Statistical estimation in varying coefficient models with surrogate data and validation sampling

Qihua Wang\textsuperscript{a,b,}\textsuperscript{*}, Riquan Zhang\textsuperscript{c,d}

\textsuperscript{a} Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, PR China
\textsuperscript{b} Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong
\textsuperscript{c} Department of Statistics, East China Normal University, Shanghai, 200062, PR China
\textsuperscript{d} Department of Mathematics, Shanxi Datong University, Datong, Shanxi, 037000, PR China

\begin{abstract}
Varying coefficient error-in-covariables models are considered with surrogate data and validation sampling. Without specifying any error structure equation, two estimators for the coefficient function vector are suggested by using the local linear kernel smoothing technique. The proposed estimators are proved to be asymptotically normal. A bootstrap procedure is suggested to estimate the asymptotic variances. The data-driven bandwidth selection method is discussed. A simulation study is conducted to evaluate the proposed estimating methods.

© 2009 Elsevier Inc. All rights reserved.
\end{abstract}

1. Introduction

Consider the varying coefficient model

\[ Y = X^T a(U) + e, \quad (1) \]

where \( Y \in \mathbb{R} \) is a response variable, \( X \in \mathbb{R}^p \) and \( U \in \mathbb{R} \) are covariates, \( a(\cdot) \) is a functional vector from \( \mathbb{R} \) to \( \mathbb{R}^p \), and \( e \) is the error variable with \( E[e|X, U] = 0 \) and \( \text{Var}(e|X, U) = \sigma^2(U) \).

The varying coefficient model is an useful extension of the classical linear regression model. Many models proposed in the literature may be viewed as special cases of (1) (see [1]). Two obvious advantages of the models are that the modeling bias can significantly be reduced and the “curse of dimensionality” can be avoided (see [2]). It addresses an issue frequently encountered by investigators in practical studies. In this case it allows one to explore the extent to which covariates affect responses over a certain random variable. In recent years, varying coefficient models have been paid considerable attention and investigated extensively. For the models, Hastie and Tibshirani [3] proposed smoothing spline and kernel methods. Wu et al. [4] modified the kernel method by allowing different smoothing parameters for different coefficient functions. Fan and Zhang [2,1] proposed two different two-step procedures, respectively. Wu and Chiang and Chiang et al. [5,6] proposed component-based kernel and smoothing spline estimators. Huang et al. [7] developed a global smoothing procedure using a basis function approximation. In [8], smoothing spline and kernel methods were studied. For nonlinear time series applications, see [9,10], among others.

\textsuperscript{*} Corresponding author at: Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, PR China.
E-mail addresses: qhwang@amss.ac.cn (Q. Wang), zhangriquan@163.com (R. Zhang).
Recently, increasing attention has been paid to studies of error-in-covariables models with validation sampling, where complete covariate information is collected only on a sample from the full study cohort. Studies of this type may arise when collecting exact covariate data for the entire cohort is difficult or expensive in time and/or cost, and hence exact measurements could be made for a relative small subset of subjects enrolled in the study to save expense. However, in such studies, usually crude or surrogate covariable information is available on every study subject. For example, in the measurement of heart muscle damage caused by a myocardial infection, peak cardiac enzyme level in the bloodstream is an easily obtained variable, but this cannot assess accurately the damage to the heart muscle. Instead, an artheroscintograph, an invasive and expensive procedure, can be used to produce a more accurate measure of the heart muscle for a small subset of subjects enrolled in the study [11]. Such phenomena also happen in the evaluation of smoking behavior. In school-based smoking prevention projects, current smoking behavior data are generally obtained through self-reporting using questionnaires; self-reported data are relatively inexpensive to obtain but may be subject to error. Expensive chemical analyses of saliva samples for the presence of nicotine can be performed for at most only a small subset of the subjects enrolled in these large-scale studies in order to yield a more accurate evaluation of smoking behavior [12]. Here, variables with error such as diagnostic data of heart damage by peak cardiac enzyme levels in the bloodstream, and self-reporting of smoking behavior are used as surrogate variables and the exact measures for a relatively small subset are used as validation variables. Other examples where validation data are available can be found in [11,13,14] among others.


In the present paper, we consider the model (1) with explanatory variable $X$ measured with error and both $Y$ and $U$ measured exactly. That is, instead of the true $X$, the surrogate variable $\tilde{X}$ is observed. Most papers assume that the measurement error in covariables is additive in the absence of validation sampling, and hence standard methodology can be used to handle this case. In practice, however, the additive model is often not appropriate. The realistic case may be that no error model structure is assumed. In this case, one solution is to use the help of validation data to capture the underlying relation between the true variables $X$ and surrogate variables $\tilde{X}$. The underlying relation between $X$ and $\tilde{X}$ can be evaluated by the regression of $X$ on $\tilde{X}$. This motivates us to use the regression calibration to establish the regression relationship between $Y$ and $(X, U)$ based on model (1). The regression relationship can be described by the standard varying coefficient model (1) with $X$ replaced by $E[X|\tilde{X}, U]$ and the error term replaced by an error term whose conditional expectation is zero given $(\tilde{X}, U)$. Without specifying any error structure equation and the distribution of $X$ given $\tilde{X}, E[X|\tilde{X}, U]$ can be estimated by a local linear method with the validation data. Estimators of the coefficient function vector $a(\cdot)$ are then developed by combining the standard local linear kernel smoothing technique and the least squares method with $X$ replaced by the local linear regression estimator of the regression of $X$ on $(\tilde{X}, U)$. We define two estimators of $a(\cdot)$ by considering the two cases where the response $Y$ is available or not in the validation set. Asymptotic results for the two estimators are derived, showing that the two proposed estimators are asymptotically normal, with bias and asymptotic variance estimated by a bootstrap approach.

A local polynomial method with order $p$ can also be used for our estimating problem. As pointed out by Fan and Gijbels [28], however, fitting polynomials of higher order leads to a possible reduction of the bias, but on the other hand also to an increase of the variability, caused by introducing more local parameters. Furthermore, Fan and Gijbels [28] also mentioned that the choice $p = v + 1$ suffices, where $v$ is the derivative order of the unknown coefficient function. We here consider the estimating problem of the coefficient function. Hence, a local linear method suffices for our estimating problems.

This paper is organized as follows. In Section 2, we describe the estimation procedures based on local linear kernel smoothing techniques. In Section 3, we develop the theoretical framework for the proposed estimators, which are proved to be asymptotically normal. Section 4 is devoted to the choices of smoothing parameters. A simulation study is conducted to evaluate the finite sample properties of the proposed estimators empirically in Section 5. Some concluding remarks are given in Section 6. Technical proofs are presented in the Appendix.

2. Estimation

Suppose that $\tilde{X}$ is a $d$-dimensional surrogate variable for $X$. Assume that we have a primary data set containing $N$ independent and identically distributed observations of $\{(Y_j, \tilde{X}_j, U_j)_{j=n+1}^{n+N}\}$ and a validation set containing $n$ independent and identically distributed observations of $\{(X_i, \tilde{X}_i, U_i)_{i=1}^{n}\}$ or $\{\tilde{Y}_j, \tilde{X}_j, U_j\}_{j=1}^{n}$. It is also assumed that the two observation subsets are independent and that the variable $(\tilde{X}, U)$ in the primary data set and the one in the validation data set are identically distributed.
Since no error equation or distribution assumption of \( X \) given \( \tilde{X} \) is specified, the estimation problem of \( a(U) \) cannot be handled directly with standard methodology. To use the information of the surrogate variable \( \tilde{X} \), we rewrite model (1) such that \( Y \) is related to \( (\tilde{X}, U) \). Let \( \tilde{V} = (\tilde{X}, U)^T \) and \( u(\tilde{V}) = E(X|\tilde{V} = \tilde{v}) \). Then the model (1) can be rewritten as
\[
Y_j = u^T(\tilde{V})a(U_j) + \epsilon_j, \quad j = n + 1, \ldots, n + N,
\]
where \( \epsilon_j = \epsilon \cdot X_j^T a(U_j) - u^T(\tilde{V})a(U_j) \). Throughout this paper, we assume
\[
E(\epsilon_j|\tilde{X}_j, X_j, U_j) = E(\epsilon_j|\tilde{X}_j, U_j).
\]
It is reasonable to make the above assumption since \( \tilde{X} \) provides no extra information on \( Y \) given \((X, U)\). This assumption together with the assumption \( E(\epsilon_j|\tilde{X}_j, X_j, U_j) = 0 \) proves \( E(\epsilon_j|\tilde{V}_j) = E(E(\epsilon_j|\tilde{X}_j, X_j, U_j)|\tilde{V}_j) = E(E(\epsilon_j|\tilde{X}_j, U_j)|\tilde{V}_j) = 0 \) for \( j = 1, 2, \ldots, n \).

