Statistics and Its Interface Volume 11 (2018) 587–602

Semiparametric varying-coefficient partially linear models with auxiliary covariates

Xiaojing Wang*,†, Yong Zhou‡, and Yang Liu

In this paper, we consider a semiparametric varying-coefficient partially linear model when some of the covariates are only measured on a selected validation set whereas auxiliary variables are observed for all study subjects. The semiparametric profile-likelihood procedure for estimating parametric and nonparametric component which incorporates information from auxiliary covariates is proposed. The resulting estimators are consistent regardless of the specification of the relationship between the covariates and the surrogate variables. Moreover, the proposed estimators are asymptotically more efficient than the validation-set-only estimators. Asymptotic properties of the proposed estimators are established. The finite sample performance is investigated and compared with alternative methods via simulation studies. The simulated results demonstrate that the asymptotic approximations of the proposed estimators are adequate for practice. We use a Boston Housing dataset to illustrate the performance of the proposed method in practice.

AMS 2000 subject classifications: Primary 62G05; secondary 62G20.

Keywords and phrases: Incomplete data, Validated samples, Profile-likelihood, Surrogate variables.

1. INTRODUCTION

To avoid the risk of introducing modeling biases in parametric regression models as well as those drawbacks of nonparametric regression models (such as the curse of dimensionality, the difficulty of interpretation and lack of extrapolation capability), various efforts have been made to balance the interpretation of linear models and flexibility of nonparametric models. The partially linear model is the most commonly used semiparametric regression model (c.f., [8], [11], [21], [31], and references therein for the applications and theoretical developments of partially linear models and their extensions). While various extensions of the partially linear model were proposed in the literature, an important extension of partially linear model is the semiparametric varying-coefficient partially linear model (SVPLM). SVPLM becomes popular and has been intensively studied recently in the literature. For example, Kai et al. [16] extended the estimation and variable selection of SVPLM in quantile regression with the application to the plasam beta-carotene level data; Tian et al. [24] focused on the variable selection of SVPLM for longitudinal data; Shao et al. [23] incorporated flexible SVPLM regression tools for interval censored data with a cured proportion; Bai et al. [1] introduced errors correlated in space and time for panel data using SVPLM.

Typically, a semiparametric varying-coefficient partially linear model (SVPLM) can be defined in the following way:

$$Y = \boldsymbol{\alpha}^T(U)\mathbf{X} + \boldsymbol{\beta}^T\mathbf{Z} + \epsilon, \tag{1.1}$$

where Y denotes the response variable, $(U, \mathbf{X}^T, \mathbf{Z}^T)$ is associated covariates for Y, and ϵ presents the noise, independent of $(U, \mathbf{X}^T, \mathbf{Z}^T)$ and satisfying $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma^2$. Define $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_q)^T$ is a q-dimensional vector of unknown parameters and $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \cdots, \alpha_p(\cdot))^T$ is a p-dimensional vector of unknown coefficient functions. Throughout the paper, we will focus on the univariate U only, but the proposed method is directly applicable to the case when U is multi-dimension.

Model (1.1) allows interaction between the covariate U and $\mathbf{X}$ in such a way that a different level of covariate U is associated with a different linear model. This enables us to examine the extent to how the effects of covariates $\mathbf{X}$ vary over different levels of the covariate U. It provides a novel and general structure, which covers many well-known semiparametric regression models, such as partially linear models proposed by [8], varying-coefficient models proposed by [13] and so on. Recently, there are extensive literature focusing on investigation of the estimation procedure for Model (1.1), such as [9], [33], [35] and so on. Especially, the profile likelihood technique introduced in the paper of Fan and Huang [9] makes the statistical inference for Model (1.1) become effective and systematic.

However, in some applications, due to the limitations of the cost, it is prohibitive to collect data on $(\mathbf{Y}, U, \mathbf{X}^T, \mathbf{Z}^T)$ for all subjects in the study. For example, it may be very difficult or expensive to measure an informative covariate U. But a surrogate W can be easily or cheaply as-

*The current work of this paper was based on part of the Master's thesis of Xiaojing Wang [28] (see Chapter 2 for details).

†Corresponding author.

‡Zhou's work was supported by the State Key Program of National Natural Science Foundation of China (NSFC) (71331006), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (91546202).

certained. Then, a double- or two-stage sampling design may be adopted by investigators as a cost-effective strategy. Generally speaking, a primary sample is drawn from the target population at the first stage of the study, wherein data on rough or proxy, but less expensive, measurements $(\mathbf{Y}, W, \mathbf{X}^T, \mathbf{Z}^T)$ for $(\mathbf{Y}, U, \mathbf{X}^T, \mathbf{Z}^T)$ are collected for all study subjects. Second, the exact information on $(\mathbf{Y}, U, \mathbf{X}^T, \mathbf{Z}^T)$ is ascertained for a validation subsample drawn from the primary sample. Thus, the data $(\mathbf{Y}, U, W, \mathbf{X}^T, \mathbf{Z}^T)$ can be regarded as incomplete data in the sense that they are incompletely observed for all subjects, except for validation members.

Formally, consider a situation when the covariates U are only exactly measured in a validation subsample but there are auxiliary information W available for U on the whole sample. Further, assume covariates $\mathbf{X}$ and $\mathbf{Z}$ as well as response Y are easily and accurately measured in all subjects. Summarize the assumptions and put them in the notations below for further usage throughout the paper:

(a) The observations $(Y_i, W_i, \mathbf{X}_i^T, \mathbf{Z}_i^T)$, $i = 1, \cdots, N$, of the primary data are independently and identically distributed (IID) with finite first and second moments, where N denotes the sample size of primary sample.
(b) The observations $(Y_i, U_i, W_i, \mathbf{X}_i^T, \mathbf{Z}_i^T)$, $i \in \mathcal{V}$, of the validation data are IID with finite first and second moments, where $\mathcal{V}$ denotes the validation subsample set with n being its sample size.
(c) The validation subsample is assumed to be a random draw from the primary sample. Let ρ denote the unknown validation fraction, where $\rho = \lim_{N\to\infty} n/N$.

Then, the data $(\mathbf{Y}, U, W, \mathbf{X}^T, \mathbf{Z}^T)$ can be regarded as incomplete data in the sense that they are incompletely observed for all subjects, except for validation members.

The incomplete data with surrogate/auxiliary covariates available is very common phenomenon in practice, especially in biomedical and economical research. There are several literature recently studying such cases of incomplete data in SVPLMs (e.g., [2], [14], [17], [29], [30], [32] and [34]). For instance, Zhou and Liang [34] considered the case when some linear covariates in SPVLMs are unobservable, but there are auxiliary variables available. As a special case of the SVPLM without parametric regression component, Li et al. [17] and Xu et al. [30] addressed statistical inference of varying-coefficient regression models in which some covariates are not observed, but anxillary variables are available to remit them; while Wang et al. [25] discussed a situation of the outcome missing with the help of auxiliary information. So far, no literature has discussed the statistical inference when the observed data of the covariate U in the nonparametric part of the SVPLM are incomplete. We aim to fill in this gap.

How to develop an effective estimating procedure to utilize both the primary and the validation samples becomes a critical issue. There are a great deal of literature studying this issue based on various models (e.g., [4], [5], [7], [15], [20], [22] and [26]). Enlightened by ideas of Chen and Chen [5] and Jiang and Zhou [15], we propose an estimating procedure by incorporating auxiliary information from the validation data sets to correct (or update) estimators of parametric and nonparametric varying-coefficient components based on semiparametric profile least-squared methods. The proposed method is robust in the sense that the consistency of the estimators do not require a correct specification of the relationship between observations U and W. Thus, it avoids the bias created by the model misspecification. Moreover, the asymptotic variances of our proposed estimators are easily obtained, making their usage convenient in the practice.

To be specific, we utilize a working model in helping of estimating parametric and nonparametric coefficients in Model (1.1), which is a novel and quite different approach than usual ideas in dealing with missing or measurement error data. The first step of our proposed method is to obtain the parametric estimators using validated and primary sample, respectively. Then, a key step is to derive the joint distribution of our parametric estimators under validated and primary sample. Although the exact joint distribution is difficult to find, their joint distribution can be proved to be asymptotically multivariate normal. Based on the fact of the conditional property of a multivariate normal distribution, we can use the conditional mean as an corrected (or updated) estimator for the parametric components in Model (1.1). As a matter of fact, if the asymptotic property of conditional normality holds in finite samples as well, we know that this conditional mean estimator is an uniformly best linear unbiased estimator for the parametric components.

The proposed procedure developed in this paper can also be applicable to situtations when the covariate U of nonparametric part has additive errors or Berkson errors with some validation data available. In comparison with the deconvolution method for dealing with measurement errors appearing in the nonparametric part of semiparametric models (e.g., [6], [10] and [18]), our proposed method, by using information from validated data, avoids the dependence on the smoothness of the errors' distribution in deriving of the asymptotic property of estimators, and does not require the linearity assumption of measurement errors.

Although our proposed method is mainly concentrated on the situation when the covariate U in the varying coefficient parts is incompletely observed for all subjects, the same logic of the proposed idea can be easily generalize to cases when other variables having incomplete values in Model (1.1). Section 2 and Section 3 discuss the estimation method and the asymptotic property of the estimators for parametric coefficients and nonparametric varying-coefficients, respectively. Section 4 introduces the construction of asymptotic variance of the estimators and how to select bandwidth. Section 5 and 6 illustrate some simulation results and apply the proposed method to a real data example from Boston Housing datasets. Section 7 concludes the results and points out the future direction for the improvement of our method. All proofs are postponed in Appendix.

2. USING AUXILIARY COVARIATES TO CORRECT (OR UPDATE) ESTIMATORS OF PARAMETRIC COMPONENTS

In this section, we will develop a method to correct (or update) the estimators of parametric components based on their profile least-square estimators using merely validated data. The key idea is to introduce a working model to combine auxiliary information with validated data for yielding a corrected (or updated) estimators of parametric coefficients in Model (1.1). We will show that our proposed estimators can significantly improve the efficiency of the estimators.

2.1 Estimation of parametric coefficients based on validation subsamples

Although there are several ways to estimate the unknown parameters $\{\boldsymbol{\beta}_j, j = 1, ..., q\}$ and unknown coefficient functions $\{\alpha_i(\cdot), i = 1, ..., p\}$ in a SVPLM, we will first adopt profile least-square estimation for unknown parameter $\boldsymbol{\beta}$ based on validated data. Fan and Huang [9] illustrated that under the setting of model (1.1), profile least square estimation is semiparametrically efficient.

Without loss of generality, suppose that $\{(U_k, X_{k1}, \cdots, X_{kp}, Z_{k1}, \cdots, Z_{kq}, Y_k), \ k = 1, \cdots, n\}$ is the validated subsample of size n for Model (1.1). Given $\boldsymbol{\beta}$ and applying the idea of the profile least squares estimation (c.f., by [9]), Model (1.1) can be written as

$$Y_k^* = \sum_{i=1}^{p} \alpha_i(U_k) X_{ki} + \epsilon_k, \ k = 1, \cdots, n, \tag{2.1}$$

where $Y_k^* = Y_k - \sum_{i=1}^{q} \boldsymbol{\beta}_j Z_{kj}$. This transforms our Model (1.1) into the varying-coefficient models. Then the estimation of coefficient functions $\{\alpha_i(\cdot), i = 1, \cdots, p\}$ is reduced to the following weighted local least-squares problem: find $\{(a_i, b_i), i = 1, \cdots, p\}$ to minimize

$$\sum_{k=1}^{n} [Y_k^* - \sum_{i=1}^{p} \{a_i + b_i(U_k - u_0)\} X_{ki}]^2 K_{h_u}(U_k - u_0),$$

where $K_{h_u}(\cdot) = K(\cdot/h_u)/h_u$, $K(\cdot)$ is a kernel function and h_u is a bandwidth. Next, denote $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)^T$, where $\mathbf{X}_i$ is a column vector of $\mathbf{X}$ with $\mathbf{X}_i = (X_{i1}, \cdots, X_{ip})^T$ for $i = 1, \cdots, n$. Similarly, let $\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n)^T$ with $\mathbf{Z}_i = (Z_{i1}, \cdots, Z_{iq})^T$. Further, to simplify notation, we use

$$\mathbf{D}_{u,h_u} = \begin{pmatrix} \mathbf{X}_1^T & \frac{U_1 - u}{h_u}\mathbf{X}_1^T \\ \vdots & \vdots \\ \mathbf{X}_n^T & \frac{U_n - u}{h_u}\mathbf{X}_n^T \end{pmatrix}, \ \mathbf{M} = \begin{pmatrix} \alpha^T(U_1)\mathbf{X}_1 \\ \vdots \\ \alpha^T(U_n)\mathbf{X}_n \end{pmatrix},$$

$\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ and $\mathbf{V}_{u,h_u} = \mathrm{diag}(K_{h_u}(U_1 - u), \cdots, K_{h_u}(U_n - u))$. Then, Model (2.1) can be rewritten as a matrix form below

$$\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{M} + \boldsymbol{\epsilon}. \tag{2.2}$$

