

Article

Not peer-reviewed version

Estimation for Longitudinal Varying Coefficient Partially Nonlinear Models Based on QR Decomposition

Jiangcui Ge , [Xiaoshuang Zhou](#) ^{*} , Cuiping Wang

Posted Date: 10 November 2025

doi: 10.20944/preprints202511.0704.v1

Keywords: varying coefficient partially nonlinear model; longitudinal data; QR decomposition; quadratic inference function



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Estimation for Longitudinal Varying Coefficient Partially Nonlinear Models Based on QR Decomposition

Jiangcui Ge ^{1,2}, Xiaoshuang Zhou ^{2,*}  and Cuiping Wang ³

¹ School of Mathematics, North University of China, Taiyuan 030051, China

² School of Mathematics and Big Data, Dezhou University, Dezhou 253023, China

³ School of Mathematics and Statistics, Shandong University of Technology, Zibo 255022, China

* Correspondence: zhouxiaoshuang@dzu.edu.cn

Abstract

Aiming at the estimation efficiency problem of multicollinearity and longitudinal data correlation in the varying coefficient partially nonlinear models, a method based on QR decomposition and quadratic inference function (QIF) is proposed to obtain the orthogonality estimation of parameter components and varying coefficient functions. QR decomposition eliminates the pathology of the design matrix, and combines the adaptive weighting of the relevant structures within the group by QIF to effectively capture the complex correlation structure of longitudinal data. The theoretical analysis proves the asymptotic nature of the estimator, and the efficiency of the estimation method proposed in this paper is verified by simulation experiments.

Keywords: varying coefficient partially nonlinear model; longitudinal data; QR decomposition; quadratic inference function

MSC: 62G05; 62G20

1. Introduction

In recent years, longitudinal data have attracted much attention due to their wide applications in fields such as biomedicine, economics and social sciences. This type of data is obtained through repeated observations of the same subject at different time points. As a result, multiple measurement results of the same subject exhibit temporal or spatial dependence, which makes the traditional independent and identically distributed assumption no longer applicable. To address this issue, the varying coefficient partially nonlinear model has become one of the research hotspots, as it combines the interpretability of parametric models and the flexibility of nonparametric models. Li and Mei [1] put forward a profile nonlinear least squares estimation approach to estimate the parameter vector and coefficient function vector of the varying coefficient partially nonlinear model, and further derived the asymptotic characteristics of the obtained estimators.

Important progress has been made in model estimation in existing studies. Yu and Zhu et al. [2] constructed robust estimators for the partially linear additive model under functional data. Yan and Tan et al. [3] explored the empirical likelihood inference targeting the partially linear errors-in-variables model under longitudinal data. Varying coefficient partially nonlinear model based on mode regression, a robust two-stage method for estimation and variable selection was proposed by Xiao and Liang [4]. Zhao and Zhou et al. [5] put forward a new orthogonality-based empirical likelihood inference method through orthogonal estimation techniques and empirical likelihood inference methods, which is used to estimate the parametric and nonparametric components in a class of varying coefficient partially linear instrumental variable models for longitudinal data. Liang and Zeger [6] applied the generalized estimating equation to longitudinal data analysis, which effectively handles the correlation issue of longitudinal data. For the varying coefficient partially nonlinear quantile regression model

under randomly left-censored data, Xu and Fan et al. [7] introduced a three-stage estimation weighting method.

However, significant challenges still exist in the existing methods for estimating the varying coefficient partially nonlinear model. For instance, the estimation of nonlinear parameters is easily disturbed by nonparametric components and longitudinal correlation structures, leading to reduced estimation efficiency; when there is multicollinearity among explanatory variables, the ill-posed nature of the design matrix will further decrease the stability of estimation. Such issues severely limit the performance and practicality of the model in practical applications.

To address the above issues, existing studies have proposed a variety of improved methods. For example, Qu and Lindsay et al. [8] used the quadratic inference function to improve the estimating equation, and this method maintains optimal performance even when the working correlation structure is misspecified. Bai and Fung et al. [9] applied the quadratic inference function to handle longitudinal data, and the results demonstrate that the proposed estimation method exhibits excellent asymptotic properties. Schumaker [10] used B-spline basis functions to approximate the varying coefficient part, which improves the computational efficiency. Wei and Wang et al. [11] proposed a model averaging method for linear models with randomly missing response variables. Jiang and Ji et al. [12] proposed an estimation method for the varying coefficient partially nonlinear model based on the exponential squared loss function. Xiao and Chen [13] proposed a procedure for local bias-corrected cross-sectional nonlinear least squares estimation. For left-censored data, Xu and Fan et al. [14] established a quantile regression estimation and variable selection method for the varying coefficient partially nonlinear model. Based on the varying coefficient partially nonlinear model where the nonparametric part has measurement errors, Qian and Huang [15] discussed a modified profile least squares estimation method. Wang and Zhao et al. [16] proposed an inverse probability weighted cross-sectional least squares method to estimate unknown parameters and nonparametric functions for the varying coefficient partially nonlinear model.

This paper proposes an orthogonal estimation framework that integrates QR decomposition and the QIF. QR decomposition can effectively eliminates the ill-posed nature of the design matrix and improves numerical stability. The QIF method avoids the limitations of traditional generalized estimating equations by adaptively weighting the intra-group correlation structure, significantly enhancing estimation efficiency. This study not only provides a new theoretical tool for longitudinal data analysis but also offers a feasible solution to complex modeling problems in practical applications.

The structure of this paper is arranged as follows: Section 2 introduces the model specification and estimation method, including the specific implementation of QR decomposition and QIF; Section 3 discusses the asymptotic properties of the estimators; Section 4 verifies the superiority of the proposed method through simulation experiments; The proof process of the main conclusions is provided in the Appendix.