Clearly, (2) is the standard varying coefficient model if \( u(\cdot) \) is a known function. In this case, the existing methods can be applied to defining estimators of \( a(\cdot) \) based on the primary data set. However, \( u(\cdot) \) is unknown in practice. The obtained estimators depend on the unknown function. Fortunately, \( u(\cdot) \) can be estimated consistently by the local linear kernel smoothing technique with validation data. Assume that \( u(\cdot) \) has continuous second partial derivatives. Then, every entry of \( u(\cdot) \) can be approximated locally at \( \tilde{v} \) by a linear function vector, i.e., \( u(\cdot) \approx \alpha + \beta(\cdot - \tilde{v}) \). Here \( \tilde{v} \) is a \((d + 1)\)-dimensional vector, \( \alpha \) is a \( p \)-dimensional vector and \( \beta \) is a \( p \times (d + 1) \) matrix. The local linear estimator of \( u(\tilde{v}) \) is then defined by \( \hat{u}(\tilde{v}) = \hat{\alpha} \), where \( \hat{\alpha} \) can be obtained by minimizing the sum of weighted squares
\[
\sum_{i=1}^{n} (X_i - \alpha - \beta(\tilde{v}_i - \tilde{v}))^T(X_i - \alpha - \beta(\tilde{v}_i - \tilde{v}))K_{h_1, h_2}(\tilde{v}_i - \tilde{v}) \tag{3}
\]
on \((\alpha, \beta)\), where \( K_{h_1, h_2}(\cdot) = K_1(\cdot/h_1)K_2(\cdot/h_2) \) is a \((d + 1)\)-dimensional kernel function, and \( h_1 = h_{1,n} > 0 \) is a bandwidth. It follows from least squares theory that
\[
\hat{u}(\tilde{v}) = e_{1 \times (d+2)}^T \sum_{i=1}^{n} [Q_i^T(\tilde{v})W_i(\tilde{v})Q_i(\tilde{v})]^{-1}Q_i^T(\tilde{v})K_{h_1, h_2}(\tilde{v}_i - \tilde{v})X_i, \tag{4}
\]
where \( e_{1 \times (d+2)} \) is a unit row vector with 1 at its first position, \( Q_i^T(\tilde{v}) = (1, (\tilde{v}_i - \tilde{v})^T/h_1)^T \), \( Q_i(\tilde{v}) \) is an \( n \times (d + 2) \) matrix with the transpose of \( Q_i^T(\tilde{v}) \) as its \( i \)-th row, and \( W_i(\tilde{v}) = \text{diag}(K_{h_1, h_2}(\tilde{v}_i - \tilde{v}), \ldots, K_{h_1, h_2}(\tilde{v}_n - \tilde{v})) \).

Assume that \( a(\cdot) \) has continuous second derivatives. Then every entry of \( a(\cdot) \) can be approximated locally at \( U \) by a linear function, i.e., \( a(\cdot) \approx a + b(\cdot - \tilde{u}) \). Here \( a \) and \( b \) are two \( p \)-dimensional vectors. If \( Y \) is only observed in the primary data set, i.e., the validation set is \( \{ (X_i, X_i, U_i) \}_{i=1}^n \), we may define the estimator of \( a(U) \) by minimizing the sum of weighted squares
\[
\sum_{j=n+1}^{n+N} \left[ Y_j - (a + b(U_j - u))^T \tilde{u}(\tilde{V}_j) \right]^2 K_{h_2, h_3}(U_j - u) \tag{5}
\]
with respect to \( a \) and \( b \), where \( K_{h_2, h_3}(\cdot) = K_2(\cdot/h_2)K_3(\cdot/h_3) \) is a kernel function, and \( h_2 = h_{2,n} > 0 \) is a bandwidth sequence. Let \( \hat{a} \) and \( \hat{b} \) be the minimizers of (5). We define \( \hat{a}(U) = \tilde{a} \) as the local linear estimator of \( a(u) \). By least squares theory, we have
\[
\hat{a}(U) = E_{p \times 2p} \sum_{i=1}^{n+N} \left[ Q_i^T(u)W_2(u)Q_2(u)^{-1}Q_2^T(u)K_{h_2, h_3}(U_j - u)Y_j \right] \tag{6}
\]
where \( E_{p \times 2p} \) is a matrix with \( e_{1 \times 2p} \) as its \( l \)-th row, \( e_{1 \times 2p} \) is a unit row vector with 1 at its \( l \)-th position, \( l = 1, 2, \ldots, p \), \( Q_2^T(u) = (\tilde{u}^T(\tilde{V}_j), (\tilde{u}^T(\til{V}_j))(U_j - u)/h_2)^T \), \( Q_2(u) \) is an \( N \times 2p \) matrix with the transpose of \( Q_2^T(u) \) as its \( (j - n) \)-th row, and \( W_2(u) = \text{diag}(K_{h_2, h_3}(U_1 - u), \ldots, K_{h_2, h_3}(U_N - u)) \).

In practice, the response variable \( Y \) may be fully observed, so that \( Y \) can be measured in the validation data set, i.e. the validation data set is \( \{ (Y_i, X_i, X_i, U_i) \}_{i=1}^n \). In this case, an alternative estimator of \( a(\cdot) \), say \( \hat{a}(\cdot) \), can be obtained by minimizing the sum of weighted squares
\[
\sum_{i=1}^{n} \left[ Y_i - (a + b(U_i - u))^T \tilde{u}(\til{V}_j) \right]^2 K_{h_3}(U_i - u) + \sum_{j=n+1}^{n+N} \left[ Y_j - (a + b(U_j - u))^T \tilde{u}(\til{V}_j) \right]^2 K_{h_3}(U_j - u) \tag{7}
\]
on \( a \) and \( b \), where \( K_{h_3}(\cdot) = K_3(\cdot/h_3)K_3(\cdot/h_3) \) is a kernel function, and \( h_3 = h_{3,n} > 0 \) is a bandwidth sequence. By least squares theory, we have
\[
\hat{a}(U) = E_{p \times 2p} \sum_{i=1}^{n} \left[ Q_i^T(u)W_3(u)Q_3(u) + Q_4^T(u)W_4(u)Q_4(u)^{-1}Q_3^T(u)K_{h_3, h_3}(U_i - u)Y_i \right]
+
E_{p \times 2p} \sum_{j=n+1}^{n+N} \left[ Q_i^T(u)W_5(u)Q_3(u) + Q_4^T(u)W_4(u)Q_4(u)^{-1}Q_3^T(u)K_{h_3, h_3}(U_j - u)Y_j \right], \tag{8}
\]
where $Q_3^{(i)}(u) = (X_i^T, X_i^T((U_i − u)/h_3))^T$, $Q_3(u)$ is an $n \times 2p$ matrix with the transpose of $Q_3^{(i)}(u)$ as its $i$-th row, and 
$W_2(u) = \text{diag}(K_{3}h_3(U_1 − u), \ldots , K_{3}h_3(U_n − u))$; $Q_4(u)$, $Q_4^{(i)}(u)$ and $W_4(u)$ are $Q_2(u)$, $Q_2^{(i)}(u)$ and $W_2(u)$, respectively, with $h_2$ replaced by $h_3$.

3. Asymptotic property

To state our results we introduce the following assumptions.

