or matrix. U is supported on Ω, a compact subset of R. $\beta = (\beta_1, \beta_2, \ldots, \beta_q)'$ is a q-dimensional unknown parameters, and

*Corresponding author

Email address: 1572762347@qq.com (Fanrong Zhao)

$g(\cdot) = (g_1(\cdot),\ g_2(\cdot),\ \ldots,\ g_p(\cdot))'$ is a p-dimensional vector of unknown smooth functions. The error term ε is assumed to be independent of $(X', Z')'$ and satisfies $\mathrm{E}(\varepsilon|U, X, Z) = 0$ and $\mathrm{Var}(\varepsilon|U, X, Z) = \sigma^2(U) > 0$. Model (1.1) includes many commonly used parametric, semiparametric and nonparametric models as its special cases. For instances, a constant vector of g corresponds to the classical linear regression model; $\beta = 0$ leads to the varying coefficient regression model; when $p = 1$, $X = 1$, (1.1) reduces to the partially linear regression model; and when X is a vector of 1's, $\beta = 0$, (1.1) becomes the well known additive model. Extensive research on these important statistical models can be found in literature.

In practice, however, often times the covariates Z cannot be observed directly, instead, one can only observe their surrogates, say W. In the measurement error literature, W and Z are often related in an additive way, that is, $W = Z + \eta$, where the q-dimensional random vector η represents the measurement error. It is well known that simply replacing the true covariates with their surrogates in some classical regression models often leads to biased estimation and inefficient testing procedures. A detailed presentation of the EV modeling can be found in Fuller (1987), Cheng and Van Ness (1999), Carroll, Ruppert, Stefanski, and Crainiceanu (2006). One can expect that same phenomenon could happen if we neglect the measurement error when doing statistical inferences in model (1.1). For some special cases of model (1.1) with measurement errors, such as the partially linear regression models, various bias-corrected estimation and testing procedures are already investigated. See Liang et al. (1999, 2000, 2007), Cui and Li (1998), Cui and Kong (2006), Wang (1999), among others, for more details. When $\sigma^2(u)$ is a constant, You and Chen (2006) proposed a modified profile least-squares approach to estimate the parametric component; Hu et al. (2009) and Wang et al. (2011) obtained a confidence region of the parametric component using the empirical likelihood method; when the parameter space is restricted, the estimation and testing procedures for the regression coefficients are considered in Zhang et al. (2011) and Wei (2012). To date, we haven't been aware of any research on model (1.1) with measurement errors when heteroscedasticity presents. This paper tries to fill this void by proposing a more efficient estimation procedure for the regression coefficients in model (1.1) via a combination of re-weighted profile least-squares method and the bias attenuation approach.

The rest of the paper is organized as follows. In Section 2, the bias-corrected profile least-squares estimates of regression and varying coefficients are briefly introduced without taking the heteroscedasticity into account, a residual vector is thus constructed which serves as the building blocks for a kernel estimate of the variance function in Section 3. together with a discussion on the uniform consistency and the asymptotic normality of the variance function estimate. The re-weighted profile least squares estimates of the regression coefficients and the improved estimate for the varying coefficients are proposed in Section 4, as well as the main results on their large sample properties. Simulation study is conducted in Section 5 to assess the finite sample performance of the proposed estimation procedures, and comparisons among different estimates of the regression coefficients, varying coefficients functions are also made there. All the proofs of the main theoretical results are relegated to the Appendix.

2. Bias-Corrected Profile Least Squares Estimation

When Z can be observed directly, we can apply the profile least-squares estimation procedure proposed by Fan and Huang (2005) to estimate the parametric component and the local polynomial smoothing to estimate the nonparametric component. To facilitate the understanding, the profile least-squares estimation procedure will be briefly introduced in this section. To be specific, let $(Y_i, X_i, Z_i, U_i),\ i = 1, 2, \ldots, n$ be a random sample from model (1.1)

$$Y_i = X_i' g(U_i) + Z_i'\beta + \varepsilon_i,\ \ i = 1, 2, \ldots, n. \tag{2.1}$$

For any given $\beta = (\beta_1,\ \beta_2,\ \ldots,\ \beta_q)'$, write

$$Y_i - \sum_{j=1}^{q} Z_{ij}\beta_j = \sum_{j=1}^{p} X_{ij} g_j(U_i) + \varepsilon_i,\ \ i = 1, 2, \ldots, n. \tag{2.2}$$

where X_{ij} and Z_{ij} are the jth elements of X_i and Z_i, respectively. Obviously, model (2.2) is transformed into the varying coefficient model. Now, we apply a local linear regression technique to estimate the functions $g_j(\cdot), j = 1, 2, \ldots, p$ in model (2.2). For u in a small neighborhood of u_0, one can approximate $g_j(\cdot)$ locally by a linear function

$$g_j(u) \approx g_j(u_0) + g_j'(u_0)(u - u_0),\ \ j = 1, 2, \ldots, p, \tag{2.3}$$

where $g_j'(u)$ is the first derivative of $g_j(u)$ with respect to u. Therefore, one can estimate β, $g(u_0)$, also $g'(u_0)$ by finding the minimizers of the locally weighted least-squares

$$\sum_{i=1}^{n} \Big[(Y_i - Z_i'\beta) - \sum_{j=1}^{p} \big(g_j(u_0) + g_j'(u_0)(U_i - u_0) \big) X_{ij} \Big]^2 K_h(U_i - u_0), \tag{2.4}$$

where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a preselected kernel function and h being the bandwidth whose optimal value can be determined by some data-driven methods such as the cross-validation methods. Clearly, routine weight least squares procedure can simultaneously produces the estimate for β, g and g'. We can also obtain the solution sequentially by first estimating g and g' by holding β as a constant vector, then estimate β by the profile least squares procedure. We shall adopt the latter to construct the estimate.

It is more convenient to use matrix form to represent model (2.1) and the solution of the least square problem (2.4). For this purpose, let

$$\mathbf{X} = \begin{pmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} Z_1' \\ Z_2' \\ \vdots \\ Z_n' \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1q} \\ Z_{21} & Z_{22} & \cdots & Z_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \cdots & Z_{nq} \end{pmatrix},$$

$$\mathbf{D_0} = \begin{pmatrix} X_1' & \frac{U_1-u_0}{h}X_1' \\ X_2' & \frac{U_2-u_0}{h}X_2' \\ \vdots & \vdots \\ X_n' & \frac{U_n-u_0}{h}X_n' \end{pmatrix}, \mathbf{M} = \begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{pmatrix} = \begin{pmatrix} X_1'g(U_1) \\ X_2'g(U_2) \\ \vdots \\ X_n'g(U_n) \end{pmatrix},$$

$\mathbf{Y} = (Y_1,\ Y_2,\ \ldots, Y_n)', \varepsilon = (\varepsilon_1,\ \varepsilon_2,\ \ldots,\ \varepsilon_n)', \mathbf{w_0} = \text{Diag}\,\{K_h(U_1-u_0), K_h(U_2-u_0), \ldots, K_h(U_n-u_0)\}$,

where $X_i = (X_{i1},\ X_{i2},\ \ldots,\ X_{ip})'$ and $Z_i = (Z_{i1},\ Z_{i2},\ \ldots,\ Z_{iq})'$ $(i = 1, 2, \ldots, n)$ are the vectors consisting of the observations of the explanatory variables $X_1,\ X_2, \ldots,\ X_n$ and $Z_1,\ Z_2,\ \ldots,\ Z_n$, respectively. Then model (2.2) can be rewritten as

$$\mathbf{Y} - \mathbf{Z}\beta = \mathbf{M} + \varepsilon, \tag{2.5}$$

Solving the weighted least-squares question in (2.4), we can obtain the local linear estimate of the varying coefficient vector $g(u) = (g_1(u), g_2(u), \ldots, g_p(u))'$ at $u = u_0$ given by

$$\breve{g}(u_0) = (\breve{g}_1(u_0), \breve{g}_2(u_0), \ldots, \breve{g}_p(u_0))' = (\mathbf{I}_p, \mathbf{0}_p)\{\mathbf{D_0'w_0D_0}\}^{-1}\mathbf{D_0'w_0}(\mathbf{Y} - \mathbf{Z}\beta), \tag{2.6}$$

where $\mathbf{I}_p$ and $\mathbf{0}_p$ are the $p \times p$ identity and null matrices. Then the initial estimate of the vector M has the form of

$$\breve{\mathbf{M}} = \begin{pmatrix} (X_1' \;\; \mathbf{0}_{1\times p})\{\mathbf{D_1'w_1D_1}\}^{-\mathbf{1}}\mathbf{D_1'w_1} \\ (X_2' \;\; \mathbf{0}_{1\times p})\{\mathbf{D_2'w_2D_2}\}^{-\mathbf{1}}\mathbf{D_2'w_2} \\ \vdots \\ (X_n' \;\; \mathbf{0}_{1\times p})\{\mathbf{D_n'w_nD_n}\}^{-\mathbf{1}}\mathbf{D_n'w_n} \end{pmatrix}(\mathbf{Y} - \mathbf{Z}\beta) = \mathbf{S}(\mathbf{Y} - \mathbf{Z}\beta), \tag{2.7}$$

where $\mathbf{D_i}$ and $\mathbf{w_i}$ are obtained by replacing u_0 in $\mathbf{D_0}$ and $\mathbf{w_0}$ with U_i $(i = 1, 2, \ldots, n)$. $\mathbf{0}_{1\times p}$ is the $1\times p$ matrix whose entities are all zeros, the matrix $\mathbf{S}$ only depends on the observations of the explanatory variables $X_1, X_2, \ldots, X_n$ and U. Substituting $\mathbf{M}$ for $\breve{\mathbf{M}}$ in model (2.5), we obtain

$$(\mathbf{I} - \mathbf{S})\mathbf{Y} = (\mathbf{I} - \mathbf{S})\mathbf{Z}\beta + \varepsilon. \tag{2.8}$$

Applying the least-squares method to fit the model (2.8), we obtain the profile least-squares estimate of the constant coefficient vector $\beta = (\beta_1, \beta_2, \ldots, \beta_q)'$ as

$$\breve{\beta} = (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-\mathbf{1}}\tilde{\mathbf{Z}}'\tilde{\mathbf{Y}}, \tag{2.9}$$

where $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S})\mathbf{Z},\ \tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S})\mathbf{Y}$.