2. Models and Methods

Consider the varying coefficient partially nonlinear model introduced by Li and Mei [1]:

$$Y = X^T \alpha(U) + g(Z, \beta) + \varepsilon, \quad (1)$$

where, $Y \in R$ is the response variable, $X \in R^p$, $U \in R$, and $Z \in R^r$ are covariates, $\alpha(U) = (\alpha_1(U), \alpha_2(U), \dots, \alpha_p(U))^T$ and $\alpha_k(U), k = 1, 2, \dots, p$ are unknown smooth functions, $g(\cdot, \cdot)$ is a nonlinear function with a known form, $\beta = (\beta_1, \beta_2, \dots, \beta_q)^T$ denotes a q -dimensional unknown parameter vector, and ε is the model error with mean zero and variance σ^2 .

2.1. Estimation of Parameter Vector

Considering the model (1) under longitudinal data, suppose the j -th observation of the i -th individual satisfies:

$$Y_{ij} = X_{ij}^T \alpha(U_{ij}) + g(Z_{ij}, \beta) + \varepsilon_{ij}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n_i, \quad (2)$$

among them, ε_{ij} has a mean of zero, and $X_{ij} \in R^p, Z_{ij} \in R^r, U_{ij}$ are covariates.

Based on the idea of Schumaker [10], the unknown function $\alpha_k(\cdot), k = 1, 2, \dots, p$ can be approximated using basis functions. The B-spline basis functions are denoted by the vector $B(u) = (B_1(u), B_2(u), \dots, B_L(u))^T$. The dimension L is defined as $L = K + M$, where K and M represent the number of interior knots and the order of the spline, respectively. Then $\alpha_k(u)$ can be approximately expressed as

$$\alpha_k(u) \approx B(u)^T \gamma_k, \quad k = 1, 2, \dots, p, \quad (3)$$

here $\gamma_k = (\gamma_{k1}, \gamma_{k2}, \dots, \gamma_{kL})^T$ are the coefficient vector for the B-spline basis functions. model (2) can be expressed as

$$Y_{ij} \approx W_{ij}^T \gamma + g(Z_{ij}, \beta) + \varepsilon_{ij}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n_i, \quad (4)$$

define $W_{ij} = I_p \otimes B(U_{ij}) \cdot X_{ij}$, where $\otimes$ is the Kronecker product, and let $\gamma = (\gamma_1^T, \gamma_2^T, \dots, \gamma_p^T)^T, Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^T, W_i = (W_{i1}, W_{i2}, \dots, W_{in_i})^T, g(Z_i, \beta) = (g(Z_{i1}, \beta), g(Z_{i2}, \beta), \dots, g(Z_{in_i}, \beta))^T, \varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in_i})^T$. Then model (4) can be expressed as

$$Y_i \approx W_i \gamma + g(Z_i, \beta) + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (5)$$

Suppose that for all $i = 1, 2, \dots, n$, the matrices W_i have full column rank, their QR decomposition can be expressed as:

$$W_i = Q_i \begin{pmatrix} R_i \\ \mathbf{0} \end{pmatrix}, \quad i = 1, 2, \dots, n, \quad (6)$$

where Q_i is an $n_i \times n_i$ orthogonal matrix, R_i is a $pL \times pL$ triangular matrix, and $\mathbf{0}$ denotes a $(n_i - pL) \times pL$ zero matrix. The matrix Q_i can be divided into two parts as $Q_i = (Q_{i1}, Q_{i2})$, where Q_{i1} is a $n_i \times pL$ matrix and Q_{i2} is a $n_i \times (n_i - pL)$ matrix. Substitute the decomposition of Q_i into (6) to obtain $W_i = Q_{i1} R_i$, from the properties of orthogonal matrices, $Q_{i2}^T Q_{i1} = 0$ can be derived, and then $Q_{i2}^T W_i = Q_{i2}^T Q_{i1} R_i = 0$ is obtained. Multiplying both sides of Equation (5) by Q_{i2}^T , and we get:

$$Q_{i2}^T Y_i \approx Q_{i2}^T g(Z_i, \beta) + Q_{i2}^T \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (7)$$

It is easy to see that model (7) is a regression model containing only unknown parameters. Following Liang [6], the generalized estimating equation for β can be defined as:

$$\sum_{i=1}^n g^{(1)}(Z_i, \beta)^T Q_{i2} V_i^{-1} Q_{i2}^T (Y_i - g(Z_i, \beta)) = 0, \quad (8)$$

among them, $g^{(1)}(Z_i, \beta) = (g^{(1)}(Z_{i1}, \beta), \dots, g^{(1)}(Z_{in_i}, \beta))^T, V_i = Q_{i2}^T \Sigma_i Q_{i2}, \Sigma_i$ is the covariance matrices of Y_i , and the structure of V_i can also be expressed as $V_i = A_i^{\frac{1}{2}} R(\rho) A_i^{\frac{1}{2}}$ according to the method of Liang [6], where A_i is a diagonal matrix, $R(\rho)$ is a working correlation matrix, and ρ is a correlation parameter. Since a consistent estimator of ρ is not always available in practice, we adopt the QIF method to approximate the working correlation matrix $R^{-1}(\rho)$ through a set of basis matrices, thereby

avoiding direct specification of the correlation structure. Let $M_1, M_2, \dots, M_s$ be a set of basis matrices and $a_1, a_2, \dots, a_s$ be coefficients satisfying:

$$R^{-1}(\rho) \approx a_1 M_1 + a_2 M_2 + \dots + a_s M_s. \quad (9)$$

Substituting (9) into (8), we obtain:

$$\sum_{i=1}^n g^{(1)}(Z_i, \beta)^T Q_{i2} A_i^{-\frac{1}{2}} (a_1 M_1 + \dots + a_s M_s) A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta)) = 0. \quad (10)$$

Define the extended score function as follows:

$$G_n(\beta) = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) \triangleq \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} g^{(1)}(Z_i, \beta)^T Q_{i2} A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta)) \\ g^{(1)}(Z_i, \beta)^T Q_{i2} A_i^{-\frac{1}{2}} M_2 A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta)) \\ \vdots \\ g^{(1)}(Z_i, \beta)^T Q_{i2} A_i^{-\frac{1}{2}} M_s A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta)) \end{pmatrix}. \quad (11)$$

