Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(1): 55–72.
DOI:10.3934/math.2024004
Received: 10 October 2023
Revised: 16 November 2023
Accepted: 22 November 2023
Published: 24 November 2023

Research article

Semi-supervised estimation for the varying coefficient regression model

Peng Lai[1,*], **Wenxin Tian**[1] **and Yanqiu Zhou**[2]

[1] School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] School of Science, Guangxi University of Science and Technology, Liuzhou 545006, China

* **Correspondence:** Email: laipeng@nuist.edu.cn.

Abstract: In many cases, the 'labeled' outcome is difficult to observe and may require a complicated or expensive procedure, and the predictor information is easy to be obtained. We propose a semi-supervised estimator for the one-dimensional varying coefficient regression model which improves the conventional supervised estimator by using the unlabeled data efficiently. The semi-supervised estimator is proposed by introducing the intercept model and its asymptotic properties are proven. The Monte Carlo simulation studies and a real data example are conducted to examine the finite sample performance of the proposed procedure.

Keywords: semi-supervised learning; varying coefficient regression model; intercept model; locally weighted regression model
Mathematics Subject Classification: 62G05, 62G20, 62R07

1. Introduction

Semi-supervised learning first appeared in machine learning literature, used to describe a situation where some data are labeled and the rest are unlabeled [10]. Conceptually, it is between supervised and unsupervised learning. It allows you to take advantage of a large amount of unlabeled data available and smaller labeled datasets in many instances. Semi-supervised learning occurs when the label variable is difficult to observe and may require complex or expensive processes. Specifically, a sample of n observations in the joint distribution $(\mathbf{X}, Y)$ is given, where Y is the label variable and $\mathbf{X}$ contains the covariates. In addition, an additional m samples are observed with only $\mathbf{X}$ given. The aim is to study the relationships between $\mathbf{X}$ and Y using additional unlabeled data.

For semi-supervised learning, much literature focuses on the case that Y takes a small number of values, which can be reduced to the case of classification tasks, such as [19]. In recent years, research in this field has focused on neural network-based models and generative learning algorithm problems. [13] introduced a simple and computationally efficient algorithm for training deep neural networks in a semi-supervised learning paradigm: interpolating Consistency Training (ICT). [8] introduced an innovative framework for federated transfer learning (FTL) known as Semi-Supervised Federated Heterogeneous Transfer Learning (SFHTL), to utilize unlabeled non-overlapping samples, addressing the challenge of model overfitting caused by the limited overlap of training samples in federated learning (FL) scenarios. Based on compatibility conditions in the semi-supervised probably approximately correct (PAC) theory, [7] demonstrated why labeled heterogeneous source data and unlabeled target data help reduce target risk. Based on this theory, two algorithms were designed as a proof of concept. One is the kernel heterogeneous domain alignment (KHDA) algorithm, which is a kernel-based algorithm, and the other is the joint mean embedded alignment (JMEA) algorithm based on neural networks. At the same time, some scholars have studied the regression problem. [14] classified different semi-supervised methods into two categories: distribution-based and marginal-based. The distribution method relied on the assumption that the conditional expectation $E(Y|\mathbf{X})$ was linked to the marginal distribution of $\mathbf{X}$. The marginal-based approach used additional information on $\mathbf{X}$. Some other studies, such as [18], took into account Y sequential values and used unlabeled data to learn the structure of $\mathbf{X}$ so that non-parametric regression could be better estimated. These efforts are very helpful where non-parametric regression is useful and unlabeled data is available. [15] proposed an estimator of the population mean using unlabeled data combined with least squares. [16] provided a semi-supervised reasoning framework focused on the mean and variance of the responses, allowing covariates to be much larger in size than the sample sizes, and provided new estimates of the mean and variance of response variables. [4] considered the linear regression problem in a semi-supervised setting and proposed a class of highly efficient adaptive semi-supervised estimators (EASE) to improve the estimation efficiency. Additionally, they applied this method to the study of electronic medical records on autoimmunity. [2] proposed a correction estimator that effectively integrates labeled and unlabeled data, called Corrected High-dimensional Inference of Variance Interpretation Estimators (CHIVE), which achieved minimax optimal convergence speed under a general semi-supervised framework. [1] studied linear regression in a semi-supervised setting and, even in a framework where $E(Y|\mathbf{X})$ was not linear, additional information about the distribution of $\mathbf{X}$ was helpful to construct better estimates than standard least squares estimates. However, there are few studies on semi-supervised learning in the varying coefficient models. Therefore, we want to extend the intercept model method of [1] to the varying coefficient model and investigate the estimation of the coefficient function under semi-supervised learning.