However, $\mathbf{Z}$ cannot be exactly observed in the current set up. Just ignoring the measurement error and replacing $\mathbf{Z}$ with $\mathbf{W}$ in (2.9) leads to an inconsistent estimate. To remove the bias caused by the measurement error, You and Chen (2006) introduced the following modified profile least-squares estimator of β,

$$\tilde{\beta} = \arg\min_{\beta\in\mathbf{R}}[(\tilde{\mathbf{Y}} - \tilde{\mathbf{W}}\beta)'(\tilde{\mathbf{Y}} - \tilde{\mathbf{W}}\beta) - n\beta'\Sigma_\eta\beta] = (\tilde{\mathbf{W}}'\tilde{\mathbf{W}} - n\Sigma_\eta)^{-1}\tilde{\mathbf{W}}'\tilde{\mathbf{Y}}, \tag{2.10}$$

where $\tilde{\mathbf{W}} = (\mathbf{I} - \mathbf{S})\mathbf{W} = (\ \tilde{W}_1',\ \tilde{W}_2',\ \ldots,\ \tilde{W}_n'\ )'$, and $\mathbf{W} = (\ W_1',\ W_2',\ \ldots,\ W_n'\ )'$. Moreover, the fact that $E(Y_i - Z_i'\beta|U_i) = E(Y_i - W_i'\beta|U_i)$ and (2.6) suggest that the estimate of $g(u)$ at $u = u_0$ can be defined as

$$\tilde{g}(u_0) = (\mathbf{I}_p, \mathbf{0}_p)\{\mathbf{D_0'w_0D_0}\}^{-1}\mathbf{D_0'w_0}(\mathbf{Y} - \mathbf{W}\tilde{\beta}). \tag{2.11}$$

3. Estimation of the error variance function

In this section, we consider the estimation of the variance function in model (1.1) with measurement errors. From (2.7), $\mathbf{M}$ can be estimated by

$$\hat{\mathbf{M}} = \mathbf{S}(\mathbf{Y} - \mathbf{W}\tilde{\beta}), \tag{3.1}$$

and (2.5) implies that the residual $e_i = Y_i - W_i^T\beta - M_i = \varepsilon_i - \eta_i^T\beta$ can be estimated by

$$\hat{e}_i = Y_i - W_i^T\tilde{\beta} - \hat{M}_i. \tag{3.2}$$

Note that $\sigma_e^2(u) = E(e^2|U = u) = \sigma^2(u) + \beta'\Sigma_\eta\beta$, so for a given point $u_0 \in \Omega$, the variance function $\sigma^2(u)$ at u_0 can be estimated by

$$\hat{\sigma}^2(u_0) = \hat{\sigma}_e^2(u_0) - \tilde{\beta}'\Sigma_\eta\tilde{\beta}, \tag{3.3}$$

where

$$\hat{\sigma}_e^2(u_0) = \frac{\sum_{j=1}^{n} \hat{e}_j^2 K_{\tilde{h}}(U_j - u_0)}{\sum_{j=1}^{n} K_{\tilde{h}}(U_j - u_0)}, \tag{3.4}$$

where $\hat{e}_j$ $(j = 1, 2, \ldots, n)$ are the residuals of the modified profile least-squares fit defined in (3.2), $K_{\tilde{h}}(\cdot)$ has the same form as $K_h(\cdot)$ in (2.4) except that the bandwidth h in $K_h(\cdot)$ is changed to a possibly different bandwidth $\tilde{h}$.

Under some regularity conditions stated in the Appendix, we have obtained the uniform convergence rate of $\hat{\sigma}^2(u_0)$, as well as its asymptotic normality. These results are summarized in the following two theorems, and their proofs are postponed to the Appendix.

Theorem 1. Under the conditions (i)-(viii) presented in the Appendix, we have

$$\sup_{u_0 \in \Omega} |\hat{\sigma}^2(u_0) - \sigma^2(u_0)| = O_p\left(\tilde{h}^2 + \left(\frac{\log(1/\tilde{h})}{n\tilde{h}}\right)^{\frac{1}{2}}\right).$$

Theorem 2. Under the conditions (i)-(viii) presented in the Appendix, as $n \to \infty$, we have

$$\sqrt{n\tilde{h}}\left(\hat{\sigma}^2(u_0) - \sigma^2(u_0) - \frac{\tilde{h}^2}{2}\mu_2\sigma^{2''}(u_0) - \frac{f'(u_0)}{f(u_0)}\sigma^{2'}(u_0)\mu_2\tilde{h}^2\right) \xrightarrow{D} N(0,\ \nu_0 f^{-1}(u_0)\tau(u_0)),$$

where $\sigma^{2'}(u_0)$ and $\sigma^{2''}(u_0)$ are the first and second order derivatives of the variance function $\sigma^2(u)$ at $u = u_0$, respectively, and $\tau(u_0) = E\Big[\big((\varepsilon - \eta'\beta)^2 - \sigma^2(u_0) - \beta'\Sigma_\eta\beta\big)^2|U = u_0\Big]$.

4. Re-weighted Estimation with Heteroscedasticity

Similar to the general linear regression, more efficient estimates of the regression parameters can be constructed via a re-weighted least square procedure if the variance functions can be estimated. Note that the bias-corrected estimate of β with g being known is the minimizer of $(Y - M - W\beta)'(Y - M - W\beta) - n\beta'\Sigma_\eta\beta$, so the appropriate weight for the i-th observation should be the reciprocal of the conditional variance of $Y_i - M_i - W_i^T\beta$ given U_i, which is simply $\sigma_e^2(U_i)$, $i = 1, 2, \ldots, n$. Taking $u_0 = U_i$ $(i = 1, 2, \ldots, n)$, the estimates of the variance function $\sigma_e^2(\cdot)$ at all designed points U_i $(i = 1, 2, \ldots, n)$ can be defined by (3.4). Similar to the general least-squares approach in classical linear models, the re-weighted estimate of the regression coefficient $\beta = (\beta_1,\ \beta_2,\ \ldots,\ \beta_q)'$ can be obtained by minimizing

$$\sum_{i=1}^{n} \left[(\tilde{Y}_i - \tilde{W}_i^T\beta)^2 - \beta^T\Sigma_\eta\beta \right] \hat{\sigma}_e^{-2}(U_i)$$

with respect to β. In matrix form, we have

$$\hat{\beta} = \left(\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-\mathbf{1}}\tilde{\mathbf{W}} - \mathrm{tr}(\hat{\mathbf{\Sigma}}^{-\mathbf{1}})\Sigma_\eta \right)^{-1} \tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-\mathbf{1}}\tilde{\mathbf{Y}}, \tag{4.1}$$

where $\hat{\mathbf{\Sigma}} = \mathrm{Diag}(\hat{\sigma}_e^2(U_1),\ \hat{\sigma}_e^2(U_2),\ \ldots,\ \hat{\sigma}_e^2(U_n))$. Then following (2.11), an improved estimate of unknown smooth coefficient functions vector $g(u) = (g_1(u),\ g_2(u),\ \ldots,\ g_p(u))'$ at $u = u_0$ can be constructed as

$$\hat{g}(u_0) = (\mathbf{I}_p, \mathbf{0}_p)\{\mathbf{D_0'w_0D_0}\}^{-1}\mathbf{D_0'w_0}(\mathbf{Y} - \mathbf{W}\hat{\beta}). \tag{4.2}$$

In the following, we present the asymptotic distributions of the re-weighted bias-corrected estimates of β, the bias-corrected profile least-squares estimate $\tilde{g}(u_0)$ and the re-weighted bias-corrected estimate $\hat{g}(u_0)$ of the varying coefficient function $g(u) = (g_1(u),\ g_2(u),\ \ldots,\ g_p(u))'$ at $u = u_0$.

The following lemma concerns about the asymptotic normality of $\tilde{\beta}$, the profile least square estimate of β without taking the heteroscedasticity into account. The proof can be found in You and Chen (2006).

Lemma 4.1. Under the conditions (i)-(vii) presented in the Appendix, the profile least-squares estimator $\tilde{\beta}$ is asymptotically normal, namely,

$$\sqrt{n}(\tilde{\beta} - \beta) \ \xrightarrow{D}\ N(0,\ \mathbf{\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}}), \quad n \to \infty,$$

where

$$\mathbf{\Sigma_1} = E(Z_1Z_1') - E(\Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1))$$

with $\Phi(U_1) = E(X_1Z_1'|U_1)$, $\Gamma(U_1) = E(X_1X_1'|U_1)$, and

$$\mathbf{\Sigma_2} = E\big[(\sigma^2(U_1) + \beta'\Sigma_\eta\beta)(Z_1Z_1' - \Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1))\big] + E(\sigma^2(U_1))\Sigma_\eta + E(\eta_1\eta_1'\beta\beta'\eta_1\eta_1') - \Sigma_\eta\beta\beta'\Sigma_\eta$$

$$+ E\big[(Z_1 - \Phi'(U_1)\Gamma^{-1}(U_1)X_1)(\eta_1'\beta'\eta_1\eta_1'\beta)\big] + E\big[(\eta_1\beta'\eta_1\eta_1'\beta)(Z_1' - X_1'\Gamma^{-1}(U_1)\Phi(U_1))\big].$$

Remark 1. If the measurement error η has a probability density symmetric around 0, $\mathbf{\Sigma_2}$ in Lemma 4.1 becomes

$$E\big[(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)(Z_1Z_1'-\Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1))\big]+E(\sigma^2(U_1))\Sigma_\eta+E(\eta_1\eta_1'\beta\beta'\eta_1\eta_1')-\Sigma_\eta\beta\beta'\Sigma_\eta.$$

Theorem 3. Under the conditions (i)-(vii) presented in the Appendix, for the re-weighted bias-corrected estimate $\hat{\beta}$ of the constant coefficient vector, we have

$$\sqrt{n}(\hat{\beta}-\beta)\ \xrightarrow{D}\ N(0,\ \mathbf{\Sigma_4^{-1}\Sigma_3\Sigma_4^{-1}}),\quad n\to\infty,$$

where

$$\begin{aligned}\mathbf{\Sigma_3}=&E\big[(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-1}(Z_1Z_1'-\Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1))\big]+E(\sigma^2(U_1)(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2})\Sigma_\eta\\&+E((\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2}\eta_1\eta_1'\beta\beta'\eta_1\eta_1')-E((\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2})\Sigma_\eta\beta\beta'\Sigma_\eta\\&+E\big[(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2}(Z_1-\Phi'(U_1)\Gamma^{-1}(U_1)X_1)(\eta_1'\beta'\eta_1\eta_1'\beta)\big]\\&+E\big[(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2}(\eta_1\beta'\eta_1\eta_1'\beta)(Z_1'-X_1'\Gamma^{-1}(U_1)\Phi(U_1))\big],\end{aligned}$$

and

$$\mathbf{\Sigma_4}=E(Z_1Z_1'(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-1})-E((\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-1}\Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1)).$$

Remark 2. If the measurement error η has a probability density symmetric around 0, $\mathbf{\Sigma_3}$ in Theorem 3 becomes

$$\begin{aligned}&E\big[(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-1}(Z_1Z_1'-\Phi'(U_1)\Gamma^{-1}(U_1)\Phi(U_1))\big]+E(\sigma^2(U_1)(\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2})\Sigma_\eta\\&+E((\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2}\eta_1\eta_1'\beta\beta'\eta_1\eta_1')-E((\sigma^2(U_1)+\beta'\Sigma_\eta\beta)^{-2})\Sigma_\eta\beta\beta'\Sigma_\eta\end{aligned}$$

The asymptotic normality of $\tilde{g}(u_0)$ and $\hat{g}(u_0)$ are summarized in the following theorems.

Theorem 4. Under the conditions (i)-(vii) presented in the Appendix, for any given $u_0\in\Omega$, the bias-corrected estimate $\tilde{g}(u_0)$ satisfies

$$\sqrt{nh}(\tilde{g}(u_0)-g(u_0)-\frac{h^2}{2}\mu_2 g''(u_0))\ \xrightarrow{D}\ N(0,\ \Sigma(u_0)),$$

where $\Sigma(u_0)=\nu_0 f^{-1}(u_0)\Gamma^{-1}(u_0)(\sigma^2(u_0)+\beta'\Sigma_\eta\beta)$.

Theorem 5. Under the conditions (i)-(vii) presented in the Appendix, for any given $u_0\in\Omega$, the re-weighted bias-corrected estimate $\hat{g}(u_0)$ satisfies

$$\sqrt{nh}(\hat{g}(u_0)-g(u_0)-\frac{h^2}{2}\mu_2 g''(u_0))\ \xrightarrow{D}\ N(0,\ \Sigma(u_0)),$$

where $\Sigma(u_0)$ is the same as the one defined in Theorem 4.

The results from Theorem 4 with Theorem 5 indicate that the bias-corrected profile least-squares and re-weighted bias-corrected estimates of the varying coefficient vector share the same asymptotic distributions. This implies that the improvement on estimating β exerts no effects on the asymptotic distribution of the estimate for the varying coefficient function.