Thus, the QIF for β can be defined as:

$$M_n(\beta) = G_n^T(\beta) \Phi_n^{-1}(\beta) G_n(\beta), \quad (12)$$

where $\Phi_n(\beta) = \frac{1}{n} \sum_{i=1}^n \eta_i(\beta) \eta_i^T(\beta)$, in this case, the estimator $\hat{\beta}$ of β is obtained by minimizing the objective function $M_n(\beta)$:

$$\hat{\beta} = \arg \min_{\beta} M_n(\beta). \quad (13)$$

2.2. Estimation of the Coefficient Functions

The QIF effectively handles the correlation in longitudinal data and improves estimation efficiency by constructing a set of estimating equations. Therefore, after obtaining the initial estimates of the parameters β , we also use the QIF method to estimate the coefficient functions $\alpha_k(u)$, $k = 1, 2, \dots, p$.

Substitute the estimator of into Equation (5), resulting in:

$$\tilde{Y}_i \approx W_i \gamma + \varepsilon_i, i = 1, 2, \dots, n, \quad (14)$$

where $\tilde{Y}_i = Y_i - g_i(Z_i, \hat{\beta})$, assuming that the covariance matrix of $\tilde{Y}_i$ is Ψ_i , and its structure is expressed following Liang[6] as $\Psi_i = A_i^{\frac{1}{2}} R(\rho) A_i^{\frac{1}{2}}$, where A_i is a diagonal matrix, $R(\rho)$ is a working correlation matrix, and ρ is a correlation parameter. Assuming V_i is known, construct the following estimating equation:

$$\sum_{i=1}^n W_i^T \Psi_i^{-1} (\tilde{Y}_i - W_i \gamma) = 0. \quad (15)$$

In practical applications, V_i is usually unknown. Based on this, we still adopt the QIF method and use $M_1, M_2, \dots, M_s$ to approximate R_i^{-1} , then we have:

$$\sum_{i=1}^n W_i^T A_i^{-\frac{1}{2}} (a_1 M_1 + a_2 M_2 + \dots + a_s M_s) A_i^{-\frac{1}{2}} (Y_i - W_i \gamma) = 0. \quad (16)$$

Define the extended score function as follows:

$$G_n(\gamma) = \frac{1}{n} \sum_{i=1}^n \varphi_i(\gamma) \triangleq \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} W_i^T A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (Y_i - W_i \gamma) \\ W_i^T A_i^{-\frac{1}{2}} M_2 A_i^{-\frac{1}{2}} (Y_i - W_i \gamma) \\ \vdots \\ W_i^T A_i^{-\frac{1}{2}} M_s A_i^{-\frac{1}{2}} (Y_i - W_i \gamma) \end{pmatrix}. \quad (17)$$

Thus, the QIF for γ can be defined as:

$$Q_n(\gamma) = G_n^T(\gamma) C_n^{-1}(\gamma) G_n(\gamma), \quad (18)$$

where $C_n(\gamma) = \frac{1}{n} \sum_{i=1}^n \varphi_i(\gamma) \varphi_i^T(\gamma)$. In this case, the estimate $\hat{\gamma}$ of γ can be obtained by minimizing the objective function $Q_n(\gamma)$:

$$\hat{\gamma} = \arg \min Q_n(\gamma). \quad (19)$$

Thus, the estimate of the coefficient function $\alpha_k(u)$ can be expressed as:

$$\hat{\alpha}_k(u) \approx B(u)^T \hat{\gamma}_k, \quad k = 1, 2, \dots, p. \quad (20)$$

where $\hat{\gamma}_k$ is the component corresponding to the k-th coefficient function in $\hat{\gamma}$.

3. Main Conclusions

This section studies the asymptotic properties of the estimators $\hat{\beta}$ and $\hat{\alpha}_k(u)$, $k = 1, 2, \dots, p$, assuming that β_0 and $\alpha_0(\cdot)$ are the true values of β and $\alpha(\cdot)$ respectively, γ_0 is the true value of γ , and $\alpha_{k0}(\cdot)$ and γ_{k0} are the k-th components of $\alpha_0(\cdot)$ and γ respectively. First, we present some common regularity conditions in longitudinal data analysis:

(C1) The random variable U has a bounded support Ω , and its probability density function $f_u(\cdot)$ is positive and has continuous second-order derivatives.

(C2) The varying coefficient function $\alpha_1(u), \dots, \alpha_p(u)$ is continuously differentiable of order r on $[0, 1]$, where $r > 2$.

(C3) For any Z , $g(Z, \beta)$ is continuous with respect to β , and $g(Z, \beta)$ has continuous partial derivatives of order r .

(C4) $\sup_i \|E(\varepsilon_i \varepsilon_i^T)\| < \infty$ holds, and there exists some $\delta > 0$ satisfying $E(\|\varepsilon_i\|^{2+\delta}) < \infty$.

(C5) The covariates X_{ij} and Z_{ij} are assumed to satisfy the following conditions: $\sup_{ij} E\{\|X_{ij}\|^4\} < \infty$, $\sup_{ij} E\{\|Z_{ij}\|^4\} < \infty$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n_i$.

(C6) Let $t_1, \dots, t_k$ be interior nodes on $[0, 1]$. Furthermore, let $t_0 = 0$, $t_{k+1} = 1$, $\xi_i = t_i - t_{i-1}$, then there exists a constant C_0 such that: $\frac{\max\{\xi_i\}}{\min\{\xi_i\}} \leq C_0$, $\max\{|\xi_{i+1} - \xi_i|\} = o(K^{-1})$

(C7) Define $P_{ik} = Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T$, then we have:

$$\frac{1}{n} \sum_{i=1}^n E\{g^{(1)}(Z_i, \beta)^T P_{ik} g^{(1)}(Z_i, \beta)\} \xrightarrow{P} D_k, \quad k = 1, 2, \dots, s.$$

$$\frac{1}{n} \sum_{i=1}^n E\{g^{(1)}(Z_i, \beta)^T P_{il}^T V_i P_{il} g^{(1)}(Z_i, \beta)\} \xrightarrow{P} \Omega_{lk}, \quad l, k = 1, 2, \dots, s.$$

Among them, D_k and Ω_{lk} are constant matrices, and $\xrightarrow{P}$ denotes convergence in probability. Define $\Omega_0 = (\Omega_{lk})_{ps \times ps}$ and $\Gamma_0 = (D_1^T, \dots, D_s^T)^T$, additionally, assume that Ω_0 and $\Gamma_0^T \Omega_0^{-1} \Gamma_0$ are both invertible.