Varying coefficient models constitute a versatile and expansive category of statistical models encompassing well-known structures like additive models, partial linear models, single-index coefficient regression models and adaptive varying coefficient partial linear models. Researchers have investigated varying coefficient models, seeking to understand and refine their applications across diverse statistical contexts. The exploration of these models extends beyond conventional frameworks, uncovering novel insights that contribute to the evolving landscape of statistical methodology. [6] introduced a robust two-step estimation approach designed for the coefficient function. The methodology involves the construction of a reliable estimator for the coefficient function, accompanied by a thorough analysis of the asymptotic mean squared error and the convergence velocity of the estimator. This method not only advances the precision of coefficient function estimation but also contributes insights into the statistical properties of the estimator under consideration. [3]

investigated the generalized varying coefficient model, focusing on both estimation and hypothesis testing. Their methodological approach centered on crafting a robust estimation for the coefficient function, leveraging the local polynomial regression technique to enhance accuracy. [9] proposed a semiparametric estimator for the heteroscedastic single-index varying coefficient model. The estimator is proved to attain the semiparametric efficiency bound. [17] considered an estimating equations approach to parameter estimation in an adaptive varying coefficient linear quantile model. They proposed estimating equations for the index vector of the model in which the unknown nonparametric functions were estimated by minimizing the check loss function, resulting in a profiled approach.

Our goal here is to study the more efficient estimates of the one-dimensional varying coefficient model with unlabeled data as a pioneer in the semi-supervised varying coefficient modeling problems. The intercept model is introduced into the varying coefficient models, and the extra information from the unlabeled data is combined to improve the performance of the estimators.

The rest of the article is organized as follows. Section 2 provides the basic setting and the proposed methods. Theoretical properties are also given in Section 2. In Section 3, we conduct Monte Carlo simulation studies to examine the finite sample performance of the proposed procedure. We also use the proposed procedure in an example with real data.

2. Materials and methods

2.1. Varying coefficient model

In the study of the demand for shared bicycles (Y), many studies focus on the impact of the temperature (X), and how the relationships between them may change with time (T). Thus the varying coefficient model is a good choice. Nevertheless, in real-world scenarios, it is common that the labeled variable Y may not be entirely observed, and only observations of the covariate (X, T) are available. In such cases, the introduction of semi-supervised learning becomes imperative. In semi-supervised learning, the data structure involves a set of n observations sampled from the joint distribution (Y, X, T), and an additional m samples are observed with only (X, T) given. The model leverages both labeled and unlabeled data, allowing us to make more informed and robust estimates by incorporating the partially observed labeled samples along with the unlabeled data. This approach is effective when dealing with practical problems where complete observations of the labeled variable Y may be limited.

To study the estimates of the varying coefficient model with unlabeled data, we introduce our basic method and consider the following model,

$$Y = g(T)X + \varepsilon, \tag{2.1}$$

where Y is the label variable, X is the one-dimensional covariant, T is the dependent variable, $g(\cdot)$ is the unknown measurable function on R and ε is the random error with $E(\varepsilon|T, X) = 0$.

