

Article

Estimation in Semi-Varying Coefficient Heteroscedastic Instrumental Variable Models with Missing Responses

Weiwei Zhang ¹, Jingxuan Luo ² and Shengyun Ma ^{1,*}

¹ College of Science, Inner Mongolia Agricultural University, Hohhot 010018, China; zhangweiwei156@imau.edu.cn

² School of Statistics, Beijing Normal University, Beijing 100875, China; luojingxuan@mail.bnu.edu.cn

* Correspondence: msy@imau.edu.cn

Abstract: This paper studies the estimation problem for semi-varying coefficient heteroscedastic instrumental variable models with missing responses. First, we propose the adjusted estimators for unknown parameters and smooth functional coefficients utilizing the ordinary profile least square method and instrumental variable adjustment technique with complete data. Second, we present an adjusted estimator of the stochastic error variance by employing the Nadaraya–Watson kernel estimation technique. Third, we apply the inverse probability-weighted method and instrumental variable adjustment technique to construct the adaptive-weighted adjusted estimators for unknown parameters and smooth functional coefficients. The asymptotic properties of our proposed estimators are established under some regularity conditions. Finally, numerous simulation studies and a real data analysis are conducted to examine the finite sample performance of the proposed estimators.

Keywords: adaptive-weighted adjusted estimation; heteroscedastic; Nadaraya–Watson kernel estimation; semi-varying coefficient instrumental variable models

MSC: 62G05; 62G20



Citation: Zhang, W.; Luo, J.; Ma, S. Estimation in Semi-Varying Coefficient Heteroscedastic Instrumental Variable Models with Missing Responses. *Mathematics* **2023**, *11*, 4853. <https://doi.org/10.3390/math11234853>

Academic Editor: Ionut Florescu

Received: 17 October 2023

Revised: 29 November 2023

Accepted: 30 November 2023

Published: 2 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an important category of statistical regression models, the varying-coefficient partially linear model possesses strong explanatory power and flexibility. It has been used widely in scientific research, such as econometrics, biomedicine and engineering technology. Its general mathematical expression is:

$$Y = \mathbf{X}^T \boldsymbol{\theta}(U) + \mathbf{Z}^T \boldsymbol{\beta} + \varepsilon, \quad (1)$$

where Y is the response variable, and $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Z} \in \mathbb{R}^q$ are covariates. To prevent the “curse of dimensionality” problem, the covariate U is confined to be one-dimensional. $\boldsymbol{\theta}(\cdot) = [\theta_1(\cdot), \dots, \theta_p(\cdot)]^T$ is a $p \times 1$ smooth functional coefficients vector, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_q)^T$ is a $q \times 1$ constant coefficients vector, and ε is the model error, which is independent of $(\mathbf{X}, \mathbf{Z}, U)$. The mean of ε is zero, and the variance is designated as a heteroscedasticity structure, which satisfies $\text{Var}(\varepsilon | \mathbf{X}, \mathbf{Z}, U) = \sigma^2(U)$. As an extension of classical linear regression models, varying-coefficient partially linear models have been investigated by many scholars with the hypothesis of homoscedasticity; please see [1–8]. However, in regression analysis, we may omit some important explanatory variables, and the sample observation data may be measured with errors. For such cases, the model error can be heteroscedastic. In recent years, statisticians have developed many statistical inference methods for model (1) with heteroscedasticity. For instance, Shen et al. [9] introduced a re-weighted estimation procedure for unknown parameters based on the generalized least squares method. Zhao et al. [10] proposed a two-stage iterative estimation method by using the orthogonal projection technique, which can individually estimate unknown

parameters and functional coefficients. Zhao et al. [11] proposed a re-weighted estimation procedure when the covariates contain an additive measurement error. Zhang and Li [12] proposed a weighted estimation and testing method when the covariates suffer from an additive measurement error. Yuan and Zhou [13] introduced an adaptive-weighted estimation method for model (1), which can increase the estimation accuracy of their proposed estimators.

However, the above works have not considered the endogenous problem of covariates. In practice, there may be endogenous explanatory variables in model (1); see [14,15]. In such cases, the above methods will promote the generation of endogenous bias, which leads to the inconsistency of obtained estimators. Therefore, the instrumental variable method will provide an effective way to eliminate endogeneity bias. In the past ten years, the semi-parametric instrumental variable model has been widely studied by many statisticians. For instance, Cai and Xiong [16] suggested a three-stage estimation approach for semi-varying coefficient models with endogenous covariates. Zhao and Li [17] developed an effective variable selection approach for the classical varying coefficient models with endogenous covariates. Zhao and Xue [18] considered the interval estimation for semi-parametric instrumental variable models by using the empirical likelihood method. Yuan et al. [19] proposed an effective method to identify important variables by combining the SCAD penalty method and instrumental variable adjustment technique for semi-varying coefficient models with endogenous covariates. Zhao et al. [20] applied the popular empirical likelihood approach to study the effective interval estimation for semi-varying coefficient instrumental variable models with the orthogonal decomposition technique. For more related research, please refer to the literature [21–24]. In this paper, the covariates U and Z are assumed to be exogenous, but the covariates X is endogenous, and $\zeta \in \mathbb{R}^r$ is an instrumental variable related to X . Similar to [16], the dimension of ζ is designated as greater than or equal to the dimension p of X for identifiability, and X and ζ are specified to satisfy the following parametric model:

$$X = \Psi\zeta + e, \quad (2)$$

where Ψ is an unknown constant matrix with dimension $p \times r$, and the error term e satisfies $\mathbb{E}(e|\zeta, Z, U) = \mathbf{0}$. $\mathbb{E}(e|X, Z, U) \neq 0$ since the covariates are endogenous, which indicates that X is associated with the model error e . Moreover, we further suppose that $\mathbb{E}(e|\zeta, Z, U) = 0$.

In applications, the missing data can occur in many areas, such as market surveys, medical research, opinion polls, and other scientific experiments. When we encounter the missing data, the classical statistical inference methods cannot be used directly. Thus, scholars have developed corresponding methods to solve the problem of missing data. The main methods included the complete sample method, inverse probability weighted technique, and imputation technique. Rubin [25] discussed the complete sample method in detail, but this method will reduce the estimation effectiveness, especially when the observed data are missing at random. Robins et al. [26] suggested an inverse probability-weighted method by assigning the weights to the observed data, which can effectively diminish the deviation caused by the missing data. Wang and Rao [27], and Wang et al. [28] developed the imputation methods for linear and semi-parametric regression models, respectively. Up to now, many scholars have studied the statistical inference for model (1) with missing data, but few scholars considered heteroscedasticity and endogeneity. For instance, Li and Xue [29] constructed an imputation estimator for unknown parameters with a missing response. When the explanatory variables are missing at random, Chen et al. [30] constructed the inverse probability weighted estimation of unknown constant and functional coefficients. For more recent works on missing data research for model (1), the reader can refer to [31–33], among others. In this paper, the response variable is specified to be missing, and other explanatory variables can be fully observed. An indicator variable δ is introduced to describe the missing mechanism. If Y is obtainable, we denote $\delta = 1$, and

otherwise $\delta = 0$. Furthermore, we suppose that the missing data mechanism is randomly missing, which is expressed as:

$$P(\delta = 1|U, X, Z, Y) = P(\delta = 1|U, X, Z) = \pi(U, X, Z), \tag{3}$$

where $\pi(\cdot)$ is referred to as the propensity score.

Although many scholars have discussed the statistical inference procedures for various semi-parametric models with endogenous covariates, the existing works have not considered the heteroscedasticity and missing data. Therefore, we consider the estimation problem for models (1) and (2) with heteroscedasticity and a missing response. The adaptive-weighted adjusted estimators of unknown parameters and functional coefficients are proposed using the profile least squares method, instrumental variables adjustment technique, Nadaraya–Watson kernel estimation, inverse probability-weighted method, and weighted least squares method, and we also establish the asymptotic properties of the proposed estimators.

The rest of the paper is organized as follows. In Section 2, we introduced an adaptive-weighted adjusted estimation method to obtain the estimators for unknown parameters and functional coefficients, and the corresponding asymptotic properties are established. In Section 3, numerous simulation studies are conducted to demonstrate the effectiveness and feasibility of the proposed estimators. A real data analysis is performed as well in Section 4. Section 5 summarizes the research results of this paper with some conclusions. The technical proofs are presented in Appendix A.

2. Estimation Methods and Main Results

2.1. Adjusted Profile Least Squares Estimation

In this subsection, we apply the local linear smoothing technique and instrumental variable adjustment technique to estimate unknown parameters β and smooth functional coefficients $\theta(\cdot)$. Assume that $\{Y_i, X_i, U_i, Z_i, \zeta_i, \delta_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) samples, which come from the semi-varying coefficient heteroscedastic instrumental variable models (1)–(3); then, we have:

$$\begin{cases} \delta_i Y_i = \delta_i X_i^\top \theta(U_i) + \delta_i Z_i^\top \beta + \delta_i \varepsilon_i, \\ X_i = \Psi \zeta_i + e_i, \quad i = 1, 2, \dots, n. \end{cases} \tag{4}$$

For any u in the small neighborhood of u_0 , each functional coefficients $\theta_j(u)$ ($j = 1, 2, \dots, p$) can be expanded by the Taylor expansion as follows:

$$\theta_j(u) \approx \theta_j(u_0) + \theta_j'(u_0)(u - u_0), \quad j = 1, 2, \dots, p.$$

If β is given, then the estimator of $\theta_j(u_0)$ is given by minimizing the following weighted least squares objective function:

$$\sum_{i=1}^n \left\{ Y_i - Z_i^\top \beta - \sum_{j=1}^p \left[\theta_j(u_0) + \theta_j'(u_0)(U_i - u_0) \right] X_{ij} \right\}^2 K_{h_1}(U_i - u_0) \delta_i, \tag{5}$$

where $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$, $K(\cdot)$ is a kernel function, which is chosen as a symmetric probability density function, and h_1 is a bandwidth. For ease of presentation, we denote:

$$Y = (Y_1, Y_2, \dots, Y_n)^\top, \quad X = (X_1, X_2, \dots, X_n)^\top, \quad Z = (Z_1, Z_2, \dots, Z_n)^\top,$$

$$\Delta_0 = \text{diag}(\delta_1, \delta_2, \dots, \delta_n), \quad M = [X_1^\top \theta(U_1), \dots, X_n^\top \theta(U_n)]^\top,$$

$$w_{h_1}(u_0) = \text{diag}[K_{h_1}(U_1 - u_0), K_{h_1}(U_2 - u_0), \dots, K_{h_1}(U_n - u_0)],$$

$$\mathbf{X}_{h_1}(u_0) = \begin{pmatrix} \mathbf{X}_1^\top & h_1^{-1}(U_1 - u_0)\mathbf{X}_1^\top \\ \vdots & \vdots \\ \mathbf{X}_n^\top & h_1^{-1}(U_n - u_0)\mathbf{X}_n^\top \end{pmatrix}.$$

Thus, the first formula of model (4) can be transformed into:

$$\Delta_0 \mathbf{Y} - \Delta_0 \mathbf{Z} \boldsymbol{\beta} = \Delta_0 \mathbf{M} + \Delta_0 \boldsymbol{\varepsilon}. \tag{6}$$