A.1. $u(\cdot)$ has continuous second partial derivatives.
A.2. $a(\cdot)$ has a continuous second derivative.
A.3. The density $f_1(\tilde{v})$ of $\tilde{V}$ is Lipschitz continuous, and satisfies $\inf_{v \in \mathcal{V}} f_1(\tilde{v}) > C$, where $C$ is a positive constant number and $\mathcal{V}$ is the compact support of $\tilde{V}$.
A.4. The density $f_{2}(u)$ of $U$ is Lipschitz continuous.
A.5. There exists a constant $\delta > 0$, such that for every component $\tilde{V}_a$ of $\tilde{V}$, $E(|\tilde{V}_a|^{2+\delta}) < \infty$, and $E(|Y|^{2+\delta}) < \infty$.
A.6. Every component of the conditional expectations $E(u(\tilde{V})u^T(\tilde{V})|U = u)$ and $E(XX^T|U = u)$ is Lipschitz continuous.
A.7. $K_1(\cdot)$ is a $(d + 1)$-dimensional, continuous and symmetric probability density function with bounded support. For simplicity, without loss of generality, let us take the product kernel $K_1(\tilde{v}) = \prod_{a=1}^{d+1} K_0(\tilde{v}_a)$, where $K_0(\cdot)$ is a univariate probability density kernel function which is symmetric about zero and Lipschitz continuous, where $\tilde{v}_a$ is the $a$-th component of $\tilde{v}$.
A.8. Both $K_2(\cdot)$ and $K_3(\cdot)$ are probability density functions which are univariate symmetric about zero, compact supported and Lipschitz continuous.

A.9. $h_1 \to 0$, $\frac{nh_1+1}{\ln(1/h_1^{2+\delta})} \to \infty$ as $n \to \infty$.
A.10. $h_2 \to 0$, $Nh_2 \to \infty$ as $N \to \infty$.
A.11. $h_1/h_2 \to 0$.
A.12. $h_3 \to 0$, $nh_3 \to \infty$, and $Nh_3 \to \infty$ as $n, N \to \infty$.
A.13. $h_1/h_3 \to 0$.

Remark 1. Assumptions A.1–A.4, A.6–A.8, A.10 and A.12 are some standard assumptions for local linear regression (see [28]). Assumptions A.5 and A.9 are to ensure that the estimator $\hat{u}(\tilde{V})$ converges to $u(\tilde{V})$ in $\mathcal{V}$ uniformly. Assumptions A.11 and A.13 are common assumptions in two-step estimation for a nonparametric setting.

Theorem 1. Under assumptions A.1–A.11, we have

$$\sqrt{Nh_2} \left( \hat{a}(u) - a(u) - \frac{\mu_2 h_2^2}{2} a''(u) \right) \overset{D}{\rightarrow} N \left( 0, \frac{v_{2,2}}{f_2(u)} \Sigma^{-1}(u) \right)$$

provided that $f_2(u) \neq 0$, where $\mu_{2,2} = \int t^2 K_2(t)dt$, $v_{2,2} = \int K_2^2(t)dt$, $\lambda = \lim(N/n)$ is a positive constant, and

$$\Sigma(u) = E[u(\tilde{V})u^T(\tilde{V})|U = u],$$
$$\Sigma_1(u) = E \left[ u(\tilde{V})u^T(\tilde{V})(Y - u(\tilde{V})a(U))^2 | U = u \right],$$
$$\Sigma_2(u) = E \left[ u(\tilde{V})u^T(\tilde{V})a^2(U)((X - u(\tilde{V}))(X - u(\tilde{V}))^T a(U)) | U = u \right].$$

Remark 2. The first component in the asymptotic covariance is the amount of information in the sample by modeling (1) as the regression relationship between $Y$ and $\tilde{V}$. The second component is the extra dispersion caused by the local linear regression estimation of the unknown mean $X$ given $\tilde{V}$ using the validation data set. If $\lambda = 0$ or $u(\tilde{V}) = X$, the second component in the asymptotic covariance disappears, and the asymptotic covariance is the same as that in [2], which implies that our result recovers that of Fan and Zhang [2].

Remark 3. Relatively, in practice, $N$ is far larger than $n$, i.e. $\lambda$ is far larger than 1. For example, $(N, n)$ can be taken to be (50, 5), (100, 10), (200, 20), (300, 30) and so on when $\lambda = 10$. One of the referees asked the question: Are the values $\lambda = 0$ or $\infty$ admissible? Theoretically, $\lambda = 0$ is admissible. It implies that the size of the primary data versus that of the validation data is very small and the proposed estimators have the same asymptotic properties as the standard estimators with complete data. Actually, in this case, the primary data can be ignored and all the true observations in the validation data set can be used to define estimators with standard methods, and hence we do not recommend our method for this special case. The case of $\lambda = \infty$ implies that the size of the validation data versus that of the primary data is too small, so that the validation data cannot provide sufficient information to evaluate the underlying relation between $X$ and $\tilde{X}$. In this case, the procedure of the proposed estimation does not work well.
Theorem 2. Under assumptions A.1–A.9, A.12 and A.13, we have

\[
\sqrt{(n + N)h_3} \left( \hat{a}(u) - a(u) - \frac{\mu_{3,2} h_3^2 (\mu_{3,2} h_3^2)'}{2} a''(u) \right) \\
\xrightarrow{D} N \left( 0, \frac{\nu_{3,2}}{f_2(u)} \Sigma_4^{-1}(u) \left[ \frac{1}{1 + \lambda} \Sigma_3(u) + \frac{\lambda}{1 + \lambda} (\lambda_1(u) + \lambda_2(u)) \right] \Sigma_4^{-1}(u) \right),
\]

(11)

provided that \( f_2(u) \neq 0 \), where \( \mu_{3,2} = \int \frac{t^2 K_3(t)}{dt}, \nu_{3,2} = \int K_2^2(t)dt \), and

\[
\begin{align*}
\Sigma_3(u) &= E[XX^T|U = u], \\
\Sigma_4(u) &= \frac{1}{1 + \lambda} \Sigma_3(u) + \frac{\lambda}{1 + \lambda} \Sigma(u), \\
\Sigma_5(u) &= E[XX^T(Y - X^T a(U))^2|U = u].
\end{align*}
\]

(12)

Remark 4. Note that

\[
\begin{align*}
\Sigma_3(u) - \Sigma(u) &= E\left[ XX^T - u(\tilde{V})u^T(\tilde{V}) | U = u \right] \\
&= E\left[ E[XX^T|\tilde{V}] - E(X|\tilde{V})E(X^T|\tilde{V}) | U = u \right] = E[\text{Var}(X|\tilde{V})| U = u] \geq 0,
\end{align*}
\]

we have \( \Sigma_3(u) \) is larger than \( \Sigma(u) \). Similarly, \( \Sigma_2 + \lambda \Sigma_2(u) \) is larger than \( \Sigma_2(u) \). Hence the asymptotic variance of \( \hat{a}(u) \) is larger than that of \( \tilde{a}(u) \) (we here make a convention that we say that a matrix \( A \) is larger than a matrix \( B \) (or \( A > B \)) if \( A - B \) is a positive definite matrix). When the primary data set is far larger than the validation data set such that \( \lambda/(1 + \lambda) \approx 1 \) and \( 1/(1 + \lambda) \approx 0 \), the variance of \( \hat{a}(u) \) is approximately equal to that of \( \tilde{a}(u) \). In this case, we recommend \( \tilde{a}(u) \) since there is little difference in the asymptotic efficiency between \( \hat{a}(u) \) and \( \tilde{a}(u) \), a verified result in our simulation study, and \( \tilde{a}(u) \) is simpler for calculation.

Clearly, the asymptotic covariances of \( \hat{a}(u) \) and \( \tilde{a}(u) \) can be estimated by combining the sample moment method and the “plug in” method. However, the covariance structures of the two estimators are complicated since some nonparametric regression functions such as \( u(\cdot), \Sigma(u) \) and so on are involved and are hence difficult to be estimated well.

Another alternative is to use the bootstrap approach to estimate the asymptotic variances. Next, we present the steps of constructing a bootstrap variance estimate of \( \hat{a}(u) \) only. The bootstrap variance estimation of \( \tilde{a}(u) \) can be constructed similarly.

Step 1: Generate bootstrap samples.

Let \( \xi_a \) be the empirical distribution of validation data \{\( (X_i, \hat{X}_i, U_i)_{i=1}^n \}. \) Given \{\( (X_i, \hat{X}_i, U_i)_{i=1}^n \}, draw an independent and identically distributed bootstrap sample \{\( (X_i^*, \hat{X}_i^*, U_i^*, v_i)_{i=1}^m \}. \) Let \( \xi_u \) be the empirical distribution of primary data set \{\( (Y_j, \hat{X}_j, U_j)_{j=m+1}^{n+m} \}. \) Given \{\( (Y_j, \hat{X}_j, U_j)_{j=m+1}^{n+m} \}, draw an independent and identically distributed bootstrap sample \{\( (Y_j^*, \hat{X}_j^*, U_j^*)_{j=m+1}^{n+m} \) from \( \xi_u \). It is assumed that the primary and validation data sets are bootstrapped independently.

