

Article

Empirical-Likelihood-Based Inference for Partially Linear Models

Haiyan Su ^{1,*} and Linlin Chen ²¹ School of Computing, Montclair State University, Montclair, NJ 07043, USA² Department of Mathematics and Statistics, Rochester Institute of Technology, Rochester, NY 14632, USA; lxcsma@rit.edu

* Correspondence: suh@montclair.edu

Abstract: Partially linear models find extensive application in biometrics, econometrics, social sciences, and various other fields due to their versatility in accommodating both parametric and nonparametric elements. This study aims to establish statistical inference for the parametric component effects within these models, employing a nonparametric empirical likelihood approach. The proposed method involves a projection step to eliminate the nuisance nonparametric component and utilizes an empirical-likelihood-based technique, along with the Bartlett correction, to enhance the coverage probability of the confidence interval for the parameter of interest. This method demonstrates robustness in handling normally and non-normally distributed errors. The proposed empirical likelihood ratio statistic converges to a limiting chi-square distribution under certain regulations. Simulation studies demonstrate that this method provides better inference in terms of coverage probabilities compared to the conventional normal-approximation-based method. The proposed method is illustrated by analyzing the Boston housing data from a real study.

Keywords: partially linear models; empirical likelihood; confidence interval**MSC:** 62G05

Citation: Su, H.; Chen, L. Empirical-Likelihood-Based Inference for Partially Linear Models. *Mathematics* **2024**, *12*, 162. <https://doi.org/10.3390/math12010162>

Academic Editor: Joseph Ngatchou-Wandji

Received: 18 November 2023

Revised: 31 December 2023

Accepted: 2 January 2024

Published: 4 January 2024



Copyright: © 2024 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

In many practical situations, linear models may not be complex enough to capture the underlying relation between the response variable and some associated covariates, especially when the response variable Y is not linearly related to all the covariates. For example, suppose one is interested in estimating the relationship between an outcome variable Y and vectors of variables X and Z . The researcher is comfortably modeling the linear function in X but hesitates to extend the linearity to Z . One example given by Engle et al. [1] is the effect of temperature on electricity consumption for four cities. They modeled the average monthly electricity consumption as the sum of a smooth function of the monthly temperatures and a linear function of the monthly price of electricity, income and 11 other monthly dummy variables. It is natural to impose linearity on the part of the regression function involving household characteristics and a nonlinear function involving temperature since electricity consumption tends to be higher at extreme temperatures but lower at moderate temperatures. A partially linear model provides a good fit for these types of data because it allows for a regression function that maintains linearity in some variables and also extends the effect of other variables to be nonlinear.

A partially linear regression model is defined as

$$Y_i = X_i\beta + g(Z_i) + \varepsilon_i, i = 1, \dots, n, \quad (1)$$

where the Y_i 's are scalar response variables, $X_i = (x_{i1}, \dots, x_{ip})$ are known p -variate covariate, Z_i is a scalar explanatory variable, and $g(\cdot)$ is the smooth part of the model, which is assumed to represent a smooth unparameterized functional relationship. $\beta = (\beta_1, \dots, \beta_p)$

is a vector of unknown parameters and $\varepsilon_1, \dots, \varepsilon_n$ are independent random errors with mean zero and finite variance σ^2 given the covariates X and Z .

The partial linear model Equation (1) is a semiparametric model since it contains both parametric and nonparametric components. The partially linear model is more flexible to interpret the effect of each linear covariate and allows one to focus on particular variables that can have nonlinear effects. It may be preferable to a completely nonparametric model because of the well-known “curse of dimensionality”. Computationally, partially linear models are remarkably easier than additive models, in which iterative approaches such as a backfitting algorithm [2] or marginal integration [3] are necessary.

Partially linear models are widely used in biometrics, econometrics, social sciences and other fields (see [1,4]), and have been studied extensively for estimating β and $g(\cdot)$. For example, Wahba [5], Engle et al. [1], and Green et al. [6] described penalized spline estimates of β and $g(\cdot)$. Heckman [7] and Rice [8] proposed the polynomial method. Speckman [9] described the kernel method. Chen and Shiao [10] used a smoothing spline method. Chen [11] proposed the projection method. For more discussions about partially linear models, we refer to Härdle et al. [12] for a summary.

In most cases, investigators are more interested in the parameter β and take $g(\cdot)$ as a nuisance parameter [13]. Estimating the confidence interval for the parametric components in partially linear models using a backfitting algorithm or marginal integration can be computationally heavy. Severini and Staniswalis [14] derived the asymptotic properties for their proposed estimators of β and $g(\cdot)$ under mild regularity conditions. These asymptotic properties serve as a foundation for constructing confidence intervals that are asymptotically accurate for the parameters. However, in practice, the finite-sample performance of these confidence intervals may be less satisfactory because of the complex structure of the covariance matrix, requiring estimates to be plugged in for multiple parameters. The linear components in the partially linear models can also be estimated using the generalized additive models [15], but the results depend on the distribution family used in the gam function in R. When a wrong distribution family is chosen, the results could be very biased. Confidence interval for the parametric components in the partially linear models can also be constructed based on the asymptotic normal distribution; however, this may not hold when the normality assumption fails or when the sample size is small.

Empirical likelihood provides a good alternative among the nonparametric methods that can be used to make statistical inference when the normality assumption fails or when the distribution is unspecified. The advantages of empirical likelihood compared to the bootstrap method and the jackknife method arise from it being a nonparametric method of inference based on a data-driven likelihood ratio function. As a combination of a nonparametric method and the likelihood method, on one hand, it does not require any specification of a family of distributions for the data; on the other hand, like parametric likelihood methods, it makes an automatic determination of the shape of confidence regions [16]. This property makes it a serious competitor with other nonparametric methods such as the bootstrap method and the jackknife method. Although empirical likelihood can be a very useful tool for deriving statistical inference, the use of a conventional empirical likelihood method or the profile empirical likelihood has limitations when constructing confidence intervals for each element of a large parameter vector.