2.2. Locally weighted linear regression estimates

In the supervised situation, where the data from model (2.1) are completely observed, the coefficient function $g(t)$ of model (2.1) can be estimated by the locally weighted linear regression estimation method [5]. For any given point t_0, use a linear function in a neighborhood of t_0,

$$g(t) \approx g(t_0) + g'(t_0)(t - t_0). \tag{2.2}$$

We can minimize the following objective functions by a and b to obtain the estimator of $g(t_0)$,

$$L(a,b)=\sum_{i=1}^{n}(Y_i-aX_i-b\frac{T_i-t_0}{h}X_i)^2K_h(T_i-t_0), \tag{2.3}$$

where $g(t_0)=a$, $hg'(t_0)=b$ and $K(\cdot)$ is some kernel function $K_h(\cdot)=K(\cdot/h)/h$ with bandwidth h. Let $Z_i=\left(X_i,X_i(\frac{T_i-t_0}{h})\right)$ and $\beta(t_0)=(a,b)^\top=(g(t_0),hg'(t_0))^\top$. Thus, $\hat{\beta}(t_0)=\left(\hat{a},\hat{b}\right)^\top=(\hat{g}(t_0),h\hat{g}'(t_0))^\top$ is

$$\hat{\beta}(t_0)=(D^\top K_wD)^{-1}(D^\top K_wY), \tag{2.4}$$

where D is the $n\times 2$ matrix and its ith line elements are $\left(X_i,X_i\frac{T_i-t_0}{h}\right)$, $Y=(Y_1,Y_2,...,Y_n)^\top$, $K_w=diag\,(K_h(T_1-t_0),K_h(T_2-t_0),...,K_h(T_n-t_0))$. Thus, the locally weighted linear regression estimate for the coefficient function $g(t_0)$ of model (2.1) is $\hat{g}_L(t_0)=e^\top\hat{\beta}(t_0)$, where $e=(1,0)^\top$.

2.3. Intercept model with total information

To take full advantage of the information from the unlabeled data, we want to discuss the form of estimation in the semi-supervised case. First, we assume that the total information is completely known.

Motivated by [1], for model (2.1), we have

$$g(T)=arg\min_{g(T)}E\left[(Y-g(T)X)|T\right]^2=\frac{E(YX|T)}{E(X^2|T)}.$$

So, we get the following model through multiplying model (2.1) by $\frac{X}{E(X^2|T)}$,

$$\tilde{Y}=g(T)\tilde{X}_{01}+a(T)\tilde{X}_{02}+\tilde{\varepsilon}=\tilde{g}(T)\tilde{\mathbf{X}}_1+\tilde{\varepsilon}, \tag{2.5}$$

where $\tilde{Y}=\frac{YX}{E(X^2|T)}$, $\tilde{g}(T)=(g(T),a(T))$, $\tilde{\mathbf{X}}_1=\left(\tilde{X}_{01},\tilde{X}_{02}\right)^\top$, $\tilde{X}_{01}=1$, $\tilde{X}_{02}=\frac{X^2}{E(X^2|T)}-1$ and $\tilde{\varepsilon}$ is the remainder term. Under the total information situation, $E(X^2|T)$ is known. The multiplication term makes the expectation of $\tilde{Y}$ to be $E(\tilde{Y}|T)=g(T)$. Therefore, by introducing this multiplication term into our estimation process, we can leverage the information from the unlabeled data set more effectively.

After introducing the intercept model, we have the locally weighted regression estimator of $\beta(t_0)$,

$$\hat{\tilde{\beta}}(t_0)=\left(\tilde{D}^\top K_w\tilde{D}\right)^{-1}\left(\tilde{D}^\top K_w\tilde{Y}\right), \tag{2.6}$$

where the ith line elements of $\tilde{D}$ are $(\tilde{X}_i^\top,\tilde{X}_i^\top(\frac{T_i-t_0}{h}))$ and $\tilde{X}_i=(\tilde{X}_{01i},\tilde{X}_{02i})^\top$, $\tilde{X}_{01i}=1$, $\tilde{X}_{02i}=\frac{X_i^2}{E(X^2|T_i)}-1$, $i=1,2,\ldots,n$, $\hat{\tilde{\beta}}(t_0)=\begin{pmatrix}\hat{\tilde{g}}(t_0)\\ h\hat{\tilde{g}}'(t_0)\end{pmatrix}$. Thus, the total information estimate of $\tilde{g}(t_0)$ is $\hat{g}_{TI}(t_0)=\tilde{e}^\top\hat{\tilde{\beta}}(t_0)$, where $\tilde{e}=(1,0,0,0)^\top$.