By minimizing the weighted least squares objective (5), the estimators of functional coefficients $\theta(u_0)$ are given by:

$$\tilde{\theta}(u_0, \boldsymbol{\beta}) = (\mathbf{I}_p, \mathbf{0}_p) \{ \mathbf{X}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 \mathbf{X}_{h_1}(u_0) \}^{-1} \mathbf{X}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 (\mathbf{Y} - \mathbf{Z} \boldsymbol{\beta}), \tag{7}$$

where \mathbf{I}_p and $\mathbf{0}_p$ denote the identity matrix and zero matrix with dimension $p \times p$. It is noteworthy that the explanatory variable X in this paper is an endogenous variable, which indicates that $\mathbb{E}(\varepsilon|X) \neq 0$. Thus, the estimators of the functional coefficients in (7) are inconsistent. Then, we make a correction for $\tilde{\theta}(u_0, \boldsymbol{\beta})$ by applying the available instrumental variables ζ . For model (2), we can easily obtain:

$$\mathbb{E}(\mathbf{X} \boldsymbol{\zeta}^\top) = \boldsymbol{\Psi} \mathbb{E}(\boldsymbol{\zeta} \boldsymbol{\zeta}^\top).$$

Therefore, a usual moment estimator of the unknown constant matrix $\boldsymbol{\Psi}$ is given by:

$$\hat{\boldsymbol{\Psi}} = \left(\sum_{i=1}^n \mathbf{X}_i \boldsymbol{\zeta}_i^\top \right) \left(\sum_{i=1}^n \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top \right)^{-1}.$$

Let $\hat{\mathbf{X}}_i = \hat{\boldsymbol{\Psi}} \boldsymbol{\zeta}_i$. Invoking (7), the adjusted estimators of functional coefficients $\theta(u)$ are given by:

$$\hat{\theta}(u_0, \boldsymbol{\beta}) = (\mathbf{I}_p, \mathbf{0}_p) \{ \hat{\mathbf{X}}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 \hat{\mathbf{X}}_{h_1}(u_0) \}^{-1} \hat{\mathbf{X}}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 (\mathbf{Y} - \mathbf{Z} \boldsymbol{\beta}), \tag{8}$$

where $\hat{\mathbf{X}}_{h_1}(u_0)$ has the same structure as $\mathbf{X}_{h_1}(u_0)$, but replacing the variable X_i with \hat{X}_i . Then, invoking (8), the estimator of \mathbf{M} is given by:

$$\hat{\mathbf{M}}(\boldsymbol{\beta}) = (\hat{\mathbf{X}}_1^\top \hat{\theta}(U_1, \boldsymbol{\beta}), \dots, \hat{\mathbf{X}}_n^\top \hat{\theta}(U_n, \boldsymbol{\beta}))^\top \triangleq \mathbf{S}(\mathbf{Y} - \mathbf{Z} \boldsymbol{\beta}) \tag{9}$$

and:

$$\mathbf{S} = \begin{pmatrix} (\hat{\mathbf{X}}_1^\top, \mathbf{0}) \{ \hat{\mathbf{X}}_{h_1}(U_1) \mathbf{w}_{h_1}(U_1) \Delta_0 \hat{\mathbf{X}}_{h_1}(U_1) \}^{-1} \hat{\mathbf{X}}_{h_1}^\top(U_1) \mathbf{w}_{h_1}^\top(U_1) \Delta_0 \\ \vdots \\ (\hat{\mathbf{X}}_n^\top, \mathbf{0}) \{ \hat{\mathbf{X}}_{h_1}(U_n) \mathbf{w}_{h_1}(U_n) \Delta_0 \hat{\mathbf{X}}_{h_1}(U_n) \}^{-1} \hat{\mathbf{X}}_{h_1}^\top(U_n) \mathbf{w}_{h_1}^\top(U_n) \Delta_0 \end{pmatrix},$$

where $\mathbf{0}$ denotes a zero vector with dimensions $1 \times p$. Replacing \mathbf{M} with $\hat{\mathbf{M}}(\boldsymbol{\beta})$ in (6), we obtain:

$$\Delta_0 (\mathbf{I} - \mathbf{S}) \mathbf{Y} = \Delta_0 (\mathbf{I} - \mathbf{S}) \mathbf{Z} \boldsymbol{\beta} + \Delta_0 \boldsymbol{\varepsilon}. \tag{10}$$

For the model (10), a least squares approach is implemented, and then the adjusted estimators of unknown parameters $\boldsymbol{\beta}$ are given by:

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{Z}}^\top \Delta_0 \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}^\top \Delta_0 \tilde{\mathbf{Y}}, \tag{11}$$

where $\tilde{\mathbf{Y}} = (\mathbf{I} - \mathbf{S}) \mathbf{Y}$, $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{S}) \mathbf{Z}$. Combining (8) and (11), the adjusted estimators of functional coefficients $\theta(u)$ at u_0 are given by:

$$\hat{\theta}(u_0, \hat{\boldsymbol{\beta}}) = (\mathbf{I}_p, \mathbf{0}_p) \{ \hat{\mathbf{X}}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 \hat{\mathbf{X}}_{h_1}(u_0) \}^{-1} \hat{\mathbf{X}}_{h_1}^\top(u_0) \mathbf{w}_{h_1}(u_0) \Delta_0 (\mathbf{Y} - \mathbf{Z} \hat{\boldsymbol{\beta}}). \tag{12}$$

2.2. Adaptive-Weighted Adjusted Profile Least Squares Estimation

In this subsection, we develop an adaptive-weighted adjusted estimation method for unknown parameters and functional coefficients based on weighted least squares estimation. First of all, according to the estimators of model residuals, we suggest an adjusted Nadaraya–Watson kernel estimation method for the variance function. By (11) and (12), we can estimate the residual error by:

$$\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n)^\top = \Delta_0(\mathbf{Y} - \widehat{\mathbf{M}}(\hat{\boldsymbol{\beta}}) - \mathbf{Z}\hat{\boldsymbol{\beta}}). \tag{13}$$

Note that $\text{Var}(\varepsilon_i | \mathbf{X}_i, \mathbf{Z}_i, U_i) = \sigma^2(U_i)$, and by using the Nadaraya–Watson kernel estimation method, an adjusted estimator of variance function $\sigma^2(u_0)$ is given by:

$$\hat{\sigma}^2(u_0) = \frac{\sum_{i=1}^n \delta_i \hat{\varepsilon}_i^2 K_{\tilde{h}}(U_i - u_0)}{\sum_{i=1}^n \delta_i K_{\tilde{h}}(U_i - u_0)}, \tag{14}$$

where $K_{\tilde{h}}(\cdot)$ and $K_{h_1}(\cdot)$ have the same structure, except that the bandwidth \tilde{h} is replaced by h_1 . Furthermore, replacing u_0 by U_i ($i = 1, \dots, n$), we can obtain:

$$\hat{\sigma}^2(U_i) = \frac{\sum_{k=1}^n \delta_k \hat{\varepsilon}_k^2 K_{\tilde{h}}(U_k - U_i)}{\sum_{k=1}^n \delta_k K_{\tilde{h}}(U_k - U_i)}, \quad i = 1, \dots, n. \tag{15}$$

The estimator $\hat{\sigma}^2(u_0)$ of the variance function $\sigma^2(u_0)$ is consistent. The proof of this property is similar to that of Theorem 1 in [9]; thus, we omit the details.

Then, we consider how to deal with missing data. In general, the selection probability function $\pi(U_i, \mathbf{X}_i, \mathbf{Z}_i)$ defined in (3) is usually unknown. We can employ nonparametric estimation methods to estimate it, such as kernel estimation and local polynomial estimation, but these methods may cause the curse of dimensionality. Therefore, similar to the method in [31], we suppose that the selection probability function satisfies:

$$\pi(U_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{w}) = \frac{\exp(w_0 + w_1 U_i + \mathbf{w}_2^\top \mathbf{X}_i + \mathbf{w}_3^\top \mathbf{Z}_i)}{1 + \exp(w_0 + w_1 U_i + \mathbf{w}_2^\top \mathbf{X}_i + \mathbf{w}_3^\top \mathbf{Z}_i)},$$

where $\mathbf{w} = (w_0, w_1, \mathbf{w}_2^\top, \mathbf{w}_3^\top)^\top$ is the unknown parameter vector, and the estimator $\hat{\mathbf{w}}$ of \mathbf{w} can be attained by the quasi-likelihood estimation method. We also assume $\mathbf{V}_i = (U_i, \mathbf{X}_i^\top, \mathbf{Z}_i^\top)^\top$, and then the estimator $\pi(U_i, \mathbf{X}_i, \mathbf{Z}_i, \hat{\mathbf{w}}) = \pi(\mathbf{V}_i, \hat{\mathbf{w}})$.

Based on the estimator of variance function $\hat{\sigma}^2(U_i)$ and selection probability function $\pi(\mathbf{V}_i, \hat{\mathbf{w}})$, the adaptive-weighted adjusted estimators for functional coefficients are given by minimizing:

$$\sum_{i=1}^n \left\{ Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \sum_{j=1}^p [\theta_j(u_0) + \theta'_j(u_0)(U_i - u_0)] X_{ij} \right\}^2 \hat{\sigma}^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \hat{\mathbf{w}})} K_{h_2}(U_i - u_0), \tag{16}$$

where $K_{h_2}(\cdot) = K(\cdot/h_2)/h_2$ with bandwidth h_2 .