It is noted that, conditions (C1) to (C5) are common conditions in varying coefficient partially non-linear models components, Condition (C6) indicates that $t_0, \dots, t_{k+1}$ is a uniform partition sequence over the interval $[0, 1]$, and Condition (C7) is used for subsequent proofs.

Theorem 1. Under conditions (C1) to (C7), and when $n \rightarrow \infty$, it follows that:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{L} N(0, \Sigma),$$

where the matrix $\Sigma = (\Gamma_0^T \Omega_0^{-1} \Gamma_0)^{-1}$, $\xrightarrow{L}$ denotes convergence in distribution.

Theorem 2. Under conditions (C1) to (C7), the number of nodes $K = O\left(n^{\frac{1}{2r+1}}\right)$, and when $n \rightarrow \infty$, it follows that:

$$\|\hat{\alpha}_k(\cdot) - \alpha_{k0}(\cdot)\| = O_p\left(n^{-\frac{r}{2r+1}}\right), \quad k = 1, 2, \dots, p.$$

where $\|\cdot\|$ denotes the L_2 norm of the function.

4. Simulation Study

This section evaluates the finite-sample performance of the proposed orthogonality estimation method based on QR decomposition and quadratic inference functions in varying coefficient partially nonlinear models through a Monte Carlo simulation study. Consider the following model:

$$Y = X^T \alpha(U) + g(Z, \beta) + \varepsilon,$$

among them, the covariates $X, Z = (Z_1, Z_2)^T$ both follow normal distributions, the nonlinear function $g(Z, \beta)$ is defined as $g(Z, \beta) = \beta_1 Z_1 + \beta_2 \exp(Z_2)$, with the parameter vector $\beta = (\beta_1, \beta_2)^T = (0.8, -0.5)^T$. Additionally, $U \sim U(0, 1)$, the coefficient function $\alpha(U) = \sin(2\pi U)$, the error term ε follows an AR(1) process, and its structure is: $\varepsilon_{i,j} = 0.5\varepsilon_{i,j-1} + v_{i,j}$, where $v_{i,j} \sim N(0, 0.5\sqrt{1 - 0.5^2})$ to ensure the error variance is 0.5.

The sample size is set to $n = 100, 150, 200$, for the i -th subject, the number of repeated measurements is $m_i \equiv 8$, and 1000 simulation runs are conducted for each case. The method combining QR decomposition and QIF proposed in this paper (OQIF) is compared with the profile nonlinear least squares method (PNLS) proposed by Li and Mei [1]. After sorting out, the following Tables 1 and 2 are obtained.

Table 1. The bias and standard deviation of $\hat{\beta}_1$ and $\hat{\beta}_2$ measured by different methods

Method	Parameter	$n = 100$		$n = 150$		$n = 200$	
		Bias	SD	Bias	SD	Bias	SD
OQIF	$\hat{\beta}_1$	0.00257	0.02012	0.00123	0.01890	0.00079	0.01568
	$\hat{\beta}_2$	-0.00390	0.03890	-0.00235	0.03568	-0.00157	0.02946
PNLS	$\hat{\beta}_1$	0.01457	0.03012	0.01235	0.02789	0.00988	0.02346
	$\hat{\beta}_2$	-0.02789	0.05890	-0.02457	0.05235	-0.02012	0.04457

It can be seen from the results in Table 1 that as the sample size increases, the bias and standard deviation of both methods decrease. However, the bias and standard deviation of the OQIF method are smaller, indicating more accurate estimation. Although a larger sample size improves the estimation accuracy of both methods, the OQIF method performs better in bias control.

The results in Table 2 show that the average length of the confidence interval of the OQIF method is significantly shorter than that of the PNLS method, demonstrating higher estimation efficiency, the coverage rate of the OQIF method is closer to the ideal 95%, while although the coverage rate of the PNLS method has improved, it is still below 95%.

Table 2. Confidence Interval Length and Coverage Probability Comparison

Method	Parameter	n=100		n=150		n=200	
		Length	Coverage	Length	Coverage	Length	Coverage
OQIF	$\hat{\beta}_1$	0.07671	0.943	0.07190	0.947	0.05971	0.951
	$\hat{\beta}_2$	0.14853	0.937	0.13551	0.941	0.11329	0.948
PNLS	$\hat{\beta}_1$	0.11329	0.911	0.10065	0.918	0.08541	0.925
	$\hat{\beta}_2$	0.20995	0.899	0.18773	0.905	0.15946	0.912

In conclusion, when the sample size increases from 100 to 200, the OQIF method demonstrates higher estimation accuracy and stronger robustness across all evaluation indicators. These trend analyses further confirm the theoretical advantages and practical application value of the OQIF method in handling nonlinear longitudinal data.

We further conducted 1000 simulations and plotted boxplots of the 1000 RMSE values for parameters $\hat{\beta}_1$ and $\hat{\beta}_2$, as shown in Figure 1 and 2. From the figures, we can observe the following: the RMSE values of $\hat{\beta}_1$ and $\hat{\beta}_2$ for both methods decrease as the sample size increases, however, the OQIF method already exhibits good performance with a small sample size (n=100), while the PNLS method requires a larger sample size to achieve similar accuracy. The boxplots of the OQIF method are more symmetric and compact, indicating that the distribution of its estimators is closer to the normal distribution and has better statistical properties. Additionally, the OQIF method has significantly fewer outliers than the PNLS method, demonstrating stronger robustness against abnormal data. This suggests that the overall performance of the OQIF method in parameter estimation is superior to that of the PNLS method, with more pronounced advantages especially in cases of finite samples.