Motivated by the above mentioned concerns, this paper develops an empirical-likelihood-based procedure which can be used to make inferences for a large parameter vector β in partially linear models in Equation (1) by incorporating the projection method. The proposed method has two main advantages. First, it does not require distribution assumptions. Second, we provide theoretical justification that the proposed method can be applied to partially linear models, and the computation requirements are relatively straightforward because it does not require an asymptotic variance estimation. After the Bartlett correction, the coverage probability of the confidence interval is improved and better than normal-approximation-based methods in most cases.

The structure of the paper is as follows. Section 2 gives the model formulation of the empirical likelihood for the parameter of interest and the Bartlett correction procedure for the proposed method. Section 3 studies the performance of the proposed methods through simulation studies and illustrates the method by a real study example. Section 4 gives the conclusion. All the proofs are given in the Appendix A.

2. Materials and Methods

2.1. Model Formulation

Since the interest in this paper is in obtaining inference for β only in the partially linear model, the nuisance parameter $g(\cdot)$ needs to be removed first. This is implemented by using the projection principle [17,18]. Y and X need to be first regressed on Z using a nonparametric regression method, where $Y = (Y_1, \dots, Y_n)$, $Z = (Z_1, \dots, Z_n)$, and X is an $n \times p$ covariate matrix. Denote the nonparametric regressions of Y on Z and X on Z by $m_Y(Z)$ and $m_X(Z)$, respectively. Here, without loss of generality, let X be a one-dimensional vector (for a multidimensional vector, $E(X_i|Z)$ can be obtained for each column X_i of X , respectively), and then, the effect of Z on Y and X can be removed by using the regression residual of Y and X given Z . For simplicity of notation, the matrix form of the partially linear model was used here:

$$Y = X\beta + g(Z) + \varepsilon. \tag{2}$$

The first step is to regress Y and X onto Z and obtain the following equation

$$m_Y(Z) = m_X(Z)\beta + g(Z). \tag{3}$$

Then, Equation (3) is subtracted from the original model (2), and the residual model is obtained as follows:

$$Y - m_Y(Z) = \{X - m_X(Z)\}\beta + \varepsilon. \tag{4}$$

Denote $A^{\otimes 2} = AA^T$ and $\tilde{\zeta} = \zeta - m_{\zeta}(Z)$. For example, $\tilde{X} = X - m_X(Z)$. Assuming \tilde{X} has full rank, based on Speckman [9], the estimator of β can then be given by Equation (5) by the least squares method if $m_X(Z)$ and $m_Y(Z)$ are known:

$$\hat{\beta} = \left[\sum_{i=1}^n (X_i - m_X(Z_i))^{\otimes 2} \right]^{-1} \left[\sum_{i=1}^n \{X_i - m_X(Z_i)\} \{Y_i - m_Y(Z_i)\} \right]. \tag{5}$$

The formula above cannot be applied directly since $m_X(Z)$ and $m_Y(Z)$ need to be estimated appropriately. There are lots of methods for estimating $m_X(Z)$ and $m_Y(Z)$, including local constant smoothers [9], higher-order local polynomial estimators [19], kernel methods with varying bandwidths, smoothing and regression splines, etc. Fan and Gijbels [19] showed that within the class of linear estimators which include kernel and spline estimates, the local linear estimates achieve the best possible rates of convergence. Due to these desirable properties, the local linear smoothers was used with fixed bandwidths for estimating the nonparametric regression of Y and X on Z . Let $\hat{m}_X(Z)$ and $\hat{m}_Y(Z)$ be the local linear nonparametric regression estimators for $m_X(Z)$ and $m_Y(Z)$, $K(\cdot)$ be a symmetric density function, h be a suitable bandwidth, and define $K_h(z) = K(z/h)/h$; then, the estimators take the form given by Fan and Gijbels [19]:

$$\hat{m}_X(Z) = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i} \text{ and } \hat{m}_Y(Z) = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}, \tag{6}$$

$$w_i = K_h(Z_i - Z) \{S_{n,2} - (Z_i - Z)S_{n,1}\},$$

where $S_{n,j} = \sum_{i=1}^n K_h(Z_i - Z)(Z_i - Z)^j$. $m_X(Z)$ and $m_Y(Z)$ are then replaced with their corresponding estimates $\hat{m}_X(Z)$, $\hat{m}_Y(Z)$ in the estimating procedure (a Gaussian kernel is an example of kernel function used in the estimating procedure) and the empirical likelihood estimator of β needs to satisfy the following estimating equation:

$$\sum_{i=1}^n \{X_i - \hat{m}_X(Z_i)\}^T [Y_i - \hat{m}_Y(Z_i) - \{X_i - \hat{m}_X(Z_i)\}\beta] = 0. \tag{7}$$

This implies that the estimator for $\hat{\beta}$ can be obtained by

$$\hat{\beta}^* = \left[\sum_{i=1}^n (X_i - \hat{m}_X(Z_i)) \otimes^2 \right]^{-1} \left[\sum_{i=1}^n \{X_i - \hat{m}_X(Z_i)\} \{Y_i - \hat{m}_Y(Z_i)\} \right]. \tag{8}$$

Next, the empirical likelihood principle was applied to construct statistical inference for β . Let p_i be the probability assigned to (X_i, Z_i, Y_i) . The empirical likelihood ratio function for β can be expressed as:

$$\mathcal{R}_n(\beta) = \sup_{p_i} \left\{ \prod_{i=1}^n n p_i \mid \sum_{i=1}^n p_i \{X_i - \hat{m}_X(Z_i)\}^T [Y_i - \hat{m}_Y(Z_i) - \{X_i - \hat{m}_X(Z_i)\}\beta] = 0, \right. \\ \left. p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \tag{9}$$

We establish the asymptotic distribution of $-2 \log\{\mathcal{R}_n(\beta)\}$ under the following assumptions:

Assumption 1. $E(\|X\|^4) < \infty$, $E(\|X\|^2 Y^2) < \infty$, and $E(X^T X)$ is nonsingular; X and Z are correlated.