Minimizing the objective Function (16), the adaptive-weighted estimators of functional coefficients can be expressed as:

$$\tilde{\boldsymbol{\theta}}_w(u_0, \boldsymbol{\beta}) = (\mathbf{I}_p, \mathbf{0}_p) \{ \mathbf{X}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Delta}} \mathbf{X}_{h_2}(u_0) \}^{-1} \mathbf{X}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Delta}} (\mathbf{Y} - \mathbf{Z}\boldsymbol{\beta}),$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}[\hat{\sigma}^2(U_1), \hat{\sigma}^2(U_2), \dots, \hat{\sigma}^2(U_n)]$, $\hat{\boldsymbol{\Delta}} = [\delta_1/\pi(\mathbf{V}_1, \hat{\mathbf{w}}), \dots, \delta_n/\pi(\mathbf{V}_n, \hat{\mathbf{w}})]$. Since the explanatory variable \mathbf{X} is endogenous, the instrumental variable adjustment technique

is used to make a correction for $\tilde{\theta}(u_0, \beta)$. Then, the adaptive-weighted adjusted estimators of functional coefficients $\theta(u_0)$ are given by:

$$\hat{\theta}_w(u_0, \beta) = (\mathbf{I}_p, \mathbf{0}_p) \{ \hat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \hat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\Sigma}^{-1} \hat{\Delta} (\mathbf{Y} - \mathbf{Z}\beta), \tag{17}$$

where $\hat{\mathbf{X}}_{h_2}(u_0)$ and $\hat{\mathbf{X}}_{h_1}(u_0)$, $\mathbf{w}_{h_2}(u_0)$ and $\mathbf{w}_{h_1}(u_0)$ have the same form except that h_1 is replaced by h_2 , respectively. Then, the estimator of \mathbf{M} is given by:

$$\hat{\mathbf{M}}_w = (\hat{\mathbf{X}}_1^\top \hat{\theta}_w(U_1, \beta), \dots, \hat{\mathbf{X}}_n^\top \hat{\theta}_w(U_n, \beta))^\top \triangleq \hat{\mathbf{S}} (\mathbf{Y} - \mathbf{Z}\beta), \tag{18}$$

where:

$$\hat{\mathbf{S}} = \begin{pmatrix} (\hat{\mathbf{X}}_1^\top, 0) \{ \hat{\mathbf{X}}_{h_2}^\top(U_1) \mathbf{w}_{h_2}(U_1) \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{X}}_{h_2}(U_1) \}^{-1} \hat{\mathbf{X}}_{h_2}^\top(U_1) \mathbf{w}_{h_2}^\top(U_1) \hat{\Sigma}^{-1} \hat{\Delta} \\ \vdots \\ (\hat{\mathbf{X}}_n^\top, 0) \{ \hat{\mathbf{X}}_{h_2}^\top(U_n) \mathbf{w}_{h_2}(U_n) \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{X}}_{h_2}(U_n) \}^{-1} \hat{\mathbf{X}}_{h_2}^\top(U_n) \mathbf{w}_{h_2}^\top(U_n) \hat{\Sigma}^{-1} \hat{\Delta} \end{pmatrix}.$$

Substituting $\hat{\mathbf{M}}_w$ into (1), we have:

$$(\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Y} = (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Z}\beta + \varepsilon. \tag{19}$$

In order to eliminate the impact of heteroscedasticity, left-multiplying the matrix $\hat{\Sigma}^{-1/2}$ to (19), we obtain:

$$\hat{\Sigma}^{-1/2} (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Y} = \hat{\Sigma}^{-1/2} (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Z}\beta + \hat{\Sigma}^{-1/2} \varepsilon, \tag{20}$$

where $\hat{\Sigma}^{-1/2} = \text{diag}[\hat{\sigma}^{-1}(U_1), \hat{\sigma}^{-1}(U_2), \dots, \hat{\sigma}^{-1}(U_n)]$.

By employing the inverse probability-weighted method and combining the idea of weighted least squares, we can derive the estimators of unknown parameters β by minimizing

$$Q(\beta) = [(\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Y} - (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Z}\beta]^\top \hat{\Sigma}^{-1} \hat{\Delta} [(\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Y} - (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Z}\beta].$$

Solving the minimum problem with respect to β , the proposed adaptive-weighted adjusted estimator of β is given by:

$$\hat{\beta}_w = (\hat{\mathbf{Z}}^\top \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{Y}}, \tag{21}$$

where $\hat{\mathbf{Y}} = (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Y}$, $\hat{\mathbf{Z}} = (\mathbf{I} - \hat{\mathbf{S}}) \mathbf{Z}$. Moreover, substituting $\hat{\beta}_w$ into (17), we give the proposed adaptive-weighted adjusted estimators of functional coefficients $\theta(u_0)$ as follows:

$$\hat{\theta}_w(u_0, \hat{\beta}_w) = (\mathbf{I}_p, \mathbf{0}_p) \{ \hat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\Sigma}^{-1} \hat{\Delta} \hat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \hat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \hat{\Sigma}^{-1} \hat{\Delta} (\mathbf{Y} - \mathbf{Z}\hat{\beta}_w). \tag{22}$$

2.3. Asymptotic Properties

In this subsection, we establish the asymptotic properties of the proposed estimators. First, some regularity conditions are needed. These conditions are mild, and similar conditions can be found in [9–11], and other varying-coefficient partially linear heteroscedasticity literature.

- (C1) The random variable U has a bounded support \mathcal{U} and its density function $f(u)$ is Lipschitz-continuous and has a second-order continuous derivative. Moreover, $f(u)$ is bounded away from zero.
- (C2) The kernel $K(\cdot)$ is a symmetric probability density function with a compact support and is Lipschitz-continuous.
- (C3) For each $u \in \mathcal{U}$, the matrix $\mathbf{\Pi}(u) = \mathbb{E}(\zeta_1 \zeta_1^\top | U = u)$ is invertible, and the matrices $\mathbf{\Pi}(u)$, $\mathbf{\Pi}^{-1}(u)$, and $\mathbf{\Xi}(u) = \mathbb{E}(\zeta_1 \mathbf{Z}_1^\top | U = u)$ are Lipschitz-continuous.
- (C4) For each $u \in \mathcal{U}$, the functional coefficients $\{\theta_j(u), j = 1, 2, \dots, p\}$ are Lipschitz-continuous and have continuous second derivatives.
- (C5) There is a $s > 2$, such that $\mathbb{E}\|\zeta_1\|^{2s} < \infty, \mathbb{E}\|\mathbf{Z}_1\|^{2s} < \infty$.

- (C6) There is a $\delta < 1 - s^{-1}$, such that $\lim_{n \rightarrow \infty} n^{2\delta-1}h_i = \infty, i = 1, 2$.
- (C7) The variance function $\sigma^2(\cdot)$ has a continuous second derivative and is uniformly bounded on the domain.
- (C8) For the bandwidth $h_i (i = 1, 2), nh_i^2 \rightarrow \infty, nh_i^8 \rightarrow 0$ and $[\log(1/h_i)]^2/(nh_i^2) \rightarrow 0$ as $n \rightarrow \infty$. In addition, h_i and \tilde{h} satisfy $O(c_{n_i})O(\tilde{c}_n) = o(n^{-1/2})$, where $c_{n_i} = h_i^2 + [\log(1/h_i)/(nh_i)]^{1/2}, i = 1, 2, \tilde{c}_n = \tilde{h}^2 + [\log(1/\tilde{h})/(n\tilde{h})]^{1/2}$.
- (C9) On the base of $(\mathbf{X}_i, \mathbf{Z}_i, U_i), \pi(\cdot)$ has a second-order continuous derivative. Moreover, $\pi(\cdot)$ is bounded away from zero.

Conditions (C1)–(C6) are quite generally required in the semi-varying coefficient model. Conditions (C7) and (C8) are mainly to obtain the consistent estimator of the variance function, which can be found in [9]. Condition (C9) provides a guarantee for the inverse probability weighted technique.

Theorem 1. Suppose that regularity conditions (C1)–(C9) hold; for each $u_0 \in \mathcal{U}$, we have:

$$\sup_{u_0 \in \mathcal{U}} |\hat{\sigma}^2(u_0) - \sigma^2(u_0)| = o_p(\tilde{c}_n).$$

Theorem 2. Suppose that regularity conditions (C1)–(C9) hold; then, the proposed adaptive-weighted adjusted estimator of β satisfies:

$$\sqrt{n}(\hat{\beta}_w - \beta) \xrightarrow{D} N(\mathbf{0}, \Lambda_1^{-1} \Lambda_2 \Lambda_1^{-1}), \quad n \rightarrow \infty,$$

where $\Lambda_1 = \mathbb{E}\{\sigma^{-2}(U_1)[\mathbf{Z}_1 - \mathbb{E}^\top(U_1)\Psi^\top(\Psi\Pi(U_1)\Psi^\top)^{-1}\Psi\zeta_1] \otimes^2\}, \Lambda_2 = \mathbb{E}\{\pi(V_1, w)^{-1} \times \sigma^{-4}(U_1)(e_1^\top \theta(U_1) + \varepsilon_1)^2[\mathbf{Z}_1 - \mathbb{E}^\top(U_1)\Psi^\top(\Psi\Pi(U_1)\Psi^\top)^{-1}\Psi\zeta_1] \otimes^2\}$, and $\mathbf{H} \otimes^2 = \mathbf{H}\mathbf{H}^\top$.

Theorem 3. Suppose that regularity conditions (C1)–(C9) hold; then, the proposed adaptive-weighted adjusted estimator of $\theta(u_0)$ satisfies:

$$\sqrt{nh_2} \left[\hat{\theta}_w(u_0, \hat{\beta}_w) - \theta(u_0) - \frac{1}{2}h_2^2\mu_2\theta''(u_0) \right] \xrightarrow{D} N(\mathbf{0}, \Lambda(u_0)), \quad n \rightarrow \infty,$$

where $\Lambda(u_0) = v_0 f^{-1}(u_0) \mathbb{E}\{\pi(V_1, w)^{-1}(e_1^\top \theta(U_1) + \varepsilon_1)^2\}(\Psi\Pi(u_0)\Psi^\top)^{-1}$,

$$\mu_2 = \int u^2 K_{h_2}(u) du, \quad v_0 = \int K_{h_2}^2(u) du.$$

Theorems 2 and 3 give the asymptotic distribution of our proposed adaptive-weighted adjusted estimators. These results can be utilized to conduct statistical inference for unknown parameters and functional coefficients. Additionally, the above theorems expand the application scale of semi-varying coefficient models to satisfy the modeling requirements of applications. By assuming that the missing responses and endogenous covariates are nonexistent, the asymptotic variance of our proposed estimators possess the same structure as that of the estimators in [13]. On the other hand, when the missing response and heteroscedasticity are not considered, the asymptotic variance of our proposed estimators is the same as that of the estimators in [16].

3. Simulation Studies

In this section, we carry out some simulations to evaluate the finite sample performance of the proposed adaptive-weighted adjusted estimation method. We generate the data from the semi-varying coefficient heteroscedastic instrumental variables model:

$$Y = Z_1\beta_1 + Z_2\beta_2 + X\theta(U) + \varepsilon,$$

where the explanatory variables Z_1 and Z_2 are both independently drawn from $N(2, 1)$, the univariate U is independently drawn from $U(0, 1)$, the explanatory variable X is an endogenous variable generated from the model $X = \zeta + k\varepsilon$, where ζ in an instrumental variable generated from normal distribution $N(1, 1)$, and k is taken as 0.2 and 0.4 to represent different levels of endogeneity. We set the parameters $\beta_1 = 1.5$, $\beta_2 = 2$, $\theta(U) = \sin(2\pi U)$. The model error $\varepsilon \sim N(0, \sigma^2(U))$ with $\sigma^2(U) = 0.25 + [c \sin(2\pi U)]^2$ for $c = 2, 4$, respectively. The Gaussian kernel $K(x) = 1/\sqrt{2\pi} \exp(-x^2/2)$ is adopted. The leave one out cross-validation (LOOCV) method is applied to choose h_1 , which is derived by minimizing

$$CV(h_1) = \frac{1}{n} \sum_{i=1}^n \delta_i [Y_i - X_i \hat{\theta}_{[-i]}(U_i) - \mathbf{Z}_i^\top \hat{\beta}_{[-i]}],$$

where $\hat{\beta}_{[-i]}$ and $\hat{\theta}_{[-i]}(\cdot)$ are the adjusted profile least-squares estimators, which are given in (11) and (12), respectively. We choose bandwidths \tilde{h} and h_2 by a similar method. To compare the performance of the proposed adaptive-weighted adjusted estimators under different missing probabilities, two selection probability functions are chosen as follows:

$$\begin{aligned} \pi_1(x, u, z) &= P(\delta = 1 | X = x, U = u, \mathbf{Z} = (z_1, z_2)) \\ &= [1 + \exp(-1 + 1.1x - z_1 - 0.4z_2 - 0.7u)]^{-1}, \\ \pi_2(x, u, z) &= P(\delta = 1 | X = x, U = u, \mathbf{Z} = (z_1, z_2)) \\ &= [1 + \exp(-1 + 1.5x - 0.8z_1 - 0.4z_2 - 0.5u)]^{-1}. \end{aligned}$$

The corresponding average response rates are about 0.9 and 0.8 when $c = 2$ and $k = 0.2$. To show the performance of our proposed adaptive-weighted adjusted profile least square estimation based on the instrumental variable adjustment technique (represented as IAWPLS), we contrast it with the two approaches below: (1) the naive adaptive-weighted profile least square estimation, denoted by NAWPLS; (2) the instrumental variable weighted profile least square estimation, denoted by IWPLS. The former omits the endogeneity of the explanatory covariate, and it is derived by combining the inverse probability-weighted method and adaptive-weighted profile least squares method in [13]. The latter ignores the heteroscedasticity of the model error and combines the inverse probability-weighted method and instrumental variable adjustment method in [16]. We set the sample size as 50, 100, 200, and 300. The results are based on 500 replications for each case. For parametric components, we use the following measures to compare the performance of different methods: (1) Mean: the average of the estimated value; (2) MSE: mean square errors for the corresponding estimators. The results are shown in Tables 1 and 2 for $\pi = \pi_1, \pi_2, c = 2, 4, k = 0.2, 0.4$ and $n = 50, 100, 200, 300$, respectively.