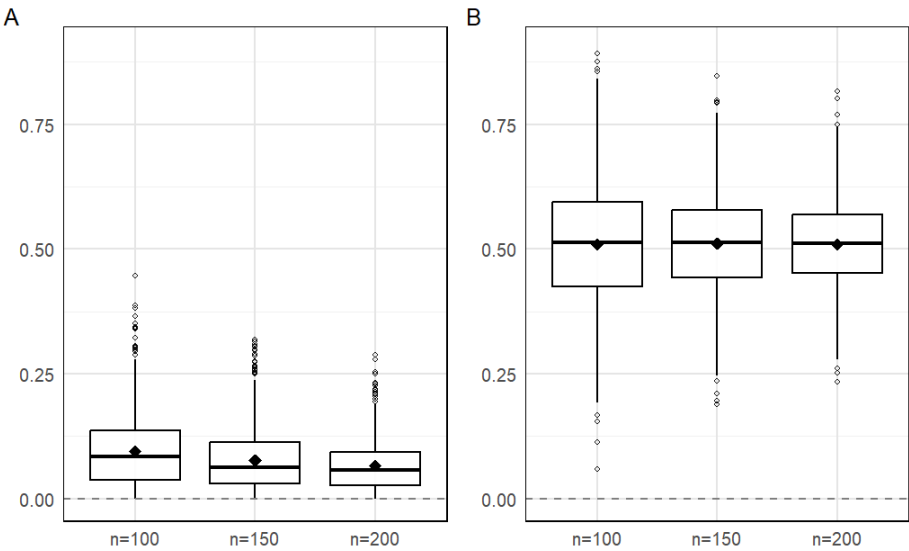


Figure 1. Box plot of 1000 RMSE values for $\hat{\beta}_1$ under the OQIF (A) and PNLS (B) methods

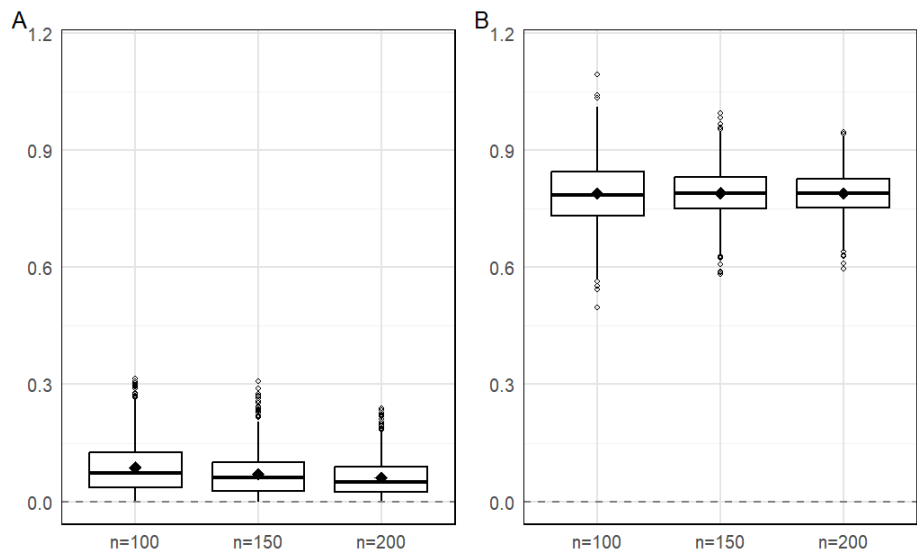


Figure 2. Box plot of 1000 RMSE values for $\hat{\beta}_2$ under the OQIF (A) and PNLs (B) methods

Next, we estimate the coefficient function $\hat{\alpha}(U)$, simulate 1000 times and draw the box plot of the RMSE of $\hat{\alpha}(U)$ under different samples, resulting in Figure 3 below.

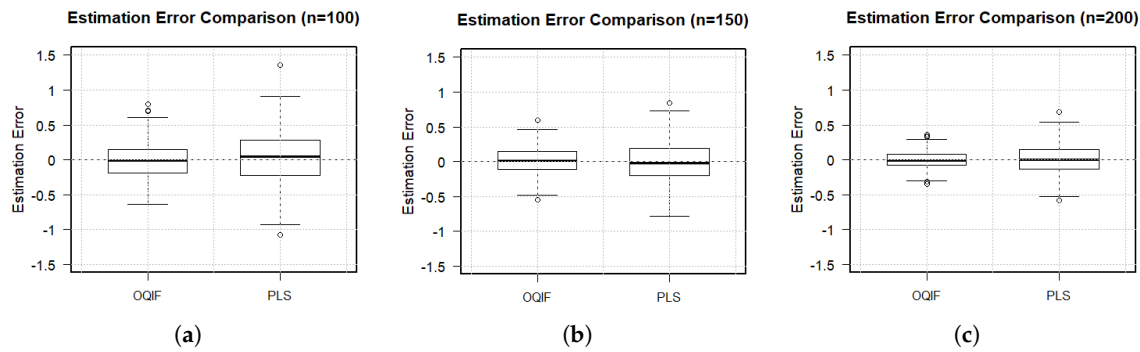


Figure 3. Box plot of 1000 RMSE for $\hat{\alpha}(U)$ with different sample sizes

As can be seen from the above boxplots: with the increase in sample size, the error distributions of both the OQIF method and the PNLs method become more concentrated, but the error of the OQIF method is significantly smaller than that of the PNLs method. This further verifies the superiority of the OQIF method in the varying coefficient partially nonlinear model.