Assumption 2. The bandwidths used in estimating $m_x(Z)$ and $m_y(Z)$ are of order $n^{-1/5}$.

Assumption 3. The function $K(\cdot)$ is a bounded symmetric density function with compact support and satisfies $\int K(u)du = 1$, $\int K(u)udu = 0$ and $\int u^2 K(u)du = 1$.

Assumption 4. The functions $m_x(Z)$ and $m_y(Z)$ have bounded and continuous second derivatives.

Assumption 5. The density function of Z , $f_z(Z)$ is bounded away from zero and has bounded continuous second derivatives.

Theorem 1. $-2 \log\{\mathcal{R}_n(\beta)\}$ converges to a chi-squared distribution with p degrees of freedom under Assumptions 1–5.

The proof of Theorem 1 is given in the Appendix A. A confidence region for β can be constructed based on Theorem 1 and further adjusted by using the Bartlett correction [20].

When β is a vector (or when X is an $n \times p$ matrix), and we are interested in a subset of the parameter vector β , say the first element β_1 , we can apply the projection method again, i.e., we regress \hat{X}_1 , the first column of \hat{X} , which is $X - \hat{m}_X(Z)$, onto the space of \hat{X}_{-1} , which is the remaining columns of \hat{X} . Similarly, we apply the same projection principle from $\hat{Y} = Y - \hat{m}_Y(Z)$ to \hat{X}_{-1} . Then, we obtain a new residual model, i.e., β_1 should satisfy the estimating equation as follows:

$$\frac{1}{n} \sum_{i=1}^n \{\hat{X}_1 - \hat{E}(\hat{X}_1 | \hat{X}_{-1})\}^T [\hat{Y} - \hat{E}(\hat{Y} | \hat{X}_{-1}) - \{\hat{X}_1 - \hat{E}(\hat{X}_1 | \hat{X}_{-1})\}\beta_1] = 0. \tag{10}$$

Let p_i be the probability assigned to (Y_i, X_i, Z_i) , where p_i could be different from the p_i in Theorem 1. The estimating equation for β_1 can be written as:

$$\mathcal{R}_n(\beta_1) = \sup_{p_i} \left\{ \prod_{i=1}^n n p_i \mid \sum_{i=1}^n p_i \{\hat{X}_1 - \hat{E}(\hat{X}_1 | \hat{X}_{-1})\}^T [\hat{Y} - \hat{E}(\hat{Y} | \hat{X}_{-1}) - \{\hat{X}_1 - \hat{E}(\hat{X}_1 | \hat{X}_{-1})\}\beta_1] = 0, \right. \\ \left. p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}. \tag{11}$$

Theorem 2. $-2 \log\{\mathcal{R}_n(\beta_1)\}$ converges to a chi-squared distribution with 1 degree of freedom under Assumptions 1–5.

The proof of Theorem 2 is given in the Appendix A. Based on Theorem 2, the $100(1 - \alpha)\%$ empirical likelihood confidence interval for β_1 can be obtained by:

$$\{\beta_1 : -2 \log\{\mathcal{R}_n(\beta_1)\} \leq c_\alpha\}. \tag{12}$$

The $100(1 - \alpha)\%$ confidence interval for other components of β can be constructed similarly.

2.2. Bartlett Correction

To further improve the accuracy of the inference, the empirical likelihood ratio may be Bartlett corrected with a higher-order error than the usual error term of order $O(n^{-1})$ [20]. The Bartlett correction can effectively control the coverage error of the confidence interval, providing more accurate estimations and reducing the chance of obtaining intervals that do not contain the true parameter value. The basic idea is to multiply the χ^2 threshold by a constant $(1 + B_c/n)$ instead of 1, where B_c is the Bartlett correction constant. Because it is very difficult to obtain an exact expression for B_c , we give an estimator of $(1 + B_c/n)$ by using the bootstrap procedure, which has successfully been applied in a more complex setting by Chen and Cui [21].

The Bartlett correction of the empirical likelihood confidence interval for a parameter of interest β_1 in a partially linear model in Equation (1) is constructed by the following procedures. The procedures for another component of β , say β_2 , would be similar.

1. First, the nonparametric regression method is used to regress Y and X on the nonparametric component Z . The reduced partial residuals follow a linear model of the form $Y - m_Y(Z) = \{X - m_X(Z)\}\beta + \varepsilon$. We use $\hat{m}_Y(Z)$ and $\hat{m}_X(Z)$ to replace $m_Y(Z)$ and $m_X(Z)$ in the estimating procedure.
2. Then, the first column of \hat{X} (denoting by \hat{X}_1) is regressed on the rest of the columns (denoting by \hat{X}_{-1}). The residual serves as the new fixed covariates of β_1 , and the residual of regressing \hat{Y} on \hat{X}_{-1} serves as the new response variable. The residual model is obtained and given by

$$\{\hat{Y} - E(\hat{Y}|\hat{X}_{-1})\} = \{\hat{X}_1 - E(\hat{X}_1|\hat{X}_{-1})\}\beta + \varepsilon.$$