The solution to Problem (2.2) is given by

$$\begin{aligned} & [\hat{a}_1(u), \cdots, \hat{a}_p(u), h\hat{b}_1(u), \cdots, h\hat{b}_p(u)]^T \\ = \ & (\mathbf{D}_{u,h_u}^T \mathbf{V}_{u,h_u} \mathbf{D}_{u,h_u})^{-1} \mathbf{D}_{u,h_u}^T \mathbf{V}_{u,h_u} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}), \end{aligned} \tag{2.3}$$

where the first p components in (2.3) are estimators of elements in $\boldsymbol{\alpha}(U)$. Define $\mathbf{S} = (\mathbf{S}_1^T, \cdots, \mathbf{S}_n^T)^T$ to be a smoothing matrix dependent on the observations $\{(U_i, \mathbf{X}_i^T), i = 1, \cdots, n\}$, where

$$\mathbf{S}_i = [\mathbf{X}_1^T \quad \mathbf{0}_p](\mathbf{D}_{u_i,h_u}^T \mathbf{V}_{u_i,h_u} \mathbf{D}_{u_i,h_u})^{-1} \mathbf{D}_{u_i,h_u}^T \mathbf{V}_{u_i,h_u}$$

for $i = 1, \cdots, n$ and $\mathbf{0}_p$ is a p-dimensional zero vector. Then, the estimator of $\mathbf{M}$ is represented as

$$\widehat{\mathbf{M}} = \mathbf{S}(\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}). \tag{2.4}$$

Substituting $\widehat{\mathbf{M}}$ into Equation (2.2), we obtain

$$(\mathbf{I} - \mathbf{S})\mathbf{Y} = (\mathbf{I} - \mathbf{S})\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.5}$$

where $\mathbf{I}$ is a $n \times n$ identity matrix. Applying least squares to the linear model (2.5), the validated estimator of $\boldsymbol{\beta}$ could be obtained as

$$\widehat{\boldsymbol{\beta}}_\nu = \{\mathbf{Z}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{Y}. \tag{2.6}$$

Based on the assumptions (A.1),(A.3), and (A.5)-(A.8) in the Appendix and an analogy to Theorem 4.1 of [9], we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_\nu - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{11}), \tag{2.7}$$

where $\Sigma_{11} = \sigma^2\Omega_u^{-1}$, $\Omega_u = \mathrm{E}(\mathbf{Z}\mathbf{Z}^T) - \mathrm{E}(H(U))$, $\Gamma(U) = \mathrm{E}(\mathbf{X}\mathbf{X}^T|U)$, $\Phi(U) = \mathrm{E}(\mathbf{X}\mathbf{Z}^T|U)$ and $H(U) = \Phi(U)^T\Gamma(U)^{-1}\Phi(U)$.

2.2 Updating the estimators of parametric coefficients via auxiliary information

Chen and Chen [5] proposed a unified approach to the estimation of regression parameters under double-sampling designs, assuming that the validation sample is a simple random subsample from the primary sample. They utilized a specific parametric model to extract the partial information contained in the primary sample but their resulting estimator is consistent even if such a model is misspecified. Jiang and Zhou [15] extended the idea of Chen and Chen [5] to an additive hazard model when some of the true covariates are measured only on a randomly selected validation set whereas auxiliary covariates are observed for all study subjects.

Thus, enlightened by the idea used in Chen and Chen [5] and Jiang and Zhou [15] for extracting partial information from the primary sample, we substitute U in Model (1.1) with its surrogate variable W and assume Y also follows an auxiliary working model

$$Y = \boldsymbol{\mu}^T(W)\mathbf{X} + \boldsymbol{\gamma}^T\mathbf{Z} + \iota, \tag{2.8}$$

where ι is independent of $(W, \mathbf{X}^T, \mathbf{Z}^T)$ and has $\mathrm{E}(\iota) = 0$ and $\mathrm{Var}(\iota) = \sigma^2$, $\boldsymbol{\mu}(\cdot) = (\mu_1(\cdot), \cdots, \mu_p(\cdot))^T$ is a p-dimensional vector of unknown coefficient functions for the working model (2.8) and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_q)^T$ is a q-dimensional vector of unknown parameters.

Similar as the derivation of (2.6), using the validation data set, we are easily get the estimator of $\boldsymbol{\gamma}$ as below

$$(2.9)\quad \widehat{\boldsymbol{\gamma}}_\nu = \{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Y},$$

where

$$\mathbf{V}_{w,h_w} = \mathrm{diag}(K_{h_w}(W_1-w), \cdots, K_{h_w}(W_n-w)),$$
$$\widetilde{\mathbf{S}}_i = [\mathbf{X}_1^T \quad 0](\mathbf{D}_{w_i,h_w}^T\mathbf{V}_{w_i,h_w}\mathbf{D}_{w_i,h_w})^{-1}\mathbf{D}_{w_i,h_w}^T\mathbf{V}_{w_i,h_w}$$

for $i = 1, \cdots, n$, $\widetilde{\mathbf{S}} = (\widetilde{\mathbf{S}}_1^T, \cdots, \widetilde{\mathbf{S}}_n^T)^T$ and $\mathbf{D}_{w,h_w}$ is defined in the same fashion as $\mathbf{D}_{u,h_u}$ with the ith row replacing by $[\mathbf{X}_i^T, \mathbf{X}_i^T(W_i-w)/h_w]$. Following Lemma A.2, we have

$$\widehat{\boldsymbol{\gamma}}_\nu \xrightarrow{p} \boldsymbol{\gamma}_0 = \Omega_w^{-1}\{E(\mathbf{Z}Y) - E(\Phi^T(W)\Gamma^{-1}(W)\Psi^T(W))\},$$

where $\Omega_w = \mathrm{E}(\mathbf{Z}\mathbf{Z}^T) - \mathrm{E}(H(W))$, $\Gamma(W) = \mathrm{E}(\mathbf{X}\mathbf{X}^T|W)$, $\Phi(U) = \mathrm{E}(\mathbf{X}\mathbf{Z}^T|U)$ and $\Psi(W) = \mathrm{E}(\mathbf{X}\mathbf{Y}|W)$.

At the meantime, $\boldsymbol{\gamma}$ can be estimated by the full data via the auxiliary working model (2.8), i.e.,

$$\overline{\boldsymbol{\gamma}} = \{\overline{\mathbf{Z}}^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Z}}\}^{-1}\overline{\mathbf{Z}}^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Y}},$$

where $\overline{\mathbf{V}}_{w,h_w} = \mathrm{diag}(K_{h_w}(W_1-w), \cdots, K_{h_w}(W_n-w))$, $\overline{\mathbf{Y}} = (Y_1, \cdots, Y_N)^T$, $\overline{\mathbf{Z}} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_N)^T$, $\overline{\mathbf{I}}$ is N by N identity matrix, $\overline{\mathbf{D}}_{w,h_w}$ and $\mathbf{D}_{w,h_w}$ has the same form but with only changing the dimension of $\mathbf{D}_{w,h_w}$ from $n \times 2p$ to $N \times 2p$, and similar fashion has been used for the definition of $\overline{\mathbf{S}}$ and $\widetilde{\mathbf{S}}$. According to Lemma A.2, $\overline{\boldsymbol{\gamma}} \xrightarrow{p} \boldsymbol{\gamma}_0$ as well. Then, both $\overline{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}_\nu$ are consistent estimators of $\boldsymbol{\gamma}_0$.

Theorem 2.1. *Under assumptions given in the Appendix,* $n^{1/2}(\widehat{\boldsymbol{\beta}}_\nu^T - \boldsymbol{\beta}_0^T, \widehat{\boldsymbol{\gamma}}_\nu^T - \boldsymbol{\gamma}_0^T)^T$ *is asymptotically normal with mean zero and covariance matrix* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, *where* $\Sigma_{11} = \sigma^2\Omega_u^{-1}$, $\Sigma_{22} = \sigma^2\Omega_w^{-1} + \Omega_w^{-1}\Omega_{\mathbf{M}}\Omega_w^{-1}$,

$$\begin{aligned}\Sigma_{12} &= \sigma^2\Omega_u^{-1}\{\mathrm{E}(\mathbf{Z}\mathbf{Z}^T) - \mathrm{E}(H(W)) - \mathrm{E}(H(U)) \\ &+ \mathrm{E}(\Phi(U)^T\Gamma(U)^{-1}\mathbf{X}\mathbf{X}^T\Gamma(W)^{-1}\Phi(W))\}\Omega_w^{-1},\end{aligned}$$

and $\Omega_{\mathbf{M}} = \mathrm{E}(\kappa\kappa^T) - \tau\tau^T$, *with*

$$\begin{aligned}\tau &= \mathrm{E}(\Phi(U)\boldsymbol{\alpha}(U)) - \mathrm{E}(\Phi(W)^T\Gamma(W)^{-1}\Upsilon(W)), \\ \kappa &= \mathbf{Z}\mathbf{X}^T\boldsymbol{\alpha}(U) - \mathbf{Z}\mathbf{X}^T\Gamma(W)^{-1}\Upsilon(W) \\ &- \Phi(W)^T\Gamma(W)^{-1}\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha}(U) \\ &+ \Phi(W)^T\Gamma(W)^{-1}\mathbf{X}\mathbf{X}^T\Gamma(W)^{-1}\Upsilon(W), \\ \Upsilon(W) &= \mathrm{E}(\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha}(U)|W),\end{aligned}$$

and $\Sigma_{21} = \{\Sigma_{12}\}^T$.

Then, by the distribution theory of multivariate normal random variables, the conditional distribution of $n^{1/2}(\widehat{\boldsymbol{\beta}}_\nu - \boldsymbol{\beta}_0)$ given $(\widehat{\boldsymbol{\gamma}}_\nu - \boldsymbol{\gamma}_0)$ is asymptotically normal with mean $\Sigma_{12}\Sigma_{22}^{-1}n^{1/2}(\widehat{\boldsymbol{\gamma}}_\nu - \boldsymbol{\gamma}_0)$. The conditional mean can be estimated by substituting consistent estimators based on the validation sample for Σ_{12} and Σ_{22}, and replacing $\boldsymbol{\gamma}_0$ with the estimator $\overline{\boldsymbol{\gamma}}$ based on the full sample. Equating $n^{1/2}(\widehat{\boldsymbol{\beta}}_\nu - \boldsymbol{\beta}_0)$ with its estimated conditional mean, an updated estimator $\overline{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ is

$$(2.10)\qquad \overline{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_\nu - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}(\widehat{\boldsymbol{\gamma}}_\nu - \overline{\boldsymbol{\gamma}}),$$

where the estimator $\overline{\boldsymbol{\beta}}$ incorporates auxiliary information available on all subjects. Since $\widehat{\boldsymbol{\beta}}_\nu$, $\widehat{\boldsymbol{\gamma}}_\nu$ and $\overline{\boldsymbol{\gamma}}$ are consistent, whether the auxiliary covariate W is informative or not, $\overline{\boldsymbol{\beta}}$ will always be a consistent estimator of $\boldsymbol{\beta}_0$.

Theorem 2.2. *Under the assumptions in the Appendix,* $n^{1/2}(\overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ *is asymptotically normal with mean zero and covariance matrix*

$$\Delta = \Sigma_{11} - (1-\rho)\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Remark 1 Since $\rho \prec 1$ and observe that the first term of Δ is the asymptotic variance for $\widehat{\boldsymbol{\beta}}_\nu$, the expression of Δ shows that asymptotically, the proposed estimator $\overline{\boldsymbol{\beta}}$ is more efficient than the estimator $\widehat{\boldsymbol{\beta}}_\nu$.

Remark 2 It is easy to verify that an consistent estimator of Δ is

$$(2.11)\qquad \widehat{\Delta} = \widehat{\Sigma}_{11} - (1-\hat{\rho})\widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}\widehat{\Sigma}_{21},$$

where $\hat{\rho} = n/N$. We will discuss how to construct the consistent estimator of $\Sigma_{12}, \Sigma_{22}, \Sigma_{21}$ in Section 4.

3. ESTIMATION FOR NONPARAMETRIC VARYING-COEFFICIENTS

Based on the updated estimator of $\boldsymbol{\beta}$, i.e., $\overline{\boldsymbol{\beta}}$, in SVPLM (1.1), we can now use validated data to derive the estimator for $\boldsymbol{\alpha}(u)$ from (2.3):

$$\begin{aligned}(3.1)\qquad \widehat{\boldsymbol{\alpha}}_\nu(u) = [\mathbf{I}_p, \mathbf{0}_p](\mathbf{D}_{u,h_u}{}^T\mathbf{V}_{u,h_u}\mathbf{D}_{u,h_u})^{-1} \\ \times \mathbf{D}_{u,h_u}{}^T\mathbf{V}_{u,h_u}(\mathbf{Y} - \mathbf{Z}\overline{\boldsymbol{\beta}}),\end{aligned}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix.

Lemma 3.1. *Under the assumptions (A.1), (A.3), (A.5)-(A.8),* $\widehat{\boldsymbol{\alpha}}_\nu(u)$ *is a consistent estimate of* $\boldsymbol{\alpha}(u)$ *and*

$$\sqrt{nh_u}\{\widehat{\boldsymbol{\alpha}}_\nu(u) - \boldsymbol{\alpha}(u) - \frac{\boldsymbol{\alpha}''(u)}{2}\mu_2 h_u^2 + o(h_u^2)\} \xrightarrow{\mathcal{D}} N(0, \Lambda_{11}),$$

where $\boldsymbol{\alpha}''(\cdot)$ *indicates the second-order derivative of* $\boldsymbol{\alpha}(\cdot)$, $\Lambda_{11} = \sigma^2\vartheta_0 f(u)^{-1}\Gamma(u)^{-1}$, $f(u)$ *denotes the density of* U, $\vartheta_0 = \int K^2(u)du$, $\mu_2 = \int u^2K(u)du$.