2.4. Intercept model with partial information

In practical problems, the overall information is not completely known, so we can only extract partial information through the observed unlabeled data. We consider the estimation under partial information (PI) i.e., the semi-supervised setting.

For semi-supervised data, consider n independent and identically distributed observations $(X_1, T_1, Y_1), (X_2, T_2, Y_2), \ldots, (X_n, T_n, Y_n)$ in a joint distribution G, and an additional set $(X_{n+1}, T_{n+1}), (X_{n+2}, T_{n+2}), \ldots, (X_{n+m}, T_{n+m})$ of m independent observations from the marginal distribution. Model (2.5) becomes

$$\check{Y} = g(T)\check{X}_{01} + a(T)\check{X}_{02} + \check{\varepsilon} = \check{g}(T)\check{\mathbf{X}}_2 + \check{\varepsilon}, \tag{2.7}$$

where $\check{Y} = \frac{YX}{\check{E}(X^2|T)}$, $\check{g}(T) = (g(T), a(T))$, $\check{\mathbf{X}}_2 = \left(\check{X}_{01}, \check{X}_{02}\right)^\top$, $\check{X}_{02} = \frac{X^2}{\check{E}(X^2|T)} - 1$, $\check{X}_{01} = 1$ and $\check{\varepsilon}$ is the remainder term. $\check{E}(X^2|T)$ is the estimated conditional expectation under T for X^2 based on the semi-supervised data, $\check{E}(X^2|T) = \frac{\sum_{i=1}^{n+m} X_i^2 K_h(T_i - T)}{\sum_{i=1}^{n+m} K_h(T_i - T)}$. Therefore, the semi-supervised estimator of $\beta(t)$ is

$$\hat{\check{\beta}}(t_0) = \left(\check{D}^\top K_w \check{D}\right)^{-1} \left(\check{D}^\top K_w \check{Y}\right), \tag{2.8}$$

where the ith line elements of $\check{D}$ are $(\check{X}_i^\top, \check{X}_i^\top(\frac{T_i - t_0}{h}))$ and $\check{X}_i = (\check{X}_{01i}, \check{X}_{02i})^\top$, $\check{X}_{01i} = 1$, $\check{X}_{02i} = \frac{X_i^2}{\check{E}(X^2|T_i)} - 1$, $i = 1, 2, \ldots, n$, $\hat{\check{\beta}}(t_0) = \begin{pmatrix} \hat{\check{g}}(t_0) \\ h\hat{\check{g}}'(t_0) \end{pmatrix}$. The semi-supervised estimator of the coefficient function $\check{g}(t_0)$ of the model (2.7) is $\hat{g}_{PI}(t_0) = \check{e}^\top \hat{\check{\beta}}(t_0)$, where $\check{e} = (1, 0, 0, 0)^\top$.

2.5. Theoretical properties

In this subsection, the following conditions are listed to study the theoretical properties of the proposed estimating procedures.

(1) The density function $f(t)$ of T is continuous over the interval [0,1] and takes values greater than 0.

(2) The varying coefficient function $g(t)$ and $E(Y|T = t)$ have continuous derivatives up to order 2, and are bounded away from zero.

(3) The kernel function $K_h(\cdot)$ is a bounded symmetric density function satisfying Lipschitz continuity in the interval (-1,1) with a window width of $0 < h$ and $h = O(n^{-\frac{1}{5}})$.

(4) $\Gamma(t) = E(X^2|T = t)$ is non-singular on the interval [0,1] and has a continuous second derivative.

(5) $E(\varepsilon^2|T = t, X = x)$ has bounded partial derivatives up to order 2 and is bounded away from zero.

(6) There exists $s > 2$ that makes $E(|X|^{2s}) < \infty$ and $E(|Y|^{2s}) < \infty$.