Step 2: Define a bootstrap version of \( \hat{a}(u) \).

Let \( \tilde{u}(\tilde{V}) \) be the bootstrap analogue of \( \hat{u}(\tilde{V}) \) defined in (4). Then, the bootstrap version of \( \hat{a}(u) \) can be given by (6) by substituting \{\( (Y_j, \hat{X}_j, U_j)_{j=m+1}^{n+m} \) with \{\( (Y_j^*, \hat{X}_j^*, U_j^*)_{j=m+1}^{n+m} \) and \( \hat{u}(\tilde{V}) \) with \( \tilde{u}(\hat{V}) \).

Step 3: Repeat the first step and second step \( B \) times. We then obtain \( B \) bootstrap estimators \( \hat{a}_l(u), l = 1, \ldots, B \). The bias of \( \hat{a}(u) \) can be estimated by

\[
\frac{1}{B} \sum_{l=1}^B \hat{a}_l(u) - \hat{a}(u)
\]

and the asymptotic covariance of \( \hat{a}(u) \) can be estimated by

\[
\left( \frac{1}{B-1} \sum_{l=1}^B (\hat{a}_l(u) - \tilde{a}_l(u))(\tilde{a}_l(u) - \tilde{a}(u))^T \right)
\]

where \( \tilde{a}_l(u) = \frac{1}{B} \sum_{l=1}^B \hat{a}_l(u) \).

4. Bandwidth selection

It is well known that one of the crucial points in applying a local linear regression estimate is the choice of the bandwidths.

In our setting, the estimator \( \tilde{a}(\cdot) \) (or \( \hat{a}(\cdot) \)) involves two smoothing parameters, bandwidths \( h_1 \) and \( h_2 \) (or \( h_1 \) and \( h_3 \)). From Theorem 1 (or Theorem 2), we can see that the variance of \( \tilde{a}(\cdot) \) (or \( \hat{a}(\cdot) \)) is of order \( O((Nh_1)^{-1/2}) \) (or \( O((n+N)h_3)^{-1/2}) \), and its bias is of order \( O((h_1^2)) \) (or \( O(h_3^2) \)). This indicates that a proper choice of \( h_1 \) specified in assumption A.11 (or A.13) does not affect the first-order terms of the mean square error, although it might affect higher-order terms. This shows that the
selection of $h_1$ might not be so critical to $\hat{a}(\cdot)$ (or $\tilde{a}(\cdot)$). Hence, here we use the optimal bandwidth $\hat{h}_{1,0}$ obtained by applying cross-validation [see, e.g. [8]] to $\hat{u}_n(\cdot)$ although $h_1$ can take an arbitrary positive value. After $h_1$ is given, the selection of $h_2$ or $h_3$ reduces to a bandwidth selection problem for a univariate nonparametric smoothing. One can apply univariate bandwidth selection procedures such as cross-validation [29], generalized cross-validation (Hoover et al. 1997), the presample substitution method [30], a plug-in bandwidth selector [31] and the empirical bias method [32] to select $h_2$ or $h_3$.

Here we use the data-driven cross-validation bandwidth selection method for $h_2$ and $h_3$, respectively, when $h_1$ is given. We select $h_2$ by minimizing

$$ CV(h_2) = \frac{1}{n} \sum_{j=n+1}^{n+N} (Y_j - \hat{u}(V_j)\tilde{a}^{-1}(U_j))^2 \omega_1(V_j), $$

where $\omega_1(\cdot)$ is the weight function which allows elimination of boundary effects by considering $\omega_1(\cdot)$ to be supported on a compact subset of the support of $V$, and $\tilde{a}^{-1}(\cdot)$ is a version of $\tilde{a}(\cdot)$ by leaving $(Y_j, V_j, U_j)$ out for $j = n + 1, n + 2, \ldots, n + N$.

We select $h_3$ by minimizing

$$ CV(h_3) = \frac{1}{n + N} \sum_{k=1}^{n+N} (Y_k - \tilde{Y}_k)^2 \omega(C_k), $$

where

$$ \tilde{Y}_k = \begin{cases} X_k^T (\tilde{a}^{-1}(U_k)) & \text{if } 1 \leq k \leq n, \\ \hat{u}(V_k) \tilde{a}^{-1}(U_k) & \text{if } n + 1 \leq k \leq n + N, \end{cases} $$

$$ \omega(C_k) = \begin{cases} \omega_2(X_k, U_k) & \text{if } 1 \leq k \leq n, \\ \omega_3(V_k) & \text{if } n + 1 \leq k \leq n + N, \end{cases} $$

$\omega_2(\cdot, \cdot)$ is a weight function which allows elimination of boundary effects by considering $\omega_2(\cdot, \cdot)$ to be supported on a compact subset of the support of $X$, $U$, and $\hat{a}^{-1}(\cdot)$ is a version of $\hat{a}(\cdot)$ by leaving $(Y_k, X_k, U_k)$ out for $1 \leq k \leq n$, or $(Y_k, \tilde{V}_k, U_k)$ out if $n + 1 \leq k \leq n + N$. In practice, for simplicity, these weighted functions can be taken to be indicator functions on some compact supports.

5. Simulation

In our simulated example, the estimators were assessed by the square root of averaged squared errors (RASE), $\text{RASE} = \sum_{r=1}^{p} \text{RASE}_r$, where

$$ \text{RASE}_r = \left( \frac{1}{n} \sum_{s=1}^{n_r} (\hat{a}_r(s) - a_r(s))^2 \right)^{1/2}, $$

and $\{s_l, l = 1, \ldots, n_r\}$ are regular grid points, $\hat{a}_r(\cdot)$ is an estimator of $a_r(\cdot)$, $r = 1, \ldots, p$, and $a_r(\cdot)$ is the $r$-th component of $a(\cdot)$.

To show the performance of the proposed estimators $\hat{a}(\cdot)$ and $\tilde{a}(\cdot)$ in Section 2, we compared them with two estimators: the naive estimator and the gold standard estimator. The naive estimator was obtained by ignoring the measurement error and applying the standard approach, and the standard estimator was obtained by using the complete data and the standard estimating approach under model (1). The standard estimator can serve as a gold standard in the simulation study even though it is practically unachievable.

We consider the varying coefficient model

$$ Y = a_1(U)X_1 + a_2(U)X_2 + \epsilon, $$

where $a_1(u) = u + 8 \exp(-16u^2)$, $a_2(u) = 2 + \sin(5\pi u)$, and $\epsilon$ has a standard normal distribution. The predictor $X = (X_1, X_2)^T$ is measured with error. The surrogate vector $\tilde{X} = (\tilde{X}_1, \tilde{X}_2)^T$ was generated such as $\tilde{X} = 1.25X^2 + 0.68X + \delta$, where $\delta$ has a bivariate standard normal distribution, $X \sim N(0, 0.5^2(2^{1/2} 2^{1/2}))$, $\epsilon \sim U(-0.5, 0.5); \delta$ is the standard deviation of the measurement error; $X, U, \xi$ and $\epsilon$ are independent. The simulations run with validation data and primary data sizes of $(n, N)$. For all kernel functions used in this simulation, we take the Epanechnikov kernel, $K(u) = 0.75(1 - u^2)$, if $|u| < 1, 0$ otherwise. For the naive estimators and the gold standard estimators, the bandwidths are chosen by the cross-validation method. For the estimators proposed in Section 2, the bandwidths are chosen by the methods mentioned in Section 4.