Applying the similiar idea as constructing $\widehat{\gamma}_\nu$ in Subsection 2.2, we can mimic the representation of the estimator for $\boldsymbol{\alpha}(u)$ using surrogate data, as

$$\widehat{\boldsymbol{\eta}}(w) = [\mathbf{I}_p, 0](\mathbf{D}_{w,h_w}^T \mathbf{V}_{w,h_w} \mathbf{D}_{w,h_w})^{-1} \times \mathbf{D}_{w,h_w}^T \mathbf{V}_{w,h_w}(\mathbf{Y} - \mathbf{Z}\overline{\boldsymbol{\beta}}). \tag{3.2}$$

It is easy to prove that $\widehat{\boldsymbol{\eta}}(w)$ converges to $\boldsymbol{\eta}_0(w)$, where

$$\boldsymbol{\eta}_0(w) = \Gamma(w)^{-1}\mathrm{E}(\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha}(U)|W = w). \tag{3.3}$$

Based on the full data, we can obtain another consistent estimator of $\boldsymbol{\eta}_0(w)$,

$$\overline{\boldsymbol{\eta}}(w) = [\mathbf{I}_p, 0](\overline{\mathbf{D}}_{w,h_w}^T \overline{\mathbf{V}}_{w,h_w} \overline{\mathbf{D}}_{w,h_w})^{-1} \times \overline{\mathbf{D}}_{w,h_w}^T \overline{\mathbf{V}}_{w,h_w}(\overline{\mathbf{Y}} - \overline{\mathbf{Z}}\,\overline{\boldsymbol{\beta}}).$$

Theorem 3.2. *Under assumptions in the appendix and without loss of generality, assume $h_u = O(h)$ and $h_w = O(h)$, then*

$$\sqrt{nh}\begin{pmatrix} \widehat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\boldsymbol{\alpha}''(u)}{2}\mu_2 h_u^2 \\ \widehat{\boldsymbol{\eta}}(w) - \boldsymbol{\eta}_0(w) - \zeta_0(w)\mu_2 h_w^2 \end{pmatrix},$$

is asymptotically normal, with zero mean and variance matrix Λ, where

$$\begin{aligned} \zeta_0(w) &= \Gamma(w)^{-1}\frac{\Upsilon''(w)}{2} \\ &+ g(w)^{-1}\Gamma(w)^{-1}[g'(w)\Upsilon'(w) + \frac{g''(w)}{2}\Upsilon(w)], \end{aligned}$$

and $\Lambda = diag(\Lambda_{11}, \Lambda_{22})$ is a block-diagonal matrix with Λ_{11} defined in Lemma 3.1, $\Lambda_{22} = \sigma^2\vartheta_0 g(w)^{-1}\Gamma(w)^{-1}$, $g(\cdot)$ being the density of W with $g'(\cdot)$, $g''(\cdot)$ being the first-order and second-order derivative of $g(\cdot)$, respectively and similarly, $\Upsilon'(\cdot)$, $\Upsilon''(\cdot)$ being the first-order and second-order derivative of $\Upsilon(\cdot)$.

Remark Theorem 3.2 indicates that the asymptotical independence of $\widehat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\boldsymbol{\alpha}''(u)}{2}\mu_2 h_u^2$ and $\widehat{\boldsymbol{\eta}}(w) - \boldsymbol{\eta}_0(w) - \zeta_0(w)\mu_2 h_w^2$ does not depend on the specific relationship between U and W. It illustrates that under general assumptions, if we try to refine the estimate of $\boldsymbol{\alpha}(u)$ using similar method as what we have done with $\widehat{\boldsymbol{\beta}}_\nu$, it would not help us to improve efficiency for the estimators of $\boldsymbol{\alpha}(u)$ based on validated sample. Thus, some new methodology is needed to study further about how to refine and improve $\widehat{\boldsymbol{\alpha}}(u)$ by utilizing auxilary information from the primary sample.

4. CONSTRUCT CONSISTENT ESTIMATORS FOR ASYMPTOTIC VARIANCE AND SELECT BANDWIDTH

To derive the updated estimators of parametric coefficients, i.e., Equation (2.10), a crucial step is to construct good estimators of variances and covariances between $\widehat{\boldsymbol{\beta}}_\nu$ and $\widehat{\boldsymbol{\gamma}}_\nu$. The quality of estimating those variance and covariance matrices will greatly impact on the corrected (or updated) estimator $\overline{\boldsymbol{\beta}}$. To better guide our simulation study and real data analysis later, a detailed discussion of how to construct those estimators is addressed in this section. Another factor that will influence the quality of the corrected (or updated) estimator $\overline{\boldsymbol{\beta}}$ is the choice of bandwidth in estimating nonparametric varying-coefficients in Model (1.1). A way using cross validation score function will be introduced here for selecting the optimal bandwidth to construct nonparametric estimators in Model (1.1).

4.1 Consistent estimators for asymptotic variances

The nature of the corrected (or updated) estimator $\overline{\boldsymbol{\beta}}$ will greatly rely on the estimations of asymptotic variance and covariance matrices, Σ_{12} and Σ_{22}, respectively. According to Theorem 2.2, the asymptotic variance of $\overline{\boldsymbol{\beta}}$, i.e., Δ is also related to Σ_{12} and Σ_{22} and besides, Δ depends on Σ_{11} as well. Also, the estimation of Δ will be the key for us to assess the performance of the corrected (or updated) estimator $\overline{\boldsymbol{\beta}}$. For instance, we can use the estimation of Δ to calculate the frequentist probability coverage of $\boldsymbol{\beta}$ in simulations shown in Section 5.

Firstly, let us take a look at constructing a consistent estimator of Σ_{11}, where $\Sigma_{11} = \sigma^2\Omega_u^{-1}$. Using the similar idea for proving Lemma A.2 in [9], we obtain

$$n^{-1}\mathbf{Z}^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{Z} \xrightarrow{p} \Omega_u.$$

Let $\widehat{\mathbf{Z}} = (\widehat{\mathbf{Z}}_1, ..., \widehat{\mathbf{Z}}_n)^T = \mathbf{S}\mathbf{Z}$, then a consistent estimator of Ω_u is

$$\widehat{\Omega}_u = \frac{1}{n}\sum_{i=1}^n(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)^T.$$

Without ambiguity, let us redefine $\widehat{\mathbf{M}} = \mathbf{S}(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}_\nu)$, which is different from Equation (2.4) by replacing $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}_\nu$. Note q is the dimension of $\widehat{\boldsymbol{\beta}}_\nu$. Then, the residuals sum of squares (RSS) in Model (1.1) can be calculated via

$$RSS = 1/(n-q)\sum_{i=1}^n\{Y_i - \widehat{\mathbf{M}}_i - \widehat{\boldsymbol{\beta}}_\nu^T\mathbf{Z}_i\}^2,$$

which is an unbiased and consistent estimator of σ^2.

To get an estimator of Δ, we further need to find suitable consistent estimators for Σ_{12} and Σ_{22}, respectively. According to Theorem 2.1, Σ_{22} consists of two parts. The first part is $\sigma^2\Omega_w^{-1}$, where σ^2 can be estimated using RSS mentioned above. Denote $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{Z}}_1, \cdots, \widetilde{\mathbf{Z}}_n)^T = \widetilde{\mathbf{S}}\mathbf{Z}$, then a consistent estimator of Ω_w is

$$\widehat{\Omega}_w = \frac{1}{n}\sum_{i=1}^n(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)^T.$$

The consistency of $\widehat{\Omega}_w$ is verified in the proof of Lemma A.3.

Meanwhile for the expression of the second part of Σ_{22}, i.e., $\Omega_w^{-1}\Omega_{\mathbf{M}}\Omega_w^{-1}$, the key to construct its consistent estimator is to develop a consistent estimator of $\Omega_{\mathbf{M}}$. Define $\widetilde{\mathbf{M}} = (\widetilde{\mathbf{M}}_1, \cdots, \widetilde{\mathbf{M}}_n)^T = \widetilde{\mathbf{S}}\widehat{\mathbf{M}}$, and replace τ by its consistent estimator $\hat{\tau} = n^{-1}\sum_{i=1}^n(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\widehat{\mathbf{M}}_i - \widetilde{\mathbf{M}}_i)$, then we could use the following estimator as a consistent estimator of $\Omega_{\mathbf{M}}$,

$$\widehat{\Omega}_{\mathbf{M}} = n^{-1}\sum_{i=1}^n\{(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)(\widehat{\mathbf{M}}_i - \widetilde{\mathbf{M}}_i)\}^{\otimes 2} - \hat{\tau}\hat{\tau}^T.$$

Here $C^{\otimes 2} = CC^T$. Similarly, the consistency of $\widehat{\Omega}_{\mathbf{M}}$ has been demonstrated in the proof of Lemma A.3.

Last, let us discuss the estimation of Σ_{12}. Notice that

$$\begin{aligned}\Sigma_{12} &= \sigma^2\Omega_u^{-1}\{\mathrm{E}(\mathbf{Z}\mathbf{Z}^T) - \mathrm{E}(H(W)) - \mathrm{E}(H(U)) \\ &+ \mathrm{E}(\Phi(U)^T\Gamma(U)^{-1}\mathbf{X}\mathbf{X}^T\Gamma(W)^{-1}\Phi(W))\}\Omega_w^{-1}.\end{aligned}$$

According the the proof of Theorem 2.1 in the Appendix,

$$\widehat{\Sigma}_{12} = \hat{\sigma}^2\widehat{\Omega}_u^{-1}\{\frac{1}{n}\sum_{i=1}^n(\mathbf{Z}_i - \widehat{\mathbf{Z}}_i)(\mathbf{Z}_i - \widetilde{\mathbf{Z}}_i)^T\}\widehat{\Omega}_w^{-1},$$

is a consistent estimator of Σ_{12}, where $\hat{\sigma}^2 = RSS$.

4.2 Selection of bandwidth

The selection of bandwidth is of great importance in nonparametric regression, which can largely impact on the accuracy of the estimations. However, according to the simulation study in [9], under the circumstance that the full data are observable, the estimation of $\boldsymbol{\beta}$ in Model (1.1) is insensitive to the choice of bandwidth. But the selection of bandwidth in the nonparametric part is still a open question. Thus, in practice, first, we often find a proper bandwidth to get an estimator of $\boldsymbol{\beta}$. Then, based on the estimation of $\boldsymbol{\beta}$, we estimate $\boldsymbol{\alpha}(u)$ by choosing an optimal bandwidth. Since in our case the part of the nonparametric covariate U is unobservable, only a few validation sample is available, the choice of bandwidth will also impact on the quality of the estimators of $\boldsymbol{\beta}$ to some extent.

Hence, first, we utilize the cross validation method to select an optimal bandwidth when estimating $\boldsymbol{\beta}$, which is widely applied in literatures of semiparametric varying coefficient models, e.g., [3], [9], and [34]. To estimate $\boldsymbol{\beta}$, it involves two bandwidths, h_u and h_w. Then, a cross validation score function of h_u and h_w is defined as below

(4.1)
$$CV(h_u, h_w) = n^{-1}\sum_{i=1}^n\left\{Y_i - \overline{\boldsymbol{\beta}}_{h_u,h_w,-i}^T\mathbf{Z}_i - \widehat{\boldsymbol{\alpha}}_{h_u,-i}^T(U_i)\mathbf{X}_i\right\}^2.$$

In (4.1), $\overline{\boldsymbol{\beta}}_{h_u,h_w,-i}$ is computed via Equation (2.10) using validated subsample $\{(\mathbf{Z}_j, \mathbf{X}_j, U_j, W_j, Y_j), j \neq i, j \in \{1, \cdots, n\}\}$ and $\{(\mathbf{Z}_j, \mathbf{X}_j, W_j, Y_j), j = n+1, \cdots, N\}$. The bandwidth h_u is used to compute $\widehat{\boldsymbol{\beta}}_{\nu,-i}$ without the ith observation in Equation (2.10), while the bandwidth h_w is applied to yield $\widehat{\boldsymbol{\gamma}}_{\nu,-i}$ and $\overline{\boldsymbol{\gamma}}_{-i}$ without the ith subject. $\widehat{\boldsymbol{\alpha}}_{h_u,-i}(\cdot)$ is obtained by replacing $\overline{\boldsymbol{\beta}}$ with $\overline{\boldsymbol{\beta}}_{h_u,h_w,-i}$ in Equation (3.1). Then the optimal bandwidths of h_u and h_w for estimating $\boldsymbol{\beta}$ are the value that minimize the cross validation scores function, i.e. $[h_{u,cv}, h_{w,cv}] = argmin_{h_u,h_w}CV(h_u, h_w)$. Second, after obtaining the estimator $\overline{\boldsymbol{\beta}}$, apply the cross validation idea again to choose the optimal bandwidth of $\widehat{\boldsymbol{\alpha}}(\cdot)$, and achieve a better estimation of $\boldsymbol{\alpha}(u)$.