3. We treat the residual model as the new linear model. The bootstrap procedure of estimating the Bartlett correction factor in the new linear model follows the procedure shown below:
 - (a). Generate bootstrap resamples of size n by sampling with replacement from the sample $\{\hat{Y} - E(\hat{Y}|\hat{X}_{-1})\}_1^n$ and $\{\hat{X}_1 - E(\hat{X}_1|\hat{X}_{-1})\}_1^n$, respectively, after the projection; then, calculate $-2 \log\{\mathcal{R}_n^*(\hat{\beta}_1)\}$ based on the resamples, where $\hat{\beta}_1$ is the global maximum empirical likelihood estimator of β_1 based on the original sample $\{\hat{Y} - E(\hat{Y}|\hat{X}_{-1})\}_1^n$ and $\{\hat{X}_1 - E(\hat{X}_1|\hat{X}_{-1})\}_1^n$.
 - (b). Repeat (a) B times to obtain $-2 \log\{\mathcal{R}_n^{*1}(\hat{\beta}_1)\}, -2 \log\{\mathcal{R}_n^{*2}(\hat{\beta}_1)\}, \dots, -2 \log\{\mathcal{R}_n^{*B}(\hat{\beta}_1)\}$ and $B^{-1} \sum_{b=1}^B -2 \log\{\mathcal{R}_n^{*b}(\hat{\beta}_1)\}$, which is the bootstrap estimator of $E[-2 \log\{\mathcal{R}_n(\hat{\beta}_1)\}]$.

The bootstrap estimator of τ is $B^{-1} \sum_{b=1}^B -2 \log\{\mathcal{R}_n^{*b}(\hat{\beta}_1)\}$. In consequence, the Bartlett corrected confidence region is constructed by

$$CI_\alpha = \{\beta_1 : -2 \log\{R_n(\beta_1)\} \leq \hat{\tau}c_\alpha\}.$$

The Bartlett corrected confidence interval for β_1 is thus constructed.

3. Results

3.1. Simulation Studies

In the simulation studies, we studied the performance of the proposed method in getting the inference of the parameter of interest β in the partially linear model (1). We first simulated Z from a $\text{Unif}(0, 1)$ distribution with sample size n . The true β value was set to be $\beta = (2, 5, 7, 4)$, and we aimed to estimate the first component of β . X was set to be the sum of two matrices Σ_1 and Σ_2 , where Σ_1 was the matrix composed of vectors $1.5\exp(1.5z)$, $5z$, $5\sqrt{z}$, and $3z + z^2$. Σ_2 was the matrix of error terms composed of n samples from the scaled multivariate normal distribution with zero mean and a compound symmetry covariance matrix with diagonal 1 and off-diagonal 0.4; the scale parameter was 0.5. The columns of the X matrix were functions of Z and were thus correlated. The nonparametric component $g(Z)$ took the function $g(Z) = \sin(Z)$. Two cases for the distribution of the error term ε were considered,

Case 1: ε follows a normal distribution with mean 0 and variance $\sigma^2 = 1$.

Case 2: ε follows the scaled log-normal distribution such that ε has mean 0 and variance $\sigma^2 = 1$.

In the simulations, the sample sizes were considered to be 50, 100, and 200. In each simulation, we generated 1000 independent data sets and constructed the 95% confidence interval for each data set. In estimating the nonparametric regression of $m_y(Z)$ and $m_x(Z)$, the direct plug-in method was used to select the bandwidth of a local linear Gaussian kernel regression estimate, as described by Ruppert, Sheather, and Wand [22]. The proposed method was compared with the normal-based method and the generalized additive model method (gam) [15].

Table 1 gives the average results from the 1000 simulations (the endpoints of the confidence intervals were obtained by the medians of the 1000 simulation results, and the confidence interval lengths were computed using the difference of the two endpoints). In Table 1, Est refers to the estimated β_1 value; Norm, Gam, EL, and ELb refer to the normal-based method, the gam function in R, the empirical-likelihood-based method without Bartlett correction, and the empirical-likelihood-based method with Bartlett correction, respectively. Length and coverage probability refer to the respective length and coverage probability of the confidence intervals constructed using the four different methods. It is worth mentioning that each confidence interval based on the normal approximation is symmetric while the confidence interval based on empirical likelihood is not symmetric. In the simulation, the Gaussian distribution was used as the distribution family within the gam function under both error cases.

Table 1. Confidence interval and coverage probability for partially linear models; $\beta = (2, 5, 7, 4)$, $\sigma = 1$, and β_1 is the parameter of interest.

	n	Est	Length				Coverage Probability			
			Norm	Gam	EL	ELb	Norm	Gam	EL	ELb
Norm	50	1.957	1.496	1.498	1.437	1.520	0.935	0.936	0.924	0.945
	100	2.047	1.052	1.085	1.016	1.049	0.963	0.963	0.949	0.953
	200	2.026	0.706	0.704	0.689	0.701	0.946	0.950	0.944	0.950
Non-norm	50	1.975	1.368	1.289	1.278	1.386	0.946	0.934	0.944	0.950
	100	2.05	1.012	1.051	0.980	1.008	0.973	0.962	0.954	0.956
	200	2.027	0.681	0.689	0.668	0.677	0.944	0.940	0.930	0.946

The simulation results from Table 1 indicates that the Bartlett correction indeed improved the statistical inference. The coverage probability was improved after the Bartlett correction, especially when the sample size is small, where the normal approximation method may not be appropriate. When the sample size is small (for example $n = 50$), our proposed method tends to enlarge the confidence interval to have a better coverage probability for the true parameter. In that case, the length of the confidence interval for the Bartlett correction is larger than that of the normal approximation, the gam method,

and the empirical likelihood without Bartlett correction, but the coverage probability is the closest to the nominal level 95%. When the sample size becomes larger, the length of the confidence interval using the proposed method tends to be close to or shorter than the confidence interval of the normal approximation method and yet still has slightly better or equally good coverage probability compared to the normal approximation method and gam method.