Remark 1. *Conditions (1)–(4) are general assumptions for varying coefficient models, which can be found in Tang and Zhou [12].*

Theorem 1. *If conditions (1)–(6) hold, and $\varepsilon(X - E(X|T))$ has a finite second order moment, then,*

$$\sqrt{nh}(\hat{g}_L(t_0) - g(t_0) - \tfrac{1}{2}\mu_2 h^2 g''(t_0)) \sim N(0, \sigma_L^2),$$

where $\sigma_L^2 = \nu_0 f^{-1}(t_0)[\Gamma^{-1}(t_0)]^2 E(X^2\varepsilon^2|T = t_0)$, $\mu_2 = \int t^2 K(t)dt$, $\nu_0 = \int K^2(t)dt$.

Proof. First, we consider the varying coefficient model (2.1), known by the properties of the varying coefficient model,

$$\frac{1}{n}D^\top K_w D \xrightarrow{P} \begin{pmatrix} \Gamma(t_0)f(t_0) & 0 \\ 0 & \mu_2\Gamma(t_0)f(t_0) \end{pmatrix}, \tag{2.9}$$

where D is the $n \times 2$ matrix and its line i elements are $\left(X_i, X_i\frac{T_i-t_0}{h}\right)$, $\Gamma(t_0) = E(X^2|T_i = t_0)$, $f(t)$ is the density function of T and $\mu_2 = \int t^2K(t)dt$. Since

$$Y_i = g(T_i)X_i + \varepsilon_i = \begin{pmatrix} X_i \\ \frac{T_i-t_0}{h}X_i \end{pmatrix}^\top \begin{pmatrix} g(t_0) \\ hg'(t_0) \end{pmatrix} + (g(T_i) - g(t_0) - g'(t_0)(T_i - t_0))X_i + \varepsilon_i,$$

then we have,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_1 & \frac{T_1-t_0}{h}X_1 \\ X_2 & \frac{T_2-t_0}{h}X_2 \\ \vdots & \vdots \\ X_n & \frac{T_n-t_0}{h}X_n \end{pmatrix}\begin{pmatrix} g(t_0) \\ hg'(t_0) \end{pmatrix} + \begin{pmatrix} (g(T_1) - g(t_0) - g'(t_0)(T_1 - t_0))X_1 \\ (g(T_2) - g(t_0) - g'(t_0)(T_2 - t_0))X_2 \\ \vdots \\ (g(T_n) - g(t_0) - g'(t_0)(T_n - t_0))X_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$= D\begin{pmatrix} g(t_0) \\ hg'(t_0) \end{pmatrix} + \begin{pmatrix} (g(T_1) - g(t_0) - g'(t_0)(T_1 - t_0))X_1 \\ (g(T_2) - g(t_0) - g'(t_0)(T_2 - t_0))X_2 \\ \vdots \\ (g(T_n) - g(t_0) - g'(t_0)(T_n - t_0))X_n \end{pmatrix} + \varepsilon.$$

Define the above as $Y = D\beta(t_0) + \Delta_g X + \varepsilon$, where $Y = (Y_1, Y_2, \ldots, Y_n)^\top$, $\Delta_g = diag\left((g(T_1) - g(t_0) - g'(t_0)(T_1 - t_0)), \ldots, (g(T_n) - g(t_0) - g'(t_0)(T_n - t_0))\right)$ and $X = (X_1, X_2, \ldots, X_n)^\top$. We have

$$\hat{\beta}(t_0) = (D^\top K_w D)^{-1}D^\top K_w Y = (D^\top K_w D)^{-1}D^\top K_w D\beta(t_0) + (D^\top K_w D)^{-1}D^\top K_w(\Delta_g X + \varepsilon). \tag{2.10}$$

Therefore,

$$\hat{\beta}(t_0) - \beta(t_0) = (\frac{1}{n}D^\top K_w D)^{-1}(\frac{1}{n}D^\top K_w \Delta_g X + \frac{1}{n}D^\top K_w \varepsilon). \tag{2.11}$$