5. Simulation experiments

To evaluate the finite sample performance of the proposed estimation procedure, a simulation study is conducted in section. The simulated data are generated from the following model:

$$\begin{aligned} &Y_i = X_{i1}g_1(U_i) + X_{i2}g_2(U_i) + Z_{i1}\beta_1 + Z_{i2}\beta_2 + \varepsilon_i, \\ &W_{i1} = Z_{i1} + \eta_{i1},\ W_{i2} = Z_{i2} + \eta_{i2},\ i = 1, 2, \ldots, n, \end{aligned} \tag{5.1}$$

where $g_1(u) = 2u,\ g_2(u) = 1 + \sin(2\pi u),\ \beta_1 = 2,\ \beta_2 = 3$, and the variance function of the error term is taken as $\sigma^2(u) = 0.5 + (c\sin(2\pi u))^2$, c is a constant. In model (5.1), the observation X_{i1}, X_{i2} and U are independently generated from the uniform distribution $U(0,\ 1)$, and $Z_{i1},\ Z_{i2}$ are independently drawn from the normal distribution $N(2, 1)$ and $N(3, 1)$, respectively. The measurement errors $\eta_{i1},\ \eta_{i2}$ are independently generated from the normal distribution $N(0,\ \sigma_\eta^2)$. To gain an insight of the effect of magnitude of the variance function and the measurement error, the distribution of the error on the performance of the proposed procedure, we choose $c = 1, 3, 5$, $\sigma_\eta = 0.1, 0.3, 0.5$, and the following three distributions will be selected as the error distribution: (1) $\varepsilon_i \sim N(0,\ \sigma^2(U_i))$, (2) $\varepsilon_i \sim U(-\sqrt{3}\sigma(U_i),\ \sqrt{3}\sigma(U_i))$, (3) $\varepsilon_i + \sigma(U_i) \sim \text{Exp}(1/\sigma(U_i)),\ i = 1, 2, \ldots, n$.

The Gauss kernel $K(x) = 1/\sqrt{2\pi}\exp(-x^2/2)$ is adopted in estimating both the varying coefficient function an the variance function. The bandwidth $h = an^{-1/5}$, and $\tilde{h} = bn^{-1/5}$ with $a = 0.3, b = 0.4$ are used to investigate the effect of bandwidth selection on the the estimation procedure. Three estimates of the regression coefficient β will be compared in the simulation study: the naive estimates by ignoring the measurement errors or setting $\Sigma_\eta = 0$ in the bias-corrected estimation procedure; the bias-corrected estimates, and the re-weighted bias-corrected estimates. For the sake of convenience, $\tilde{\beta}_N$, $\hat{\beta}_N$, $\hat{\sigma}_N^2(u)$, $\hat{g}_N(u)$ denote the naive estimates of β, $\sigma^2(u)$ and $g(u)$ corresponding to (2.10), (4.1), (3.3) and (4.2), respectively, by letting $\Sigma_\eta = 0$. Several sample sizes such as $n = 200, 300$ are used in the simulation. For each setup, the simulation is repeated 1000 times.

The simulation results for the regression parameter estimates are presented in Table 1 through 3 for $c = 1, 3, 5$ and $n = 200, 300$. For any fixed c values, the performance of the re-weighted bias-corrected estimates generally performs better than the bias-corrected estimates, which in turn outperforms the naive estimates. The MSEs increase as σ_η increases, and decrease as n becomes larger, as it should be. It is also clear that when c gets larger, the MSEs increase, while the superiority of the re-weighted estimation procedure over the other two estimation procedures become clearer.

To visualize the estimates of variance function and the varying coefficient functions, $\hat{\sigma}^2(u)$ and $\hat{\sigma}_N(u)$ are calculated at $n = 300$ equally space values $u_i = i/n \in [0, 1]$. The average of 500 estimates at each u_i is used as the final estimate of the variance function and the varying coefficient functions, respectively, $i = 1, 2, \ldots, n$. Figure 1 presents the plots of the naive estimates and the re-weighted bias-corrected estimates of $\sigma^2(u)$ versus the u at $u = i/n, i = 1, 2, \ldots, n$ for three different error types. In all subsequent figures, the solid curve denotes the true variance function and the varying coefficient functions, and the dashed and dash-dotted curves represent their re-weighted bias-corrected estimates and the naive estimates, respectively. Clearly, the re-weighted bias-corrected estimates perform very well in all scenarios, while this is not the case for naive estimate, which are systematically deviated from the truth. This is not beyond of our expectation in that such a phenomenon is very common in measurement error modeling. A striking feature found in Figure 1 is that given the same amount of measurement error, increasing the magnitude of the variance function might be able to alleviate the measurement error effects. Figure 2 shows the plots of the estimated varying coefficient functions. It can be see from Figure 2 that the proposed re-weighted bias-corrected estimates work satisfactorily in all cases, while the naive one leads to biased estimates. Since the patterns are all similar in all the cases considered, only the results of $n = 300,\ \sigma_\eta = 0.3$ are reported here for the sake of brevity.

We also conduct the simulation study for other choices of a and b when selecting the bandwidths. Our finding is that the estimates for β are pretty robust to the bandwidth selection, but it is not the case for estimating the nonparametric components. As in other nonparametric smoothing literature, the performance of nonparametric function estimate is very sensitive to the choices of bandwidth. In the real application, we will certainly recommend to use some data driven bandwidth selection procedures, such as various cross validation schemes. More investigation on the bandwidth selection in the current set up, in particular for computationally efficient procedures, will be pursued in the future research.

Appendix

We begin with a list of mild technical conditions required for deriving the main theoretical results.

(i). The random variable U has a compact support Ω, and its density function $f(u)$ has continuous second order derivative and bounded away from 0 on its support.

(ii). The $p \times p$ matrix $E(XX'|U = u)$ is nonsingular for each $u \in \Omega$, $E(XX'|U = u), E(ZZ'|U = u)$, $E(XZ'|U = u)$ and $E(X|U = u)$ are all Lipschitz continuous.

(iii). There exists an $s > 2$ such that $E||X||^{2s} < \infty, E||Z||^{2s} < \infty, E||\varepsilon||^{2s} < \infty$, and $E||\eta||^{2s} < \infty$. and for some $\delta < 2 - s^{-1}$ there is $n^{2\delta - 1}h \to \infty$ as $n \to \infty$.

(iv). All of the coefficient functions $g_j(\cdot)$ $(j = 1, 2, \ldots, p)$ are Lipschitz continuous and have continu-

β	Error	n	σ_η	$\tilde{\beta}_N$		$\tilde{\beta}$		$\hat{\beta}$	
				Mean	MSE	Mean	MSE	Mean	MSE
$\beta_1 = 2$	$N(0, \sigma^2(u))$	200	0.1	1.9949	0.0051	2.0020	0.0050	2.0025	0.0047
			0.3	1.9409	0.0124	1.9954	0.0100	1.9986	0.0099
			0.5	1.8375	0.0408	2.0056	0.0214	2.0132	0.0226
		300	0.1	1.9932	0.0033	2.0022	0.0030	2.0028	0.0028
			0.3	1.9338	0.0099	2.0061	0.0073	2.0080	0.0072
			0.5	1.8431	0.0337	2.0008	0.0148	2.0057	0.0154
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	1.9899	0.0057	1.9971	0.0047	1.9956	0.0042
			0.3	1.9408	0.0124	1.9975	0.0107	1.9990	0.0106
			0.5	1.8457	0.0385	2.0148	0.0223	2.0204	0.0228
		300	0.1	1.9919	0.0033	2.0018	0.0031	1.9991	0.0028
			0.3	1.9366	0.0097	2.0002	0.0063	2.0010	0.0062
			0.5	1.8391	0.0350	2.0053	0.0151	2.0106	0.0158
$\beta_2 = 3$	$N(0, \sigma^2(u))$	200	0.1	2.9894	0.0035	3.0006	0.0034	3.0009	0.0033
			0.3	2.9069	0.0144	3.0078	0.0068	3.0114	0.0069
			0.5	2.7604	0.0670	3.0144	0.0152	3.0253	0.0158
		300	0.1	2.9881	0.0024	2.9991	0.0023	2.9994	0.0021
			0.3	2.9096	0.0118	3.0022	0.0048	3.0051	0.0047
			0.5	2.7541	0.0669	3.0053	0.0105	3.0133	0.0110
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	2.9904	0.0038	2.9996	0.0033	2.9963	0.0028
			0.3	2.9112	0.0137	3.0124	0.0071	3.0136	0.0070
			0.5	2.7610	0.0672	3.0056	0.0147	3.0148	0.0157
		300	0.1	2.9898	0.0023	2.9987	0.0021	2.9964	0.0019
			0.3	2.9116	0.0118	3.0041	0.0044	3.0055	0.0042
			0.5	2.7571	0.0655	3.0111	0.0100	3.0174	0.0106

Table 1: Means and MSEs of $\tilde{\beta}_N, \tilde{\beta}, \hat{\beta}$ $(c = 1)$

.

β	Error	n	σ_η	$\tilde{\beta}_N$		$\tilde{\beta}$		$\hat{\beta}$	
				Mean	MSE	Mean	MSE	Mean	MSE
$\beta_1 = 2$	$N(0, \sigma^2(u))$	200	0.1	1.9985	0.0241	2.0001	0.0231	2.0008	0.0170
			0.3	1.9431	0.0297	2.0030	0.0286	2.0056	0.0242
			0.5	1.8387	0.0561	1.9972	0.0431	2.0065	0.0406
		300	0.1	1.9923	0.0163	1.9999	0.0143	1.9996	0.0099
			0.3	1.9307	0.0209	2.0040	0.0188	2.0055	0.0151
			0.5	1.8460	0.0421	2.0118	0.0285	2.0149	0.0261
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	1.9879	0.0283	1.9988	0.0217	1.9947	0.0152
			0.3	1.9440	0.0307	2.0002	0.0303	1.9967	0.0242
			0.5	1.8447	0.0564	2.0123	0.0450	2.0157	0.0423
		300	0.1	1.9915	0.0162	2.0025	0.0160	1.9989	0.0103
			0.3	1.9354	0.0225	1.9998	0.0180	1.9996	0.0140
			0.5	1.8395	0.0455	2.0076	0.0294	2.0121	0.0260
$\beta_2 = 3$	$N(0, \sigma^2(u))$	200	0.1	2.9883	0.0164	2.9997	0.0146	3.0019	0.0109
			0.3	2.9057	0.0273	3.0102	0.0184	3.0132	0.0151
			0.5	2.7594	0.0780	3.0161	0.0298	3.0248	0.0284
		300	0.1	2.9868	0.0120	2.9977	0.0100	2.9981	0.0067
			0.3	2.9101	0.0193	3.0025	0.0126	3.0051	0.0102
			0.5	2.7523	0.0751	3.0019	0.0195	3.0102	0.0180
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	2.9906	0.0189	3.0044	0.0161	2.9977	0.0109
			0.3	2.9126	0.0256	3.0058	0.0182	3.0047	0.0147
			0.5	2.7655	0.0775	3.0071	0.0299	3.0128	0.0273
		300	0.1	2.9896	0.0112	3.0051	0.0093	2.9978	0.0062
			0.3	2.9139	0.0203	3.0041	0.0123	3.0015	0.0095
			0.5	2.7568	0.0734	3.0056	0.0171	3.0092	0.0161

Table 2: Means and MSEs of $\tilde{\beta}_N, \tilde{\beta}, \hat{\beta}$ ($c = 3$)

.