3.2. A Real Study Example

The proposed method is illustrated by an application to the Boston housing data set, which was obtained from the StatLib archive and has been extensively used in regression analysis. The data set consists of the median value of owner-occupied homes in 506 US census tracts in the Boston area in 1970, as well as several variables which might explain the variation in housing values. Based on the correlations and multicollinearity analysis, we fit a partially linear model with the variable of interest MEDV (median value of owner-occupied home in USD 1000) linearly related with predictor PTRATIO (pupil–teacher ratio by town), RM (number of rooms per dwelling), and nonlinearly related with variable LSTAT (% lower status of the population). The partially linear model has the following form:

$$MEDV = \beta_0 + \beta_1 PTRATIO + \beta_2 RM + g(LSTAT) + \varepsilon.$$

The proposed method was used to construct the 95% confidence interval for β_1 . The proposed empirical-likelihood-based Bartlett corrected 95% confidence interval for β_1 was (2.375, 4.656), and the normal-based 95% confidence interval was (2.406, 4.502). Both methods indicated a positive linear relationship between PTRATIO and MEDV, with the proposed method's confidence interval slightly wider than the normal-based confidence interval. Based on our simulation results for the coverage probability under a large sample size, the confidence interval obtained from the proposed method was comparable with the normal-based confidence interval and was trustworthy.

4. Discussion

In this paper, an empirical-likelihood-based method to construct the confidence interval for the linear components in partially linear models was proposed. Simulation studies showed that the length of the confidence interval for the proposed empirical likelihood with Bartlett correction method was larger than the normal approximation when the sample size was small, but the coverage probability was the closest to the nominal 95% level. When the sample size was larger, the confidence interval for the proposed empirical likelihood with Bartlett correction method had a slightly shorter length and a similar coverage probability as the normal-based method and gam method, which indicated the confidence interval constructed by the proposed method was more desirable in estimating the parameter of interest. The above findings are mostly true under both normally distributed error and non-normally distributed error terms. This ensures the robustness of our proposed test numerically, which also makes the proposed method a practically useful tool in real studies where we usually do not know the distribution of the data. The trade-off of the proposed method is that it requires more computation than the normal-approximation method.

In summary, this proposed method gives better inference in terms of the length and coverage probabilities of the confidence intervals compared to the normal-approximation-based method. It does not impose any restrictions on the data distribution, and the computations are relatively straightforward for partially linear models. This proposed method is recommended for estimating and constructing confidence intervals for the linear components in partially linear models, particularly when the sample size is small.

Author Contributions: Methodology, H.S.; software, H.S. and L.C.; formal analysis, H.S. and L.C.; writing—original draft, revision, H.S.; writing—review and editing, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The boston housing data used in the paper was obtained from the StatLib archive (<http://lib.stat.cmu.edu/datasets/boston>, accessed on 1 November 2023).

Acknowledgments: The authors would like to thank the editor and three referees for their insightful comments that significantly improved an earlier version of this paper.

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

Appendix A

Proof of Theorem 1. First, we give the following fact, which is used later in the proof of the theorem. Its proof can be shown by Assumptions 2–5:

$$\widehat{m}_x(Z) - m_x(Z) = o_p(n^{-1/4}), \widehat{m}_y(Z) - m_y(Z) = o_p(n^{-1/4}) \tag{A1}$$

From $Y = X\beta + g(Z) + \varepsilon$, we have $Y - m_y(Z) = \{X - m_x(Z)\beta\} + \varepsilon$. In the estimating process, it is rewritten as $\hat{Y} = \hat{X}\beta + \varepsilon$ with the notation that $\widehat{m}_\zeta(T)$ is the local linear kernel regression estimator (for example, the kernel function can be the Gaussian kernel function, the bandwidth h can be determined by using the direct plug-in method by Ruppert, Sheather, and Wand [22] of a local linear Gaussian kernel regression estimate) of $m_\zeta(T)$ and $\hat{\zeta} = \zeta - \widehat{m}_\zeta(T)$.

Let $\Omega_i = \hat{X}_i^\top(\hat{Y}_i - \hat{X}_i\beta)$, $\tilde{\Omega}_i = \tilde{X}_i^\top(\tilde{Y}_i - \tilde{X}_i\beta)$. A standard simplification as in Owen [16] yields

$$p_i = \frac{1}{n(1 + a^\top \tilde{\Omega}_i)}, i = 1, \dots, n,$$

where a is the solution of the equation

$$n^{-1} \sum_{i=1}^n \frac{\tilde{\Omega}_i}{1 + a^\top \tilde{\Omega}_i} = 0.$$

A direct calculation yields

$$\begin{aligned} \tilde{\Omega}_i - \Omega_i &= \tilde{X}_i^\top(\tilde{Y}_i - \tilde{X}_i\beta) - (\hat{X}_i - \tilde{X}_i + \tilde{X}_i)^\top(\hat{Y}_i - \tilde{Y}_i + \tilde{Y}_i - \hat{X}_i\beta + \tilde{X}_i\beta - \tilde{X}_i\beta) \\ &= \tilde{X}_i^\top(\tilde{Y}_i - \hat{Y}_i) + \tilde{X}_i^\top(\hat{X}_i - \tilde{X}_i)\beta - (\hat{X}_i - \tilde{X}_i)^\top(\hat{X}_i - \tilde{Y}_i) \\ &\quad + (\hat{X}_i - \tilde{X}_i)^\top(\hat{X}_i - \tilde{X}_i)\beta - (\hat{X}_i - \tilde{X}_i)^\top(\tilde{Y}_i - \tilde{X}_i\beta) \\ &= o_p(1), \end{aligned} \tag{A2}$$

where $o_p(1)$ is independent of the index i . The above equation is of order $o_p(1)$ because \tilde{X}_i is $O_p(1)$. $\tilde{Y}_i - \hat{Y}_i$ and $\tilde{X}_i - \hat{X}_i$ are of order $o_p(n^{-1/4})$ by (A1).