According to Tables 1 and 2, we have the following results.

- (1) The IAWPLS and IWPLS estimators are asymptotically unbiased, but the NAWPLS estimators are biased. For fixed π, c, k , with the increase of n , the MSEs of all three estimators decrease.
- (2) For fixed π, c, k , when the sample size $n = 50$, the MSEs of our proposed IAWPLS estimators are slightly larger than those of NAWPLS in some cases, but obviously smaller than those of the NAWPLS and IWPLS estimators when the sample size is greater than 100.
- (3) For fixed c, k, n , with the increase in the missing probability, the MSEs of all three estimators increase.
- (4) For fixed π, k, n , with the increase in c , the MSEs of all three estimators increase.
- (5) For fixed π, c, n , with the increase in k , the MSEs of all three estimators increase.

Subsequently, we further consider the behavior of the adaptive-weighted adjusted estimation method for the variance functions and functional coefficients. The corresponding estimated values are computed at $n = 200$ equally spaced values $U_i = i/n \in [0, 1]$, and the mean value of 500 simulations at every point U_i is taken for the ultimate estimated values.

Due to the similarity of the estimated curves for different sample sizes and missing probabilities, we only plot the estimated curves of the variance functions and functional coefficients when $c = 2, 4, k = 0.2, 0.4, n = 200$, and $\pi = \pi_1$. The estimated curves are shown in Figures 1 and 2. To demonstrate the effectiveness of the proposed estimation method for the variance function, two methods are taken for contrast: (1) the adjusted Nadaraya–Watson kernel estimation method based on instrumental variable adjustment techniques; (2) the naive Nadaraya–Watson kernel estimation method, which ignores the endogeneity of explanatory variables and uses the standard Nadaraya–Watson kernel estimation.

Table 1. Sample means and MSEs for β_1 based on the IAWPLS, NAWPLS, and IWPLS methods.

π	c	k	n	IAWPLS		NAWPLS		IWPLS	
				Mean	MSE	Mean	MSE	Mean	MSE
π_1	2	0.2	50	1.5107	0.0355	1.4335	0.0329	1.5153	0.0607
			100	1.5070	0.0128	1.4379	0.0141	1.5124	0.0261
			200	1.5062	0.0056	1.4366	0.0089	1.5081	0.0121
		0.4	300	1.5019	0.0036	1.4424	0.0064	1.5023	0.0094
			50	1.5182	0.0429	1.3808	0.0377	1.5289	0.0778
			100	1.5129	0.0164	1.3898	0.0232	1.5192	0.0383
	4	0.2	200	1.5091	0.0074	1.3920	0.0165	1.5115	0.0203
			300	1.5030	0.0046	1.3954	0.0140	1.5041	0.0113
			50	1.5231	0.1056	1.2857	0.2360	1.5226	0.2382
		0.4	100	1.5058	0.0367	1.3025	0.0716	1.5214	0.1063
			200	1.4963	0.0143	1.3116	0.0510	1.5146	0.0481
			300	1.5019	0.0094	1.3172	0.0430	1.5062	0.0323
π_2	2	0.2	50	1.5274	0.1598	1.2252	0.1324	1.5264	0.3682
			100	1.5110	0.0449	1.2361	0.0917	1.5255	0.1377
			200	1.5088	0.0182	1.2468	0.0749	1.5227	0.0813
		0.4	300	1.5053	0.0105	1.2517	0.0686	1.5204	0.0510
			50	1.5187	0.0352	1.4274	0.0364	1.5245	0.0630
			100	1.5050	0.0169	1.4289	0.0186	1.5112	0.0346
	4	0.2	200	1.4974	0.0072	1.4316	0.0109	1.5009	0.0151
			300	1.5027	0.0049	1.4386	0.0077	1.5028	0.0112
			50	1.4929	0.0535	1.3581	0.0502	1.4994	0.1011
		0.4	100	1.5083	0.0202	1.3785	0.0264	1.5187	0.0438
			200	1.5044	0.0079	1.3871	0.0184	1.5140	0.0221
			300	1.5032	0.0058	1.3882	0.0163	1.5093	0.0153
4	0.2	50	1.5438	0.1646	1.2972	0.1531	1.5601	0.3148	
		100	1.5180	0.0417	1.3103	0.0723	1.5446	0.1141	
		200	1.5159	0.0192	1.3054	0.0538	1.5262	0.0625	
	0.4	300	1.5062	0.0124	1.3210	0.0442	1.5165	0.0491	
		50	1.5455	0.2016	1.2323	0.1424	1.5895	0.4071	
		100	1.5397	0.0555	1.2566	0.0871	1.5802	0.1955	
4	0.4	200	1.5226	0.0271	1.2469	0.0784	1.5559	0.1111	
		300	1.5223	0.0181	1.2482	0.0725	1.5511	0.1043	

Figure 1 shows that the proposed adjusted Nadaraya–Watson kernel estimators are asymptotically unbiased, but the naive Nadaraya–Watson kernel estimators are biased, and the deviation increases with the increase of c or k . Note that the performance of our proposed adjusted Nadaraya–Watson kernel estimators may be affected by larger c and k . From Figure 2, we find that the estimated curves obtained by the IAWPLS and IWPLS methods both approach the true curves, but the estimated curves obtained by the NAWPLS method are biased, and the deviation increases with the increase of c or k .

Table 2. Sample means and MSEs for β_2 based on the IAWPLS, NAWPLS, and IWPLS methods.

π	c	k	n	IAWPLS		NAWPLS		IWPLS	
				Mean	MSE	Mean	MSE	Mean	MSE
π_1	2	0.2	50	1.9900	0.0350	1.9224	0.0344	1.9810	0.0573
			100	1.9911	0.0134	1.9292	0.0157	1.9882	0.0288
			200	1.9962	0.0058	1.9374	0.0092	1.9927	0.0121
		0.4	300	2.0009	0.0039	1.9423	0.0067	2.0015	0.0093
			50	1.9875	0.0475	1.8815	0.0384	1.9836	0.0820
			100	1.9898	0.0171	1.8880	0.0234	1.9885	0.0384
	4	0.2	200	1.9960	0.0070	1.8877	0.0171	1.9971	0.0200
			300	2.0014	0.0045	1.8896	0.0152	2.0079	0.0117
			50	2.0024	0.1140	1.7800	0.1347	2.0073	0.2454
		0.4	100	2.0017	0.0354	1.8183	0.0623	1.9927	0.1010
			200	2.0148	0.0159	1.8275	0.0447	2.0154	0.0525
			300	2.0065	0.0089	1.8270	0.0392	2.0043	0.0336
π_2	2	0.2	50	1.9688	0.1469	1.7602	0.1441	1.9604	0.3306
			100	1.9893	0.0488	1.7568	0.0829	1.9933	0.1606
			200	2.0001	0.0195	1.7594	0.0702	2.0082	0.0887
		0.4	300	1.9996	0.0118	1.7596	0.0645	1.9995	0.0554
			50	1.9764	0.0349	1.9124	0.0362	1.9805	0.0602
			100	1.9994	0.0181	1.9349	0.0189	1.9985	0.0339
	4	0.2	200	2.0021	0.0073	1.9414	0.0098	1.9999	0.0155
			300	1.9979	0.0047	1.9396	0.0075	2.0023	0.0105
			50	2.0004	0.0520	1.8870	0.0409	2.0089	0.0970
		0.4	100	1.9956	0.0177	1.8896	0.0242	2.0035	0.0406
			200	1.9970	0.0087	1.8872	0.0183	2.0002	0.0230
			300	2.0024	0.0056	1.8919	0.0157	2.0018	0.0164
4	0.2	50	1.9750	0.1558	1.7766	0.1635	1.9782	0.3179	
		100	1.9977	0.0405	1.8048	0.0739	1.9958	0.1295	
		200	1.9880	0.0178	1.8085	0.0529	1.9887	0.0647	
	0.4	300	1.9898	0.0120	1.8103	0.0472	1.9911	0.0497	
		50	1.9523	0.2195	1.7392	0.1803	1.9487	0.4049	
		100	1.9748	0.0688	1.7389	0.0978	1.9570	0.2404	
300	200	1.9781	0.0241	1.7510	0.0763	1.9552	0.1095		
	300	1.9933	0.0171	1.7532	0.0695	1.9933	0.0171		

Owing to the estimated curves obtained by the IAWPLS and IWPLS methods being close to each other, we further utilize the root mean squared error (RMSE) to evaluate the effectiveness for functional coefficients:

$$RMSE = \left\{ \frac{1}{N} \sum_{k=1}^N \|\hat{\theta}(U_k) - \theta(U_k)\|^2 \right\}^{1/2},$$

where U_k ($k = 1, 2, \dots, N$) are the selected split points defined on the bounded support \mathcal{U} . In this case, we set $N = 100$ and assume that U_k takes equal intervals on the interval $[0, 1]$. We calculate RMSE of the functional coefficients estimators. Results are presented in Table 3 for $\pi = \pi_1, \pi_2, c = 2, 4, k = 0.2, 0.4$ and $n = 50, 100, 200, 300$.

According to Table 3, we find that the proposed IAWPLS estimation method for functional coefficients has smaller RMSEs than those of NAWPLS and IWPLS methods for given π, k, c, n . The RMSEs of all three estimators decrease with the increase of n . Nevertheless, the RMSEs increase with the increase of c, k or missing probability.