To show the effects of the rate of the size of primary data to that of the validation data and that of the standard deviation of the measurement error on the proposed estimators, four cases are studied, which are $n = 20, N = 200$, $\delta = 0.35$ or $\delta = 0.75$, and $n = 100, N = 200$, $\delta = 0.35$ or $\delta = 0.75$. For every case, we replicated the simulation 400 times. Figures (a) and (b) in Figs. 1–4 present the four estimates of $a_1(\cdot)$ and $a_2(\cdot)$, respectively, where the true lines are shown by solid
lines, the gold standard estimators by dashed lines, the proposed estimators $\hat{a}(\cdot)$ by dotted lines, $\tilde{a}(\cdot)$ by dash–dotted lines, and the naive estimators by starred lines. The boxplots for the 400 RASEs are presented in Figures (c) in Figs. 1–4, where $R_{N1}$ and $R_{N2}$ are the RASEs of the naive estimators of $a_1(\cdot)$ and $a_2(\cdot)$, respectively, $R_{g1}$ and $R_{g2}$ are the RASEs of the gold standard estimators of $a_1(\cdot)$ and $a_2(\cdot)$, respectively, $R_{t1}$ and $R_{t2}$ are the RASEs of $\hat{a}_1(\cdot)$ and $\hat{a}_2(\cdot)$, respectively, $R_{w1}$ and $R_{w2}$ are the RASEs of $\tilde{a}_1(\cdot)$ and $\tilde{a}_2(\cdot)$, respectively, and $R_N = R_{N1} + R_{N2}$, $R_g = R_{g1} + R_{g2}$, $R_t = R_{t1} + R_{t2}$, and $R_w = R_{w1} + R_{w2}$. Simultaneously, 95% confidence bands for $(\hat{a}_1(\cdot), \hat{a}_2(\cdot))^\top$ are given by the bootstrap in Figs. 1(d)–4(d) and 1(e)–4(e), respectively, where these estimated curves are shown by solid lines, and the confidence bands by dash–dotted lines.

From Figs. 1–4, we can see that the naive estimators have much larger RASEs than the gold standard estimators and the proposed estimators in all cases. This suggests that the naive estimators perform poorly. The proposed estimators have a slight larger RASEs than the gold standard estimators, which implies that the proposed estimators $\hat{a}(\cdot)$ and $\tilde{a}(\cdot)$ work well.

**Fig. 1.** Simulation results for the case $n = 20$, $N = 200$, $\delta = 0.35$.  

From Figs. 1(a)–4(a) and 1(b)–4(b), it is difficult to distinguish the lines of \( \hat{\alpha}(\cdot) \) and \( \tilde{\alpha}(\cdot) \) because the two lines are very close. From Figs. 1(c)–4(c), the RASE of \( \tilde{\alpha}(\cdot) \) is very slightly larger than that of \( \hat{\alpha}(\cdot) \). We suggest using \( \tilde{\alpha}(\cdot) \) because it is much simpler.

Comparing Fig. 1 with Fig. 3 and Fig. 2 with Fig. 4, it can be seen that these RASEs increase with \( \lambda \) increasing. Comparing Fig. 1 with Fig. 2 and Fig. 3 with Fig. 4, we can see that these RASEs increase as the standard deviation of the measurement error increases.

6. Concluding remarks

When the dimension of \( X \) and hence of \( \tilde{V} \) is large, the curse of dimensionality may occur for estimating \( E(X|\tilde{V}) \). In this case, a more appealing approach is to consider a dimension-reduction method by assuming that \( E(X|\tilde{V}) = (g_1(\beta_1^T\tilde{V}), \ldots, g_p(\beta_p^T\tilde{V}))^T \). The parametric vectors \( \beta_1, \ldots, \beta_p \) can first be estimated by sliced inverse regression techniques.
We can then estimate \( g_1(\cdot), \ldots, g_p(\cdot) \) by the local linear regression method. After obtaining the estimate of \( \mathbb{E}(X|\tilde{V}) \), we can use a local linear smooth technique similar to (5) and (6), or (7) and (8) to estimate the functional coefficient vector \( a(\cdot) \) in the model (1), and obtain results similar to those in previous sections.

**Acknowledgments**

This research was supported by National Science Fund for Distinguished Young Scholars (10725106), National Natural Science Foundation of China (10671198 and 10871072), National Science Fund for Creative Research Groups and a grant from the Research Grants Council of the Hong Kong, China (HKU 7050/06P).

Our sincere thanks are due to the editor and the two referees for their valuable comments and helpful suggestions which have led to substantial improvements in the presentation of the manuscript.
Appendix

Lemma 1. Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. random vectors, where \(Y_i\) is a scalar random variable and \(X_i\) is a \(q\)-dimensional random vector. Assume further that \(E|y|^s < \infty\) and \(\sup_x |y|^s f(x, y)dy < \infty\), where \(f(\cdot, \cdot)\) denotes the joint density of \((X, Y)\). Let \(K(\cdot)\) be a bounded positive function with a bounded support satisfying a Lipschitz condition. Then

\[
\sup_{x \in \mathcal{X}} \left| \frac{1}{n} \sum_{i=1}^{n} [K_h(X_i - x)Y_i - E(K_h(X_i - x)Y_i)] \right| = O_p \left( \left\{ \frac{\ln(1/h^q)}{nh^{q-s}} \right\}^{1/2} \right),
\]

provided that \(n^{2\varepsilon-1}h^q \to \infty\) for some \(\varepsilon < 1 - s^{-1}\), where \(\mathcal{X}\) is the compact support of \(X\).
Similar to [2], Lemma 1 follows immediately from the result obtained by [35].

**Lemma 2.** If assumptions A.1–A.5 hold, we have, uniformly for \( \tilde{v} \in \mathcal{V} \),

\[
S_1 \equiv \frac{1}{n} \sum_{j=1}^{N} Q_j^T(\tilde{v}) W_j(\tilde{v}) Q_j(\tilde{v}) = f_1(\tilde{v}) \text{diag}(1, \mu_{0,2}, \ldots, \mu_{0,2}) + O_p(h_1),
\]

(A.1)

where \( \mu_{0,2} = \int t^2 K_0(t) dt \) with \( K_0(\cdot) \) defined in assumption A.7.

**Proof.** Notice that

\[
S_1 = \begin{pmatrix}
\frac{1}{n} \sum_{i=1}^{n} K_{1,h_1}(\tilde{v}_i - \tilde{v}) & \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right)^T K_{1,h_1}(\tilde{v}_i - \tilde{v}) \\
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right) K_{1,h_1}(\tilde{v}_i - \tilde{v}) & \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right)^T \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right) K_{1,h_1}(\tilde{v}_i - \tilde{v})
\end{pmatrix}.
\]

The elements of \( S_1 \) have the form of \( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right)^{\otimes \lambda} K_{1,h_1}(\tilde{v}_i - \tilde{v}) \), \( \lambda = 0, 1, 2 \), and

\[
E \left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\tilde{v}_i - \tilde{v}}{h_1} \right)^{\otimes \lambda} K_{1,h_1}(\tilde{v}_i - \tilde{v}) \right) = f_1(\tilde{v}) \int t^{\otimes \lambda} K_1(t) dt + O(h_1),
\]

uniformly \( \tilde{v} \in \mathcal{V} \), where \( D^{\otimes 0} = 1, D^{\otimes 1} = D, D^{\otimes 2} = DD^T \) for every matrix \( D \), so we have (A.1). \( \square \)

**Lemma 3.** Under assumptions A.1–A.5, we have

\[
\hat{u}(\tilde{v}) - u(\tilde{v}) = f_1^{-1}(\tilde{v}) \{ M_{1,n}(\tilde{v}) + R_{1,n}(\tilde{v}) \} \left( 1 + o_p(1) \right),
\]

(A.2)

where \( M_{1,n}(\tilde{v}) = \frac{1}{n} \sum_{i=1}^{n} K_{1,h_1}(\tilde{v}_i - \tilde{v}) \left( u(\tilde{v}_i) - (u_1(\tilde{v}), \ldots, u_{d+1}(\tilde{v}))^T (\tilde{v}_i - \tilde{v}) \right) \) and \( R_{1,n}(\tilde{v}) = \frac{1}{n} \sum_{i=1}^{n} K_{1,h_1}(\tilde{v}_i - \tilde{v}) (X_i - u(\tilde{v}_i)). \)

Furthermore,

\[
\sup_{\tilde{v} \in \mathcal{V}} |\hat{u}(\tilde{v}) - u(\tilde{v})| = O_p(d_1),
\]

(A.3)

where \( d_1 = h_1^2 + \left\{ \frac{\ln(1/h_1^{d+1})}{m^{d+1}} \right\}^{1/2} \). \( u_1'(\tilde{v}) \) is the first partial derivative vector.