5. SIMULATION STUDIES

In this section, we present some results of Monte Carlo simulations to show the finite sample performance of our proposed estimators. We are going to consider two situations to generate the validated subsamples about U. One is just a simple random draw from the whole sample, which can be regarded as missing completely at random. The other depends on the probability of the observed response Y and the surrogate variable W. The purpose is to investigate finite sample performance of our proposed estimator $\overline{\boldsymbol{\beta}}$ in comparison with the naive estimator $\widehat{\boldsymbol{\beta}}_\nu$, which is based on only the validation subsample and the benchmark estimator $\widehat{\boldsymbol{\beta}}_b$, which is based on the data where U is assumed to be exactly measured.

Example 5.1 A special case is supposed that the validation sample is a simple random subsample from the primary sample. Consider a SVPLM

$$Y = \sin(2\pi U)X_1 + \exp(U)X_2 + \beta Z + \epsilon,$$

where the covariate U is uniformly distributed on [0,1], $X_1 \sim N(1,1)$, $X_2 \sim N(-1, 2^2)$, $Z \sim N(-1, 3.5^2)$ and $\epsilon \sim N(0,1)$. Let $\beta = 2$ and we design the ratio to select validation subsample from the primary sample are 0.20 and 0.40, respectively. Consider three kinds of relationship between the surrogate variable W and covariate U:

(1) additive error: $W = U + \xi$, where $\xi \sim N(0,1)$ and is independent of ϵ.
(2) nonlinear error: $W = \exp(U) + U + \xi$, where $\xi \sim N(0,1)$ and is independent of ϵ.
(3) noninformative: $W = U' + \xi$, where $U' \sim N(0,1)$ and $\xi \sim N(0,1)$, respectively, and further, U' is independent of U and ξ is independent of U' and ϵ.

We simulate the data 1000 times for the setting mentioned above and assume in each simulation, the sample size is $N = 200$ and 400. Choose the kernel function to be

$$K_h(\cdot) = \frac{1}{h\sqrt{2\pi}}\exp(\frac{-(\cdot)^2}{2h^2}),$$

and use the cross-validation method to select an optimal bandwidth for $\overline{\beta}$, $\widehat{\beta}_\nu$ and $\widehat{\beta}_b$. For $\rho = 0.2$, the results of $\overline{\beta}$,

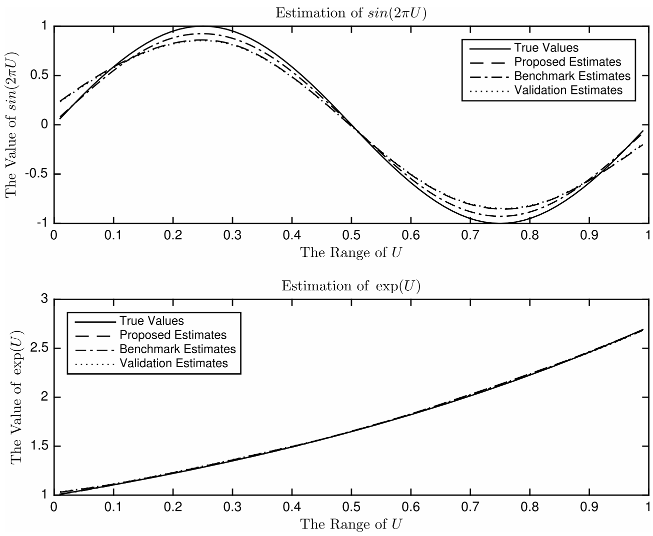

Figure 1. Example 5.1: $Y = sin(2\pi U)X_1 + exp(U)X_2 + \beta Z + \epsilon$ with $\rho = 0.2$ and $N = 200$. U and W satisfies $W = U + \xi$. The solid line represents the original curve, the dotted line represents estimation by the naive method, the dot dash line is estimated by the benchmark method, the dashed line is obtained by the proposed method.

$\widehat{\beta}_\nu$ and $\widehat{\beta}_b$ of $N = 200$ is listed in Table 1, while the results of $N = 400$ are shown in Table 7 in the Appendix B. According to Theorem 2.2, the proposed method is more asymptotic efficiency than the näive method. Table 1 supports this theoretic finding by showing that the proposed procedure is better than the näive estimator $\widehat{\beta}_\nu$ in finite samples: 1) The proposed estimator $\overline{\beta}$ has smaller standard deviation (SD), estimated standard error (SE) and mean square error (MSE) than the näive estimator $\widehat{\beta}_\nu$; and 2) the proposed method has higher frequentist coverage (Cov) of the truth within 95% confidence interval than that of the näive method. In addition, by comparing SD, SE, MSE and Cov with the benchmark estimator $\widehat{\beta}_b$, the proposed estimator $\overline{\beta}$ is also doing a good job. Similar conclusion can be made from Table 7 in Appendix B, though the overall performance of both the näive method and our proposed methods are better when the sample size is larger. For $\rho = 0.4$, the results of $\overline{\beta}$, $\widehat{\beta}_\nu$ and $\widehat{\beta}_b$ of $N = 200$ is listed in Table 11 in the Appendix B, while the results of $N = 400$ are shown in Table 8 in the Appendix B. The conclusions are similar as $\rho = 0.2$ and we skip the detailed discussion here.

Figure 1 listed as an example to illustrate the comparison for estimators of nonparametric varying-coefficients for the nonlinear relationship, i.e., $W = U + \xi$, when $\rho = 0.2$ and $N = 200$. The proposed method used Equation (3.1) to estimate $\boldsymbol{\alpha}(u)$, while the näive method replaced $\bar{\beta}$ in Equation (3.1) with $\widehat{\beta}_\nu$ to compute $\boldsymbol{\alpha}(u)$. The two estimates of $\boldsymbol{\alpha}(u)$ do not have obvious differences in the estimating nonparametric varying-coefficients in SVPLM models. It is reasonable

Table 1. The simulated results for the selection probability equal to 0.2 when $N = 200$

	$W = U + \xi$			$W = \exp(U)+U+\xi$			$W = U' + \xi$		
	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$
Bias	0.000	0.001	-0.001	0.000	0.001	-0.001	0.000	0.001	-0.001
SE	0.044	0.053	0.021	0.042	0.053	0.021	0.050	0.063	0.021
SD	0.042	0.051	0.021	0.040	0.051	0.021	0.046	0.059	0.022
MSE	0.002	0.003	0.000	0.002	0.003	0.000	0.003	0.004	0.000
Cov	0.945	0.937	0.946	0.944	0.937	0.946	0.940	0.933	0.964

Table 2. The simulated results for $\rho = 1/[1 + \exp(-1 + W - Y)]$ when $N = 200$ and $W = \exp(U) + U + \xi$

	$\overline{\beta}_1$	$\overline{\beta}_2$	$\overline{\beta}_3$	$\widehat{\beta}_{\nu,1}$	$\widehat{\beta}_{\nu,2}$	$\widehat{\beta}_{\nu,3}$	$\widehat{\beta}_{b,1}$	$\widehat{\beta}_{b,2}$	$\widehat{\beta}_{b,3}$
Bias	0.009	-0.004	0.013	-0.035	0.045	-0.046	0.002	-0.000	0.001
SE	0.050	0.054	0.147	0.070	0.075	0.187	0.021	0.023	0.075
SD	0.050	0.052	0.142	0.067	0.069	0.189	0.022	0.024	0.079
MSE	0.003	0.003	0.020	0.006	0.007	0.038	0.001	0.001	0.006
Cov	0.951	0.948	0.954	0.921	0.883	0.950	0.958	0.955	0.961

since the biases for both estimates are not large. For example, in Figure 1, the MSE for the estimation of $sin(2\pi U)$ relative to its true value on the interval of $[0, 1]$ are 0.011, 0.011 and 0.002 for the proposed method, the näive method and the benchmark, respectively. Those differences between the estimate and the truth of $\exp(U)$ are even invisible, all of their MSE are less than 0.000. Similar results are obtained for the nonlinear relationship of $W = \exp(U) + U + \xi$ and noninformative relationship $W = U' + \xi$ and also with the different sample size, i.e., $N = 400$ and with different selection probability, i.e., $\rho = 0.4$ shown in the Appendix B. To save the space, we omit other figures here.

Example 5.2 Another example is illustrated when the selection of validated subsamples is not uniform among the primary sample. In another word, the probability of a sample being selected will generally depend on observed values. Consider a covariate set $\mathbf{Z}$ has three predictors. Then, the coefficient of $\boldsymbol{\beta}$ is a three-dimensional vector. We select the validation subsample according to the scheme: 1) generate a uniform random variable m on $[0, 1]$, 2) if $m \le 1/[1+\exp(-1+W-Y)]$, we include the corresponding sample in validated subset. Employ a SVPLM model below,

$$Y = \cos(\sqrt{2}\pi U)X_1 + \sin(2\pi U)X_2 + \boldsymbol{\beta}^T\mathbf{Z} + \epsilon,$$

where the covariate U is uniformly distributed on $[0, 1]$, $X_1 \sim N(1, 1)$, $X_2 \sim N(-1, 2^2)$, $Z_1 \sim N(-1, 3.5^2)$, $Z_2 \sim N(2, 3^2)$, $Z_3 \sim N(0, 1)$, and $\epsilon \sim N(0, 1)$. Assume the true value of $\boldsymbol{\beta} = (2, -1, 3)'$ and we consider two relationships for W and U in this situation:

(1) nonlinear error: $W = \exp(U)+U+\xi$, where $\xi \sim N(0, 1)$ and is independent of ϵ.

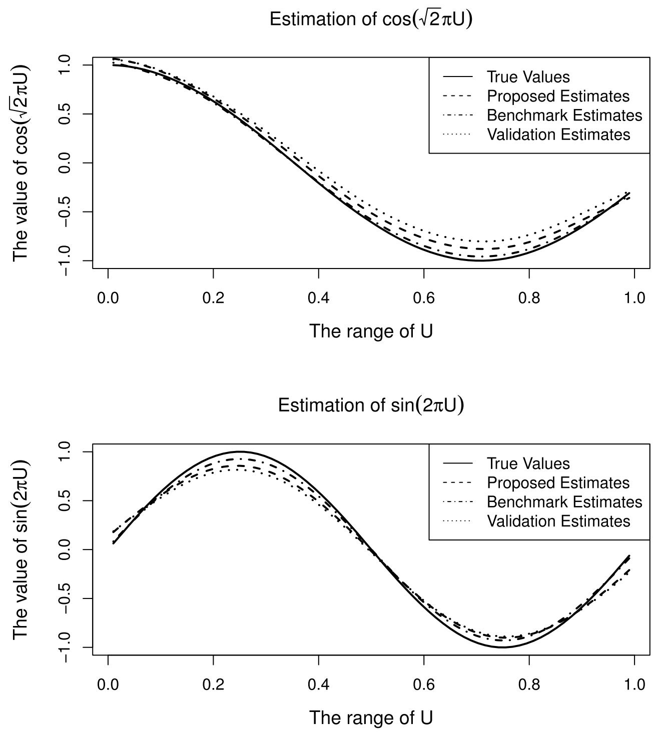

Figure 2. Example 5.2: $Y = \cos(\sqrt{2}\pi U)X_1 + \sin(2\pi U)X_2 + \boldsymbol{\beta}^T\mathbf{Z} + \epsilon$ *with* $\rho = 1/[1 + \exp(-1 + W - Y)]$. U *and* W *satisfies* $W = \exp(U) + U + \xi$. *The solid line represents the original curve, the dotted line represents estimation by the naive method, the dot dash line is estimated by the benchmark method, the dashed line is obtained by the proposed method.*

(2) noninformative: $W = U' + \xi$, where $U' \sim N(0,1)$ and $\xi \sim N(0,1)$, respectively, and further, U' is independent of U and ξ is independent of ξ.