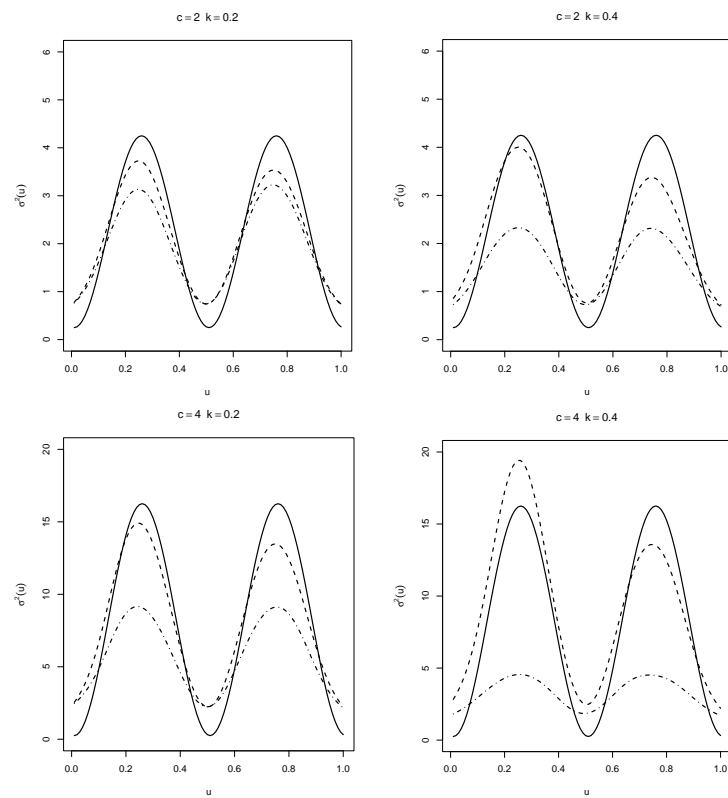


Figure 1. Plot of the variance function estimates through the use of the adjusted Nadaraya–Watson kernel estimation method (denoted using dashed lines) and the naive Nadaraya–Watson kernel estimation method (denoted using dot-dashed lines); the solid lines represent the true curves.

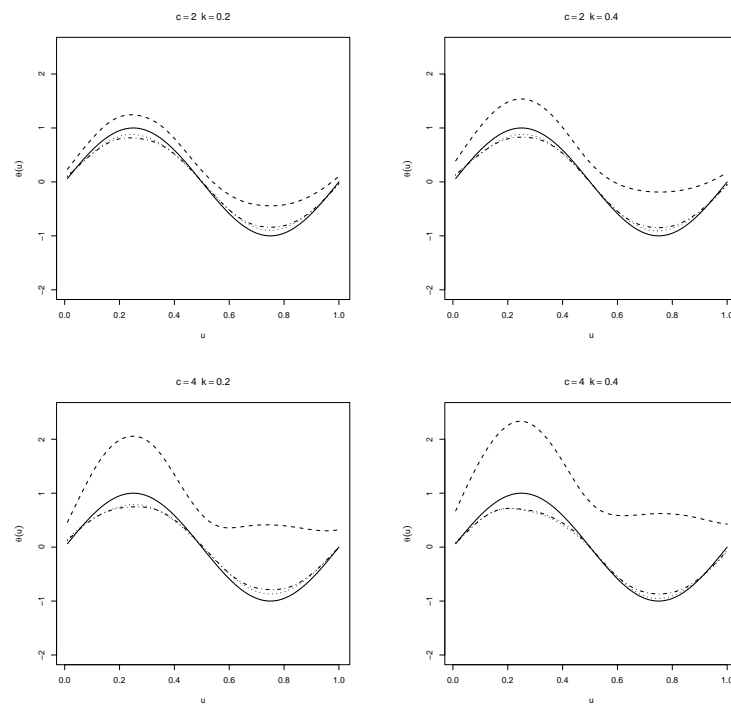


Figure 2. Plot of the functional coefficient estimates through the use of the IAWPLS method (denoted using dot-dashed lines), the NAWPLS method (denoted using dashed lines), and the IWPLS method (denoted using dotted lines); the solid lines represent the true curves.

Table 3. Sample RMSEs for $\theta(\cdot)$ based on the IAWPLS, NAWPLS, and IWPLS methods.

π	c	k	n	IAWPLS	NAWPLS	IWPLS
π_1	2	0.2	50	0.4113	0.5122	0.4348
			100	0.2703	0.4154	0.2845
			200	0.1974	0.3665	0.2043
		0.4	300	0.1666	0.3657	0.1710
			50	0.4413	0.6413	0.4713
			100	0.2843	0.5921	0.3020
	4	0.2	200	0.2121	0.5615	0.2205
			300	0.1775	0.5600	0.1819
			50	0.7796	1.1731	0.8454
		0.4	100	0.4833	1.0768	0.5332
			200	0.3495	1.0242	0.3827
			300	0.2854	1.0083	0.3107
π_2	2	0.2	50	0.8237	1.3033	0.9124
			100	0.5442	1.2269	0.6099
			200	0.3816	1.2144	0.4253
		0.4	300	0.3208	1.2050	0.3562
			50	0.5148	0.5856	0.5424
			100	0.3174	0.4392	0.3376
	4	0.2	200	0.2380	0.3876	0.2473
			300	0.2005	0.3746	0.2066
			50	0.5494	0.7295	0.5908
		0.4	100	0.3509	0.6042	0.3740
			200	0.2532	0.5756	0.2674
			300	0.2176	0.5659	0.2296
4	0.2	50	0.9109	1.2606	0.9901	
		100	0.6003	1.1050	0.6560	
		200	0.4252	1.0332	0.4671	
	0.4	300	0.3603	1.0185	0.3962	
		50	1.0109	1.3693	1.1332	
		100	0.6934	1.2575	0.7847	
4	0.4	200	0.5296	1.2201	0.5875	
		300	0.4517	1.2134	0.5030	

4. Real Data Analysis

We applied our adaptive-weighted adjusted estimation method to the National Longitudinal Survey of Young Men (NLSYM) dataset, which includes 3010 samples from 1976. This dataset has been used widely to analyze the endogeneity issues for parametric and semi-parametric models, such as in [18,21,23,34]. We aim to study the potential relationship between the log of hourly wage in cents (Y) and six other explanatory variables: the years of schooling (Z_1), the dummy variable black (Z_2), south (Z_3), the standard metropolitan statistical area (Z_4 , smsa), age (U), and work experience (X), constructed as $U - Z_1 - 6$; further details regarding the variables in the dataset can be found in [34]. Similar to [20,34], we constructed the following semi-varying coefficient model:

$$Y = X\theta(U) + Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4 + \varepsilon.$$

Based on the idea of [34], we took a four-year university as the corresponding instrumental variable since the years of schooling are not randomly assigned or endogenous. For the missing data, we used the following selection probability model to randomly delete approximately 11% of the responses:

$$\begin{aligned} \pi(x, u, z) &= P(\delta = 1 | X = x, U = u, \mathbf{Z} = (z_1, z_2, z_3, z_4)) \\ &= [1 + \exp(-1 + 1.5x - 0.4z_1 - 0.4z_2 - 0.5z_3 - 0.7z_4 - 0.5u)]^{-1}. \end{aligned}$$

In addition, the Gaussian kernel function was chosen in this case, and the bandwidths $h_1, h_2,$ and \hat{h} were all selected as 0.3 for ease of calculation. We first applied the adjusted profile least-squares method to obtain the initial estimators for unknown parameters and functional coefficients. Then, the model fitting values \hat{Y} and residuals $Y - \hat{Y}$ could be derived using a simple calculation, and a scatter plot of \hat{Y} to $Y - \hat{Y}$ is presented in Figure 3.

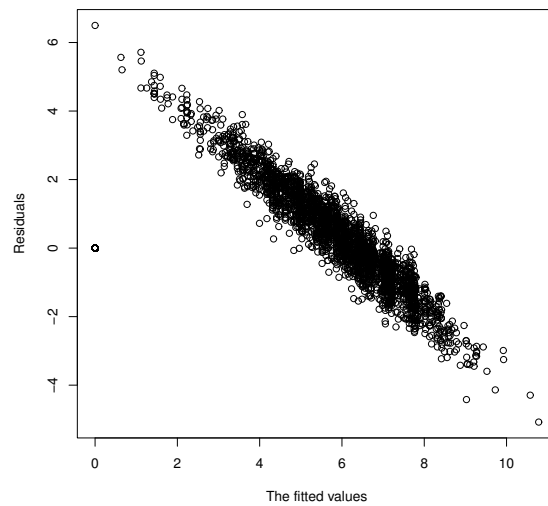


Figure 3. Plot of the model residuals for NLSYM data.

Figure 3 demonstrates that the model residuals show a certain linear trend instead of random variation. Therefore, we concluded that there should be a heteroscedasticity structure. To further reveal the heteroscedasticity, an adjusted Nadaraya–Watson kernel estimation method is suggested for estimating the variance function, and the estimated curve is presented in Figure 4.

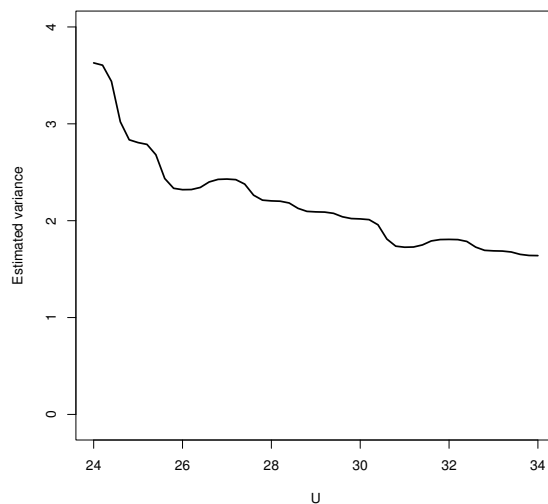


Figure 4. Plot of the variance function estimate for NLSYM data.

From Figure 4, we found that the variance function shows a significant downward trend with the increase of age U , indicating the existence of heteroscedasticity in this specified model. Then, we employed the proposed adaptive-weighted adjusted estimation method to estimate unknown parameters and functional coefficients. For comparison, the three estimation methods IAWPLS, NAWPLS, and IWPLS were included, and the simulation results for unknown parameters and functional coefficients are shown in Table 4 and Figure 5, respectively.

Table 4. The estimates for the parameters’ vector based on the IAWPLS, IWPLS, and NAWPLS methods for NLSYM data.

	β_1	β_2	β_3	β_4
IAWPLS	0.0806	−0.6135	1.5279	0.4864
IWPLS	0.0841	−0.5959	1.6003	0.5702
NAWPLS	0.3069	−0.1681	0.1772	−0.1093

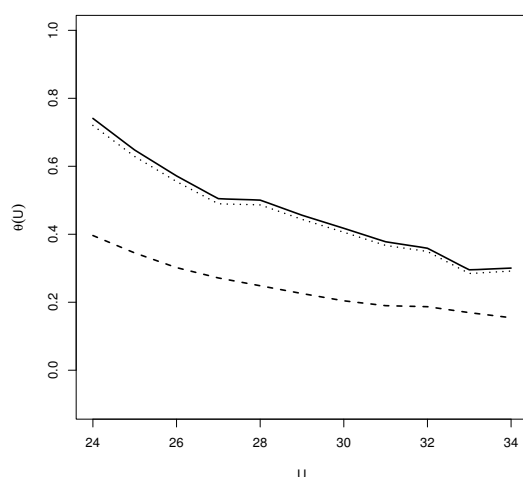


Figure 5. Plot of the functional coefficient estimates through the use of the IAWPLS method (solid lines), the NAWPLS method (dashed lines), and the IWPLS method (dotted lines) for NLSYM data.