**Proof.** Let \( \hat{u}_r(\tilde{v}), u_r(\tilde{v}) \) and \( X_{ir} \) be the \( r \)-th components of \( \hat{u}(\tilde{v}), u(\tilde{v}) \) and \( X_i \), respectively. For every \( r \), we have, by Lemma 2,

\[
\hat{u}_r(\tilde{v}) - u_r(\tilde{v}) = e_{1 \times (d+2)} \sum_{i=1}^{n} \left[ Q_i^T(\tilde{v}) W_i(\tilde{v}) Q_i(\tilde{v}) \right]^{-1} Q_i^T(\tilde{v}) K_{1,h_1}(\tilde{v}_i - \tilde{v}) X_{ir}
\]

\[
- e_{1 \times (d+2)} \left[ Q_i^T(\tilde{v}) W_i(\tilde{v}) Q_i(\tilde{v}) \right]^{-1} Q_i^T(\tilde{v}) W_i(\tilde{v}) Q_i(\tilde{v}) \left[ (u_r(\tilde{v}), (u_r'(\tilde{v}))^T h_1) \right]^T
\]

\[
= f_1^{-1}(\tilde{v}) \frac{1}{n} \sum_{i=1}^{n} K_{1,h_1}(\tilde{v}_i - \tilde{v}) (X_{ir} - u_r(\tilde{v}) - (\tilde{v}_i - \tilde{v})^T u_r'(\tilde{v}))(1 + o_p(1)),
\]

which implies (A.2).

By the Taylor series of \( u(\tilde{v}) \) at \( \tilde{v} \) in \( M_{1,n}(\tilde{v}) \) and Lemma 1, (A.3) is obtained easily. \( \square \)

**Lemma 4.** Under assumptions A.1–A.10, we have

\[
S_2(u) = \frac{1}{N} \left( Q_2^T(u) W_2(u) Q_2(u) \right) = f_2(u) \text{diag} \left( \Sigma(u), \Sigma(u) \mu_{2,2} \right) + O_p(1).
\]

(A.4)

**Proof.** Note that \( S_2(u) \) can be rewritten as

\[
\frac{1}{N} \left( \sum_{j=m+1}^{n+N} \tilde{u}^{\otimes 2}(\tilde{v}_j) K_{2,h_2}(U_j - u) + \sum_{j=m+1}^{n+N} \tilde{u}^{\otimes 2}(\tilde{v}_j) \left( \frac{U_j - u}{h_2} \right)^2 K_{2,h_2}(U_j - u) \right).
\]
which implies that all arrays of $S_2(u)$ share the form of
\[
S_{2,\lambda}(u) = \frac{1}{N} \sum_{j=n+1}^{n+N} \tilde{u}^{\otimes 2}(\tilde{V}) \left( \frac{U_j - u}{h_2} \right)^{\lambda} K_{2,h_2}(U_j - u)
\]
for $\lambda = 0$, 1, or 2. Next we discuss $S_{2,\lambda}(u)$. By Lemma 3, we have
\[
S_{2,\lambda}(u) = \frac{1}{N} \sum_{j=n+1}^{n+N} \left( u(\tilde{V}) + \tilde{u}(\tilde{V}) - u(\tilde{V}) \right) \otimes 2 \left( \frac{U_j - u}{h_2} \right)^{\lambda} K_{2,h_2}(U_j - u)
\]
\[
= \frac{1}{N} \sum_{j=n+1}^{n+N} u(\tilde{V}) \left( \frac{U_j - u}{h_2} \right)^{\lambda} K_{2,h_2}(U_j - u) + \frac{1}{N} \sum_{j=n+1}^{n+N} u(\tilde{V})(\tilde{u}(\tilde{V}) - u(\tilde{V}))\bigg( \frac{U_j - u}{h_2} \bigg)^{\lambda} K_{2,h_2}(U_j - u)
\]
\[
+ \frac{1}{N} \sum_{j=n+1}^{n+N} \left( \tilde{u}(\tilde{V}) - u(\tilde{V}) \right) u(\tilde{V}) \left( \frac{U_j - u}{h_2} \right)^{\lambda} K_{2,h_2}(U_j - u)
\]
\[
= f_2(u) \Sigma(u) \int t^\lambda K_2(t)dt + o_p(1).
\]
The Lemma is proved. \qed

**Lemma 5.** (1) Under assumptions A.7–A.9 and A.12, we have
\[
S_3(u) = \frac{1}{n} Q_3^T(u)W_3(u)Q_3(u) = f_2(u)\operatorname{diag}(\Sigma_3(u), \mu_{3,2} \Sigma_3(u)) + o_p(1) \tag{A.5}
\]
\[
S_4(u) = \frac{1}{n} Q_4^T(u)W_4(u)Q_4(u) = f_2(u)\operatorname{diag}(\Sigma(u), \mu_{3,2} \Sigma(u)) + o_p(1), \tag{A.6}
\]
where $f_2(\cdot)$ is defined in assumption 4, $\Sigma(u)$ in Theorem 1 and $\Sigma_3(u)$ in Theorem 2.

The proof of this Lemma is similar to that of Lemmas 2 and 4.

**Lemma 6.** Let $\{\xi_i\}$, $i = 1, \ldots, N$ be an i.i.d. sample, and $\{\zeta_j\}$, $j = 1, \ldots, n$ be another i.i.d. sample which is independent of $\{\xi_i\}$. The functions $\psi_N(\xi, t, h)$ is a sequence of random function with a bandwidth $h$. Suppose that
\begin{enumerate}
  \item there exist square integrable functions $q_1(z)$ and $q_2(t)$ such that
  \[ |E[\psi_N(\xi, \zeta, h)|z] \leq q_1(z) \quad \text{and} \quad |E[\psi_N(\zeta, \zeta, h)|\xi] \leq q_2(t) |. \]
  \item $\lim_{N \to \infty} E[\psi_N(\xi, \zeta, h)|\zeta] = p_1(\xi)$, a.e. and $\lim_{N \to \infty} E[\psi_N(\xi, \zeta, h)|\zeta] = p_2(\zeta)$, a.e. for some measurable functions $p_1(z)$ and $p_2(t)$, and
  \item $\lim_{N \to \infty} \sqrt{N}E[\psi_N(\xi, \zeta, h)] = 0.
\end{enumerate}

Then
\[
\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} \psi_N(\xi_i, \xi_j, h) \xrightarrow{D} N(0, \lambda \operatorname{Cov}(p_1(\xi)) + \operatorname{Cov}(p_2(\xi))),
\]
where $\lambda = \lim(N/N)$ is assumed to be finite.

This Lemma is Theorem B.1 of [20].

**Proof of Theorem 1.** Notice that, by Lemma 4, $\hat{a}(u) - a(u)$ can be rewritten as
\[
E_{p \times 2p}(Q_2^T(u)W_2(u)Q_2(u))^{-1} \left\{ \sum_{j=n+1}^{n+N} Q_2^{(j)}(u)K_{2,h_2}(U_j - u)Y_j - (Q_2^T(u)W_2(u)Q_2(u))(a^T(u), h_2(a'(u)))^T \right\}
\]
\[
= E_{p \times 2p}(Q_2^T(u)W_2(u)Q_2(u))^{-1} \left\{ \sum_{j=n+1}^{n+N} Q_2^{(j)}(u)K_{2,h_2}(U_j - u) \left\{ Y_j - \tilde{u}^T(\tilde{V})(a(u) + a'(u)(U_j - u)) \right\} \right\}
\]
\[
= \frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} \tilde{u}(\tilde{V})K_{2,h_2}(U_j - u) \left\{ Y_j - \tilde{u}^T(\tilde{V})(a(u) + a'(u)(U_j - u)) \right\} (1 + o_p(1))
\]
\[
= \frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} \left( \tilde{u}(\tilde{V}_j)K_{2,b_2}(U_j - u) \left( u^T(\tilde{V}_j)a(U_j) - \tilde{u}^T(\tilde{V}_j)[a(u) + a'(u)(U_j - u)] + \varepsilon \right) 1 + o_p(1) \right)
\]

\[
= \frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} \left( u^T(\tilde{V}_j)[a(u) + a'(u)(U_j - u)] + \varepsilon \right) 1 + o_p(1) \right)
\]

\[
\equiv (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5 + \Delta_6)(1 + o_p(1)), \tag{A.7}
\]

where \( U_j^* \) lies between \( U_j \) and \( u \), and

\[
\Delta_1 = \frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} [\tilde{u}(\tilde{V}_j) - u(\tilde{V}_j)]K_{2,b_2}(U_j - u)\varepsilon_j,
\]

\[
\Delta_2 = \frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} u(\tilde{V}_j)K_{2,b_2}(U_j - u)\varepsilon_j,
\]

\[
\Delta_3 = \frac{\Sigma^{-1}(u)}{2f_2(u)N} \sum_{j=n+1}^{n+N} [\tilde{u}(\tilde{V}_j) - u(\tilde{V}_j)]u^T(\tilde{V}_j)a''(U_j^*)((U_j - u)^2)K_{2,b_2}(U_j - u),
\]

\[
\Delta_4 = \frac{\Sigma^{-1}(u)}{2f_2(u)N} \sum_{j=n+1}^{n+N} u(\tilde{V}_j)u^T(\tilde{V}_j)a''(U_j^*)((U_j - u)^2)K_{2,b_2}(U_j - u),
\]

\[
\Delta_5 = -\frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} [\tilde{u}(\tilde{V}_j) - u(\tilde{V}_j)]^{\otimes 2}[a(u) + a'(u)(U_j - u)]K_{2,b_2}(U_j - u),
\]

\[
\Delta_6 = -\frac{\Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} u(\tilde{V}_j)[\tilde{u}(\tilde{V}_j) - u(\tilde{V}_j)]^T[a(u) + a'(u)(U_j - u)]K_{2,b_2}(U_j - u).
\]

Next we discuss \( \Delta_i \) for \( i = 1, \ldots, 6 \), respectively.