As Example 5.1, we simulate the data 1000 times for the setting mentioned above and assume in each simulation, the sample size is $N = 200$ and 400. The estimated ρ of the nonlinear error situation (1) is equal to 0.266 for $N = 200$ in Table 2 and 0.267 for $N = 400$ in Table 10 of the Appendix B, while the estimated ρ of the noninformative situation (2) are 0.359 for $N = 200$ in Table 3 and 0.356 for $N = 400$ in Table 11 of the Appendix B. In Table 2, the bias of our proposed estimators $\overline{\beta}_1, \overline{\beta}_2$, and $\overline{\beta}_3$ are much smaller comparing to the näive estimators $\widehat{\beta}_{\nu,1}$, $\widehat{\beta}_{\nu,2}$ and $\widehat{\beta}_{\nu,3}$, respectively. Also, in view of SD and SE, the proposed estimators tend to have smaller values in relative to the näive estimators. Moreover, the SD of the proposed estimators are very close to the corresponding SE, which implies the theoretical derivations of the asymptotic variance of $\overline{\boldsymbol{\beta}}$ in Theorem 2.2 is consistent with the empirical findings. Noticeably, the Cov of the proposed estimator are much higher than that of the näive estimator, which shows the superiority of $\overline{\boldsymbol{\beta}}$. In term of MSE criterion, our proposed methods are

Table 3. The simulated results for $\rho = 1/[1 + \exp(-1 + W - Y)]$ *when* $N = 200$ *and* $W = U' + \xi$

	$\overline{\beta}_1$	$\overline{\beta}_2$	$\overline{\beta}_3$	$\widehat{\beta}_{\nu,1}$	$\widehat{\beta}_{\nu,2}$	$\widehat{\beta}_{\nu,3}$	$\widehat{\beta}_{b,1}$	$\widehat{\beta}_{b,2}$	$\widehat{\beta}_{b,3}$
Bias	0.008	-0.005	0.021	-0.046	0.045	-0.058	0.001	0.001	0.002
SE	0.048	0.047	0.134	0.059	0.058	0.161	0.021	0.023	0.072
SD	0.049	0.048	0.133	0.059	0.057	0.162	0.022	0.024	0.079
MSE	0.002	0.002	0.018	0.006	0.005	0.029	0.000	0.001	0.006
Cov	0.949	0.953	0.949	0.875	0.875	0.933	0.971	0.955	0.964

Table 4. First 10 data points of subsample by seed 1000

	MEDV	NOX	RM	TAX	PTRATIO	$\sqrt{\text{LSTAT}}$
1	34.700	0.469	7.185	242.000	17.800	2.007
2	28.700	0.458	6.430	222.000	18.700	2.283
3	19.900	0.538	5.834	307.000	21.000	2.910
4	20.200	0.538	5.456	307.000	21.000	3.419
5	13.600	0.538	5.570	307.000	21.000	4.585
6	14.500	0.538	5.813	307.000	21.000	4.459
7	18.400	0.538	6.495	307.000	21.000	3.578
8	21.000	0.538	6.674	307.000	21.000	3.461
9	14.500	0.538	6.072	307.000	21.000	3.611
10	20.000	0.499	5.841	279.000	19.200	3.378

more comparable to the benchmark estimation. Regarding to the estimation of the nonparametric part shown in Figure 2, since we obtain better estimates of $\boldsymbol{\beta}$, our proposed method performs better in estimating $\boldsymbol{\alpha}(u)$ than the näive method. Similar results are obtained for the noninformative relationship $W = U' + \xi$ see Table 3 for more details. Also, we investigate the performance of our proposed methods in comparison to the näive and benchmark methods with the different sample size, i.e., $N = 400$ in both situations. Similar conclusions can be drawn from Table 10 and Table 11, respectively, in the Appendix B. However, to save the space, we omit all other figures in the display.

6. REAL EXAMPLES

In this section, we illustrate the proposed method by investigating the Boston housing dataset from UCI machine learning repository, which has been also studied in [9]. The whole dataset can be downloaded from the website https://archive.ics.uci.edu/ml/datasets/Housing. This dataset constitutes the response variable, i.e., median value of owner-occupied homes (MEDV) in 506 US census tracts in the Boston area in 1970, and several covariates which can explain the variation in housing values [12]. For example, in our study, we are going to choose NOX (nitric oxide concentration in parts per 10 million), RM (average number of rooms per dwelling), TAX (full-value property-tax rate per \$ 10,000), PTRATIO (pupil-teacher ratio by town), LSTAT (percentage of lower status of the population) as explanatory variables. Moreover, when we do the exploratory data analysis, we have found that CRIM (per capita crime rate

Table 5. Estimators by different random seeds using Scheme 1

Scenario	$\overline{\beta}(T_n)$	$\widehat{\beta}_\nu(T_n)$	$\widehat{\beta}_b(T_n)$
1	-0.573(21.74)	-0.564(21.70)	-0.628(80.44)
2	-0.590(60.83)	-0.655(61.44)	-0.628(80.44)
3	-0.588(33.85)	-0.502(34.53)	-0.628(80.44)
4	-0.575(29.48)	-0.725(29.60)	-0.628(80.44)
5	-0.443(33.89)	-0.468(33.86)	-0.628(80.44)
6	-0.635(30.49)	-0.701(30.34)	-0.628(80.44)

Table 6. Estimators by different random seeds using Scheme 2

Scenario	$\overline{\beta}(T_n)$	$\widehat{\beta}_\nu(T_n)$	$\widehat{\beta}_b(T_n)$	$\hat{\rho}$
1	-0.618(34.85)	-0.607(34.85)	-0.628(80.44)	0.306
2	-0.464(66.28)	-0.478(66.33)	-0.628(80.44)	0.340
3	-0.420(29.13)	-0.460(29.21)	-0.628(80.44)	0.326
4	-0.594(31.86)	-0.686(31.84)	-0.628(80.44)	0.354
5	-0.659(49.75)	-0.641(49.89)	-0.628(80.44)	0.332
6	-0.311(39.85)	-0.258(39.67)	-0.628(80.44)	0.385

by town) and LSTAT are highly correlated. To investigate the benefits of our proposed methods, we are going to use CRIM as our surrogate variable, which is fully observed and presume LSTAT be partially observed. In practice, it is also more often to collect information on criminal rate than to observe lower status of the population. Thus, our assumption makes sense here.

To simplify the notations, we denote MEDV by Y, and denote the covariates NOX, RM, TAX, PTRATIO by X_1, X_2, X_3, Z, and set $U = \sqrt{\text{LSTAT}}$, and $W = \log(\text{CRIM})$. We consider the SVPLM model below for the analysis of the Boston housing data, i.e.,

$$Y = \alpha_1(U) + \alpha_2(U)X_1 + \alpha_3(U)X_2 + \alpha_4(U)X_3 + \beta Z + \epsilon. \tag{6.1}$$

A similar SVPLM model was used in Fan and Huang [9]. Further, we consider two situations for generating validated subsamples.

Scheme 1 : Assume the observed validated subsamples be randomly drawn from the whole population of Boston housing with the selection probability being 0.2.

Scheme 2 : Assume the observed validated subsamples be randomly drawn from the whole population of Boston housing with the selection probability $1/[3 + \exp(2 + W - Y)]$, which implies the validated subsamples are more difficult to collect when the criminal rate is increasing and the median value of owner-occupied homes is higher.

Using different random seeds, we will get different validated subsample according to schemes above. Thus, if different random seeds were picked, the selected validated subsamples in Boston housing datasets would be different, which would yield different results. For example, one of the validated subsample selected by Scheme 1 was shown in Table 4. To save the space, we only illustrate the first 10 observations in this validated subsample.

We display 6 different scenarios of random seeds with two schemes mentioned above to generate validated subsamples, the inferences of the parametric estimators are shown in Table 5 and Table 6, respectively. Moreover, in Table 6, we illustrate the estimated validation fraction $\hat{\rho}$ for the reference to the readers since it is expected that there are variations among the selection probabilities in Scheme 2 since they depend on the criminal rate and median value of owner-occupied homes. If we regard the benchmark estimator $\hat{\beta}_b$ as our best guess for the truth, then from Table 5 and Table 6, we can see our corrected (updated) estimator $\bar{\beta}$ is uniformly closer to the benchmark estimator $\hat{\beta}_b$ in most scenarios. Further, we examine the significance of $\beta \neq 0$ in the SVPLM model (6.1) using the profile likelihood ratio (PLR) test introduced in Fan and Huang [9]. For each scenario listed in Table 5 and Table 6, T_n is the corresponding PLR test statistic, and since $2T_n > \chi^2_{0.05} = 3.84$, the PLR test of $\beta \neq 0$ is significant at the nominal level 0.05 in all cases. This indicates the house value tends to be lower in the tracts where pupil-teacher ratio by town are higher, which is persistent with the findings in Fan and Huang [9].

7. CONCLUSION

In this paper, we have developed an effective estimating procedure to utilize available data from auxiliary covariates to estimate parametric coefficients in SVPLM models when we only have small proportional sample of covariates in the nonparametric component observed as validated samples. One great advantage of our resulting estimators is that their consistency does not depend on the specific relationship between the true covariates and the surrogate variables. Besides, they are more efficient than the validation-set-only estimators either shown in asymptotical properties or in the performance of finite sample simulations.

One of our next goal is to generalize our methodology to focus on the incorporation of the missing mechanism of validated subset into the estimation. If we can introduce this missing mechanism via an inverse probability weighted scheme in Equation (2.6), we envision we can get a better estimator for $\widehat{\boldsymbol{\beta}}_\nu$, which can further improve our corrected (updated) estimators $\overline{\boldsymbol{\beta}}$ in Equation (2.10). Similarly, by incorporating missing mechanism via an inverse probability weighted scheme, we are expected to improve the efficiency of the estimation of the nonparametric components $\boldsymbol{\alpha}(u)$.

APPENDIX A. ASYMPTOTIC PROPERTIES AND THEIR PROOFS

Without ambiguity, we suspend the subscript of h_u and h_w in the notation $\mathbf{D}_{u,h_u}$, $\mathbf{D}_{w,h_w}$ and similarly, for the notation $\mathbf{V}_{u,h_u}$ and $\mathbf{V}_{w,h_w}$. Also, we will use h, instead of distinguishing h_u and h_w, in the context. For example, the assumption (A.8) will be applicable to both h_u and h_w, so we use h to present as this common property applied to any bandwidth we discussed in the paper. Before we give out the outline of the proofs for those theorems in the text, the following assumptions are needed for deriving our results which could be found in many SVPLM references such as [9] and [34]:

(A.1) The random variable U has a bounded support $\mathcal{U}$. Its density function $f(\cdot)$ is Lipschitz continuous and bounded away from 0 on its support.
(A.2) The surrogate variable W has a bounded support $\mathcal{W}$ with $g(\cdot)$ being Lipschitz continuous density and bounded away from 0 on its support.
(A.3) The $p\times p$ matrix $\mathrm{E}(\mathbf{X}\mathbf{X}^T|U=u)$, are non-singular for each $u\in\mathcal{U}$. $\mathrm{E}(\mathbf{X}\mathbf{X}^T|U)$, $\mathrm{E}(\mathbf{X}\mathbf{X}^T|U)^{-1}$, and $\mathrm{E}(\mathbf{X}\mathbf{Z}^T|U)$ are all Lipschitz continuous.
(A.4) The $p\times p$ matrix $\mathrm{E}(\mathbf{X}\mathbf{X}^T|W)$, are non-singular for each $W\in\mathcal{W}$. $\mathrm{E}(\mathbf{X}\mathbf{X}^T|W)$, $\mathrm{E}(\mathbf{X}\mathbf{X}^T|W)^{-1}$, $\mathrm{E}(\mathbf{X}\mathbf{Z}^T|W)$, $\mathrm{E}(\mathbf{X}Y|W)$ and $\mathrm{E}(\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha}(u)|W)$ are all Lipschitz continuous.
(A.5) There is an $s>2$ such that $\mathrm{E}\|\mathbf{X}\|^{2s}<\infty$ and $\mathrm{E}\|\mathbf{Z}\|^{2s}<\infty$ for some $\varsigma<2-s^{-1}$ such that $N^{2\varsigma-1}h\to\infty$.
(A.6) $\{\alpha_i(\cdot), i=1,\cdots,p\}$ have the continuous second derivative in $U\in\mathcal{U}$ as well as $W\in\mathcal{W}$.
(A.7) The function $K(\cdot)$ is a symmetric density function with compact support and has finite second moment.
(A.8) $nh^8\to 0$ and $nh^2/(log(n))^2\to\infty$.

where (A.1), (A.2), (A.6), (A.7), (A.8) are common assumptions to make the estimate of nonparametric component to converge, whereas (A.3), (A.4), (A.5) are necessary conditions for proving the asymptotic normality of profile least-squares estimator of parametric component. And the following notation will be used in the proof of the lemmas and theorems below. Let $\mu_i=\int u^iK(u)du$, $\vartheta_i=\int u^iK^2(u)du$, $c_n=\{\frac{log(1/h)}{nh}\}^{1/2}+h^2$ and set $\Gamma(u)=\mathrm{E}(\mathbf{X}\mathbf{X}^T|U=u)$, $\Phi(u)=\mathrm{E}(\mathbf{X}\mathbf{Z}^T|U=u)$, $\Gamma(w)=\mathrm{E}(\mathbf{X}\mathbf{X}^T|W=w)$, $\Phi(w)=\mathrm{E}(\mathbf{X}\mathbf{Z}^T|W=w)$, $\Psi(w)=\mathrm{E}(\mathbf{X}Y|W=w)$, $\Upsilon(w)=\mathrm{E}(\mathbf{X}\mathbf{X}^T\boldsymbol{\alpha}(u)|W=w)$.