According to Table 4 and Figure 5, we found that although the IAWPLS estimators and IWPLS estimators are relatively close to each other, the IWPLS estimation method slightly overestimated the parameters vector. In addition, there is a significant deviation in the NAWPLS estimators compared with the other estimators. Overall, our proposed estimation method can effectively eliminate the endogeneity and heteroscedasticity for semi-varying coefficient models with missing data.

5. Conclusions

In this paper, we study an estimation problem for semi-varying coefficient instrumental variable models with missing data, and the model error is subject to heteroscedasticity simultaneously. An adaptive-weighted adjusted estimation procedure is proposed based on the instrumental variable adjustment technique. The consistency of the variance function estimator is established, and the asymptotic distribution of unknown parameters and functional coefficient estimators is also demonstrated under some regular conditions. Moreover, numerous simulation studies and a NLSYM data analysis further demonstrate the effectiveness of the proposed estimation method. However, we only discuss the estimation problem of the semi-varying coefficient instrumental variables model with heteroscedasticity and missing data in this study. More interesting research topics can be explored in the future, including variable selection and model averaging issues. In addition, high-dimensional data have become a focus in statistical research. How to develop statistical inference methods and theories for a semi-varying coefficient heteroscedastic instrumental variables model with high-dimensional data is an interesting research direction. These issues will be studied in future work.

Author Contributions: Conceptualization, W.Z., S.M. and J.L.; Methodology, W.Z., S.M. and J.L.; Validation, W.Z., S.M. and J.L.; Formal analysis, W.Z.; Investigation, W.Z.; Data curation, W.Z.; Writing—original draft preparation, W.Z.; Writing—review and editing, W.Z., S.M. and J.L.; Supervision, W.Z., S.M. and J.L.; Project administration, W.Z. and S.M.; Funding acquisition, W.Z. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Inner Mongolia Autonomous Region of China (Nos. 2023QN01001, 2022MS07014), the National Natural Science Foundation of China (No. 12271046 and 71661027), and the Research Program of Humanities and Social Sciences at Universities of Inner Mongolia Autonomous Region of China (No. NJSY22497).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available to protect sensitive information.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Lemma A1 (Mack and Silverman [35]). *Suppose that $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ are i.i.d. random vectors, where Y_i ($i = 1, \dots, n$) are scalar random variables. Assume that $\mathbb{E}|Y_i|^s < \infty$ and $\sup_x \int |y|^s f(x, y) dy < \infty$, where $f(x, y)$ denotes the joint probability density of random variables X and Y , and $K(\cdot)$ is a bounded positive function and defined on a bounded support. Moreover, $K(\cdot)$ satisfies a Lipschitz condition. Given $n^{2r-1}h \rightarrow \infty$ for some $r < 1 - s^{-1}$, we have:*

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

Lemma A2 (Shi and Lau [36]). *Let T_1, \dots, T_n be i.i.d. random variables. If $\mathbb{E}|T_i|^s$ is bounded for $s > 1$, then $\max_{1 \leq i \leq n} |T_i| = o_p(n^{1/s})$, a.s.*

Lemma A3 (Chen et al. [31]). *Let $\boldsymbol{\tau}_i = (1, \mathbf{X}_i^\top, U_i, \mathbf{Z}_i^\top)^\top$, λ_{\min} denotes the least eigenvalue of $\sum_{i=1}^n \boldsymbol{\tau}_i \boldsymbol{\tau}_i^\top$, and assume that $\sup_{i \geq 1} \|\boldsymbol{\tau}_i\| < \infty$ and $\lambda_{\min} \rightarrow \infty$; then, the quasi-likelihood estimation $\hat{\boldsymbol{w}} = (\hat{w}_0, \hat{w}_1, \hat{\boldsymbol{w}}_2^\top, \hat{\boldsymbol{w}}_3^\top)^\top$ of $\boldsymbol{w} = (w_0, w_1, \boldsymbol{w}_2^\top, \boldsymbol{w}_3^\top)^\top$ satisfies:*

$$\sqrt{n}(\hat{\boldsymbol{w}} - \boldsymbol{w}) = \mathbf{A}^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \boldsymbol{\tau}_i (\delta_i - \pi_i) + o_p(1),$$

where $\mathbf{A} = \mathbb{E}[\boldsymbol{\tau}_1 \boldsymbol{\tau}_1^\top \pi_1 (1 - \pi_1)]$.

Lemma A4. *Under regularity conditions (C1)–(C9), we have:*

$$\max_{1 \leq i \leq n} \left| \frac{\delta_i}{\pi(\mathbf{V}_i, \hat{\boldsymbol{w}})} - \frac{\delta_i}{\pi(\mathbf{V}_i, \boldsymbol{w})} \right| = o_p(n^{-\frac{1}{2} + \frac{1}{2s}}).$$

Proof of Lemma A4. According to Lemma A3, we use the first-order Taylor expansion $\delta_i / \pi(\mathbf{V}_i, \hat{\boldsymbol{w}})$ at \boldsymbol{w} ,

$$\begin{aligned} \frac{\delta_i}{\pi(\mathbf{V}_i, \hat{\boldsymbol{w}})} &= \frac{\delta_i}{\pi(\mathbf{V}_i, \boldsymbol{w})} + \left[\frac{\delta_i}{\pi(\mathbf{V}_i, \boldsymbol{w})} \right]' (\hat{\boldsymbol{w}} - \boldsymbol{w}) + o_p(|\hat{\boldsymbol{w}} - \boldsymbol{w}|) \\ &= \frac{\delta_i}{\pi(\mathbf{V}_i, \boldsymbol{w})} - \frac{\delta_i \pi'(\mathbf{V}_i, \boldsymbol{w})}{\pi^2(\mathbf{V}_i, \boldsymbol{w})} (\hat{\boldsymbol{w}} - \boldsymbol{w}) + o_p(1). \end{aligned}$$

By Condition (C9) and Lemma A2, we have:

$$\left| \frac{\delta_i}{\pi(\mathbf{V}_i, \hat{\boldsymbol{w}})} - \frac{\delta_i}{\pi(\mathbf{V}_i, \boldsymbol{w})} \right| \leq \max \left| \boldsymbol{\tau}_i^\top \right| \left| \frac{-\delta_i [1 - \pi(\mathbf{V}_i, \boldsymbol{w})]}{\pi^2(\mathbf{V}_i, \boldsymbol{w})} \right| |(\hat{\boldsymbol{w}} - \boldsymbol{w})| + o_p(1) = o_p(n^{-\frac{1}{2} + \frac{1}{2s}}).$$

This completes the proof of Lemma A4. □

Lemma A5. *Under regularity conditions (C1)–(C9), as $n \rightarrow \infty$, it holds that:*

$$\frac{1}{n} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{X}}_{h_2}^\top(u_0) = \sigma^{-2}(u_0) f(u_0) \mathbf{\Psi} \mathbf{\Pi}(u_0) \mathbf{\Psi}^\top \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \{1 + O_p(c_{n_2})\}, \tag{A1}$$

$$\frac{1}{n} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \mathbf{Z} = \sigma^{-2}(u_0) f(u_0) \mathbf{\Psi} \mathbf{\Xi}(u_0) \otimes (1 \ 0)^\top \{1 + O_p(c_{n_2})\}, \tag{A2}$$

$$\frac{1}{n} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \mathbf{M} = \sigma^{-2}(u_0) f(u_0) \mathbf{\Psi} \mathbf{\Pi}(u_0) \mathbf{\Psi}^\top \boldsymbol{\theta}(u_0) \otimes (1 \ 0)^\top \{1 + O_p(c_{n_2})\}, \tag{A3}$$

$$\frac{1}{n} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \boldsymbol{\varepsilon} = O_p(c_{n_2}), \tag{A4}$$

where \otimes denotes the Kronecker product.

Proof of Lemma A5. For any $u_0 \in \mathcal{U}$, by some simple calculation, we have:

$$\frac{1}{n} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{X}}_{h_2}^\top(u_0) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \mathbf{B}_i & \sum_{i=1}^n \mathbf{B}_i h_2^{-1}(U_i - u_0) \\ \sum_{i=1}^n \mathbf{B}_i h_2^{-1}(U_i - u_0) & \sum_{i=1}^n \mathbf{B}_i h_2^{-2}(U_i - u_0)^2 \end{pmatrix},$$

where $\mathbf{B}_i = [\delta_i / \pi(\mathbf{V}_i, \widehat{\mathbf{w}})] \widehat{\sigma}^{-2}(U_i) \widehat{\mathbf{X}}_i \widehat{\mathbf{X}}_i^\top K_{h_2}(U_i - u_0)$.

We first consider the term of the upper-left corner of the matrix. It is noteworthy that $\widehat{\Psi}$ is the usual moment estimator of Ψ . According to [16], we obtain $\widehat{\Psi} = \Psi + O_p(n^{-1/2})$. Therefore, combining Theorem 1, Lemma A4 and condition (C8), we can obtain:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \widehat{\sigma}^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \widehat{\mathbf{w}})} \widehat{\mathbf{X}}_i \widehat{\mathbf{X}}_i^\top K_{h_2}(U_i - u_0) \\ &= \frac{1}{n} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \mathbf{\Psi} \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top \mathbf{\Psi}^\top K_{h_2}(U_i - u_0) \\ & \quad + \frac{1}{n} \sum_{i=1}^n [\widehat{\sigma}^{-2}(U_i) - \sigma^{-2}(U_i)] \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \mathbf{\Psi} \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top \mathbf{\Psi}^\top K_{h_2}(U_i - u_0) \\ & \quad + \frac{1}{n} \sum_{i=1}^n \sigma^{-2}(U_i) \left[\frac{\delta_i}{\pi(\mathbf{V}_i, \widehat{\mathbf{w}})} - \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \right] \mathbf{\Psi} \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top \mathbf{\Psi}^\top K_{h_2}(U_i - u_0) \\ & \quad + \frac{1}{n} \sum_{i=1}^n [\widehat{\sigma}^{-2}(U_i) - \sigma^{-2}(U_i)] \left[\frac{\delta_i}{\pi(\mathbf{V}_i, \widehat{\mathbf{w}})} - \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \right] \mathbf{\Psi} \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top \mathbf{\Psi}^\top K_{h_2}(U_i - u_0) + O_p(n^{-1/2}) \\ &= \sigma^{-2}(u_0) f(u_0) \mathbf{\Psi} \mathbf{\Pi}(u_0) \mathbf{\Psi}^\top + o_p(\tilde{c}_n) + o_p(n^{-\frac{1}{2} + \frac{1}{2s}}) + o_p(\tilde{c}_n) o_p(n^{-\frac{1}{2} + \frac{1}{2s}}) + O_p(n^{-1/2}) \\ &= \sigma^{-2}(u_0) f(u_0) \mathbf{\Psi} \mathbf{\Pi}(u_0) \mathbf{\Psi}^\top + O_p(\tilde{c}_{n_2}). \end{aligned}$$

By using the same argument as above, other terms in the matrix can be proven, which lead to the result of (A1). The proofs of (A2)–(A4) are similar to those of (A1). We omit the details here. \square

Lemma A6. Under regularity conditions (C1)–(C9), as $n \rightarrow \infty$, we have:

$$\frac{1}{n} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{Z}}^\top \rightarrow \boldsymbol{\Lambda}_1, \text{ a.s.,}$$

where $\boldsymbol{\Sigma} = \text{diag}[\sigma^2(U_1), \sigma^2(U_2), \dots, \sigma^2(U_n)]$, $\boldsymbol{\Delta} = \text{diag}\left[\frac{\delta_1}{\pi(\mathbf{V}_1, \widehat{\mathbf{w}}_1)}, \dots, \frac{\delta_n}{\pi(\mathbf{V}_n, \widehat{\mathbf{w}}_n)}\right]$, $\boldsymbol{\Lambda}_1$ is given in Theorem 2.