Using arguments similar to those in the proof of Theorem 3.2 of Owen [16], we have

$$\|a\| = O_p(n^{-1/2}) \text{ and } \max_{1 \leq i \leq n} \|\tilde{\Omega}_i\| = o_p(n^{1/2}).$$

With (A2), we have $\max_{1 \leq i \leq n} \|\Omega_i\| \leq \max_{1 \leq i \leq n} \|\tilde{\Omega}_i\| + o_p(1) = o_p(n^{1/2})$. Using the same argument as those in the proof of Theorem 4 in Liang et al. [23], we have

$$\begin{aligned} -2 \log\{\mathcal{R}_n(\beta)\} &= \sum_{i=1}^n a^\top \tilde{\Omega}_i \tilde{\Omega}_i^\top a + o_p(1) \\ &= \left(n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i \right)^\top \left(n^{-1} \sum_{i=1}^n \tilde{\Omega}_i \tilde{\Omega}_i^\top \right)^{-1} \left(n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i \right) + o_p(1). \end{aligned}$$

Now, we need to show that by replacing $\tilde{\Omega}$ with Ω_i , the above equation still holds. To show that, we first show that $n^{-1/2} \sum_{i=1}^n \tilde{X}_i^\top (\tilde{Y}_i - \hat{Y}_i) = o_p(1)$. Given $\epsilon > 0$ and some constant c , let $a_i = \tilde{Y}_i - \hat{Y}_i = \hat{m}_y(Z_i) - m_y(Z_i)$. We have

$$\begin{aligned} & P\left(n^{-1/2} \left| \sum_{i=1}^n \tilde{X}_i^\top (\tilde{Y}_i - \hat{Y}_i) \right| > \epsilon\right) \\ & \leq P\left(n^{-1/2} \left| \sum_{i=1}^n \tilde{X}_i^\top a_i \right| > \epsilon, |a_i| \leq cn^{-1/4}\right) + P(|a_i| > cn^{-1/4}) \\ & \leq P\left(n^{-1/2} \cdot cn^{-1/4} \left| \sum_{i=1}^n \tilde{X}_i^\top \right| > \epsilon\right) + o_p(1) \\ & = P\left(n^{-1/2} \left| \sum_{i=1}^n \tilde{X}_i^\top \right| > cn^{1/4} \epsilon\right) + o_p(1) \\ & = o_p(1) + o_p(1) \\ & = o_p(1). \end{aligned}$$

In the above equations, $P\left(n^{-1/2} \sum_{i=1}^n |\tilde{X}_i^\top| > cn^{1/4} \epsilon\right) = o_p(1)$ because $n^{-1/2} \sum_{i=1}^n \tilde{X}_i^\top \sim N(0, v(X|Z))$, where $v(X|Z)$ is the covariance matrix of $X - E(X|Z)$. Similarly, we can show

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i^\top (\tilde{X}_i - \hat{X}_i) \beta = o_p(1). \\ & \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - \tilde{X}_i)^\top (\tilde{Y}_i - \hat{Y}_i) \beta = o_p(1). \end{aligned}$$

To show $\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - \tilde{X}_i)^\top (\hat{Y}_i - \tilde{Y}_i) = o_p(1)$, note that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - \tilde{X}_i)^\top (\hat{Y}_i - \tilde{Y}_i) = n^{-1/2} o_p(n^{-1/4}) \cdot o_p(n^{-1/4}) \cdot n = o_p(1). \tag{A3}$$

The first equal sign in (A3) holds because $\sup_{\|Z_i\|} |\hat{m}_w(Z_i) - m_w(Z_i)| = o_p(n^{-1/4})$, where $w = X$ or $w = Y$, so that we can take $o_p(n^{-1/4})$ out of the summation. With the same procedure, we can also show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{X}_i - \tilde{X}_i)^\top (\hat{X}_i - \tilde{X}_i) \beta = o_p(1).$$

These arguments imply that $n^{-1/2} \sum_{i=1}^n \Omega_i$ and $n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i$ asymptotically have the same limiting normal distribution, and $n^{-1} \sum_{i=1}^n \Omega_i \Omega_i^\top$ and $n^{-1} \sum_{i=1}^n \tilde{\Omega}_i \tilde{\Omega}_i^\top$ have the same limiting value. Since

$$\left(n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i\right)^\top \left(n^{-1} \sum_{i=1}^n \tilde{\Omega}_i \tilde{\Omega}_i^\top\right) \left(n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i\right) \sim \chi_p^2,$$

we have

$$\left(n^{-1/2} \sum_{i=1}^n \Omega_i\right)^\top \left(n^{-1} \sum_{i=1}^n \Omega_i \Omega_i^\top\right) \left(n^{-1/2} \sum_{i=1}^n \Omega_i\right) \sim \chi_p^2.$$

The proof is thus complete. \square

Proof of Theorem 2. We continue to use the notations $\tilde{\xi} = \xi - \xi(Z)$, $\hat{\xi} = \xi - \hat{m}_\xi(Z)$ for any random vector ξ . First, denote

$$\Omega_i = \{\hat{X}_{1i} - \hat{E}(\hat{X}_1|\hat{X}_{-1})_i\}^\top[\hat{Y}_i - \hat{E}(\hat{Y}_i|\hat{X}_{-1}) - \{\hat{X}_{1i} - \hat{E}(\hat{X}_1|\hat{X}_{-1})_i\beta_1\}].$$

$$\hat{\Omega}_i = \{\hat{X}_{1i} - E(\hat{X}_1|\hat{X}_{-1})_i\}^\top[\hat{Y}_i - E(\hat{Y}_i|\hat{X}_{-1}) - \{\hat{X}_{1i} - E(\hat{X}_1|\hat{X}_{-1})_i\beta_1\}].$$

$$\tilde{\Omega}_i = \{\tilde{X}_{1i} - E(\tilde{X}_1|\tilde{X}_{-1})_i\}^\top[\tilde{Y}_i - E(\tilde{Y}_i|\tilde{X}_{-1}) - \{\tilde{X}_{1i} - E(\tilde{X}_1|\tilde{X}_{-1})_i\beta_1\}].$$