Let $\frac{1}{n}D^\top K_w \Delta_g X = A_1^{(1)}$, $\frac{1}{n}D^\top K_w \varepsilon = A_2^{(2)}$, $A_1 = e_1^\top A_1^{(1)}$, $A_2 = e_1^\top A_2^{(2)}$, where e_1 is a unit vector with elements $(1, 0)^\top$. For A_1,

$$E(A_1) = E(\frac{1}{n}\sum_{i=1}^{n} X_i K_h(T_i - t_0)X_i(g(T_i) - g(t_0) - g'(t_0)(T_i - t_0))) = \Gamma(t_0)f(t_0)\frac{1}{2}g''(t_0)h^2\mu_2 + o(h^2).$$

Since $\hat{\beta}(t_0) - \beta(t_0) = \begin{pmatrix} \hat{g}(t_0) \\ h\hat{g}'(t_0) \end{pmatrix} - \begin{pmatrix} g(t_0) \\ hg'(t_0) \end{pmatrix}$, we have

$$\hat{g}_L(t_0) - g(t_0) = (\Gamma(t_0)f(t_0))^{-1}(\Gamma(t_0)f(t_0)\frac{1}{2}g''(t_0)h^2\mu_2 + o(h^2) + \frac{1}{n}\sum_{i=1}^{n} X_i K_h(T_i - t_0)\varepsilon_i). \tag{2.12}$$

It should be noted that $\frac{1}{n}\sum_{i=1}^{n} X_i K_h(T_i - t_0)\varepsilon_i \sim N(0, \frac{1}{nh}E(X_i^2\varepsilon_i^2|t_0)f(t_0)\nu_0)$, where $\nu_0 = \int K^2(t)dt$. Thus,

$$\hat{g}_L(t_0) - g(t_0) = \frac{1}{2}g''(t_0)h^2\mu_2 + o(h^2) + (\Gamma(t_0)f(t_0))^{-1}\frac{1}{n}\sum_{i=1}^{n} X_i K_h(T_i - t_0)\varepsilon_i, \tag{2.13}$$

and it follows that

$$\sqrt{nh}(\hat{g}_L(t_0) - g(t_0) - \frac{1}{2}\mu_2 h^2 g''(t_0)) \sim N(0, \nu_0 f^{-1}(t_0)(\Gamma^2(t_0))^{-1} E(X_i^2 \varepsilon_i^2 | t_0)). \tag{2.14}$$

□

Theorem 2. *Assuming that* $\left(\tilde{Y}, \tilde{X}_{01}, \tilde{X}_{02}\right)$ *has a finite second-order moment, the total information estimator of the intercept model has*

$$\sqrt{nh}(\hat{g}_{TI}(t_0) - g(t_0) - \tfrac{1}{2}\mu_2 h^2 g''(t_0)) \sim N(0, \sigma_{TI}^2),$$

where $\sigma_{TI}^2 = \nu_0 f^{-1}(t_0) E(\tilde{\varepsilon}^2 | t_0)$, $\sigma_L^2 = \sigma_{TI}^2 + \sigma_{diff1}^2$, $\sigma_{diff1}^2 = nhE\{\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)(\frac{\varepsilon_i X_i}{E(X^2|T_i)} - \tilde{\varepsilon}_i)\}^2$.

Proof. Before discussing the form of estimation in the semi-supervised case, we assume that all information is completely known and consider the estimation with complete data information. Consider the intercept model (2.5) under the complete data. We have the objective function

$$L(\tilde{\beta}) = \sum_{i=1}^{n}(\tilde{Y}_i - \tilde{g}(t_0)\tilde{X}_i - h\tilde{g}'(t_0)\frac{T_i - t_0}{h}\tilde{X}_i)^2 K_h(T_i - t_0), \tag{2.15}$$

where, $\tilde{X}_i = (\tilde{X}_{01i}, \tilde{X}_{02i})^\top$.

Let the line i elements of $\tilde{D}$ be $(\tilde{X}_i^\top, \tilde{X}_i^\top(\frac{T_i - t_0}{h}))$, $\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, \cdots, \tilde{Y}_n)^\top$, $\hat{\tilde{\beta}} = \begin{pmatrix} \hat{\tilde{g}}(t_0) \\ h\hat{\tilde{g}}'(t_0) \end{pmatrix}$, so locally weighted linear regression by varying coefficients shows that $\hat{\tilde{\beta}}(t_0) = (\tilde{D}^\top K_w \tilde{D})^{-1}(\tilde{D}^\top K_w \tilde{Y})$, $\hat{\tilde{g}}(t_0) = e_1^\top \hat{\tilde{\beta}}(t_0)$, where the first element of e_1 is 1 and the others are 0.