β	Error	n	σ_η	$\tilde{\beta}_N$		$\tilde{\beta}$		$\hat{\beta}$	
				Mean	MSE	Mean	MSE	Mean	MSE
$\beta_1 = 2$	$N(0, \sigma^2(u))$	200	0.1	2.0025	0.0631	2.0068	0.0592	2.0048	0.0394
			0.3	1.9460	0.0660	2.0083	0.0649	2.0103	0.0471
			0.5	1.8400	0.0869	2.0103	0.0885	2.0199	0.0725
		300	0.1	1.9913	0.0425	2.0018	0.0361	1.9981	0.0213
			0.3	1.9265	0.0433	2.0063	0.0419	2.0069	0.0282
			0.5	1.8491	0.0614	1.9972	0.0580	2.0026	0.0441
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	1.9856	0.0742	2.0019	0.0523	1.9951	0.0316
			0.3	1.9468	0.0692	2.0052	0.0646	2.0016	0.0446
			0.5	1.8439	0.0927	2.0152	0.0854	2.0160	0.0648
		300	0.1	1.9911	0.0427	1.9993	0.0382	1.9942	0.0221
			0.3	1.9338	0.0493	2.0006	0.0408	1.9994	0.0265
			0.5	1.8408	0.0672	2.0074	0.0545	2.0090	0.0406
$\beta_2 = 3$	$N(0, \sigma^2(u))$	200	0.1	2.9870	0.0424	3.0022	0.0412	3.0026	0.0279
			0.3	2.9037	0.0542	2.9955	0.0425	3.0015	0.0301
			0.5	2.7585	0.1010	3.0235	0.0586	3.0313	0.0485
		300	0.1	2.9854	0.0315	3.0001	0.0255	3.0032	0.0153
			0.3	2.9110	0.0351	3.0067	0.0267	3.0064	0.0179
			0.5	2.7495	0.0918	3.0073	0.0362	3.0148	0.0290
	$\sigma(u) \cdot (\text{Exp}(1) - 1)$	200	0.1	2.9908	0.0495	2.9985	0.0377	2.9872	0.0228
			0.3	2.9138	0.0514	3.0129	0.0444	3.0033	0.0300
			0.5	2.7696	0.1012	3.0112	0.0590	3.0131	0.0452
		300	0.1	2.9893	0.0293	2.9990	0.0270	2.9929	0.0157
			0.3	2.9165	0.0381	3.0040	0.0298	2.9984	0.0188
			0.5	2.7560	0.0901	3.0096	0.0366	3.0094	0.0276

Table 3: Means and MSEs of $\tilde{\beta}_N, \tilde{\beta}, \hat{\beta}$ ($c = 5$)

.

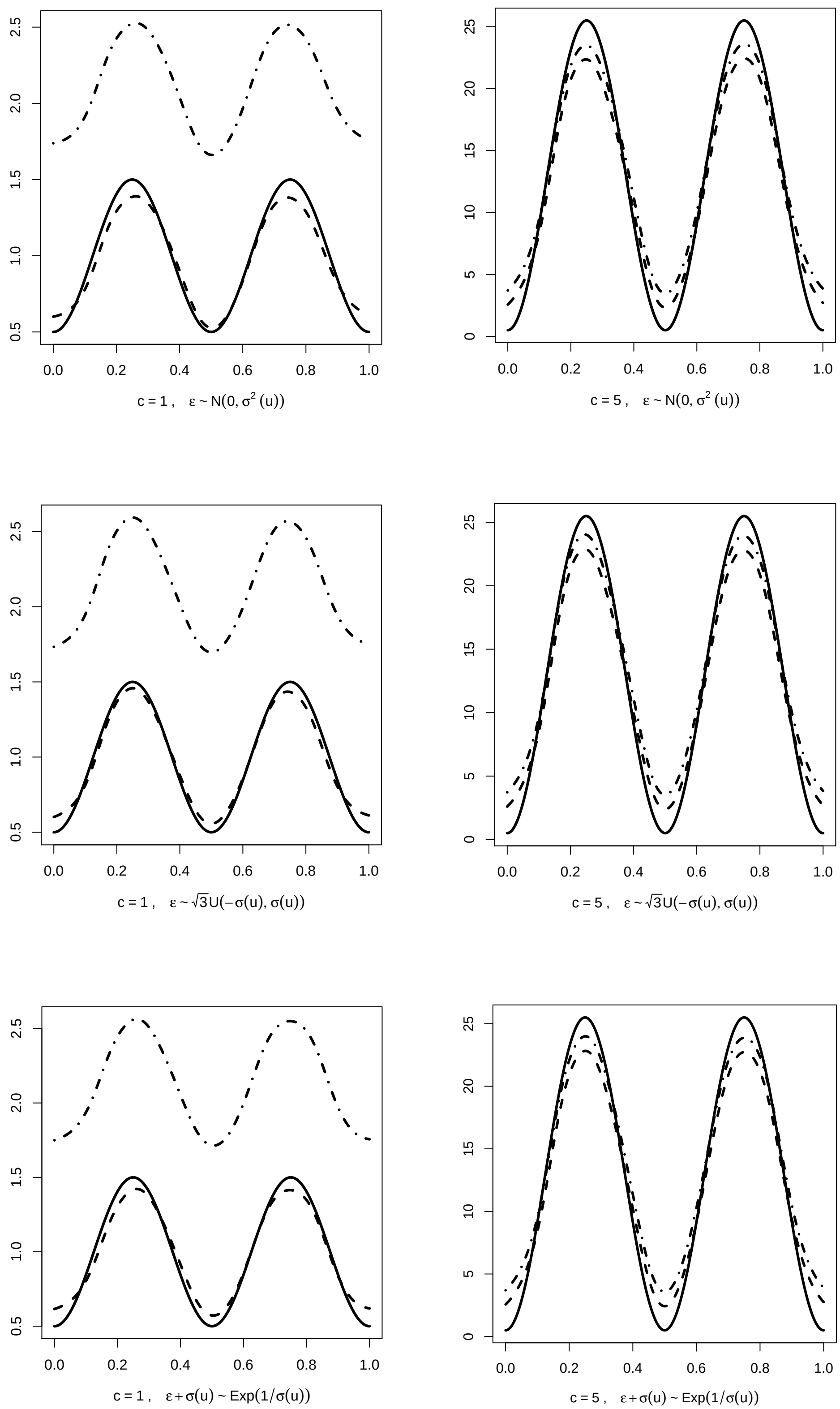

Figure 1. Estimates of the variance functions

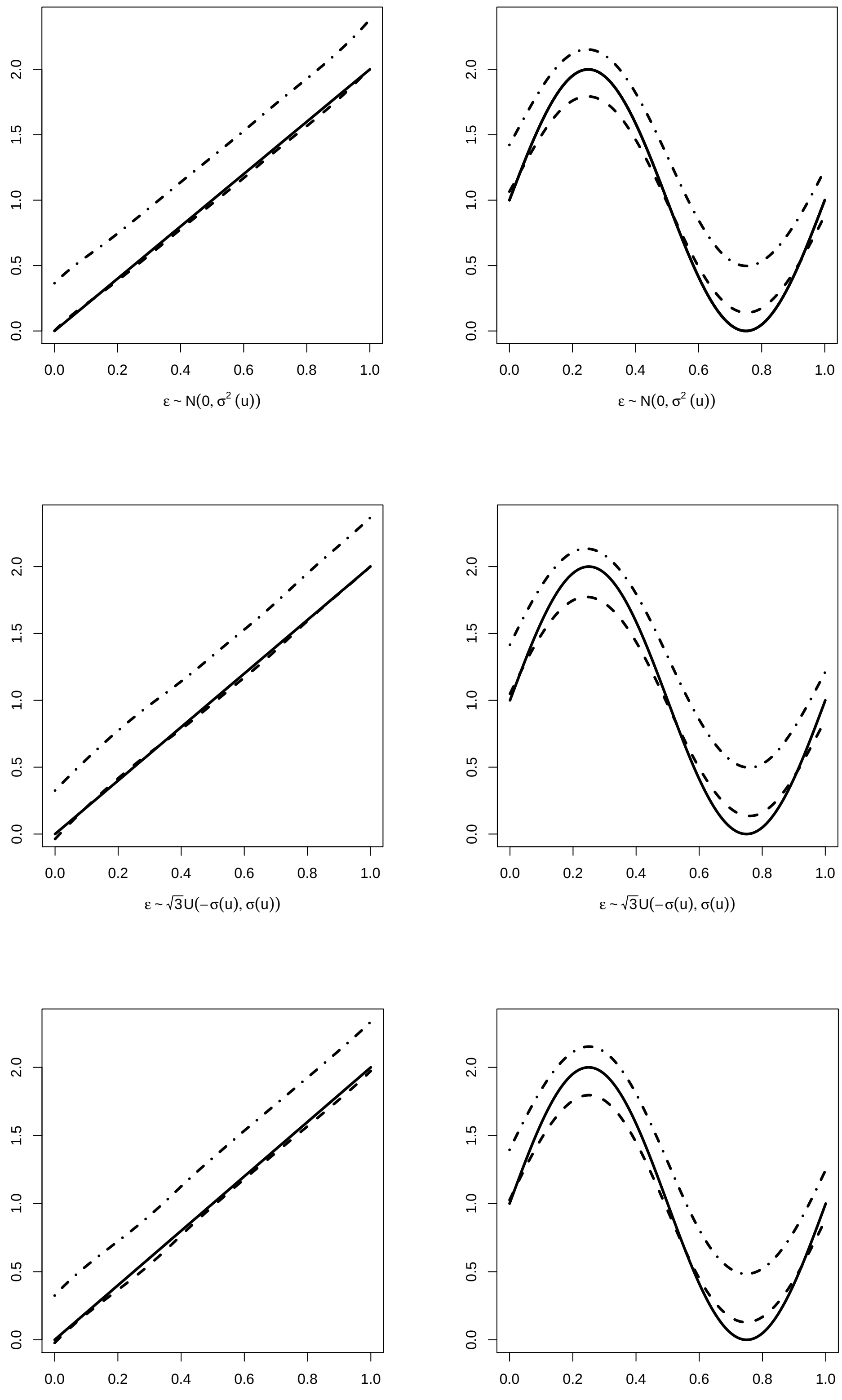

Figure 2. Estimates of the varying coefficient functions, $c = 1$

ous second order derivatives.

(v). The kernel function $K(\cdot)$ is a symmetric probability density function with a compact support. Also, it is Lipschitz continuous.

(vi). The variance function $\sigma^2(\cdot) > 0$ has continuous second order derivative and is uniformly bounded on its domain.

(vii). The sample size n and the smoothing parameter h are assumed to satisfy $\lim_{n\to\infty} h = 0,\ \lim_{n\to\infty} nh^2 = \infty,\ \lim_{n\to\infty} nh^8 = 0$ and $\lim_{n\to\infty} [\log(1/h)]^2/(nh^2) = 0$.

(viii). For the bandwidth $\tilde{h}$, we assume $\lim_{n\to\infty} \tilde{h} = 0,\ \lim_{n\to\infty} n\tilde{h}^2 = \infty,\ \lim_{n\to\infty} n\tilde{h}^8 = 0,\ \lim_{n\to\infty} \frac{[\log(1/\tilde{h})]^2}{n\tilde{h}^2} = 0$, and h and $\tilde{h}$ satisfy

$$O\left[h^2 + \left(\frac{\log(1/h)}{nh}\right)^{1/2}\right] O\left[\tilde{h}^2 + \left(\frac{\log(1/\tilde{h})}{n\tilde{h}}\right)^{1/2}\right] = o(n^{-1/2}).$$

Condition (i) guarantees the kernel estimate of the density function of U stays away from 0 in order to stabilize the nonparametric estimates of the varying coefficient and variance function; Conditions (ii), (iii) and (iv) are the typical assumptions in the varying coefficient regression modeling; Conditions (v), (vi) are very common in nonparametric smoothing; and Conditions (vii), (viii) are required to ensure proper convergence rates needed for deriving large sample properties for the proposed estimates. These conditions are also assumed in You and Chen (2006) and Shen et al. (2014).