Proof of Lemma A6. Invoking (A1) and (A2), it is easy to show that:

$$\begin{aligned} & (\widehat{\mathbf{X}}^\top \mathbf{0}) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \mathbf{Z} \\ &= (\Psi \zeta)^\top (\Psi \Pi(u_0) \Psi^\top)^{-1} \Psi \Xi(u_0) \{ 1 + O_p(c_{n_2}) \}. \end{aligned}$$

Then, we have:

$$\widehat{\mathbf{S}}\mathbf{Z} = \begin{pmatrix} (\Psi \zeta_1)^\top (\Psi \Pi(U_1) \Psi^\top)^{-1} \Psi \Xi(U_1) \\ \vdots \\ (\Psi \zeta_n)^\top (\Psi \Pi(U_n) \Psi^\top)^{-1} \Psi \Xi(U_n) \end{pmatrix} \{ 1 + O_p(c_{n_2}) \}. \tag{A5}$$

Then, invoking (A5), some calculations yield:

$$\frac{1}{n} \widehat{\mathbf{Z}}^\top \Sigma^{-1} \Delta \widehat{\mathbf{Z}} = \frac{1}{n} (\mathbf{Z} - \widehat{\mathbf{S}}\mathbf{Z})^\top \Sigma^{-1} \Delta (\mathbf{Z} - \widehat{\mathbf{S}}\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^\top + O_p(c_{n_2}),$$

where $\boldsymbol{\eta}_i = \mathbf{Z}_i - \Xi(U_i) \Psi^\top (\Psi \Pi(U_i) \Psi^\top)^{-1} \Psi \zeta_i$. Then, by a law of large numbers, Lemma A6 can be easily proven. \square

Lemma A7. Under regularity conditions (C1)–(C9), then as $n \rightarrow \infty$, we can obtain:

$$n^{-1/2} \widehat{\mathbf{Z}}^\top \Sigma^{-1} \Delta (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}) \xrightarrow{D} N(\mathbf{0}, \Lambda_2), \quad n \rightarrow \infty,$$

where $\widehat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \widehat{\mathbf{S}})\boldsymbol{\varepsilon}$, $\widehat{\mathbf{M}}_w = (\mathbf{I} - \widehat{\mathbf{S}})\mathbf{M}$ is given in Theorem 2.

Proof of Lemma A7. Invoking (A1) and (A3), it is easy to check that:

$$\begin{aligned} & (\widehat{\mathbf{X}}^\top \mathbf{0}) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \mathbf{M} \\ &= (\Psi \zeta)^\top \boldsymbol{\theta}(u_0) \{ 1 + O_p(c_{n_2}) \}. \end{aligned}$$

Then, invoking (A5), some calculations yield:

$$\begin{aligned} & n^{-1/2} \widehat{\mathbf{Z}}^\top \Sigma^{-1} \Delta \widehat{\mathbf{M}}_w = n^{-1/2} \widehat{\mathbf{Z}}^\top (\mathbf{I} - \widehat{\mathbf{S}})^\top \Sigma^{-1} \Delta (\mathbf{I} - \widehat{\mathbf{S}}) \mathbf{M} \\ &= n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i (1 + O_p(c_{n_2})) [\mathbf{X}_i^\top \boldsymbol{\theta}(U_i) - (\Psi \zeta_i)^\top \boldsymbol{\theta}(U_i) (1 + O_p(c_{n_2}))] \\ &= n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i (\Psi \zeta_i)^\top \boldsymbol{\theta}(U_i) O_p(c_{n_2}) + n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i \boldsymbol{\varepsilon}_i^\top \boldsymbol{\theta}(U_i). \end{aligned}$$

Note that $\mathbb{E}[\boldsymbol{\eta}_i (\Psi \zeta_i)^\top \boldsymbol{\theta}(U_i) | U_i] = 0$, and then we can prove:

$$n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i (\Psi \zeta_i)^\top \boldsymbol{\theta}(U_i) O_p(c_{n_2}) = O_p(n^{1/2} c_{n_2}^2).$$

Therefore, we derive:

$$n^{-1/2} \widehat{\mathbf{Z}}^\top \Sigma^{-1} \Delta \widehat{\mathbf{M}}_w = n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i \boldsymbol{\varepsilon}_i^\top \boldsymbol{\theta}(U_i) + O_p(n^{1/2} c_{n_2}^2). \tag{A6}$$

In addition, invoking (A1) and (A4), we have:

$$(\widehat{\mathbf{X}}^\top \mathbf{0}) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\Sigma}^{-1} \widehat{\Delta} \boldsymbol{\varepsilon} = O_p(c_{n_2}).$$

Then, invoking (A5), some calculations yield:

$$\begin{aligned} n^{-1/2}\widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\boldsymbol{\varepsilon}} &= n^{-1/2}\widehat{\mathbf{Z}}^\top (\mathbf{I} - \widehat{\mathbf{S}})^\top \boldsymbol{\Sigma}^{-1} \Delta (\mathbf{I} - \widehat{\mathbf{S}}) \boldsymbol{\varepsilon} \\ &= n^{-1/2} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w}_i)} \boldsymbol{\eta}_i \boldsymbol{\varepsilon}_i + O_p(c_{n_2}). \end{aligned} \tag{A7}$$

Hence, combining (A6), (A7), and condition (C8), with the help of the central limit theorem, the result of Lemma A7 can be obtained. \square

The proof of Theorem 1 is similar to that of Theorem 1 in [9]. We omit the details here.

Proof of Theorem 2. Note that:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_w - \boldsymbol{\beta}) = \sqrt{n}(\widehat{\boldsymbol{\beta}}_w - \widehat{\boldsymbol{\beta}}_v) + \sqrt{n}(\widehat{\boldsymbol{\beta}}_v - \boldsymbol{\beta}),$$

where $\widehat{\boldsymbol{\beta}}_v = (\widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Y}}$. Then, we need to complete the proof of

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_v - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Lambda}_1^{-1} \boldsymbol{\Lambda}_2 \boldsymbol{\Lambda}_1^{-1}), \quad n \rightarrow \infty \tag{A8}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_w - \widehat{\boldsymbol{\beta}}_v) = o_p(1). \tag{A9}$$

Multiplying both sides of model (1) by $(\mathbf{I} - \widehat{\mathbf{S}})$, we obtain:

$$\widehat{\mathbf{Y}} = \widehat{\mathbf{Z}} \boldsymbol{\beta} + \widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}.$$

Then, we have:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_v - \boldsymbol{\beta}) = (n^{-1} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} n^{-1/2} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}).$$

Invoking Lemma A6 and Lemma A7, we can derive the result of (A8) using the Slutsky Theorem. Since:

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\beta}}_w - \widehat{\boldsymbol{\beta}}_v) &= \sqrt{n}[(\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Y}} - (\widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Y}}] \\ &= \sqrt{n}[(\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} - (\widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Z}})^{-1}] \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}) \\ &\quad + \sqrt{n}(\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Z}})^{-1} (\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta - \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta) (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}). \end{aligned}$$

Thus, to obtain the result of (A9), we have to prove:

$$\begin{aligned} n^{-1}(\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Z}} - \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta \widehat{\mathbf{Z}}) &= o_p(1), \quad n^{-1/2} \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}) = O_p(1), \\ n^{-1} \widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{Z}} &= O_p(1), \quad n^{-1/2} (\widehat{\mathbf{Z}}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \Delta - \widehat{\mathbf{Z}}^\top \boldsymbol{\Sigma}^{-1} \Delta) (\widehat{\mathbf{M}}_w + \widehat{\boldsymbol{\varepsilon}}) = o_p(1). \end{aligned}$$

Using a similar method to the proof of Theorem 3 in [9], it is easy to prove the above conclusions under regular conditions (C1)–(C9), so the details are omitted here. Then, combining (A8) and (A9) yields the result in Theorem 2. \square

Proof of Theorem 3. Recall the definition of $\widehat{\boldsymbol{\theta}}_w(u_0, \widehat{\boldsymbol{\beta}}_w)$ in (21); we have:

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_w(u_0, \widehat{\boldsymbol{\beta}}_w) &= (\mathbf{I}_p, \mathbf{0}_p) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta (\mathbf{Y} - \mathbf{Z} \widehat{\boldsymbol{\beta}}_w) \\ &= (\mathbf{I}_p, \mathbf{0}_p) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \mathbf{M}_\zeta \\ &\quad + (\mathbf{I}_p, \mathbf{0}_p) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \mathbf{Z} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_w) \\ &\quad + (\mathbf{I}_p, \mathbf{0}_p) \{ \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta \widehat{\mathbf{X}}_{h_2}(u_0) \}^{-1} \widehat{\mathbf{X}}_{h_2}^\top(u_0) \mathbf{w}_{h_2}(u_0) \widehat{\boldsymbol{\Sigma}}^{-1} \Delta (\boldsymbol{\varepsilon} + \mathbf{e}) \\ &\triangleq K_1 + K_2 + K_3, \end{aligned}$$

where:

$$\begin{aligned} \mathbf{M}_\zeta &= [(\Psi\zeta_1)^\top \boldsymbol{\theta}(U_1), (\Psi\zeta_2)^\top \boldsymbol{\theta}(U_2), \dots, (\Psi\zeta_n)^\top \boldsymbol{\theta}(U_n)]^\top, \\ \mathbf{e} &= (\mathbf{e}_1^\top \boldsymbol{\theta}(U_1), \mathbf{e}_2^\top \boldsymbol{\theta}(U_2), \dots, \mathbf{e}_n^\top \boldsymbol{\theta}(U_n))^\top. \end{aligned}$$

Let us consider K_1 first; for any point U_i in the neighborhood of u_0 , each functional coefficients $\boldsymbol{\theta}(U_i)$ can be approximated by:

$$\boldsymbol{\theta}(U_i) = \boldsymbol{\theta}(u_0) + h_2 \boldsymbol{\theta}'(u_0) \frac{U_i - u_0}{h_2} + \frac{h_2^2}{2} \boldsymbol{\theta}''(u_0) \left(\frac{U_i - u_0}{h_2}\right)^2 + o_p(h_2^2).$$