Author Contributions: Conceptualization, J.G., X.Z. and C.W.; methodology, J.G., X.Z. and C.W.; software, J.G. and C.W.; validation, J.G., X.Z. and C.W.; data curation, J.G. and X.Z.; writing—original draft preparation, J.G.; writing—review and editing, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Shandong Province (Grant No.ZR2022MA065).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

QIF	Quadratic inference function
OQIF	The combination of QR decomposition and QIF
PNLS	Profile nonlinear least squares

Appendix A

To prove Theorems 1 and 2, we first present the following lemmas:

Lemma A1. Let c be a positive constant, and let $\xi_i, i = 1, 2, \dots, n$, be a sequence of independent and identically distributed random variables. If it satisfies conditions $E(\xi_i) = 0$ and $E(\xi_i^2) < c < \infty$, then we have

$$\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \xi_i \right| = O_p(\sqrt{n} \log n).$$

Proof. The result has been proven in Lemma A.2 of Zhao and Xue [17]. $\square$

Lemma A2. Assume that the conditions (C1) to (C7) hold, when $n \rightarrow \infty$, we have:

$$\sqrt{n}G_n(\beta_0) \xrightarrow{L} N(0, \Omega_0),$$

where Ω_0 is defined in Condition (C7).

Proof. Let $R(U_i) = (X_{i1}^T r(U_{i1}), \dots, X_{in_i}^T r(U_{in_i}))^T$, $r(u) = (r_1(u), \dots, r_p(u))^T$, $r_k(u) = \alpha_k(u) - B(u)^T \gamma_k$, combining with $Q_{i2}^T W_i = 0$, we can obtain:

$$\begin{aligned} Q_{i2}^T Y_i &= Q_{i2}^T g(Z_i, \beta_0) + Q_{i2}^T W_i \gamma_0 + Q_{i2}^T R(U_i) + Q_{i2}^T \varepsilon_i \\ &= Q_{i2}^T g(Z_i, \beta_0) + Q_{i2}^T R(U_i) + Q_{i2}^T \varepsilon_i. \end{aligned} \quad (A1)$$

Let $G_{n,k}(\beta)$ denote the k -th component of $G_n(\beta)$, combining with Equation (11), we obtain:

$$\begin{aligned} \sqrt{n}G_{n,k}(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta_0)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T R(U_i) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T \varepsilon_i. \end{aligned} \quad (A2)$$

Conclusion $\|R(U_i)\| = O(K^{-r})$ can be derived from Conditions (C2), (C5), and Corollary 6.21 of Schumaker [10], further combining with Lemma A1, we obtain:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T R(U_i) = O_p(n^{-\frac{1}{2}} n^{\frac{1}{2}} \log n K^{-r}) = o_p(1). \quad (A3)$$

Let $\zeta_{ik} = g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T (Y_i - g(Z_i, \beta_0))$, combining with $E(\varepsilon_i | X_i, Z_i) = 0$ and $\text{cov}(\varepsilon_i | X_i, Z_i) = V_i$, we can thus conclude $E(\zeta_{ik} | X_i, Z_i) = 0$ and $\text{cov}(\zeta_{ik} | X_i, Z_i) = g^{(1)}(Z_i, \beta_0)^T P_{ik} V_i P_{ik}^T g^{(1)}(Z_i, \beta_0)$, where P_{ik} is defined in Condition (C7). We further let $\zeta_i = (\zeta_{i1}^T, \zeta_{i2}^T, \dots, \zeta_{is}^T)^T$, it follows that $E(\zeta_i | X_i, Z_i) = 0$ and

$$\text{cov}(\zeta_i | X_i, Z_i) = \begin{pmatrix} N_{i11} & \cdots & N_{i1s} \\ \vdots & \ddots & \vdots \\ N_{is1} & \cdots & N_{iss} \end{pmatrix}, \quad (A4)$$

where $N_{ilk} = g^{(1)}(Z_i, \beta_0)^T P_{il} V_i P_{ik}^T g^{(1)}(Z_i, \beta_0)$, $l, k = 1, 2, \dots, s$. combining with Condition (C7) and the Law of Large Numbers, we conclude that:

$$\frac{1}{n} \sum_{i=1}^n N_{ilk} \xrightarrow{P} \Omega_{lk}. \quad (\text{A5})$$

Furthermore, for any constant vector $a \in R^{sp}$ that satisfying condition $a^T a = 1$, we have $E\{a^T \zeta_i\} = 0$ and $\sup_i E\|a^T \zeta_i\|^{2+\delta} \leq \|a\|^{2+\delta} \sup_i E\|\zeta_i\|^{2+\delta} \leq C \sup_i E\|\varepsilon_i\|^{2+\delta} < \infty$, where C is a positive constant. Therefore, $a^T \zeta_i$ satisfies the Lyapunov condition of the Central Limit Theorem. Thus, we obtain:

$\frac{\sum_{i=1}^n a^T \zeta_i}{(a^T \sum_{i=1}^n \text{cov}(\zeta_i) a)^{\frac{1}{2}}} \xrightarrow{L} N(0, 1)$. Further combining with Equations (A4) and (A5), we can obtain:

$$\frac{1}{n} \sum_{i=1}^n \text{cov}(\zeta_i) \xrightarrow{P} \Omega_0. \quad (\text{A6})$$

Further combining with Equations (A2), (A3), and the Central Limit Theorem, we can obtain:

$$\sqrt{n} G_n(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i + o_p(1) \xrightarrow{L} N(0, \Omega_0). \quad (\text{A7})$$

The proof of Lemma A2 is completed. $\square$

Lemma A3. Assume that the conditions (C1) to (C7) hold, we have $\dot{G}_n(\beta_0) \xrightarrow{P} -\Gamma_0$ and $\Omega_n(\beta_0) \xrightarrow{P} \Omega_0$, where Γ_0 and Ω_0 are defined in Condition (C7), and $\dot{G}_n(\beta_0)$ is the first-order derivative of $G_n(\beta_0)$ with respect to β_0 .