The following lemma is needed to facilitate the proof of the main theorems. Let $\mu_i = \int u^i K(u)du,\ \nu_i = \int u^i K^2(u)du$ and $c_n = \left(\frac{\log(1/h)}{nh}\right)^{1/2} + h^2,\ \tilde{c}_n = \left(\frac{\log(1/\tilde{h})}{n\tilde{h}}\right)^{1/2} + \tilde{h}^2$.

Lemma 1. Under the conditions (i)-(vii), as $n \to \infty$, it holds

$$\sup_{u_0\in\Omega} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{U_i - u_0}{h}\right) \left(\frac{U_i - u_0}{h}\right)^k X_{ij_1} X_{ij_2} = f(u_0)\Gamma_{j_1 j_2}(u_0)\mu_k + O(c_n)\ a.s.$$

and

$$\sup_{u_0\in\Omega} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{U_i - u_0}{h}\right) \left(\frac{U_i - u_0}{h}\right)^k X_{ij_1} \varepsilon_i = O\left\{\left(\frac{\log(1/h)}{nh}\right)^{\frac{1}{2}}\right\}\ a.s., \qquad (A.1)$$

where $j,\ j_1,\ j_2 = 1, 2, \ldots, p,\ k = 0, 1, 2$ and $\Gamma_{j_1 j_2}(u_0)$ is the (j_1, j_2)-th element of $\Gamma(u_0)$.

The proof of Lemma 1 is similar to that of Lemma 2 in Xia and Li (1999), hence it is omitted here for the sake of brevity.

Lemma 2. Under the conditions (i)-(viii), we have

$$\sup_{u_0\in\Omega}\left|\sigma^2(u_0)-\sum_{i=1}^n\tilde{k}_i(u_0)\sigma^2(U_i)\right|=O_p(\tilde{c}_n), \tag{A.2}$$

where

$$\tilde{k}_i(u_0)=\frac{K_{\tilde{h}}(U_i-u_0)}{\sum\limits_{j=1}^n K_{\tilde{h}}(U_j-u_0)},\ \ i=1,2,\ldots,n.$$

This Lemma can be obtained in You, Chen and Zhou (2007).

Lemma 3. Under the conditions (i)-(v) and (vi), as $n\to\infty$, then

$$\frac{1}{n}\mathbf{D_0'w_0D_0}=\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\otimes f(u_0)\Gamma(u_0)\{1+O_p(c_n)\}, \tag{A.3}$$

$$\frac{1}{n}\mathbf{D_0'w_0W}=(1,0)'\otimes f(u_0)\Phi(u_0)\{1+O_p(c_n)\}, \tag{A.4}$$

and

$$\frac{1}{n}\mathbf{D_0'w_0M}=(1,0)'\otimes f(u_0)\Gamma(u_0)g(u_0)\{1+O_p(c_n)\}, \tag{A.5}$$

hold uniformly in $u_0\in\Omega$.

Proof. For any given $u_0\in\Omega$, according to the definitions of $\mathbf{D_0}$ and $\mathbf{w_0}$, we can calculate easily that

$$\frac{1}{n}\mathbf{D_0'w_0D_0}=\begin{pmatrix}\frac{1}{n}\sum\limits_{i=1}^n X_iX_i'K_h(U_i-u_0) & \frac{1}{n}\sum\limits_{i=1}^n X_iX_i'\left(\frac{U_i-u_0}{h}\right)K_h(U_i-u_0)\\ \frac{1}{n}\sum\limits_{i=1}^n X_iX_i'\left(\frac{U_i-u_0}{h}\right)K_h(U_i-u_0) & \frac{1}{n}\sum\limits_{i=1}^n X_iX_i'\left(\frac{U_i-u_0}{h}\right)^2K_h(U_i-u_0)\end{pmatrix}.$$

Each element of the above matrix is in the form of a kernel regression. By Lemma 1, we have

$$\frac{1}{n}\mathbf{D_0'w_0D_0}=\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\otimes f(u_0)\Gamma(u_0)\{1+O_p(c_n)\}.$$

The proofs of (A.4) and (A.5) are similar to the proof of (A.3).

Lemma 4. Under the conditions (i)-(vii), as $n\to\infty$, we have that

$$\frac{1}{n}\mathbf{\tilde{W}'\tilde{W}}\to\mathbf{\Sigma_1}+\Sigma_\eta,\ \ \text{a.s.} \tag{A.6}$$

and

$$\frac{1}{n}\left(\mathbf{\tilde{W}'\Sigma^{-1}\tilde{W}}-\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\right)\to\mathbf{\Sigma_4}.\ \ a.s. \tag{A.7}$$

where $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_4}$ are defined in Lemmma 4.1 and Theorem 3, respectively.

$\mathbf{proof}$. By Lemma 3, we have

$$\mathbf{SW} = \begin{pmatrix} X_1'\Gamma^{-1}(U_1)\Phi(U_1) \\ \vdots \\ X_n'\Gamma^{-1}(U_n)\Phi(U_n) \end{pmatrix} \{1 + O_p(c_n)\}.$$

Now, using the results above, it is easy to show that

$$\frac{1}{n}\tilde{\mathbf{W}}'\tilde{\mathbf{W}} = \frac{1}{n}\sum_{i=1}^{n}[W_i - \Phi'(U_i)\Gamma^{-1}(U_i)X_i][W_i' - X_i'\Gamma^{-1}(U_i)\Phi(U_i)]\{1 + O_p(c_n)\},$$

and

$$\frac{1}{n}\tilde{\mathbf{W}}'\mathbf{\Sigma}^{-1}\tilde{\mathbf{W}} = \frac{1}{n}\sum_{i=1}^{n}\{(\sigma^{-2}(U_i)+\beta'\Sigma_\eta\beta)[W_i-\Phi'(U_i)\Gamma^{-1}(U_i)X_i][W_i'-X_i'\Gamma^{-1}(U_i)\Phi(U_i)]\}\{1+O_p(c_n)\}.$$

By the strong law of large numbers, we complete the proof of Lemma 4.

$\mathbf{Lemma\ 5}$. Under the conditions (i)- (vii), we have

$$\frac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{M}} = O_p(n^{1/2}c_n^2),$$

where $\tilde{\mathbf{M}} = (\mathbf{I} - \mathbf{S})\mathbf{M}$.

$\mathbf{Proof}$: Similar to the proof of Lemma.3, we have the following equation

$$(X_i', \mathbf{0}_{1\times p})\{\mathbf{D_i'w_iD_i}\}^{\mathbf{-1}}\mathbf{D_i'w_iM} = X_i'g(U_i)\{1 + O_p(c_n)\}, \ \ i = 1, 2, \ldots, n.$$

Therefore

$$\begin{aligned}
&\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{M}} = \tfrac{1}{\sqrt{n}}(\mathbf{W} - \mathbf{SW})'\mathbf{\Sigma^{-1}}(\mathbf{M} - \mathbf{SM}) \\
= &\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[(\sigma^2(U_i) + \beta'\Sigma_\eta\beta)\big(W_i - \mathbf{W'w_iD_i}(\mathbf{D_i'w_iD_i})^{\mathbf{-1}}(X_i', \mathbf{0})'\big)\big(X_i'g(U_i) \\
&- (X_i', \mathbf{0})(\mathbf{D_i'w_iD_i})^{\mathbf{-1}}\mathbf{D_i'w_iM}\big)\Big] \\
= &\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[(\sigma^2(U_i) + \beta'\Sigma_\eta\beta)\big(W_i - \Phi'(U_i)\Gamma^{-1}(U_i)X_i\big)\big(X_i'g(U_i) - (X_i', \mathbf{0})(\mathbf{D_i'w_iD_i})^{\mathbf{-1}}\mathbf{D_i'w_iM}\big)\Big] \\
&+ \tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[(\sigma^2(U_i) + \beta'\Sigma_\eta\beta)\big(\Phi'(U_i)\Gamma^{-1}(U_i)X_i - \mathbf{W'w_iD_i}(\mathbf{D_i'w_iD_i})^{\mathbf{-1}}(X_i', \mathbf{0})'\big)\big(X_i'g(U_i) \\
&- (X_i', \mathbf{0})(\mathbf{D_i'w_iD_i})^{\mathbf{-1}}\mathbf{D_i'w_iM}\big)\Big] \\
= &O_p(\tfrac{1}{\sqrt{n}}c_n) + O_p(n^{\frac{1}{2}}c_n^2) = O_p(n^{\frac{1}{2}}c_n^2).
\end{aligned}$$

Proof of Theorem 1. For any given $u_0 \in \Omega$, let $\mathbf{K_0} = \mathrm{Diag}(k_1(u_0),\ k_2(u_0),\ \ldots,\ k_n(u_0))$. According to (3.3) and (3.4), we have

$$\begin{aligned}
&\hat{\sigma}^2(u_0) = \hat{\sigma}_e^2(u_0) - \tilde{\beta}'\Sigma_\eta\tilde{\beta} = \hat{e}'\mathbf{K_0}\hat{e} - \tilde{\beta}'\Sigma_\eta\tilde{\beta} \\
&= \big(\mathbf{Y} - \mathbf{W}\tilde{\beta} - \hat{\mathbf{M}}\big)'\mathbf{K_0}\big(\mathbf{Y} - \mathbf{W}\tilde{\beta} - \hat{\mathbf{M}}\big) - \tilde{\beta}'\Sigma_\eta\tilde{\beta} \\
&= \big[\mathbf{Y} - \mathbf{Z}\beta + \mathbf{Z}\beta - \mathbf{W}\tilde{\beta} - \hat{\mathbf{M}} - \mathbf{M} + \mathbf{M}\big]'\mathbf{K_0}\big[\mathbf{Y} - \mathbf{Z}\beta + \mathbf{Z}\beta - \mathbf{W}\tilde{\beta} - \hat{\mathbf{M}} - \mathbf{M} + \mathbf{M}\big] - \tilde{\beta}'\Sigma_\eta\tilde{\beta} \\
&= \big[\varepsilon + (\mathbf{M} - \hat{\mathbf{M}}) + (\mathbf{Z}\beta - \mathbf{W}\tilde{\beta})\big]'\mathbf{K_0}\big[\varepsilon + (\mathbf{M} - \hat{\mathbf{M}}) + (\mathbf{Z}\beta - \mathbf{W}\tilde{\beta})\big] - \tilde{\beta}'\Sigma_\eta\tilde{\beta} \\
&= \varepsilon'\mathbf{K}_0\varepsilon + \big(\mathbf{M} - \hat{\mathbf{M}}\big)'\mathbf{K_0}\big(\mathbf{M} - \hat{\mathbf{M}}\big) + (\mathbf{Z}\beta - \mathbf{W}\tilde{\beta})'\mathbf{K_0}(\mathbf{Z}\beta - \mathbf{W}\tilde{\beta}) + 2\varepsilon'\mathbf{K_0}\big(\mathbf{M} - \hat{\mathbf{M}}\big) \\
&\quad + 2\varepsilon'\mathbf{K_0}(\mathbf{Z}\beta - \mathbf{W}\tilde{\beta}) + 2\big(\mathbf{M} - \hat{\mathbf{M}}\big)'\mathbf{K_0}(\mathbf{Z}\beta - \mathbf{W}\tilde{\beta}) - \tilde{\beta}'\Sigma_\eta\tilde{\beta} \\
&= B_1 + B_2 + B_3 + B_4 + B_5 + B_6 + B_7,
\end{aligned}$$

where $\hat{e} = (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n)',\ \hat{e}_i$ is defined in (3.2), $i = 1, 2, \ldots, n.$ Since $B_1 = \varepsilon'\mathbf{K_0}\varepsilon = \sum_{i=1}^{n} \tilde{k}_i(u_0)\varepsilon_i^2$, so we can obtain based on Lemma 1 and 2 that

$$\begin{aligned}
\sup_{u_0\in\Omega} |B_1 - \sigma^2(u_0)| &= \sup_{u_0\in\Omega} \Big|B_1 - \sum_{i=1}^{n} \tilde{k}_i(u_0)\sigma^2(U_i) + \sum_{i=1}^{n} \tilde{k}_i(u_0)\sigma^2(U_i) - \sigma^2(u_0)\Big| \\
&\le \sup_{u_0\in\Omega} \Big|B_1 - \sum_{i=1}^{n} \tilde{k}_i(u_0)\sigma^2(U_i)\Big| + \sup_{u_0\in\Omega} \Big|\sum_{i=1}^{n} \tilde{k}_i(u_0)\sigma^2(U_i) - \sigma^2(u_0)\Big| = O_p(\tilde{c}_n).
\end{aligned}$$