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

$$\begin{aligned}&\sup_x|\ \frac{1}{n}\sum_{i=1}^n[K_h(\mathbf{X}_i-x)Y_i-E\{K_h(\mathbf{X}_i-x)Y_i\}]\ |\\ =\ & O_p(\{\frac{\log(1/h)}{nh}\}^{1/2}).\end{aligned}$$

Proof. This follows immediately from Proposition 4 obtained by Mack and Silverman [19]. □

Lemma A.2. *Under the conditions (A.2),(A.4),(A.5)-(A.8), we can show that*

$$\widehat{\gamma}_\nu=\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Y}\xrightarrow{p}\boldsymbol{\gamma}_0$$

where $\boldsymbol{\gamma}_0=\Omega_w^{-1}\{\mathrm{E}(\mathbf{Z}Y)-E(\Phi^T(W)\Gamma^{-1}(W)\Psi(W))\}$ and $\Omega_w=\mathrm{E}(\mathbf{Z}\mathbf{Z}^T)-\mathrm{E}(\Phi(W)^T\Gamma(W)^{-1}\Phi(W))$. Similarly,

$$\overline{\boldsymbol{\gamma}}=\{\overline{\mathbf{Z}}^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Z}}\}^{-1}\overline{\mathbf{Z}}^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})^T(\overline{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Y}}\xrightarrow{p}\boldsymbol{\gamma}_0$$

Proof. Following the same idea of Lemma A.2 in [9], using Lemma A.1 and Assumptions (A.2), (A.4), we have

$$\text{(A.1)}\quad \frac{1}{n}\mathbf{D}_w^T\mathbf{V}_w\mathbf{D}_w=g(w)\Gamma(w)\otimes\begin{pmatrix}1&0\\0&\mu_2\end{pmatrix}\{1+O_p(c_n)\},$$

holding uniformly for any $w\in\mathcal{W}$ and note $g(\cdot)$ is the density of W. Similarly,

$$\text{(A.2)}\qquad \frac{1}{n}\mathbf{D}_w^T\mathbf{V}_w\mathbf{Z}=g(w)\Phi(w)\otimes(1,0)^T\{1+O_p(c_n)\},$$

holds uniformly for any $w\in\mathcal{W}$. Combining Equations (A.1) and (A.2), then the equation below holds uniformly for any $w\in\mathcal{W}$,

$$\begin{aligned}&[\mathbf{X}^T,\mathbf{0}_p](\mathbf{D}_w^T\mathbf{V}_w\mathbf{D}_w)^{-1}\mathbf{D}_w^T\mathbf{V}_w\mathbf{Z}\\ =\ &\mathbf{X}^T\Gamma(w)^{-1}\Phi(w)\{1+O_p(c_n)\}.\end{aligned}$$

This implies

$$\text{(A.3)}\qquad \begin{aligned}&(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\\ =&\begin{pmatrix}\mathbf{Z}_1^T-\mathbf{X}_1^T\Gamma(W_1)^{-1}\Phi(W_1)\{1+O_p(c_n)\}\\ \vdots\\ \mathbf{Z}_n^T-\mathbf{X}_n^T\Gamma(W_n)^{-1}\Phi(W_n)\{1+O_p(c_n)\}\end{pmatrix}.\end{aligned}$$

Therefore,

$$\begin{aligned}&n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\\ =\ &n^{-1}\sum_{i=1}^n\mathbf{Z}_i\mathbf{Z}_i^T\\ -\ &\Bigg(2n^{-1}\sum_{i=1}^n\mathbf{Z}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Phi(W_i)\\ +\ &n^{-1}\sum_{i=1}^n\Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Phi(W_i)\Bigg)\\ \times\ &\{1+O_p(c_n)\}.\end{aligned}$$

Using the Law of Large Numbers, we have

$$\begin{aligned}&n^{-1}\sum_{i=1}^n\mathbf{Z}_i\mathbf{Z}_i^T\xrightarrow{p}\mathrm{E}(\mathbf{Z}\mathbf{Z}^T),\\ &n^{-1}\sum_{i=1}^n\mathbf{Z}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Phi(W_i)\xrightarrow{p}\mathrm{E}(H(W)),\end{aligned}$$

$$n^{-1}\sum_{i=1}^{n}\Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Phi(W_i)$$
$$\xrightarrow{p}\quad \mathrm{E}(H(W)).$$

Therefore,

$$\text{(A.4)}\qquad n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\xrightarrow{p}\Omega_w.$$

Applying the same idea, we can prove

$$\begin{aligned}&n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Y}\\&\xrightarrow{p}\mathrm{E}(\mathbf{Z}Y)-\mathrm{E}(\Phi(W)^T\Gamma(W)^{-1}\Psi(W)).\end{aligned}$$

Then, we can conclude that

$$\widehat{\boldsymbol{\gamma}}_\nu=\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Y}\xrightarrow{p}\boldsymbol{\gamma}_0.$$

Similarly, we can show

$$\overline{\boldsymbol{\gamma}}=\{\overline{\mathbf{Z}}^T(\bar{\mathbf{I}}-\overline{\mathbf{S}})^T(\bar{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Z}}\}^{-1}\overline{\mathbf{Z}}^T(\bar{\mathbf{I}}-\overline{\mathbf{S}})^T(\bar{\mathbf{I}}-\overline{\mathbf{S}})\overline{\mathbf{Y}}\xrightarrow{p}\boldsymbol{\gamma}_0$$

holds. □

Lemma A.3. *Under the conditions (A.2),(A.4),(A.5)-(A.8),*

$$\widehat{\boldsymbol{\gamma}}_\nu=\{\mathbf{Z}^T(\mathbf{I}-\tilde{\mathbf{S}})^T(\mathbf{I}-\tilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\tilde{\mathbf{S}})^T(\mathbf{I}-\tilde{\mathbf{S}})\mathbf{Y}$$

is asymptotically normal, i.e.,

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}_\nu-\boldsymbol{\gamma}_0^*)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Sigma_{22}),$$

where $\Sigma_{22}=\sigma^2\Omega_w^{-1}+\Omega_w^{-1}\Omega_{\mathbf{M}}\Omega_w^{-1}$, $\Omega_{\mathbf{M}}=\mathrm{E}(\kappa\kappa^T)-\tau\tau^T$ *with* τ, κ *defined in Theorem 2.1 and* $\boldsymbol{\gamma}_0^*=\boldsymbol{\beta}_0+\Omega_w^{-1}\tau$. *Combining with the result* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_\nu-\boldsymbol{\beta}_0)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Sigma_{11})$, *it is easy to see that* $\boldsymbol{\gamma}_0^*=\boldsymbol{\gamma}_0$.

Proof. If we substitute the model $Y=\boldsymbol{\alpha}^T(U)\mathbf{X}+\boldsymbol{\beta}^T\mathbf{Z}+\epsilon$ into the expression of $\widehat{\boldsymbol{\gamma}}_\nu$ (i.e.(2.9)), then we have

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}_\nu=\boldsymbol{\beta}_0&+\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}\\&+\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\boldsymbol{\epsilon}.\end{aligned}$$

The second term might create bias when we use $\widehat{\boldsymbol{\gamma}}_\nu$ to estimate $\boldsymbol{\beta}_0$. From conditions (A.2), (A.4) and Lemma A.2, we can first simplify,

$$\frac{1}{n}\mathbf{D}_w^T\mathbf{V}_w\mathbf{M}=g(w)\Upsilon(w)\otimes(1,0)^T\{1+O_p(c_n)\}.$$

Together with (A.1), the following equation holds uniformly in $w\in\mathcal{W}$,

$$\begin{aligned}&[\mathbf{X}^T,\mathbf{0}_p]\{\mathbf{D}_w^T\mathbf{V}_w\mathbf{D}_w\}^{-1}\mathbf{D}_w^T\mathbf{V}_w\mathbf{M}\\=\;&\mathbf{X}^T\Gamma(W)^{-1}\Upsilon(w)\{1+O_p(c_n)\}.\end{aligned}$$

Then, using (A.3), we have

$$\begin{aligned}&n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}\\=\;&n^{-1}\sum_{i=1}^{n}[\mathbf{Z}_i-\Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\{1+O_p(c_n)\}]\\\times\;&[\mathbf{X}_i^T\alpha(U_i)-\mathbf{X}_i^T\Gamma(W_i)^{-1}\Upsilon(W_i)\{1+O_p(c_n)\}].\end{aligned}$$

As before, using the Law of Large Numbers and the results of (A.4), we have

$$\begin{aligned}\text{(A.5)}\quad&\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{Z}\}^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}\\\xrightarrow{p}\quad&\Omega_w^{-1}\{\mathrm{E}(\Phi(U)\boldsymbol{\alpha}(U))-\mathrm{E}(\Phi(W)^T\Gamma(W)^{-1}\Upsilon(W))\}.\end{aligned}$$

Then, we can rewrite the expression of $\sqrt{n}(\widehat{\boldsymbol{\gamma}}_\nu-\boldsymbol{\gamma}_0^*)$ as

$$\begin{aligned}&\sqrt{n}(\widehat{\boldsymbol{\gamma}}_\nu-\boldsymbol{\gamma}_0^*)\\=\;&\sqrt{n}(\widehat{\boldsymbol{\gamma}}_\nu-\boldsymbol{\beta}_0-\Omega_w^{-1}\tau)\\=\;&\sqrt{n}\Omega_w^{-1}\frac{1}{n}\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}-n\tau\}\\+\;&\sqrt{n}\Omega_w^{-1}\frac{1}{n}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\boldsymbol{\epsilon}+o_p(1)\\\triangleq\;&I_1+I_2+o_p(1).\end{aligned}$$

First, let us consider I_2. Notice any element in $\widetilde{\mathbf{S}}\boldsymbol{\epsilon}$ satisfies

$$[\mathbf{X}^T,\mathbf{0}_p](\mathbf{D}_w^T\mathbf{V}_w\mathbf{D}_w)^{-1}\mathbf{D}_w^T\mathbf{V}_w\boldsymbol{\epsilon}=O_p(c_n).$$

Applying (A.3), $n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\boldsymbol{\epsilon}$ can be written as

$$\begin{aligned}\frac{1}{n}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\boldsymbol{\epsilon}\;&=\;\frac{1}{n}\sum_{i=1}^{n}\{\mathbf{Z}_i-\Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\}\\&\times\;\epsilon_i\{1+o_p(1)\}.\end{aligned}$$

By the Central Limit Theorem (CLT), we could obtain the asymptotic normality of I_2, i.e.,

$$\text{(A.6)}\quad n^{-1/2}\Omega_w^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\boldsymbol{\epsilon}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma^2\Omega_w^{-1}).$$

Next, consider I_1, where $n^{-1}\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}-n\tau\}$ can be written as

$$\begin{aligned}&n^{-1}\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}-\tau\\=\;&n^{-1}\sum_{i=1}^{n}\big\{\mathbf{Z}_i\mathbf{X}_i^T\alpha(U_i)-\mathbf{Z}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Upsilon(W_i)\\-\;&\Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\mathbf{X}_i^T\alpha(U_i)+\Phi(W_i)^T\Gamma(W_i)^{-1}\\\times\;&\mathbf{X}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Upsilon(W_i)-\tau\big\}\{1+O_p(c_n)\}\\\triangleq\;&n^{-1}\sum_{i=1}^{n}\{\kappa_i-\tau\}\{1+O_p(c_n)\},\end{aligned}$$

where

$$\kappa_i\;=\;\mathbf{Z}_i\mathbf{X}_i^T\alpha(U_i)-\mathbf{Z}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Upsilon(W_i)$$

$$-\quad \Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\mathbf{X}_i^T\alpha(U_i)$$
$$+\quad \Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\mathbf{X}_i^T\Gamma(W_i)^{-1}\Upsilon(W_i).$$

Following the assumption in the Introduction that $(W_i, U_i, \mathbf{X}_i, \mathbf{Z}_i)$ are i.i.d random vectors for any $i = 1, \cdots, n$ and using the CLT, then I_1 is asymptotic normal, i.e.,

$$\text{(A.7)}\qquad n^{-1/2}\Omega_w^{-1}\{\mathbf{Z}^T(\mathbf{I}-\widetilde{\mathbf{S}})^T(\mathbf{I}-\widetilde{\mathbf{S}})\mathbf{M}-n\tau\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_w^{-1}\Omega_{\mathbf{M}}\Omega_w^{-1}).$$

Moreover, from Model (1.1), we know ϵ and $(W, U, \mathbf{X}^T, \mathbf{Z}^T)$ are independent and $\mathrm{E}(\epsilon) = 0$. Thus, the cross section term of I_1 and I_2 are zero, and then

$$\begin{aligned}&\sqrt{n}(\widehat{\gamma}_\nu - \gamma_0^*)\\ =\;& n^{-1/2}\Omega_w^{-1}\sum_{i=1}^n\{\kappa_i-\tau\}\\ +\;& n^{-1/2}\Omega_w^{-1}\sum_{i=1}^n\{\mathbf{Z}_i - \Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\}\epsilon_i + o_p(1).\end{aligned}$$

Using asymptotic expressions (A.6) and (A.7), we have

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}_\nu - \boldsymbol{\gamma}_0^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{22}).$$

Thus, it is easy to derive that $\widehat{\gamma}_\nu \xrightarrow{p} \boldsymbol{\gamma}_0^*$, From Lemma A.2, we have $\widehat{\gamma}_\nu \xrightarrow{p} \boldsymbol{\gamma}_0$. Hence, for the uniqueness of limits, $\boldsymbol{\gamma}_0^*$ must be equal to $\boldsymbol{\gamma}_0$. □

Lemma A.4. *Under the conditions (A.2),(A.4),(A.5)-(A.8), the estimator of $\boldsymbol{\gamma}_0$ based on full samples, i.e., $\overline{\boldsymbol{\gamma}}$, is asymptotically normal,*

$$\sqrt{N}(\overline{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{22}),$$

where Σ_{22} is the same as defined in Lemma A.3.

Proof. The outline of the proof of this lemma is directly follow the proof of Lemma A.3. We just need to replace n with N and use notations of full data set in the proof of Lemma A.3. □

Proof of Theorem 2.1 in Section 2.