We first need to show $\Omega_i = \tilde{\Omega}_i + o_p(1)$, and $n^{-1/2} \sum_{i=1}^n \Omega_i$ and $n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i$ asymptotically have the same limiting distribution.

Since in the linear model case, we have proved that $\Omega_i - \hat{\Omega}_i = o_p(1)$ and $n^{-1/2} \sum_{i=1}^n \hat{\Omega}_i$ and $n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i$ asymptotically have the same limiting distribution, now we only need to show

$$\hat{\Omega}_i - \tilde{\Omega}_i = o_p(1)$$

and

$$n^{-1/2} \sum_{i=1}^n \hat{\Omega}_i - n^{-1/2} \sum_{i=1}^n \tilde{\Omega}_i = o_p(1). \tag{A4}$$

Assume

$$E(\tilde{Y}|\tilde{X}_{-1}) = \tilde{X}_{-1}\eta,$$

$$E(\tilde{X}_1|\tilde{X}_{-1}) = \tilde{X}_{-1}\gamma.$$

Recall that in the estimating procedures, we replaced \tilde{Y}, \tilde{X} with \hat{Y}, \hat{X} , so we have

$$E(\hat{Y}|\hat{X}_{-1}) = \hat{X}_{-1}\eta,$$

$$E(\hat{X}_1|\hat{X}_{-1}) = \hat{X}_{-1}\gamma.$$

We first show that with this replacement, $\tilde{\Omega}_i = \hat{\Omega}_i + o_p(1)$ holds. Note that by using Equation (A1) where $m_v(Z) - \hat{m}_v(Z) = o_p(1)$ and $v = x$ or y , because $\tilde{X}_{1i} - \tilde{X}_{-1i}\gamma$, $\tilde{Y}_i - \tilde{X}_{-1i}\eta$ and $\hat{X}_{1i} - \hat{X}_{-1i}\gamma$ are random variables, we have

$$\begin{aligned} & \tilde{\Omega}_i - \hat{\Omega}_i \\ &= (\tilde{X}_{1i} - \tilde{X}_{-1i}\gamma)^\top \{ \tilde{Y}_i - \hat{Y}_i - (\tilde{X}_{-1i} - \hat{X}_{-1i})\eta - (\tilde{X}_{1i} - \hat{X}_{1i})\beta_1 + (\tilde{X}_{-1i} - \hat{X}_{-1i})\gamma\beta_1 \} \\ & \quad - \{ (\hat{X}_{1i} - \tilde{X}_{1i}) - (\hat{X}_{-1i} - \tilde{X}_{-1i})\gamma \}^\top [\hat{Y}_i - \tilde{Y}_i - (\hat{X}_{-1i} - \tilde{X}_{-1i})\eta + \hat{Y}_i - \tilde{X}_{-1i}\eta \\ & \quad - \{ \hat{X}_{1i} - \tilde{X}_{1i} - (\hat{X}_{-1i} - \tilde{X}_{-1i})\gamma + \tilde{X}_{1i} - \tilde{X}_{-1i}\gamma \} \beta_1] \\ &= O_p(1) \{ o_p(n^{-1/4}) - o_p(n^{-1/4})\eta - o_p(n^{-1/4})\beta_1 + o_p(n^{-1/4})\gamma\beta_1 \} \\ & \quad - \{ o_p(n^{-1/4}) - o_p(n^{-1/4})\gamma \} [o_p(n^{-1/4}) - o_p(n^{-1/4})\eta + O_p(1) \\ & \quad - \{ o_p(n^{-1/4}) - o_p(n^{-1/4})\gamma - O_p(1) \} \beta_1] \\ &= o_p(1). \end{aligned}$$

To show Equation (A4), we first need to show

$$n^{-1/2} \sum_{i=1}^n (\tilde{X}_{1i} - \tilde{X}_{-1i}\gamma)(\tilde{Y}_i - \hat{Y}_i) = o_p(1).$$

For a given ϵ and a certain constant c , we have

$$\begin{aligned}
 & P\left\{n^{-1/2}\left|\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)(\tilde{Y}_i-\hat{Y}_i)\right|>\epsilon\right\} \\
 \leq & P\left(n^{-1/2}\left|\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)(\tilde{Y}_i-\hat{Y}_i)\right|>\epsilon,|\tilde{Y}_i\hat{Y}_i|\leq cn^{-1/4}\right) \\
 & +P\left(|\tilde{Y}_i-\hat{Y}_i|>cn^{-1/4}\right) \\
 \leq & P\left(n^{-1/2}cn^{-1/4}\left|\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)\right|>\epsilon\right)+o_p(1) \\
 = & o_p(1).
 \end{aligned}
 \tag{A5}$$