Proof. After simple calculations, we obtain:

$$\dot{G}_n(\beta_0) = -\frac{1}{n} \sum_{i=1}^n \begin{pmatrix} g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} Q_{i2}^T g^{(1)}(Z_i, \beta_0) \\ \vdots \\ g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_s A_i^{-\frac{1}{2}} Q_{i2}^T g^{(1)}(Z_i, \beta_0) \end{pmatrix}. \quad (\text{A8})$$

Combining with Condition (C7) and the Law of Large Numbers, we have that:

$$\frac{1}{n} \sum_{i=1}^n g^{(1)}(Z_i, \beta_0)^T Q_{i2} A_i^{-\frac{1}{2}} M_k A_i^{-\frac{1}{2}} Q_{i2}^T g^{(1)}(Z_i, \beta_0) \xrightarrow{P} D_k, \quad k = 1, 2, \dots, s. \quad (\text{A9})$$

Further combining with Equation (A8), we obtain: $\dot{G}_n(\beta_0) \xrightarrow{P} -\Gamma_0$. The proof of Lemma 2 shows that:

$$\Omega_n(\beta_0) = \frac{1}{n} \sum_{i=1}^n \zeta_i \zeta_i^T + o_p(1) = \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \zeta_{i1} \zeta_{i1}^T & \cdots & \zeta_{i1} \zeta_{is}^T \\ \vdots & \ddots & \vdots \\ \zeta_{is} \zeta_{i1}^T & \cdots & \zeta_{is} \zeta_{is}^T \end{pmatrix} + o_p(1). \quad (\text{A10})$$

Similar to the proof of Equation (A5), we can obtain:

$$\frac{1}{n} \sum_{i=1}^n \zeta_{il} \zeta_{ik}^T \xrightarrow{P} \Omega_{lk}. \quad (\text{A11})$$

Therefore, $\Omega_n(\beta_0) \xrightarrow{P} \Omega_0$, The proof of Lemma 3 is completed. $\square$

Lemma A4. Assume that the conditions (C1) to (C7) hold, we obtain:

$$\left\| \dot{M}_n(\beta_0) - 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) \right\| = O_p(n^{-1}),$$

and

$$\left\| \ddot{M}_n(\beta_0) - 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{G}_n(\beta_0) \right\| = o_p(1).$$

Where $\dot{M}_n(\beta_0)$ and $\ddot{M}_n(\beta_0)$ are the first-order derivative and the second-order derivative of $M_n(\beta_0)$, respectively.

Proof. According to the definition of $M_n(\beta)$, we can obtain the following result through calculation:

$$\dot{M}_n(\beta_0) = 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) + G_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0). \quad (\text{A12})$$

It follows from Lemma A2 that: $G_n(\beta_0) = O_p(n^{-\frac{1}{2}})$. Further, from Condition (C7) and Lemma A3, we can obtain:

$$G_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) = O_p(n^{-1}). \quad (\text{A13})$$

Combining the Equations (A12) and (A13), we can obtain:

$$\dot{M}_n(\beta_0) = 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) + O_p(n^{-1}). \quad (\text{A14})$$

Similarly, we can obtain the following through calculation:

$$\ddot{M}_n(\beta_0) = 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{G}_n(\beta_0) + R_n, \quad (\text{A15})$$

where,

$$\begin{aligned} R_n &= 2\ddot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) \\ &\quad - 4\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) \\ &\quad + 2G_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\dot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)\dot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) \\ &\quad - G_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\ddot{\Omega}_n(\beta_0)\Omega_n^{-1}(\beta_0)G_n(\beta_0) \\ &= I_1 + I_2 + I_3 + I_4. \end{aligned} \quad (\text{A16})$$

Combining Lemma A2 and Lemma A3, we can obtain the following: I_1 and I_2 are both $O_p(n^{-\frac{1}{2}})$, while I_3 and I_4 are both $O_p(n^{-1})$, thus, $R_n = o_p(1)$. The proof of Lemma A4 is completed. $\square$

Proof of Theorem 1. Noting that $\hat{\beta} = \arg \min_{\beta} M_n(\beta)$, thus $\dot{M}_n(\hat{\beta}) = 0$, according to the Lagrange Mean Value Theorem, we have:

$$\dot{M}_n(\beta_0) + \ddot{M}_n(\beta^*)(\hat{\beta} - \beta_0) = 0, \quad (\text{A17})$$

where β^* lies between $\hat{\beta}$ and β_0 , thus, we have:

$$\sqrt{n}(\hat{\beta} - \beta_0) = \sqrt{n}\ddot{M}_n^{-1}(\beta^*)\dot{M}_n(\beta_0). \quad (\text{A18})$$

Combining Lemma A3 and Lemma A4, we can obtain:

$$\ddot{M}_n^{-1}(\beta^*) = \left[2\dot{G}_n^T(\beta^*)\Omega_n^{-1}(\beta^*)\dot{G}_n(\beta^*) + o_p(1) \right]^{-1} = \frac{1}{2} \left(\Gamma_0^T \Omega_0^{-1} \Gamma_0 \right)^{-1} + o_p(1), \quad (\text{A19})$$

and

$$\sqrt{n}\dot{M}_n(\beta_0) = 2\dot{G}_n^T(\beta_0)\Omega_n^{-1}(\beta_0)\sqrt{n}G_n(\beta_0) + o_p(1) = 2\Gamma_0\Omega_0^{-1}\sqrt{n}G_n(\beta_0) + o_p(1). \quad (\text{A20})$$

Combining Equations (A18) to (A20) and Slutsky's Theorem, we can obtain:

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta_0) &= \left[\frac{1}{2} \left(\Gamma_0^T \Omega_0^{-1} \Gamma_0 \right)^{-1} + o_p(1) \right] \left[2\Gamma_0 \Omega_0^{-1} \sqrt{n} G_n(\beta_0) + o_p(1) \right] \\ &= \left(\Gamma_0^T \Omega_0^{-1} \Gamma_0 \right)^{-1} \Gamma_0 \Omega_0^{-1} \sqrt{n} G_n(\beta_0) + o_p(1) \\ &\xrightarrow{L} N(0, \Sigma),\end{aligned}\tag{A21}$$

where, $\Sigma = \left(\Gamma_0^T \Omega_0^{-1} \Gamma_0 \right)^{-1}$, the proof of Theorem 1 is completed. $\square$