Proof. Let $\widehat{\boldsymbol{\theta}}_\nu = \begin{pmatrix}\widehat{\boldsymbol{\beta}}_\nu\\ \widehat{\boldsymbol{\gamma}}_\nu\end{pmatrix}$ and $\boldsymbol{\theta}_0 = \begin{pmatrix}\boldsymbol{\beta}_0\\ \boldsymbol{\gamma}_0\end{pmatrix}$, then the goal becomes to derive

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\nu - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma).$$

By an analogy to Theorem 4.1 in [9] and from Lemma A.3, we have

$$\begin{aligned}&\sqrt{n}(\widehat{\boldsymbol{\theta}}_\nu - \boldsymbol{\theta}_0)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\begin{pmatrix}\Omega_u^{-1}\{\mathbf{Z}_i - \Phi(U_i)^T\Gamma(U_i)^{-1}\mathbf{X}_i\}\epsilon_i\\ \Omega_w^{-1}[\{\mathbf{Z}_i - \Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i\}\epsilon_i + \{\kappa_i-\tau\}]\end{pmatrix}\\ &\quad\times\{1+o_p(1)\}\end{aligned}$$

Then using the Slutsky Theorem and the CLT, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\nu - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}.$$

Noticing that ϵ_i and κ_i are mutually independent, then

$$\begin{aligned}\Sigma_{12} = \sigma^2\Omega_u^{-1}\{&\mathrm{E}(\mathbf{Z}\mathbf{Z}^T) - \mathrm{E}(H(W)) - \mathrm{E}(H(U))\\ &+ \mathrm{E}(\Phi(U)^T\Gamma(U)^{-1}\mathbf{X}\mathbf{X}^T\Gamma(W)^{-1}\Phi(W))\}\Omega_w^{-1}.\end{aligned}$$

This ends the proof of Theorem 2.1. □

Proof of Theorem 2.2 in Section 2.

Proof. From (2.10), we know that the expression of $\overline{\boldsymbol{\beta}}$ is

$$\overline{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_\nu - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}(\widehat{\gamma}_\nu - \overline{\gamma}).$$

Then $\overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ could be rewritten as

$$\begin{aligned}\text{(A.8)}\quad &\overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\\ &= (\widehat{\boldsymbol{\beta}}_\nu - \boldsymbol{\beta}_0) - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}(\widehat{\boldsymbol{\gamma}}_\nu - \boldsymbol{\gamma}_0) + \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}(\overline{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0).\end{aligned}$$

Set $\widehat{\Theta} = (\widehat{\boldsymbol{\beta}}_\nu^T, \widehat{\boldsymbol{\gamma}}_\nu^T, \overline{\boldsymbol{\gamma}}^T)^T$ and $\Theta_0 = (\boldsymbol{\beta}_0^T, \boldsymbol{\gamma}_0^T, \boldsymbol{\gamma}_0^T)^T$. If we can derive the asymptotical normality of $\sqrt{n}(\widehat{\Theta} - \Theta_0)$, then the asymptotical normality of $\sqrt{n}(\overline{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is a linear combination of $\sqrt{n}(\widehat{\Theta} - \Theta_0)$ using $(1, -\widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}, \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1})$ (c.f., Theorem 3.3.4 in [27]).

Similar as the derivation in the Proof of Theorem 2.1, we have

$$\sqrt{n}(\widehat{\Theta} - \Theta_0) \triangleq A_1 + A_2 + o_p(1),$$

where

$$A_1 = \sqrt{n}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n\Omega_u^{-1}[t(U_i)]\epsilon_i\\ \frac{1}{n}\sum_{i=1}^n\Omega_w^{-1}\{t(W_i)\epsilon_i + [\kappa_i-\tau]\}\\ \frac{1}{n}\sum_{i=1}^n\rho\Omega_w^{-1}\{t(W_i)\epsilon_i + [\kappa_i-\tau]\}\end{pmatrix},$$

$$A_2 = \sqrt{n}\begin{pmatrix}0\\ 0\\ \frac{\sqrt{\rho(1-\rho)}}{\sqrt{N-n}}\sum_{i=n+1}^N\Omega_w^{-1}\{t(W_i)\epsilon_i + [\kappa_i-\tau]\}\end{pmatrix},$$

where $t(W_i) = \mathbf{Z}_i - \Phi(W_i)^T\Gamma(W_i)^{-1}\mathbf{X}_i$ and $t(U_i) = \mathbf{Z}_i - \Phi(U_i)^T\Gamma(U_i)^{-1}\mathbf{X}_i$. Noticing that $\{\epsilon_i\}_{i=1}^n$ and $\{\kappa_i\}_{i=1}^n$ are mutually independent, using the CLT,

$$\begin{aligned}A_1 &\xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12} & \rho\Sigma_{12}\\ \Sigma_{12} & \Sigma_{22} & \rho\Sigma_{22}\\ \rho\Sigma_{12} & \rho\Sigma_{22} & \rho\Sigma_{22}\end{pmatrix}\right),\\ A_2 &\xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & \rho(1-\rho)\Sigma_{22}\end{pmatrix}\right).\end{aligned}$$

Since $\{\epsilon_i\}_{i=1}^n$ and $\{\kappa_i\}_{i=1}^n$ are i.i.d., combining the asymptotic normality of A_1 and A_2, we obtain

$$\sqrt{n}(\widehat{\Theta}-\Theta_0)\xrightarrow{\mathcal{D}}\mathcal{N}\left(\mathbf{0},\begin{pmatrix}\Sigma_{11} & \Sigma_{12} & \rho\Sigma_{12}\\ \Sigma_{12} & \Sigma_{22} & \rho\Sigma_{22}\\ \rho\Sigma_{12} & \rho\Sigma_{22} & \rho\Sigma_{22}\end{pmatrix}\right).$$

As $\overline{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=(1,-\widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1},\widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1})(\widehat{\Theta}-\Theta_0)$ and $\widehat{\Sigma}_{12}$, $\widehat{\Sigma}_{22}$ are consistent estimators of Σ_{12} and Σ_{22}, respectively, then using Theorem 3.3.4 in [27] again, we get the result

$$\sqrt{n}(\overline{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Delta)$$

where $\Delta=\Sigma_{11}-(1-\rho)\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, and here $\Sigma_{21}=\Sigma_{12}^T$. □

Proof of Lemma 3.1 in Section 3.

Proof. Based on the proof of Lemma A.2 in [9],

$$\frac{1}{n}\mathbf{D}_u^T\mathbf{V}_u\mathbf{D}_u=f(U)\Gamma(U)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\{1+O_p(c_n)\},$$

and representing Y using Model (1.1), then the estimator of $\boldsymbol{\alpha}(u)$ in Equation (3.1) can be rewritten as

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}(u)&=[\mathbf{I}_p,\mathbf{0}_p](\mathbf{D}_u^T\mathbf{V}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{V}_u(\mathbf{Y}-\mathbf{Z}\overline{\boldsymbol{\beta}})\\
&=[\mathbf{I}_p,\mathbf{0}_p](\mathbf{D}_u^T\mathbf{V}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{V}_u(\boldsymbol{\alpha}^T(U)\mathbf{X}+\boldsymbol{\epsilon})\\
&+[\mathbf{I}_p,\mathbf{0}_p](\mathbf{D}_u^T\mathbf{V}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{V}_u\mathbf{Z}(\boldsymbol{\beta}_0-\overline{\boldsymbol{\beta}})\\
&=f(u)^{-1}\Gamma(u)^{-1}\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i\epsilon_iK_h(U_i-u)\{1+O_p(c_n)\}\\
&+f(u)^{-1}\Gamma(u)^{-1}\frac{1}{n}\\
&\times\sum_{i=1}^n\mathbf{X}_i\mathbf{X}_i^T\boldsymbol{\alpha}(U_i)K_h(U_i-u)\{1+O_p(c_n)\}\\
&+f(u)^{-1}\Gamma(u)^{-1}\\
&\times\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i\mathbf{Z}_i^T(\boldsymbol{\beta}_0-\overline{\boldsymbol{\beta}})K_h(U_i-u)\{1+O_p(c_n)\}.\end{aligned}$$

For any $U_i\in(u-h,u+h)$, using Taylor's expansion we have

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

Plugging the above expression into

$$\sum_{i=1}^n\mathbf{X}_i\mathbf{X}_i^T\alpha(U_i)K_h(U_i-u),$$

we have

$$\begin{aligned}\text{(A.9)}\quad&\widehat{\boldsymbol{\alpha}}(u)-\boldsymbol{\alpha}(u)-\frac{\boldsymbol{\alpha}''(u)}{2}\mu_2h^2+o_p(h^2)\\
=\;&f(u)^{-1}\Gamma(u)^{-1}\frac{1}{n}\sum_{i=1}^n\mathbf{X}_i\epsilon_iK_h(U_i-u)\times\{1+O_p(c_n)\}\\
+\;&\Gamma(U)^{-1}\Phi(u)(\boldsymbol{\beta}_0-\overline{\boldsymbol{\beta}})+o_p(n^{-\frac{1}{2}}).\end{aligned}$$

Then, according to Theorem 2.1, $\sqrt{n}(\boldsymbol{\beta}_0-\overline{\boldsymbol{\beta}})\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Delta)$, which implies $(\boldsymbol{\beta}_0-\overline{\boldsymbol{\beta}})=o_p(n^{-\frac{1}{2}})$. Combining this into (A.9) and using CLT, we have $\sqrt{nh}\{\widehat{\boldsymbol{\alpha}}(u)-\boldsymbol{\alpha}(u)-\frac{\boldsymbol{\alpha}''(u)}{2}\mu_2h^2+o_p(h^2)\}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Lambda_{11})$. □

Lemma A.5. *Under assumption (A.2),(A.4), (A.5)-(A.8), the estimator of nonparametric function $\boldsymbol{\eta}(w)$ based on the validation subsample, $\widehat{\boldsymbol{\eta}}(w)$, has the following asymptotic normality*

$$\sqrt{nh}\{\widehat{\boldsymbol{\eta}}(w)-\boldsymbol{\eta}_0(w)-\zeta_0(w)\mu_2h^2+o_p(h^2)\}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Lambda_{22}),$$

where $\Lambda_{22}=\sigma^2\vartheta_0g(w)^{-1}\Gamma(W)^{-1}$ and

$$\begin{aligned}\zeta_0(w)&=\Gamma(w)^{-1}\frac{\Upsilon''(w)}{2}+g(w)^{-1}\Gamma(W)^{-1}[g'(w)\Upsilon'(w)\\
&+\frac{g''(w)}{2}\Upsilon(w).\end{aligned}$$

Lemma A.6. *Under assumption (A.2),(A.4), (A.5)-(A.8), the estimator of nonparametric function $\boldsymbol{\eta}(w)$ based on the full sample, $\overline{\boldsymbol{\eta}}(w)$, has the following asymptotic normality*

$$\sqrt{Nh}\{\overline{\boldsymbol{\eta}}(w)-\boldsymbol{\eta}_0(w)-\zeta_0(w)\mu_2h^2+o_p(h^2)\}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Lambda_{22}),$$

where $\Lambda_{22}=\sigma^2\vartheta_0g(w)^{-1}\Gamma(w)^{-1}$.

Note: The proof of Lemma A.5 is the similar as that of Lemma 3.1, while the proof of Lemma A.6 is the same as that of Lemma A.5 except replacing n with N. Thus, we omit the details here.

Proof of Theorem 3.2 in Section 3.

Proof. In this proof, we will reuse h_u and h_w to distinguish the bandwidth used for estimating $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\eta}(\cdot)$, respectively. From Lemma A.5 and Lemma 3.1, it is easy to verify that

$$\begin{aligned}&\sqrt{nh}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}(u)-\boldsymbol{\alpha}(u)-\frac{\boldsymbol{\alpha}''(u)}{2}\mu_2h_u^2\\ \widehat{\boldsymbol{\eta}}(w)-\boldsymbol{\eta}_0(w)-\zeta_0(w)\mu_2h_w^2\end{pmatrix}\\
=\;&\frac{1}{\sqrt{nh}}\sum_{i=1}^n\begin{pmatrix}f(u)^{-1}\Gamma(U)^{-1}X_i\epsilon_iK(\frac{U_i-u}{h_u})\\ g(w)^{-1}\Gamma(W)^{-1}X_i\epsilon_iK(\frac{W_i-w}{h_w})\end{pmatrix}\\
\times\;&\{1+o_p(1)\}.\end{aligned}$$

with the assumption $h_u=O(h)$ and $h_w=O(h)$. Since $\{\epsilon_i,i=1,2,...,n\}$ are i.i.d random variable, then following the CLT,

$$\sqrt{nh}\begin{pmatrix}\widehat{\boldsymbol{\alpha}}(u)-\boldsymbol{\alpha}(u)-\frac{\boldsymbol{\alpha}''(u)}{2}\mu_2h_u^2\\ \widehat{\boldsymbol{\eta}}(w)-\boldsymbol{\eta}_0(w)-\zeta_0(w)\mu_2h_w^2\end{pmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Lambda).$$

where $\Lambda\equiv\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix}$. Definition of Λ_{11} and Λ_{22} are given in Lemma A.5 and Lemma 3.1, and $\Lambda_{12}=\Lambda_{21}^T$, where

$$\Lambda_{12}=\frac{1}{h}f(u)^{-1}\Gamma(u)^{-1}\sigma^2$$

Table 7. The simulated results for the selection probability equal to 0.2 when $N = 400$