Notice that \( \Delta_2 \) is a sum of independent random variables with \( \text{E}(\Delta_2) = 0 \) and

\[
\text{Cov}(\Delta_2) = f_2^{-2}(u)\frac{1}{N} \Sigma^{-1}(u)E\left[ u(\tilde{V})u^T(\tilde{V})K_{2,b_2}^2(U - u)e^2 \right] \Sigma^{-1}(u)
\]

\[
= f_2^{-2}(u)\frac{1}{N} \Sigma^{-1}(u)E\left[ u(\tilde{V})u^T(\tilde{V})K_{2,b_2}^2(U - u)e^2|U \right] \Sigma^{-1}(u)
\]

\[
= f_2^{-2}(u)\frac{1}{N} \Sigma^{-1}(u) \int \left[ E\left[ u(\tilde{V})u^T(\tilde{V})e^2|U = v \right] K_{2,b_2}^2(v - u)f_2(v)dv \right] \Sigma^{-1}(u).
\]

By letting \( (v - u)/h_2 = t \) and Taylor expansions, it can be shown that \( \text{Cov}(\Delta_2) \) is

\[
f_2^{-2}(u)\frac{1}{Nh_2} \Sigma^{-1}(u) \int \left[ E\left[ u(\tilde{V})u^T(\tilde{V})e^2|U = u + h_2t \right] K_{2,b_2}^2(t)f_2(u + h_2t)dt \right] \Sigma^{-1}(u)
\]

\[
= f_2^{-1}(u)\frac{\nu_{2,2}}{Nh_2} \Sigma^{-1}(u)E\left[ u(\tilde{V})u^T(\tilde{V})(Y - u^T(\tilde{V})a(U))^2|U = u \right] \Sigma^{-1}(u)(1 + o(1)),
\]

where \( \nu_{2,2} = \int f_2^2(t)dt \). So we have

\[
\sqrt{Nh_2}\Delta_2 \xrightarrow{D} N\left(0, \frac{\nu_{2,2}}{f_2(u)} \Sigma^{-1}(u)\Sigma_1(u) \Sigma^{-1}(u) \right). \tag{A.8}
\]

For \( \Delta_4 \), notice that \( a''(\cdot) \) is continuous; we can see easily that

\[
\sqrt{Nh_2}\Delta_4 = \sqrt{Nh_2} \left( \frac{1}{2f_2(u)} \Sigma^{-1}(u)E\left[ u(\tilde{V})u^T(\tilde{V})a''(U - u)^2K_{2,b_2}(U - u) \right] \right) + \sqrt{Nh_2}O_p(h_2^2/\sqrt{Nh_2})
\]

\[
= \sqrt{Nh_2} \frac{\mu_{2,2}h_2^2}{2} a''(u) + o_p(1). \tag{A.9}
\]

Note that, with (A.2), \( \sqrt{Nh_2}\Delta_6 \) can be divided into

\[
\sqrt{Nh_2}\Delta_6 = -(\Delta_61 + \Delta_62 + \Delta_63 + \Delta_64)(1 + o_p(1)). \tag{A.10}
\]
where

\[ A_{61} = \frac{\sqrt{Nh_2} \Sigma^{-1}(u)}{Nf_2(u)} \sum_{j=n+1}^{n+N} u(\tilde{\nu}_j) f_{1}^{-1}(\tilde{\nu}_j) R_{1,n}^T(\tilde{\nu}_j) a(u) K_{2,h_2}(U_j - u), \]

\[ A_{62} = \frac{\sqrt{Nh_2} \Sigma^{-1}(u)}{Nf_2(u)} \sum_{j=n+1}^{n+N} u(\tilde{\nu}_j) f_{1}^{-1}(\tilde{\nu}_j) M_{1,n}^T(\tilde{\nu}_j) a(u) K_{2,h_2}(U_j - u), \]

\[ A_{63} = \frac{\sqrt{Nh_2} \Sigma^{-1}(u)}{Nf_2(u)} \sum_{j=n+1}^{n+N} u(\tilde{\nu}_j) f_{1}^{-1}(\tilde{\nu}_j) R_{1,n}^T(\tilde{\nu}_j) a'(u)(U_j - u) K_{2,h_2}(U_j - u), \]

\[ A_{64} = \frac{\sqrt{Nh_2} \Sigma^{-1}(u)}{Nf_2(u)} \sum_{j=n+1}^{n+N} u(\tilde{\nu}_j) f_{1}^{-1}(\tilde{\nu}_j) M_{1,n}^T(\tilde{\nu}_j) a'(u)(U_j - u) K_{2,h_2}(U_j - u), \]

with \( M_{1,n}(\cdot) \) and \( R_{1,n}(\cdot) \) defined in Lemma 3. For every \( i \neq j \), let

\[ \psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j) = \sqrt{h_2 f_{1}^{-1}(\tilde{\nu}_j)} u(\tilde{\nu}_j)(X_i - u(\tilde{\nu}_j))^T a(u) K_{1,h_1}(\tilde{\nu}_i - \tilde{\nu}_j) K_{2,h_2}(U_j - u), \]

and \( T_{n,N} = \frac{1}{n \sqrt{N}} \sum_{i=1}^{n} \sum_{j=n+1}^{n+N} \psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j) \),

then

\[ A_{61} = \frac{\Sigma^{-1}(u)}{f_2(u)} T_{n,N}. \]  

Note that \( E[\psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j)] = 0 \), which implies that

\[ \lim_{N \to \infty} \sqrt{N} E[\psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j)] = 0, \]

\[ \text{Cov}(E[\psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j)]) = 0, \]  

(A.13)

and, by letting \( (\tilde{\nu} - \tilde{\nu}_i)/h_1 = t \), where \( \tilde{\nu}_{d+1} \) denotes the \( d + 1 \)-th entry of \( \tilde{\nu} \), and noting assumption A.11, as \( n \to \infty \),

\[ E[\psi_{n,N}(X_i, \tilde{\nu}_i; \tilde{\nu}_j) | X_i, \tilde{\nu}_i] = \sqrt{h_2} \int u(\tilde{\nu}) (X_i - u(\tilde{\nu}))^T a(u) K_{1,h_1}(\tilde{\nu}_i - \tilde{\nu}_j) K_{2,h_2}(\tilde{\nu}_{d+1} - u) d\tilde{\nu} \]

\[ = \frac{1}{\sqrt{h_2}} \int u(\tilde{\nu}) (X_i - u(\tilde{\nu}))^T a(u) K_{1,h_1}(t) K_{2} \left( \frac{U_i - u + h_1 t_{d+1}}{h_2} \right) dt \]

\[ = \frac{1}{\sqrt{h_2}} u(\tilde{\nu}_i) (X_i - u(\tilde{\nu}_i))^T a(u) K_{2} \left( \frac{U_i - u}{h_2} \right) (1 + o_p(1)) \]

\[ \doteq p_1(X_i, \tilde{\nu}_i; u)(1 + o_p(1)). \]

It is shown easily that \( E(p_1(X, \tilde{\nu}; u)) = 0 \), and

\[ \text{Cov}(p_1(X, \tilde{\nu}; u)) = v_{2,2} f_2(u) \Sigma_2(u) + o(1). \]  

(A.14)