Equation (A5) holds because $n^{-1/2}\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)$ converges to $N(0, v(\tilde{X}_1|\tilde{X}_{-1}))$, where $v(\tilde{X}_1|\tilde{X}_{-1})$ is the variance of $(\tilde{X}_1-\tilde{X}_{-1}\gamma)$. Using a similar proof, we have

$$\begin{aligned}
 n^{-1/2}\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)(\tilde{X}_{-1i}-\hat{X}_{-1i})\eta &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)(\tilde{X}_{1i}-\hat{X}_{1i})\beta_1 &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma)(\tilde{X}_{-1i}-\hat{X}_{-1i})\gamma\beta_1 &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})(\tilde{Y}_i-\tilde{X}_{-1i}\eta) &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma) &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma(\tilde{Y}_i-\tilde{X}_{-1i}\eta) &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma(\tilde{X}_{1i}-\tilde{X}_{-1i}\gamma) &= o_p(1).
 \end{aligned}$$

With the same proof as in Equation (A3), we have

$$\begin{aligned}
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})^\top(\hat{Y}_i-\tilde{Y}_i) &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})^\top(\hat{X}_{-1i}-\tilde{X}_{-1i})\eta &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})^\top(\hat{X}_{1i}-\tilde{X}_{1i})\beta_1 &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{1i}-\tilde{X}_{1i})^\top(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma\beta_1 &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma^\top(\hat{Y}_i-\tilde{Y}_i) &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma^\top(\hat{X}_{-1i}-\tilde{X}_{-1i})\eta &= o_p(1). \\
 n^{-1/2}\sum_{i=1}^n(\hat{X}_{-1i}-\tilde{X}_{-1i})\gamma^\top(\hat{X}_{1i}-\tilde{X}_{1i})\beta_1 &= o_p(1).
 \end{aligned}$$

$$n^{-1/2} \sum_{i=1}^n (\hat{X}_{-1i} - \tilde{X}_{-1i}) \gamma^{\top} (\hat{X}_{-1i} - \tilde{X}_{-1i}) \gamma \beta_1 = o_p(1).$$

With the above equations, Equation (A4) holds. The proof is thus completed following the same procedure as in proving Theorem 1. \square

References

- Engle, R.; Granger, C.; Rice, J.; Weiss, A. Nonparametric estimates of the relation between weather and electricity sales. *J. Am. Stat. Assoc.* **1986**, *81*, 310–320. [[CrossRef](#)]
- Hastie, T.; Tibshirani, R. *Generalized Additive Models*; Chapman and Hall: London, UK, 1990.
- Linton, O.; Nielsen, J. A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika* **1995**, *82*, 93–100. [[CrossRef](#)]
- Gray, R. Spline-based test in survival analysis. *Biometrika* **1994**, *50*, 640–652. [[CrossRef](#)]
- Wahba, G. Cross validated spline methods for the estimation of multivariate functions from data on functionals. In Proceedings of the Iowa State University Statistical Laboratory 50th Anniversary Conference, Ames, IA, USA, 13–15 June 1984; David, H.A., David, H.T., Eds.; The Iowa State University Press: Ames, IA, USA, 1984; pp. 205–235.
- Green, P.; Jennison, C.; Seheult, A. Analysis of field experiments by least squares smoothing. *J. R. Stat. Soc. Ser.* **1985**, *47*, 299–315. [[CrossRef](#)]
- Heckman, N. Smoothing Spline in partly linear models. *J. R. Stat. Ser. B* **1986**, *48*, 244–248. [[CrossRef](#)]
- Rice, J. Convergence rates for partially splined models. *Stat. Probab. Lett.* **1994**, *4*, 203–208. [[CrossRef](#)]
- Speckman, P. Kernel Smoothing in Partial Linear Models. *J. R. Stat. Soc. Ser. B* **1988**, *50*, 413–436. [[CrossRef](#)]
- Chen, H.; Shiau, J.J.H. Data-Driven Efficient Estimators for a Partially Linear Model. *Ann. Stat.* **1994**, *22*, 211–237. [[CrossRef](#)]
- Chen, H. Convergence rates for parametric components in a partly linear model. *Ann. Stat.* **1988**, *16*, 136–146. [[CrossRef](#)]
- Härdle, W.; Liang, H.; Gao, J. *Partially Linear Models*; Springer Physica: Heidelberg, Germany, 2000.
- Liang, H. Estimation in Partially Linear Models and Numerical Comparisons. *Comput. Stat. Data Anal.* **2006**, *50*, 675–687. [[CrossRef](#)] [[PubMed](#)]
- Severini, T.; Staniswalis, J. Quasilikelihood estimation in semiparametric models. *J. Am. Stat. Assoc.* **1994**, *89*, 501–511. [[CrossRef](#)]
- Hastie, T.J. *Generalized Additive Models*; Wadsworth and Brooks/Cole: Pacific Grove, CA, USA, 1992.
- Owen, A. *Empirical Likelihood*; Chapman and Hall/CRC: London, UK, 2001.
- Robinson, P.M. Root- n -consistent semiparametric regression. *Econometrica* **1988**, *56*, 931–954. [[CrossRef](#)]
- Su, H.; Liang, H. An empirical likelihood-based method for comparison of treatment effects—test of equality of coefficients in linear models. *Comput. Stat. Data Anal.* **2010**, *54*, 1079–1088. [[CrossRef](#)] [[PubMed](#)]
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; Chapman and Hall/CRC: London, UK, 1996.
- DiCiccio, T.; Hall, P.; Romano, J. Empirical likelihood is bartlett-correctable. *Ann. Stat.* **1991**, *19*, 1053–1061. [[CrossRef](#)]
- Chen, S.; Cui, H. On bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **2006**, *93*, 215–220. [[CrossRef](#)]
- Ruppert, D.; Sheather, S.J.; Wand, M.P. An effective bandwidth selector for local least squares regression. *J. Am. Med. Assoc.* **1995**, *90*, 1257–1270. [[CrossRef](#)]
- Liang, H.; Wang, S.; Carroll, R. Partially linear models with missing response variables and error-prone covariates. *Biometrika* **2007**, *94*, 185–198.

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.