Then, we have:

$$\mathbf{M}_\zeta = \begin{pmatrix} (\Psi\zeta_1)^\top \boldsymbol{\theta}(U_1) \\ \vdots \\ (\Psi\zeta_n)^\top \boldsymbol{\theta}(U_n) \end{pmatrix} = \widehat{\mathbf{X}}_{h_2}^\top(u_0) \begin{pmatrix} \boldsymbol{\theta}(u_0) \\ h_2 \boldsymbol{\theta}'(u_0) \end{pmatrix} + \frac{h_2^2}{2} \begin{pmatrix} \widehat{\mathbf{X}}_1^\top \left(\frac{U_1 - u_0}{h_2}\right)^2 \\ \vdots \\ \mathbf{X}_n^\top \left(\frac{U_n - u_0}{h_2}\right)^2 \end{pmatrix} \boldsymbol{\theta}''(u_0) + o_p(h_2^2).$$

By Lemma 1, we obtain:

$$K_1 = \boldsymbol{\theta}(u_0) + \frac{h_2^2}{2} \mu_2 \boldsymbol{\theta}''(u_0) + o_p(h_2^2). \tag{A10}$$

For K_2 , combining (A1), (A2), Theorem 2, and condition (C8), we can obtain that:

$$\sqrt{nh_2} K_2 = \sqrt{nh_2} (\Psi\Pi(u_0)\Psi^\top)^{-1} \Psi\Xi(u_0) \{1 + O_p(c_{n_2})\} O_p(n^{-1/2}) = o_p(1). \tag{A11}$$

Now, we consider K_3 . Combining (A1), Theorem 1, and Lemma 4, we can derive:

$$\begin{aligned} \sqrt{nh_2} K_3 &= \sigma^2(u_0) f^{-1}(u_0) (\Psi\Pi(u_0)\Psi^\top)^{-1} \\ &\quad \times \sqrt{nh_2} \frac{1}{n} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \widehat{\mathbf{X}}_i K_{h_2}(U_i - u_0) (\varepsilon_i + \mathbf{e}_i^\top \boldsymbol{\theta}(U_i)) + o_p(1). \end{aligned}$$

Since $\sqrt{nh_2} \frac{1}{n} \sum_{i=1}^n \sigma^{-2}(U_i) \frac{\delta_i}{\pi(\mathbf{V}_i, \mathbf{w})} \widehat{\mathbf{X}}_i K_{h_2}(U_i - u_0) (\varepsilon_i + \mathbf{e}_i^\top \boldsymbol{\theta}(U_i))$ follows an asymptotic normal distribution with mean zero and the variance function

$$v_0 f(u_0) \sigma^{-4}(u_0) \mathbb{E}\{\pi(\mathbf{V}_1, \mathbf{w})^{-1} (\mathbf{e}_1^\top \boldsymbol{\theta}(U_1) + \varepsilon_1)^2\} \Psi\Pi(u_0)\Psi^\top.$$

Then, by the Slutsky Theorem, we have:

$$\sqrt{nh_2} K_3 \xrightarrow{D} v_0 f^{-1}(u_0) \mathbb{E}\{\pi(\mathbf{V}_1, \mathbf{w})^{-1} (\mathbf{e}_1^\top \boldsymbol{\theta}(U_1) + \varepsilon_1)^2\} (\Psi\Pi(u_0)\Psi^\top)^{-1}. \tag{A12}$$

Combining (A10)–(A12) together with the Slutsky Theorem yields the result in Theorem 3. \square

References

1. Zhang, W.Y.; Lee, S.Y.; Song, X.Y. Local polynomial fitting in semi-varying coefficient models. *J. Multivar. Anal.* **2002**, *82*, 166–188. [[CrossRef](#)]
2. Zhou, X.; You, J.H. Wavelet estimation in varying-coefficient partially linear regression model. *Stat. Probab. Lett.* **2004**, *68*, 91–104. [[CrossRef](#)]
3. Fan, J.Q.; Huang, T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **2005**, *11*, 1031–1057. [[CrossRef](#)]
4. Zhao, P.X.; Xue, L.G. Variable selection for semiparametric varying coefficient partially linear models. *Stat. Probab. Lett.* **2009**, *79*, 2148–2157. [[CrossRef](#)]
5. Kai, B.; Li, R.Z.; Zou, H. New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Stat.* **2011**, *39*, 305–332. [[CrossRef](#)] [[PubMed](#)]
6. Yang, J.; Lu, F.; Yang, H. Quantile regression for robust estimation and variable selection in partially linear varying-coefficient models. *Stat. J. Theor. Appl. Stat.* **2017**, *51*, 1–21. [[CrossRef](#)]

7. Li, Y.J.; Li, G.R.; Lian, H.; Tong, T.J. Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models. *J. Multivar. Anal.* **2017**, *155*, 133–150. [[CrossRef](#)]
8. Zhao, P.X.; Yang, Y.P. A new orthogonality-based estimation for varying-coefficient partially linear models. *J. Korean Stat. Soc.* **2019**, *48*, 29–39. [[CrossRef](#)]
9. Shen, S.L.; Cui, J.L.; Mei, C.L.; Wang, C.L. Estimation and inference of semi-varying coefficient models with heteroscedastic errors. *J. Multivar. Anal.* **2014**, *124*, 70–93. [[CrossRef](#)]
10. Zhao, Y.Y.; Lin, J.G.; Xu, P.R.; Ye, X.G. Orthogonality-projection-based estimation for semi-varying coefficient models with heteroscedastic errors. *Comput. Stat. Data Anal.* **2015**, *89*, 204–221. [[CrossRef](#)]
11. Zhao, F.R.; Song, W.X.; Shi, J.H. Statistical inference for heteroscedastic semi-varying coefficient EV models. *Commun. Stat.-Theory Methods* **2018**, *48*, 2432–2455. [[CrossRef](#)]
12. Zhang, W.W.; Li, G.R. Weighted bias-corrected restricted statistical inference for heteroscedastic semiparametric varying-coefficient errors-in-variables model. *J. Korean Stat. Soc.* **2021**, *50*, 1098–1128. [[CrossRef](#)]
13. Yuan, Y.Z.; Zhou, Y. Adaptive-weighted estimation of semi-varying coefficient models with heteroscedastic errors. *J. Stat. Comput. Simul.* **2021**, *91*, 3029–3047. [[CrossRef](#)]
14. Greenland, S. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* **2000**, *29*, 722–729. [[CrossRef](#)] [[PubMed](#)]
15. Fan, J.Q.; Liao, Y. Endogeneity in dimensions. *Ann. Stat.* **2014**, *42*, 872–917. [[CrossRef](#)] [[PubMed](#)]
16. Cai, Z.W.; Xiong, H.Y. Partially varying coefficient instrumental variables models. *Stat. Neerl.* **2012**, *66*, 85–110. [[CrossRef](#)]
17. Zhao, P.X.; Li, G.R. Modified SEE variable selection for varying coefficient instrumental variable models. *Stat. Methodol.* **2013**, *12*, 60–70. [[CrossRef](#)]
18. Zhao, P.X.; Xue, L.G. Empirical likelihood inferences for semiparametric instrumental variable models. *J. Appl. Math. Comput.* **2013**, *43*, 75–90. [[CrossRef](#)]
19. Yuan, J.Y.; Zhao, P.X.; Zhang, W.G. Semiparametric variable selection for partially varying coefficient models with endogenous variables. *Comput. Stat.* **2016**, *31*, 693–707. [[CrossRef](#)]
20. Zhao, P.X.; Zhou, X.S.; Wang, X.L.; Huang, X.S. A new orthogonality empirical likelihood for varying coefficient partially linear instrumental variable models with longitudinal data. *Commun. Stat. Simul. Comput.* **2020**, *49*, 3328–3344. [[CrossRef](#)]
21. Yao, F. Efficient semiparametric instrumental variable estimation under conditional heteroskedasticity. *J. Quant. Econ.* **2012**, *10*, 32–55.
22. Yang, Y.P.; Chen, L.F.; Zhao, P.X. Empirical likelihood inference in partially linear single-index models with endogenous covariates. *Commun. Stat.-Theory Methods* **2017**, *46*, 3297–3307. [[CrossRef](#)]
23. Huang, J.T.; Zhao, P.X. Orthogonal weighted empirical likelihood-based variable selection for semiparametric instrumental variable models. *Commun. Stat.-Theory Methods* **2018**, *47*, 4375–4388. [[CrossRef](#)]
24. Tang, X.R.; Zhao, P.X.; Yang, Y.P.; Yang, W.M. Adjusted empirical likelihood inferences for varying coefficient partially non linear models with endogenous covariates. *Commun. Stat.-Theory Methods* **2022**, *51*, 953–973. [[CrossRef](#)]
25. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [[CrossRef](#)]
26. Robins, J.M.; Rotnitzky, A.; Zhao, L.P. Estimation of regression coefficient when some regressors are not always observed. *J. Am. Stat. Assoc.* **1994**, *89*, 846–866. [[CrossRef](#)]
27. Wang, Q.H.; Rao, J.N.K. Empirical likelihood-based inference in linear models with missing response data. *Scand. J. Stat.* **2002**, *29*, 563–576. [[CrossRef](#)]
28. Wang, Q.H.; Linton, O.; Härdle, W. Semiparametric regression analysis with missing response at random. *J. Am. Stat. Assoc.* **2004**, *99*, 334–345. [[CrossRef](#)]
29. Li, Z.Q.; Xue, L.G. The imputation estimators of semiparametric varying-coefficient models with missing data. *Acta Math. Appl. Sin.* **2009**, *32*, 422–430. (In Chinese)
30. Chen, P.P.; Feng, S.Y.; Xue, L.G. Statistical inference for semiparametric varying coefficient partially linear model with missing data. *Acta Math. Sci.* **2015**, *35A*, 345–358. (In Chinese)
31. Xu, H.X.; Fan, G.L.; Wu, C.X. Statistical inference for varying-coefficient partially linear errors-in-variables models with missing data. *Commun. Stat.-Theory Methods* **2019**, *48*, 5621–5636. [[CrossRef](#)]
32. Xiao, Y.T.; Li, F.X. Estimation in partially linear varying-coefficient errors-in-variables models with missing response variables. *Comput. Stat.* **2020**, *35*, 1637–1658. [[CrossRef](#)]
33. Yan, Y.X.; Lan, S.H.; Zhang, C.Y. Statistical inference for partially linear varying coefficient quantile models with missing responses. *Symmetry* **2022**, *14*, 2258. [[CrossRef](#)]
34. Card, D. *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*; Nber Working Papers; University of Toronto Press: Toronto, ON, Canada, 1993; pp. 1127–1160.
35. Mack, Y.P.; Silverman, B.W. Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie Verwandte Geb.* **1982**, *61*, 405–415. [[CrossRef](#)]
36. Shi, J.; Lau, T.S. Empirical likelihood for partially linear models. *J. Multivar. Anal.* **2000**, *72*, 132–148. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.