Proof of Theorem 2. Let $\tau = n^{-\frac{r}{2r+1}}$, $\gamma = \gamma_0 + \tau s$. That is, we need to prove that for any given $\varepsilon > 0$, there exists a constant C such that the following holds:

$$P\left\{ \sup_{\|s\|=C} Q_n(\gamma) > Q_n(\gamma_0) \right\} < 1 - \varepsilon.\tag{A22}$$

Let $\Delta(\gamma) = nK^{-1}[Q_n(\gamma) - Q_n(\gamma_0)]$. By Taylor's Formula, we can obtain:

$$\begin{aligned}\Delta(\gamma) &= nK^{-1}[Q_n(\gamma_0 + \tau s) - Q_n(\gamma_0)] \\ &= nK^{-1}\tau s^T Q_n'(\gamma_0) + \frac{1}{2}nK^{-1}\tau^2 s^T Q_n''(\gamma_0)s + \frac{1}{6}nK^{-1}\tau^3 \left(\frac{\partial}{\partial \gamma} \{s^T Q_n''(\gamma^*)s\} \right)^T s \\ &= J_1 + J_2 + J_3,\end{aligned}\tag{A23}$$

where γ^* lies between γ and γ_0 . Noting that $K = O\left(n^{\frac{1}{2r+1}}\right)$, we can obtain the following through the assumption conditions and some calculations: $J_1 = O_p(n\tau K^{-1})\|s\| = o_p(1)\|s\|$, $J_2 = O_p(n\tau^2 K^{-1})\|s\|^2 = O_p(1)\|s\|^2$, $J_3 = O_p(n\tau^3 K^{-1})\|s\|^3 = o_p(1)\|s\|^3$. Thus, there exists a sufficiently large constant C such that when $\|s\| = C$, J_2 can control J_1 and J_3 ; therefore, Equation (A8) holds, and there further exists a maximum point $\hat{\gamma}$ satisfying:

$$\|\hat{\gamma} - \gamma_0\| = O_p(\tau) = O_p(n^{-\frac{r}{2r+1}}),\tag{A24}$$

then we can obtain:

$$\|\hat{\alpha}_k(\cdot) - \hat{\alpha}_{k0}(\cdot)\| = O_p(n^{-\frac{r}{2r+1}}), \quad k = 1, 2, \dots, p.\tag{A25}$$

The proof of Theorem 2 is completed. $\square$

References

1. Li, T.; Mei, C. Estimation and inference for varying coefficient partially nonlinear models. *Journal of Statistical Planning and Inference* **2013**, *143*, 2023–2037.
2. Yu, P.; Zhu, Z.; Shi, J.; Ai, X. Robust estimation for partial functional linear regression model based on modal regression. *Journal of Systems Science and Complexity* **2020**, *33*, 527–544.
3. Yan, L.; Tan, X.y.; Chen, X. Empirical likelihood for partially linear errors-in-variables models with longitudinal data. *Acta Mathematicae Applicatae Sinica, English Series* **2022**, *38*, 664–683.
4. Xiao, Y.; Liang, L. Robust estimation and variable selection for varying-coefficient partially nonlinear models based on modal regression. *Journal of the Korean Statistical Society* **2022**, *51*, 692–715.
5. Zhao, P.; Zhou, X.; Wang, X.; Huang, X. A new orthogonality empirical likelihood for varying coefficient partially linear instrumental variable models with longitudinal data. *Communications in Statistics-Simulation and Computation* **2020**, *49*, 3328–3344.
6. Liang, K.Y.; Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, *73*, 13–22.
7. Xu, H.X.; Fan, G.L.; Liang, H.Y. Quantile regression for varying-coefficient partially nonlinear models with randomly truncated data. *Statistical Papers* **2024**, *65*, 2567–2604.

8. Qu, A.; Lindsay, B.G.; Li, B. Improving generalised estimating equations using quadratic inference functions. *Biometrika* **2000**, *87*, 823–836.
9. Bai, Y.; Fung, W.K.; Zhu, Z.Y. Penalized quadratic inference functions for single-index models with longitudinal data. *Journal of Multivariate Analysis* **2009**, *100*, 152–161.
10. Schumaker, L. Spline Functions: Basic Theory. Wiley **1981**.
11. Wei, Y.; Wang, Q.; Liu, W. Model averaging for linear models with responses missing at random. *Annals of the Institute of Statistical Mathematics* **2021**, *73*, 535–553.
12. Jiang, Y.; Ji, Q.; Xie, B. Robust estimation for the varying coefficient partially nonlinear models. *Journal of Computational and Applied Mathematics* **2017**, *326*, 31–43.
13. Xiao, Y.T.; Chen, Z.S. Bias-corrected estimations in varying-coefficient partially nonlinear models with measurement error in the nonparametric part. *Journal of Applied Statistics* **2018**, *45*, 586–603.
14. Xu, H.X.; Fan, G.L.; Liang, H.Y. Quantile regression for varying-coefficient partially nonlinear models with randomly truncated data. *Statistical Papers* **2024**, *65*, 2567–2604.
15. Qian, Y.; Huang, Z. Statistical inference for a varying-coefficient partially nonlinear model with measurement errors. *Statistical Methodology* **2016**, *32*, 122–130.
16. Wang, X.; Zhao, P.; Du, H. Statistical inferences for varying coefficient partially non linear model with missing covariates. *Communications in Statistics-Theory and Methods* **2021**, *50*, 2599–2618.
17. Zhao, P.; Xue, L. Empirical likelihood inferences for semiparametric varying-coefficient partially linear errors-in-variables models with longitudinal data. *Journal of Nonparametric Statistics* **2009**, *21*, 907–923.

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.