Furthermore, according to the result in Lv and Zhang (2007), we can get $\sup_{u\in\Omega} |\tilde{g}(u) - g(u)| = O_p(c_n)$. Then

$$\begin{aligned}
\sup_{u_0\in\Omega} |B_2| &= \sup_{u_0\in\Omega} \left|\left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)(g(U_i) - \tilde{g}(U_i))'X_iX_i'(g(U_i) - \tilde{g}(U_i))\right| \\
&= O_p(c_n^2) \sup_{u_0\in\Omega} \left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)|X_iX_i'| = O_p(\tilde{c}_n).
\end{aligned}$$

Since

$$\begin{aligned}
&\sup_{u_0\in\Omega} |B_3 - \beta'\Sigma_\eta\beta| = \sup_{u_0\in\Omega} |(\beta - \tilde{\beta})'\mathbf{W}'\mathbf{K_0}\mathbf{W}(\beta - \tilde{\beta}) + 2(\beta - \tilde{\beta})'\mathbf{W}'\mathbf{K_0}\eta\beta + \beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta| \\
&\le \sup_{u_0\in\Omega} |(\beta - \tilde{\beta})'\mathbf{W}'\mathbf{K_0}\mathbf{W}(\beta - \tilde{\beta})| + 2\sup_{u_0\in\Omega} |(\beta - \tilde{\beta})'\mathbf{W}'\mathbf{K_0}\eta\beta| + \sup_{u_0\in\Omega} |\beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta| \\
&= \sup_{u_0\in\Omega} \left|(\beta - \tilde{\beta})'\left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K(\frac{U_i-u_0}{\tilde{h}})W_iW_i'(\beta - \tilde{\beta})\right| \\
&\quad + \sup_{u_0\in\Omega} \left|(\beta - \tilde{\beta})'\left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)W_i\eta_i'\beta\right| + \sup_{u_0\in\Omega} |\beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta| \\
&\le \ \parallel \beta - \tilde{\beta} \parallel^2 \sup_{u_0\in\Omega} \left|\left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)W_iW_i'\right| \\
&\quad + \parallel \beta - \tilde{\beta} \parallel \sup_{u_0\in\Omega} \left|\left(\frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1} \frac{1}{n\tilde{h}}\sum_{i=1}^{n} K\left(\frac{U_i-u_0}{\tilde{h}}\right)W_i\eta_i'\beta\right| + \sup_{u_0\in\Omega} |\beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta| \\
&= O_p(\frac{1}{n}) + O_p(\frac{1}{\sqrt{n}})O_p\left(\sqrt{\frac{\log 1/\tilde{h}}{n\tilde{h}}}\right) + \sup_{u_0\in\Omega} |\beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta| \\
&= O_p(\tilde{c}_n) + \sup_{u_0\in\Omega} |\beta'\eta'\mathbf{K_0}\eta\beta - \beta'\Sigma_\eta\beta|.
\end{aligned}$$

According to the Lemma 1, we can get $\sup_{u_0\in\Omega}|\beta'\eta'\mathbf{K_0}\eta\beta-\beta'\Sigma_\eta\beta|=O_p(\tilde{c}_n)$. This implies that $\sup_{u_0\in\Omega}|B_3-\beta'\Sigma_\eta\beta|=O_p(\tilde{c}_n)$. And according to the Cauchy inequality, we can derive that $\sup_{u_0\in\Omega}|B_4|=O_p(\tilde{c}_n)$. For B_5 and B_6, we have

$$\begin{aligned}\sup_{u_0\in\Omega}|B_5| &= 2\sup_{u_0\in\Omega}\left|\varepsilon'\mathbf{K_0}(\mathbf{W}(\beta-\tilde{\beta})-\eta\beta)\right| \le 2\sup_{u_0\in\Omega}\left|\varepsilon'\mathbf{K_0}\mathbf{W}(\beta-\tilde{\beta})\right| + 2\sup_{u_0\in\Omega}\left|\varepsilon'\mathbf{K_0}\eta\beta\right| \\ &= O_p(\tilde{c}_n) + 2\sup_{u_0\in\Omega}\left|\left(\frac{1}{n\tilde{h}}\sum_{i=1}^n K\left(\frac{U_i-u_0}{\tilde{h}}\right)\right)^{-1}\frac{1}{n\tilde{h}}\sum_{i=1}^n K\left(\frac{U_i-u_0}{\tilde{h}}\right)\varepsilon_i\eta_i'\beta\right| \\ &= O_p(\tilde{c}_n) + O_p\left(\sqrt{\frac{\log 1/\tilde{h}}{n\tilde{h}}}\right) = O_p(\tilde{c}_n),\end{aligned}$$

and

$$\begin{aligned}\sup_{u_0\in\Omega}|B_6| &= 2\sup_{u_0\in\Omega}\left|(\mathbf{M}-\hat{\mathbf{M}})'\mathbf{K_0}(\mathbf{W}(\beta-\tilde{\beta})-\eta\beta)\right| \\ &\le 2\sup_{u_0\in\Omega}\left|(\mathbf{M}-\hat{\mathbf{M}})'\mathbf{K_0}\mathbf{W}(\beta-\tilde{\beta})\right| + 2\sup_{u_0\in\Omega}\left|(\mathbf{M}-\hat{\mathbf{M}})'\mathbf{K_0}\eta\beta\right| \\ &\le 2\sup_{u_0\in\Omega}\left[(\mathbf{M}-\hat{\mathbf{M}})'\mathbf{K_0}(\mathbf{M}-\hat{\mathbf{M}})\right]^{\frac{1}{2}}\left[(\beta-\tilde{\beta})'\mathbf{K_0}(\beta-\tilde{\beta})\right]^{\frac{1}{2}} \\ &\quad + 2\sup_{u_0\in\Omega}\left[(\mathbf{M}-\hat{\mathbf{M}})'\mathbf{K_0}(\mathbf{M}-\hat{\mathbf{M}})\right]^{\frac{1}{2}}\left[\beta'\eta'\mathbf{K_0}\eta\beta\right]^{\frac{1}{2}} \\ &= O_p(\tilde{c}_n)O_p\big(\tfrac{1}{\sqrt{n}}\big) + O_p(\tilde{c}_n)O_p(1) = O_p(\tilde{c}_n).\end{aligned}$$

So

$$\begin{aligned}\sup_{u_0\in\Omega}|\hat{\sigma}^2(u_0)-\sigma^2(u_0)| &= \sup_{u_0\in\Omega}\left|\hat{e}'\mathbf{k_0}\hat{e}-\tilde{\beta}'\Sigma_\eta\tilde{\beta}-\sigma^2(u_0)-\beta'\Sigma_\eta\beta+\beta'\Sigma_\eta\beta\right| \\ &\le \sup_{u_0\in\Omega}\left|B_1-\sigma^2(u_0)\right| + \sup_{u_0\in\Omega}\left|B_3-\beta'\Sigma_\eta\beta\right| + \sup_{u_0\in\Omega}\left|\beta'\Sigma_\eta\beta-\tilde{\beta}\Sigma_\eta\tilde{\beta}\right| \\ &= O_p(\tilde{c}_n) + \sup_{u_0\in\Omega}\left|(\beta-\tilde{\beta})'\Sigma_\eta(\beta-\tilde{\beta})-2\beta'\Sigma_\eta\beta+\beta\Sigma_\eta\tilde{\beta}+\tilde{\beta}'\Sigma_\eta\beta\right| \\ &\le O_p(\tilde{c}_n) + \sup_{u_0\in\Omega}\left|(\beta-\tilde{\beta})'\Sigma_\eta(\beta-\tilde{\beta})\right| + \sup_{u_0\in\Omega}\left|\beta'\Sigma_\eta(\beta-\tilde{\beta})\right| \\ &\quad + \sup_{u_0\in\Omega}\left|(\beta-\tilde{\beta})'\Sigma_\eta\beta\right| \\ &= O_p(\tilde{c}_n).\end{aligned}$$

Then we obtain the desired result.

Proof of Theorem 2. From the proof of Theorem 1, we have

$$\sqrt{n\tilde{h}}B_2 = o_p(1), \quad \sqrt{n\tilde{h}}\big(B_3-\beta'\eta'\mathbf{K_0}\eta\beta\big) = o_p(1),$$

$$\sqrt{n\tilde{h}}B_4 = o_p(1), \quad \sqrt{n\tilde{h}}\big(B_5+2\varepsilon'\mathbf{K_0}\eta\beta\big) = o_p(1),$$

and

$$\sqrt{n\tilde{h}}B_6 = o_p(1), \quad \sqrt{n\tilde{h}}\big(B_7+\beta'\Sigma_\eta\beta\big) = o_p(1).$$

Therefore we obtain

$$
\begin{aligned}
\sqrt{n\tilde{h}}\big(\hat{\sigma}^2(u_0)-\sigma^2(u_0)\big) &= \sqrt{n\tilde{h}}\big(\varepsilon'\mathbf{K_0}\varepsilon-\sigma^2(u_0)-\beta'\Sigma_\eta\beta+\beta'\eta'\mathbf{K_0}\eta\beta-2\varepsilon'\mathbf{K_0}\eta\beta\big)+o_p(1)\\
&= \sqrt{n\tilde{h}}\frac{\xi}{\frac{1}{n}\sum\limits_{i=1}^{n}K_{\tilde{h}}(U_i-u_0)}+o_p(1)=\sqrt{n\tilde{h}}\frac{\xi}{\hat{f}(u_0)}+o_p(1)\\
&= \sqrt{n\tilde{h}}\frac{\xi}{f(u_0)}+\sqrt{n\tilde{h}}\Big(\frac{1}{\hat{f}(u_0)}-\frac{1}{f(u_0)}\Big)\xi+o_p(1),
\end{aligned}
$$

where $\xi=\frac{1}{n}\sum\limits_{i=1}^{n}K_{\tilde{h}}(U_i-u_0)\big(\varepsilon_i^2-\sigma^2(u_0)+\beta'\eta_i\eta_i'\beta-\beta'\Sigma_\eta\beta-2\varepsilon_i\eta_i'\beta\big)$.