By Lemma 6 and conjoining (A.11) and (A.13) with (A.14), we have

\[ T_{n,N} \overset{d}{\to} N \left( 0, \lambda v_{2,2} f_2(u) \Sigma_2(u) \right). \]  

(A.15)

This together with (A.12) proves

\[ A_{61} \overset{d}{\to} N \left( 0, \lambda v_{2,2} f_2^{-1}(u) \Sigma^{-1}(u) \Sigma_2(u) \Sigma^{-1}(u) \right). \]  

(A.16)

By a method similar to that for \( A_{61} \), we can show that \( A_{6i} \) converges to a random vector with zero asymptotic mean and zero asymptotic covariance for \( i = 2, 3, 4 \), which implies that

\[ A_{62} = o_p(1), \quad A_{63} = o_p(1), \quad A_{64} = o_p(1). \]  

(A.17)

Combining (A.10), (A.16) and (A.17), we have

\[ \sqrt{Nh_2} \Delta_6 \overset{d}{\to} N \left( 0, \lambda v_{2,2} f_2^{-1}(u) \Sigma^{-1}(u) \Sigma_2(u) \Sigma^{-1}(u) \right). \]  

(A.18)

By using an argument similar to that for \( \Delta_6 \), we have

\[ \sqrt{Nh_2} \Delta_1 = o_p(1), \]  

(A.19)
and
\[ \sqrt{N\lambda_2} \Delta_3 = o_p(1). \] (A.20)

By (A.2), it is shown that
\[ \sqrt{N\lambda_2} \Delta_5 = \frac{-\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} [R_{1,n}(u\tilde{V}_j) + M_{1,n}(u\tilde{V}_j)]^{\otimes 2}(a(u) + a'(u)(U_j - u))K_{2,h_2}(U_j - u) \]
\[ = -D_1 - D_2 - \ldots - D_8, \] (A.21)

where
\begin{align*}
D_1 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} M_{1,n}(u\tilde{V}_j)R_{1,n}^T(u\tilde{V}_j)a(u)K_{2,h_2}(U_j - u), \\
D_2 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} M_{1,n}(u\tilde{V}_j)R_{1,n}^T(u\tilde{V}_j)a'(u)(U_j - u)K_{2,h_2}(U_j - u), \\
D_3 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} M_{1,n}(u\tilde{V}_j)M_{1,n}^T(u\tilde{V}_j)a(u)K_{2,h_2}(U_j - u), \\
D_4 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} M_{1,n}(u\tilde{V}_j)M_{1,n}^T(u\tilde{V}_j)a'(u)(U_j - u)K_{2,h_2}(U_j - u), \\
D_5 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} R_{1,n}(u\tilde{V}_j)M_{1,n}^T(u\tilde{V}_j)a(u)K_{2,h_2}(U_j - u), \\
D_6 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} R_{1,n}(u\tilde{V}_j)M_{1,n}^T(u\tilde{V}_j)a'(u)(U_j - u)K_{2,h_2}(U_j - u), \\
D_7 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} R_{1,n}(u\tilde{V}_j)R_{1,n}^T(u\tilde{V}_j)a(u)K_{2,h_2}(U_j - u), \\
D_8 & = \frac{\sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} R_{1,n}(u\tilde{V}_j)R_{1,n}^T(u\tilde{V}_j)a'(u)(U_j - u)K_{2,h_2}(U_j - u).
\end{align*}

Next we discuss \( D_1 \). Let \( A_i(\tilde{V}) \) denote the Hessian matrix of \( u_r(\tilde{V}), r = 1, \ldots, d + 1, M_{1,n}(\tilde{V}_j) \) can be rewritten as
\[ M_{1,n}(\tilde{V}_j) = \frac{1}{2n} \sum_{i=1}^{n} K_{1,h_1}(V_i - \tilde{V}_j)(V_i - \tilde{V}_j)^T A_1(V_i - \tilde{V}_j), \ldots, (V_i - \tilde{V}_j)^T A_{d+1}(V_i - \tilde{V}_j) \]
\[ = A(\tilde{V}_j)h_1^2(1 + o_p(1)), \]
uniformly in \( j \), where \( V_i^r \) lies between \( V_i \) and \( \tilde{V}_j, r = 1, \ldots, d + 1, A(\tilde{V}_j) \) is a matrix. So we have
\[ D_1 = \frac{h_1^2 \sqrt{N\lambda_2} \Sigma^{-1}(u)}{f_2(u)N} \sum_{j=n+1}^{n+N} A(\tilde{V}_j)R_{1,n}^T(\tilde{V}_j)a(u)K_{2,h_2}(U_j - u)(1 + o_p(1)). \]

By using an argument similar to that for \( A_{r+1} \), we have \( D_1 = o_p(1) \).

By using an argument similar to that for \( D_1 \), we have \( D_r = o_p(1), r = 2, \ldots, 8 \). Furthermore, we have
\[ \sqrt{N\lambda_2} \Delta_3 = o_p(1). \] (A.22)

It is easily shown that
\[ \text{Cov}(\Delta_2, \Delta_3) = 0. \] (A.23)

Conjoining (A.7)–(A.9) with (A.18)–(A.23), we complete the proof of the Theorem. \( \square \)

**Proof of Theorem 2.** Note that, by Lemma 6,
\[ \tilde{a}(u) - a(u) = E_{p \times 2p} \sum_{i=1}^{n} \left( Q_1^T \tilde{W}_3 Q_3 + Q_1^T \tilde{W}_4 Q_4 \right)^{-1} Q_1^T \tilde{K}_{3,h_2}(U_i - u)(Y_i - \tilde{X}_i^T(a(u) + a'(u)(U_i - u))) \]
\[ + E_{p \times 2p} \sum_{j=n+1}^{n+N} \left( Q_1^T \tilde{W}_3 Q_3 + Q_1^T \tilde{W}_4 Q_4 \right)^{-1} Q_1^T \tilde{K}_{3,h_2}(U_j - u)(Y_j - \tilde{u}_j^T(\tilde{V}_j)(a(u) + a'(u)(U_j - u))) \]
\[ f_2^{-1}(u) \sum_i^{N} (u_i - u) (Y_i - X_i^T (a(u) + a'(u)(U_i - u))) (1 + o_p(1)) \]
\[ + f_2^{-1}(u) \sum_i^{N} (u_i - u) (a(u) + a'(u)(U_i - u))) (1 + o_p(1)) \]
\[ \Rightarrow (A_1 + A_2)(1 + o_p(1)). \]  
\[ \text{(A.24)} \]
Replacing \( Y_i \) in \( A_1 \) by \( X_i^T a(U_i) + e_i \), and using the Taylor series of \( a(U_i) \) at \( u \), we have
\[ A_1 = \frac{1}{1 + \lambda} \sum_i^{N} (u) \sum_j^{N} (u) \frac{\lambda u}{2} a''(u) (1 + o_p(1)) + A_{12}. \]  
\[ \text{(A.25)} \]
where \( A_{12} = f_2^{-1}(u) \sum_i^{N} (u_i - u) \sum_j^{N} (u) X_iK_{3,h_3}(U_i - u)e_i. \)
It is easily shown that
\[ \sqrt{(n + N)h_3} A_{12} \xrightarrow{d} N \left( 0, \frac{1}{1 + \lambda} \sum_i^{N} (u) \sum_j^{N} (u) \frac{\lambda u}{2} a''(u) \right). \]  
\[ \text{(A.26)} \]
Using a proof similar to that of Theorem 1 (only replacing \( K_2(\cdot) \) and \( h_2 \) by \( K_3(\cdot) \) and \( h_3 \) respectively), we have
\[ \sqrt{(n + N)h_3} \left( A_2 - \frac{\lambda}{1 + \lambda} \sum_i^{N} (u) \sum_j^{N} (u) \frac{\lambda u}{2} a''(u) \right) \xrightarrow{d} N \left( 0, \frac{\lambda}{1 + \lambda} \sum_i^{N} (u) \sum_j^{N} (u) \frac{\lambda u}{2} a''(u) \right). \]  
\[ \text{(A.27)} \]
Recalling the proof of Theorem 1, we find that \( \text{Cov}(A_{12}, A_{21}) = 0, \text{Cov}(A_{12}, A_{21}) = 0 \), which implies that
\[ \text{Cov}(A_1, A_2) = 0. \]  
\[ \text{(A.28)} \]
Conjoining (A.24)–(A.28), the proof of Theorem 2 is completed. \( \square \)

References