	$W = U + \xi$			$W = \exp(U)+U+\xi$			$W = U' + \xi$		
	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$
Bias	0.000	0.001	-0.000	0.001	-0.000	0.001	-0.001	-0.003	-0.000
SE	0.033	0.040	0.015	0.033	0.041	0.015	0.033	0.038	0.015
SD	0.033	0.040	0.015	0.031	0.040	0.015	0.033	0.040	0.015
MSE	0.001	0.002	0.000	0.001	0.002	0.000	0.001	0.002	0.000
Cov	0.951	0.949	0.953	0.944	0.947	0.955	0.960	0.960	0.951

Table 8. The simulated results for the selection probability equal to 0.4 when $N = 200$

	$W = U + \xi$			$W = \exp(U)+U+\xi$			$W = U' + \xi$		
	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$
Bias	0.001	0.001	0.000	-0.000	0.001	-0.000	0.000	0.001	0.001
SE	0.033	0.037	0.021	0.031	0.037	0.021	0.033	0.036	0.021
SD	0.031	0.036	0.021	0.030	0.036	0.021	0.031	0.035	0.021
MSE	0.001	0.001	0.000	0.001	0.001	0.000	0.001	0.001	0.000
Cov	0.941	0.935	0.956	0.937	0.930	0.955	0.942	0.938	0.941

$$\times \quad \mathrm{E}[\mathbf{X}\mathbf{X}^T K(\frac{U-u}{h_u})K(\frac{W-w}{h_w})]g(w)^{-1}\Gamma(w)^{-1}.$$

If we assume the joint density of (U, W) exists, and denote it as $l(u, w)$, which is Lipchitz continuous as well as twice differentiable. Further, define $\varphi(u, w) = \mathrm{E}(\mathbf{X}\mathbf{X}^T|U = u, W = w)$, then

$$\begin{aligned}
& \mathrm{E}[\mathbf{X}\mathbf{X}^T K(\frac{U-u}{h_u})K(\frac{W-w}{h_w})] \\
= \; & \mathrm{E}[\varphi(u,w)K(\frac{U-u}{h_u})K(\frac{W-w}{h_w})] \\
= \; & \int\int \varphi(U,W)K(\frac{U-u}{h_u})K(\frac{W-w}{h_w})l(U,W)dUdW \\
= \; & h_u \int K(\frac{W-w}{h_w}) \\
\times \; & \left\{\int \varphi(th_u+u,W)l(th_u+u,W)K(t)dt\right\}dW \\
\approx \; & h_u \int K(\frac{W-w}{h_w})\varphi(u,W)l(u,W)dW \\
= \; & h_u h_w \int K(s)\varphi(u,sh_w+w)l(u,sh_w+w)ds \\
\approx \; & O(h^2)\varphi(u,w)l(u,w).
\end{aligned}$$

Therefore, we have $\Lambda_{12} = O(h)$. Similarly, $\Lambda_{21} = O(h)$. This shows when $h \to 0$, $\widehat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}(u) - \frac{\boldsymbol{\alpha}''(u)}{2}\mu_2 h_u^2$ and $\widehat{\boldsymbol{\eta}}(w) - \boldsymbol{\eta}_0(w) - \zeta_0(w)\mu_2 h_w^2$ are asymptotically independent. □

APPENDIX B. ADDITIONAL SIMULATION RESULTS FOR SECTION 5

In this section, we present several different simulation results for Example 5.1 and Example 5.2. Table 7 to Table 9 illustrate the results yielded from different scenarios for Example 5.1. In Table 7, it shows our proposed method and the näive method are both comparable to the benchmark estimates when $N = 400$ under the selection probability equal to 0.2. It makes sense since in this situation even the näive method will have large enough sample size to estimate the unknown parameter. Table 8 and Table 9 display the results for the selection probability equal to 0.4 for $N = 200$ and $N = 400$, respectively. Table 8 shows that the performance of our proposed method is outweigh the näive method in term of comparing SD, SE, MSE and Cov. Moreover, in comparison to the benchmark estimation, our proposed method is doing a good job in estimating β in the SVPLM model. However, both our proposed method and the näive method are producing good estimates of β when the sample size is larger as shown in Table 9. Table 10 and Table 11 display the results for Example 5.2 when the sample size increases to $N = 400$ for the nonlinear relationship $W = \exp(U) + U + \xi$ and the noninformative relationship $W = U' + \xi$, respectively. In either $N = 400$ or $N = 200$ situation, our proposed method is doing much better in comparison to the näive method for both nonlinear and noninformative rela-

Table 9. The simulated results for the selection probability equal to 0.4 when $N = 400$

	$W = U + \xi$			$W = \exp(U)+U+\xi$			$W = U' + \xi$		
	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$	$\overline{\beta}$	$\widehat{\beta}_\nu$	$\widehat{\beta}_b$
Bias	-0.000	0.000	-0.000	0.001	0.001	-0.001	0.001	0.001	0.001
SE	0.033	0.039	0.015	0.032	0.039	0.015	0.032	0.037	0.014
SD	0.033	0.040	0.015	0.031	0.041	0.015	0.033	0.040	0.015
MSE	0.001	0.002	0.000	0.001	0.002	0.000	0.001	0.002	0.000
Cov	0.959	0.954	0.961	0.946	0.953	0.955	0.953	0.963	0.967

Table 10. The simulated results for $\rho = 1/[1 + \exp(-1 + W - Y)]$ when $N = 400$ and $W = \exp(U) + U + \xi$

	$\overline{\beta}_1$	$\overline{\beta}_2$	$\overline{\beta}_3$	$\widehat{\beta}_{\nu,1}$	$\widehat{\beta}_{\nu,2}$	$\widehat{\beta}_{\nu,3}$	$\widehat{\beta}_{b,1}$	$\widehat{\beta}_{b,2}$	$\widehat{\beta}_{b,3}$
Bias	0.010	-0.010	0.025	-0.033	0.038	-0.031	-0.000	-0.001	0.002
SE	0.032	0.034	0.096	0.043	0.046	0.122	0.015	0.016	0.052
SD	0.034	0.036	0.097	0.045	0.047	0.128	0.015	0.017	0.054
MSE	0.001	0.001	0.010	0.003	0.004	0.017	0.000	0.000	0.003
Cov	0.955	0.954	0.945	0.892	0.866	0.953	0.961	0.963	0.960

Table 11. The simulated results for $\rho = 1/[1 + \exp(-1 + W - Y)]$ when $N = 400$ and $W = U' + \xi$

	$\overline{\beta}_1$	$\overline{\beta}_2$	$\overline{\beta}_3$	$\widehat{\beta}_{\nu,1}$	$\widehat{\beta}_{\nu,2}$	$\widehat{\beta}_{\nu,3}$	$\widehat{\beta}_{b,1}$	$\widehat{\beta}_{b,2}$	$\widehat{\beta}_{b,3}$
Bias	0.011	-0.009	0.028	-0.042	0.042	-0.046	-0.000	0.000	0.000
SE	0.031	0.031	0.090	0.038	0.037	0.104	0.014	0.015	0.052
SD	0.033	0.033	0.091	0.040	0.039	0.110	0.015	0.017	0.054
MSE	0.001	0.001	0.009	0.003	0.003	0.014	0.000	0.000	0.003
Cov	0.951	0.956	0.950	0.830	0.830	0.950	0.965	0.961	0.950

tionship when the selection probability depends on observed values. Our method tends to have smaller bias and higher frequentist coverage.

Received 20 July 2016

REFERENCES

[1] Bai, Y., Hu, J., You, J., 2015. Panel data partially linear varying-coefficient model with errors correlated in space and time. Statistica Sinica 25, 275–294. MR3328815

[2] Bravo, F., 2013. Partially linear varying coefficient models with missing at random responses. Annals of the Institute of Statistical Mathematics 65 (4), 721–762. MR3094954

[3] Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. Journal of the American Statistical Association 95, 888–902. MR1804446

[4] Chen, Y. H., 2002. Cox regression in cohort studies with validation sampling. Journal of Royal Statistical Society B 64, 51–62. MR1881844

[5] Chen, Y. H., Chen, H., 2000. A unified approach to regression analysis under double-sampling designs. Journal of Royal Statistical Society B 62, 449–460. MR1772408

[6] Delaigle, A., Hall, P., Qiu, P. H., 2006. Nonparametric methods for solving the berkson errors-in-variables problem. Journal of Royal Statistical Society B 68, 201–220. MR2188982

[7] Du, L., Zou, C., Wang, Z., 2011. Nonparametric regression function estimation for errors-in-variables models with validation data. Statistica Sinica 21 (3), 1093. MR2827516

[8] Engle, R. F., Granger, W. J., Rice, J., Weiss, A., 1986. Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association 80, 310–319.

[9] Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031–1057. MR2189080

[10] Fan, J. Q., Truong, Y. K., 1993. Nonparametric regression with errors in variables. The Annals of Statistics 21, 1900–1925. MR1245773

[11] Härdle, W., Liang, H., 2007. Partially linear models. Springer. MR2376405

[12] Harrison, D., Rubinfeld, D. L., 1978. Hedonic housing prices and the demand for clean air. Journal of environmental economics and management 5 (1), 81–102.

[13] Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. Journal of Royal Statistical Society B 55, 757–796. MR1229881

[14] Hogan, J. W., Lin, X., Herman, B., 2004. Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. Biometrics 60 (4), 854–864. MR2133537

[15] Jiang, J. C., Zhou, H. B., 2007. Additive hazard regression with auxiliary covariates. Biometrika 97, 359–369. MR2380565

[16] Kai, B., Li, R., Zou, H., 2011. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Annals of statistics 39 (1), 305. MR2797848

[17] Li, X. L., You, J. H., Zhou, Y., 2011. Statistical inference for varying-coefficient models with error-prone covariates. Journal of Statistical Computation and Simulation 81 (12), 1755–1771. MR2864171

[18] Liang, H., 2000. Asymptotic normality of parametric part in partially linear models with measurement error in the nonparametric part. Journal of Statistical Planning and Inference 86, 51–62. MR1763180

[19] Mack, Y. P., Silverman, B. W., 1982. Weak and strong uniform consistency of kernel regression estimates. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 61, 405–415. MR0679685

[20] Pepe, M., 1992. Inference using surrogate outcome data and a validation sample. Biometrika 79, 355–365. MR1185137

[21] Ruppert, D., Wand, M. P., Carroll, R. J., 2003. Semiparametric regression. Cambridge university press. MR1998720

[22] Sepanski, J. H., Lee, L. F., 1995. Semiparametric estimation of nonlinear errors-in-variables models with validation study. Journal of Nonparametric Statisitcs 4, 365–394. MR1366782

[23] Shao, F., Li, J., Ma, S., Lee, M.-L. T., 2014. Semiparametric varying-coefficient model for interval censored data with a cured proportion. Statistics in medicine 33 (10), 1700–1712. MR3246689

[24] Tian, R., Xue, L., Liu, C., 2014. Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. Journal of Multivariate Analysis 132, 94–110. MR3266262

[25] Wang, L., Rotnitzky, A., Lin, X., 2010. Nonparametric regression with missing outcomes using weighted kernel estimating equations. Journal of the American Statistical Association 105 (491), 1135–1146. MR2752609

[26] Wang, Q. H., 2006. Nonparametric regression function estimation with surrogate data and validation sampling. Journal of Multivariate Analysis 97, 1142–1161. MR2276152

[27] Wang, S. G., Shi, J. H., Yin, S. J., Wu, M. X., 2004. Introduction to linear models. Beijing: Science Press. MR0876675

[28] Wang, X., 2008. Statistical analysis for semiparametric varying-coefficient partially linear models with incomplete data. Chinese Master's Theses Full-text Database 7 (9).

[29] Wei, C., 2012. Statistical inference in partially linear varying-coefficient models with missing responses at random. Communications in Statistics-Theory and Methods 41 (7), 1284–1298.

[30] Xu, Y., Li, X., Chen, G., 2014. Estimation and inference for varying-coefficient regression models with error-prone covariates. Journal of Systems Science and Complexity 27 (6), 1263–1285. MR3284350

[31] Yatchew, A., 2003. Semiparametric regression for the applied econometrician. Cambridge University Press.

[32] You, J. H., Chen, G. M., 2006. Estimation of a semiparametric varying-coefficient partially linear errors-in-varialbes model. Journal of Multivariate Analysis 97, 324–341. MR2234026

[33] Zhang, W., Lee, X. S., Song, X., 2002. Local polynomial fitting in semivarying coefficient models. Journal of Multivariate Analysis 82, 166–188. MR1918619

[34] Zhou, Y., Liang, H., 2009. Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. The Annals of Statistics 37 (1), 427–458. MR2488358

[35] Zhou, Y., You, J., Wang, X. J., 2009. Strong convergence rates of several estimators in semiparametric varying-coefficient partially linear models. Acta Mathematica Scientia Series B, English Edition, In press. MR2567102

Xiaojing Wang
Department of Statistics
University of Connecticut
215 Glenbrook Rd, U-4120
Storrs, Connecticut 06250
USA
E-mail address: xiaojing.wang@uconn.edu

Yong Zhou
Institute of Statistics and Interdisciplinary Sciences,
Faculty of Economics and Managment,
East China Normal University Shanghai, 200241
P. R. China E-mail address: yzhou@amss.ac.cn

Yang Liu
Department of Statistics
University of Connecticut
215 Glenbrook Rd, U-4120
Storrs, Connecticut 06250
USA
E-mail address: yang.5.liu@uconn.edu