It is known that the estimate $\hat{f}(u)$ of the density function of U is consistent. We shall only prove that $\sqrt{n\tilde{h}}\xi$ is asymptotically normal. Since ξ is a sum of i.i.d. random variables, so by routine arguments in kernel regression, the asymptotic normality can be easily derived, we only have to identify the asymptotic mean and variance. Note that

$$
\begin{aligned}
E\xi &= E\Big[\frac{1}{n}\sum_{i=1}^{n}K_{\tilde{h}}(U_i-u_0)\big(\varepsilon_i^2-\sigma^2(u_0)+\beta'\eta_i\eta_i'\beta-\beta'\Sigma_\eta\beta-2\varepsilon_i\eta_i'\beta\big)\Big]\\
&= E\Big[K_{\tilde{h}}(U_1-u_0)\big(\varepsilon_1^2-\sigma^2(u_0)+\beta'\eta_1\eta_1'\beta-\beta'\Sigma_\eta\beta-2\varepsilon_1\eta_1'\beta\big)\Big]\\
&= E\Big[K_{\tilde{h}}(U_1-u_0)E\big(\varepsilon_1^2-\sigma^2(u_0)+\beta'\eta_1\eta_1'\beta-\beta'\Sigma_\eta\beta-2\varepsilon_1\eta_1'\beta\big|U_1,X_1\big)\Big]\\
&= E\Big[K_{\tilde{h}}(U_1-u_0)\big(\sigma^2(U_1)-\sigma^2(u_0)\big)\Big]\\
&= \int\frac{1}{\tilde{h}}K\big(\frac{u-u_0}{\tilde{h}}\big)\big(\sigma^2(u)-\sigma^2(u_0)\big)f(u)du\\
&= \int K(v)\big(\sigma^2(u_0+v\tilde{h})-\sigma^2(u_0)\big)f(u_0+v\tilde{h})dv\\
&= \frac{\tilde{h}^2}{2}f(u_0)\sigma^{2''}(u_0)\mu_2+f'(u_0)\sigma^{2'}(u_0)\mu_2\tilde{h}^2+O(\tilde{h}^3)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(\xi) &= \frac{1}{n}\mathrm{Var}\Big[K_{\tilde{h}}(U_1-u_0)\big(\varepsilon_1^2-\sigma^2(u_0)+\beta'\eta_1\eta_1'\beta-\beta'\Sigma_\eta\beta-2\varepsilon_1\eta_1'\beta\big)\Big]\\
&= \frac{1}{n}\mathrm{Var}\Big[K_{\tilde{h}}(U_1-u_0)\big((\varepsilon_1^2-\eta_1'\beta)^2+(\sigma^2(u_0)+\beta'\Sigma_\eta\beta)\big)\Big]\\
&= \frac{1}{n}E\Big[K_{\tilde{h}}^2(U_1-u_0)\big((\varepsilon_1^2-\eta_1'\beta)^2+(\sigma^2(u_0)+\beta'\Sigma_\eta\beta)\big)^2\Big]-\frac{1}{n}(E\xi)^2\\
&= \frac{1}{n}E\Big[K_{\tilde{h}}^2(U_1-u_0)E\Big(\big((\varepsilon_1^2-\eta_1'\beta)^2+(\sigma^2(u_0)+\beta'\Sigma_\eta\beta)\big)^2\big|U_1\Big)\Big]-\frac{1}{n}(E\xi)^2\\
&\triangleq \frac{1}{n}E\Big[K_{\tilde{h}}^2(U_1-u_0)\tau(U_1)\Big]-\frac{1}{n}(E\xi)^2=\frac{1}{n}\int\frac{1}{\tilde{h}^2}K^2\big(\frac{u-u_0}{\tilde{h}}\big)\tau(u)f(u)du+O\big(\frac{\tilde{h}^4}{n}\big)\\
&= \frac{1}{n\tilde{h}}\int K^2(v)\tau(u_0+v\tilde{h})f(u_0+v\tilde{h})dv+O\big(\frac{\tilde{h}^4}{n}\big)=\frac{1}{n\tilde{h}}f(u_0)\tau(u_0)\nu_0+O\big(\frac{\tilde{h}}{n}\big)+O\big(\frac{\tilde{h}^4}{n}\big),
\end{aligned}
$$

so, we obtain

$$
\sqrt{n\tilde{h}}\Big(\xi-\frac{\tilde{h}^2}{2}f(u_0)\sigma^{2''}(u_0)\mu_2-f'(u_0)\sigma^{2'}(u_0)\mu_2\tilde{h}^2\Big)\stackrel{D}{\longrightarrow}N(0,\ \nu_0f(u_0)\tau(u_0)).
$$

This completes the proof of Theorem 2.

Proof of Theorem 3. Since $\sqrt{n}(\hat\beta-\beta)=\sqrt{n}(\hat\beta-\tilde\beta_T)+\sqrt{n}(\tilde\beta_T-\beta)$, where $\tilde\beta_T=(\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{W}}-\mathrm{tr}(\mathbf{\Sigma}^{-1})\Sigma_\eta)^{-1}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{Y}}$. To show Theorem 4, it suffices to show $\sqrt{n}(\hat\beta-\tilde\beta_T)=o_p(1)$, and $\sqrt{n}(\tilde\beta_T-\beta)\xrightarrow{D}N(0,\mathbf{\Sigma_4^{-1}\Sigma_3\Sigma_4^{-1}}),\ \ n\to\infty$. According to the expression of the estimate $\tilde\beta_T$, let

$$\nabla=\frac{1}{n}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{W}}-\frac{\mathrm{tr}(\mathbf{\Sigma}^{-1})}{n}\Sigma_\eta.$$

Then we have

$$\begin{aligned}\sqrt{n}(\tilde\beta_T-\beta)&=\sqrt{n}\big[\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{Y}}-\beta\big]=\sqrt{n}\big[\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}(\tilde{\mathbf{M}}+\tilde{\mathbf{Z}}\beta+\tilde\varepsilon)-\beta\big]\\&=\nabla^{-1}\tfrac{\mathrm{tr}(\mathbf{\Sigma}^{-1})}{\sqrt{n}}\Sigma_\eta\beta+\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}\tilde{\mathbf{M}}+\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}(\tilde\varepsilon-\tilde\eta\beta)\\&=o_p(1)+\nabla^{-1}\tfrac{1}{\sqrt{n}}\Big[\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}(\tilde\varepsilon-\tilde\eta\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big].\end{aligned}$$

Now, we consider $\frac{1}{\sqrt{n}}\Big[\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}(\tilde\varepsilon-\tilde\eta\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]$. Since

$$\begin{aligned}&\tfrac{1}{\sqrt{n}}\Big[\tilde{\mathbf{W}}'\mathbf{\Sigma^{-1}}(\tilde\varepsilon-\tilde\eta\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]\\&=\tfrac{1}{\sqrt{n}}\Big[\sum_{i=1}^n(\sigma^2(U_i)+\beta'\Sigma_\eta\beta)^{-1}\tilde W_i(\tilde{\varepsilon_i}-\tilde\eta_i'\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]\\&=\tfrac{1}{\sqrt{n}}\Big[\sum_{i=1}^n(\sigma^2(U_i)+\beta'\Sigma_\eta\beta)^{-1}(W_i-\mathbf{W'w_iD_i(D_i'w_iD_i)^{-1}}(X_i,\mathbf{0})')(\varepsilon_i\\&\quad-(X_i,\mathbf{0})(\mathbf{D_i'w_iD_i})^{-1}\mathbf{D_iw_i}\varepsilon-\eta_i'\beta+(X_i,\mathbf{0})(\mathbf{D_i'w_iD_i})^{-1}\mathbf{D_iw_i}\eta\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]\\&=\tfrac{1}{\sqrt{n}}\Big[\sum_{i=1}^n(\sigma^2(U_i)+\beta'\Sigma_\eta\beta)^{-1}(W_i-\mathbf{W'w_iD_i(D_i'w_iD_i)^{-1}}(X_i,\mathbf{0})')(\varepsilon_i-\eta_i'\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]\\&\quad+\tfrac{1}{\sqrt{n}}\sum_{i=1}^n(\sigma^2(U_i)+\beta'\Sigma_\eta\beta)^{-1}(W_i-\mathbf{W'w_iD_i(D_i'w_iD_i)^{-1}}(X_i,\mathbf{0})')(X_i,\mathbf{0})(\mathbf{D_i'w_iD_i})^{-1}\mathbf{D_iw_i}(\eta_i\beta\\&\quad-\varepsilon_i)\\&=\tfrac{1}{\sqrt{n}}\Big[\sum_{i=1}^n(\sigma^2(U_i)+\beta'\Sigma_\eta\beta)^{-1}(W_i-\Phi'(U_i)\Gamma^{-1}(U_i)X_i)(\varepsilon_i-\eta_i'\beta)+\mathrm{tr}(\mathbf{\Sigma^{-1}})\Sigma_\eta\beta\Big]+\tfrac{1}{\sqrt{n}}\Big[\sum_{i=1}^n(\sigma^2(U_i)\\&\quad+\beta'\Sigma_\eta\beta)^{-1}(\Phi'(U_i)\Gamma^{-1}(U_i)X_i-\mathbf{W'w_iD_i(D_i'w_iD_i)^{-1}}(X_i,\mathbf{0})')(\varepsilon_i-\eta_i'\beta)\Big]+\tfrac{1}{\sqrt{n}}\sum_{i=1}^n(\sigma^2(U_i)\\&\quad+\beta'\Sigma_\eta\beta)^{-1}(W_i-\mathbf{W'w_iD_i(D_i'w_iD_i)^{-1}}(X_i,\mathbf{0})')(X_i,\mathbf{0})(\mathbf{D_i'w_iD_i})^{-1}\mathbf{D_iw_i}(\eta_i\beta-\varepsilon_i)\\&=A_1+A_2+A_3\end{aligned}$$

We can easily show that $A_i=o_p(\frac{1}{\sqrt{n}}),\ i=2,3$. Applying central limit theorem, we can derive the asymptotic normality of A_1. Combining Lemma 4, we complete the proof of the asymptotic normality of $\sqrt{n}(\tilde\beta_T-\beta)$.

To show that $\sqrt{n}(\hat\beta-\tilde\beta_T)=o_p(1)$, let

$$\hat\nabla=\frac{1}{n}\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{\mathbf{-1}}\tilde{\mathbf{W}}-\frac{\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})}{n}\Sigma_\eta,$$

then we have

$$
\begin{aligned}
\sqrt{n}(\hat\beta-\tilde\beta_T) &= \hat\nabla^{-1}\tfrac{1}{\sqrt{n}}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}\tilde\eta\big]\beta-\nabla^{-1}\tfrac{1}{\sqrt{n}}\big[\mathrm{tr}(\mathbf{\Sigma}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\mathbf{\Sigma}^{-1}\tilde\eta\big]\beta\\
&\quad+\hat\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{M}}+\tilde\varepsilon)-\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{M}}+\tilde\varepsilon)\\
&=(\hat\nabla^{-1}-\nabla^{-1})\tfrac{1}{\sqrt{n}}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}\tilde\eta\big]\beta+\nabla^{-1}\tfrac{1}{\sqrt{n}}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}\tilde\eta\\
&\quad-\mathrm{tr}(\mathbf{\Sigma}^{-1})\Sigma_\eta+\tilde{\mathbf{W}}'\mathbf{\Sigma}^{-1}\tilde\eta\big]\beta+(\hat\nabla^{-1}-\nabla^{-1})\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{M}}+\tilde\varepsilon)\\
&\quad+\nabla^{-1}\tfrac{1}{\sqrt{n}}\tilde{\mathbf{W}}'(\hat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1})(\tilde{\mathbf{M}}+\tilde\varepsilon).
\end{aligned}
$$

Under conditions (i)-(viii), it is also easy to prove

$$\hat\nabla-\nabla=o_p(1),\quad \frac{1}{\sqrt{n}}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}\tilde\eta\big]=O_p(1),$$

$$\nabla=O_p(1),\quad \frac{1}{\sqrt{n}}\big[\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}\tilde\eta-\mathrm{tr}(\mathbf{\Sigma}^{-1})\Sigma_\eta-\tilde{\mathbf{W}}'\mathbf{\Sigma}^{-1}\tilde\eta\big]=o_p(1),$$

and

$$\frac{1}{\sqrt{n}}\tilde{\mathbf{W}}'\hat{\mathbf{\Sigma}}^{-1}(\tilde{\mathbf{M}}+\tilde\varepsilon)=O_p(1),\quad \frac{1}{\sqrt{n}}\tilde{\mathbf{W}}'(\hat{\mathbf{\Sigma}}^{-1}-\mathbf{\Sigma}^{-1})(\tilde{\mathbf{M}}+\tilde\varepsilon)=o_p(1).$$

Then the proof of Theorem 3 is finished.

Proof of Theorem 4. Denote

$$P_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_iX_i'\left(\frac{U_i-u_0}{h}\right)^k K_h(U_i-u_0)$$

$$Q_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_i(Y_i-W_i'\tilde\beta)\left(\frac{U_i-u_0}{h}\right)^k K_h(U_i-u_0)$$

$$N(U_i)=g(U_i)-g(u_0)-(U_i-u_0)g'(u_0)-\frac{1}{2}(U_i-u_0)^2g''(u_0)$$

$$R_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_iX_i'N(U_i)\left(\frac{U_i-u_0}{h}\right)^2 K_h(U_i-u_0)$$

$$S_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_i\varepsilon_i\left(\frac{U_i-u_0}{h}\right)^2 K_h(U_i-u_0)$$

$$T_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_iW_i'\left(\frac{U_i-u_0}{h}\right)^2 K_h(U_i-u_0)(\beta-\tilde\beta)$$

$$L_{n,k}(u_0)=\frac{1}{n}\sum_{i=1}^n X_i\eta_i'\left(\frac{U_i-u_0}{h}\right)^2 K_h(U_i-u_0)\beta$$

$$d_{n,k}(u_0)=\frac{1}{2}P_{n,k+2}(u_0)h^2g''(u_0).$$

From (2.11), we have

$$\begin{aligned}\tilde{g}(u_0) &= (\mathbf{I}_p, \mathbf{0}_p)\{\mathbf{D_0'}\mathbf{w_0}\mathbf{D_0}\}^{-1}\mathbf{D_0'}\mathbf{w_0}(\mathbf{Y}-\mathbf{W}\tilde{\beta}) \\ &= (\mathbf{I}_p, \mathbf{0}_p)\{\tfrac{1}{n}\mathbf{D_0'}\mathbf{w_0}\mathbf{D_0}\}^{-1}\tfrac{1}{n}\mathbf{D_0'}\mathbf{w_0}(\mathbf{Y}-\mathbf{W}\tilde{\beta}) \\ &= (\mathbf{I}_p, \mathbf{0}_p)\begin{pmatrix} P_{n,0}(u_0) & P_{n,1}(u_0) \\ P_{n,1}(u_0) & P_{n,2}(u_0)\end{pmatrix}^{-1}\begin{pmatrix} T_{n,0}(u_0) \\ T_{n,1}(u_0)\end{pmatrix}.\end{aligned} \qquad (A.8)$$

For k=0,1,

$$\begin{aligned}Q_{n,k}(u_0) &= \tfrac{1}{n}\sum_{i=1}^{n} X_i(Y_i - W_i'\tilde{\beta})\left(\tfrac{U_i-u_0}{h}\right)^k K_h(U_i-u_0) \\ &= \tfrac{1}{n}\sum_{i=1}^{n} X_i\Big[(Y_i - W_i'\beta + W_i'(\beta-\tilde{\beta})\Big]\left(\tfrac{U_i-u_0}{h}\right)^k K_h(U_i-u_0) \\ &= \tfrac{1}{n}\sum_{i=1}^{n} X_i\Big[(\varepsilon_i + X_i'\beta(U_i) - \eta_i'\beta + W_i'(\beta-\tilde{\beta})\Big]\left(\tfrac{U_i-u_0}{h}\right)^k K_h(U_i-u_0) \\ &= S_{n,k}(u_0) + P_{n,k}(u_0)g(u_0) + P_{n,k+1}(u_0)hg'(u_0) + \tfrac{1}{2}P_{n,k+2}(u_0)(u_0)h^2g''(u_0) \\ &\quad + R_{n,k}(u_0) - L_{n,k}(u_0) + T_{n,k}(u_0) \\ &= S_{n,k}(u_0) + P_{n,k}(u_0)g(u_0) + P_{n,k+1}(u_0)hg'(u_0) + d_{n,k}(u_0) + R_{n,k}(u_0) - L_{n,k}(u_0) \\ &\quad + T_{n,k}(u_0).\end{aligned} \qquad (A.9)$$

Combining (A.8) and (A.9), we have

$$\tilde{g}(u_0) = g(u_0) + (\mathbf{I}_p, \mathbf{0}_p)P_n^{-1}(u_0)\Big[S_n(u_0) + d_n(u_0) + R_n(u_0) - L_n(u_0) + T_n(u_0)\Big],$$

where

$$S_n(u_0) = \begin{pmatrix} S_{n,0}(u_0) \\ S_{n,1}(u_0)\end{pmatrix},\ d_n(u_0) = \begin{pmatrix} d_{n,0}(u_0) \\ d_{n,1}(u_0)\end{pmatrix}, R_n(u_0) = \begin{pmatrix} R_{n,0}(u_0) \\ R_{n,1}(u_0)\end{pmatrix},$$

$$L_n(u_0) = \begin{pmatrix} L_{n,0}(u_0) \\ L_{n,1}(u_0)\end{pmatrix}, T_n(u_0) = \begin{pmatrix} T_{n,0}(u_0) \\ T_{n,1}(u_0)\end{pmatrix}, P_n(u_0) = \begin{pmatrix} P_{n,0}(u_0) & P_{n,1}(u_0) \\ P_{n,1}(u_0) & P_{n,2}(u_0)\end{pmatrix}.$$

Similar to the proof of Lemma 3, we can obtain

$$R_n(u_0) + d_n(u_0) = \frac{h^2}{2}\begin{pmatrix} \mu_2 f(u_0)\Gamma(u_0)g''(u_0) \\ \mathbf{0}_{p\times 1}\end{pmatrix}\{1 + o_p(h^2)\}.$$

According to the result of Lemma 4.1, we obtain

$$\tilde{\beta} - \beta = \mathbf{1}_{q\times 1}O_p(n^{-\frac{1}{2}}).$$

We also have

$$T_{n,k}(u_0) = \mathbf{1}_{p\times 1} O_p(n^{-\frac{1}{2}}). \quad k = 0, 1.$$

Now we show the asymptotic normality of $S_n(u_0)+L_n(u_0)$. For this purpose, let α_1 and α_2 are two column vectors of length p. We have to show the asymptotic normality of $\alpha_1'(S_{n,0}(u_0)+L_{n,0}(u_0))+\alpha_2'(S_{n,1}(u_0)+L_{n,1}(u_0))$. Clearly its expectation is 0, and

$$\begin{aligned}
&\mathrm{Var}\big(\alpha_1'(S_{n,0}(u_0) + L_{n,0}(u_0)) + \alpha_2'(S_{n,1}(u_0) + L_{n,1}(u_0))\big) \\
&= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\alpha_1' + \alpha_2'\left(\frac{U_i-u_0}{h}\right)\right) X_i K_h(U_i - u_0)(\varepsilon_i - \eta_i'\beta)\right) \\
&= \frac{1}{n}\mathrm{Var}\left(\left(\alpha_1' + \alpha_2'\left(\frac{U_1-u_0}{h}\right)\right) X_1 K_h(U_1 - u_0)(\varepsilon_1 - \eta_1'\beta)\right) \\
&= \frac{1}{n}E\left[\left(\alpha_1' + \alpha_2'\left(\frac{U_1-u_0}{h}\right)\right) X_1 K_h(U_1 - u_0)(\varepsilon_1 - \eta_1'\beta)\right]^2 \\
&= \frac{1}{n}\begin{pmatrix} \alpha_1' & \alpha_2' \end{pmatrix} E\left[\left(\mathbf{I}_p, \frac{U_1-u_0}{h}\mathbf{I}_p\right)' X_1 X_1' \left(\mathbf{I}_p, \frac{U_1-u_0}{h}\mathbf{I}_p\right) K_h^2(U_1 - u_0)(\varepsilon_1 - \eta_1'\beta)^2\right] \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \\
&= \begin{pmatrix} \alpha_1' & \alpha_2' \end{pmatrix} \begin{pmatrix} \frac{1}{nh}\nu_0\Gamma(u_0)f(u_0)(\sigma^2(u_0) + \beta'\Sigma_\eta\beta) & O(\frac{1}{n}) \\ O(\frac{1}{n}) & \frac{1}{nh}\nu_2\Gamma(u_0)f(u_0)(\sigma^2(u_0) + \beta'\Sigma_\eta\beta) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.
\end{aligned}$$

Combining Lemma 3, eventually we have

$$\sqrt{nh}(\tilde{g}(u_0) - g(u_0) - \frac{h^2}{2}\mu_2 g''(u_0)) \xrightarrow{D} N\big(0,\ \nu_0 f^{-1}(u_0)\Gamma^{-1}(u_0)(\sigma^2(u_0) + \beta'\Sigma_\eta\beta)\big), \quad n \to \infty.$$

The proof of Theorem 5 is similar to that of Theorem 4, hence omitted here for the sake of brevity.

Acknowledgments

Jianhong Shi's research is supported by the Natural Science Foundation of Shanxi Province, China (2013011002-1), and Weixing Song's research is partly supported by the NSF DMS 1205276.

References

[1] Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models, second ed. Chapman and Hall, New York.

[2] Cheng, C.L., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Arnold Publishers, London.

[3] Cui, H.J., Kong, E.F., 2006. Empirical likelihood confidence region for parameters in semi-linear errors-in-variables models. Scand. J. Statist. 33, 153-168.

[4] Cui, H.J., Li, R.C., 1998. On parameter estimation for semi-linear errors-in-variables models. J. Multi. Anal. 64, 1-24.

[5] Fan, J.Q., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli. 11, 1031-1057.

[6] Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.

[7] Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer-Verlag, New York.

[8] Hu, X.M., Wang, Z.Z., Zhao, Z.Z., 2009. Empirical likelihood for semiparametric varying-coefficient partially linear errors-in-variables models. Statist. Probab. Lett. 79, 1044-1052.

[9] Liang, H., Härdle, W., Carroll, R.J., 1999. Estimation in a semiparametric partially linear errors-in-variables model. Ann. Statist. 27, 1519-1535.

[10] Liang, H., Wang, S.J., Carroll, R.J., 2007. Partially linear models with missing response variables and error-prone covariates. Biometrika 94, 185-198.

[11] Lv, S.Q., Zhang, R.Q., Lu, Z.W., 2007. The asymptotic normality of profile least-squares estimation on semivarying coefficients model. Math. Appl. 20, 322-327.

[12] Shen, S.L., Cui, J.L., Mei, C.L., Wang, C.W., 2014. Estimation and inference of Semi-varying coefficient models with heteroscedastic error. J. Multi. Anal. 124, 70-93.

[13] Wang, Q.H., 1999. Estimation of partial linear error-in-variables models with validation data. J. Multi. Anal. 69, 30-64.

[14] Wang, X.L., Li, G.R., Lin, L., 2011. Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models. Metrika. 73, 171-185.

[15] Wei, C.H., 2012. Statistical inference for restricted partially linear varying coefficient errors-in-variables models. J. Statist. Plann. Inference. 142, 2464-2472.

[16] Xia, Y., Li, W.K., 1999. On the estimation and testing of functional-coefficient linear models. Statist. Sinica. 9, 737-757.

[17] You, J.H., Chen, G.M., 2006. Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. J. Multi. Anal. 97, 324-341.

[18] You, J.H., Chen, G.M., Zhou, Y., 2007. Statistical inference of partially linear regression models with heteroscedastic errors. J. Multi. Anal. 98, 1539-1557.

[19] Zhang, W.W., Li, G.R., Xue, L.G., 2011. Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition. Comp. Statist. and Da. Anal. 55, 3027-3040.