In addition, we have

$$\frac{1}{n}\tilde{D}^\top K_w \tilde{D} \xrightarrow{P} \begin{pmatrix} \tilde{\Gamma}(t_0) f(t_0) & 0 \\ 0 & \mu_2 \tilde{\Gamma}(t_0) f(t_0) \end{pmatrix}, \tag{2.16}$$

where $\tilde{\Gamma}(t_0) = E(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^\top | t_0)$ and $\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{X}_1^\top \\ \vdots \\ \tilde{X}_n^\top \end{pmatrix}$.

Consider $\frac{1}{n}\tilde{D}^\top K_w \tilde{Y} = \begin{pmatrix} \frac{1}{n}\tilde{\mathbf{X}}^\top K_w \tilde{Y} \\ \frac{1}{n}\tilde{\mathbf{X}}^\top T_h^\top K_w \tilde{Y} \end{pmatrix}$ and

$$\tilde{Y}_i = \tilde{g}(T_i)^\top \tilde{X}_i + \tilde{\varepsilon}_i = \begin{pmatrix} \tilde{X}_i^\top & (\frac{T_i - t_0}{h})\tilde{X}_i^\top \end{pmatrix}\begin{pmatrix} \tilde{g}(t_0) \\ h\tilde{g}'(t_0) \end{pmatrix} + \tilde{X}_i^\top \begin{pmatrix} g(T_i) - g(t_0) - g'(t_0)(T_i - t_0) \\ a(T_i) - a(t_0) - a'(t_0)(T_i - t_0) \end{pmatrix} + \tilde{\varepsilon}_i.$$

We can define

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \vdots \\ \tilde{Y}_n \end{pmatrix} = \tilde{D}\tilde{\beta}(t_0) + \tilde{\mathbf{X}}\Delta_{ga} + \tilde{\varepsilon}, \tag{2.17}$$

where $\Delta_{ga} = diag\left(\begin{pmatrix} g(T_1) - g(t_0) - g'(t_0)(T_1 - t_0) \\ a(T_1) - a(t_0) - a'(t_0)(T_1 - t_0) \end{pmatrix}, \ldots, \begin{pmatrix} g(T_n) - g(t_0) - g'(t_0)(T_n - t_0) \\ a(T_n) - a(t_0) - a'(t_0)(T_n - t_0) \end{pmatrix}\right)$.

Therefore,

$$\hat{\tilde{\beta}}(t_0) = (\tilde{D}^\top K_w \tilde{D})^{-1}\tilde{D}^\top K_w \tilde{Y} = \tilde{\beta}(t_0) + (\frac{1}{n}\tilde{D}^\top K_w \tilde{D})^{-1}(\frac{1}{n}\tilde{D}^\top K_w \tilde{\mathbf{X}}\Delta_{ga} + \frac{1}{n}\tilde{D}^\top K_w \tilde{\varepsilon}).$$

Note that

$$\tilde{X}_i\tilde{X}_i^\top = \begin{pmatrix} \tilde{X}_{01i} \\ \tilde{X}_{02i} \end{pmatrix}\begin{pmatrix} \tilde{X}_{01i} \\ \tilde{X}_{02i} \end{pmatrix}^\top = \begin{pmatrix} 1 & \frac{X_i^2}{E(X^2|T_i)} - 1 \\ \frac{X_i^2}{E(X^2|T_i)} - 1 & [\frac{X_i^2}{E(X^2|T_i)} - 1]^2 \end{pmatrix}.$$

We have the following form,

$$\frac{1}{n}\tilde{D}^\top K_w \tilde{D} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top K_h(T_i - t_0) & \frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top K_h(T_i - t_0)\frac{T_i - t_0}{h} \\ \frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top K_h(T_i - t_0)\frac{T_i - t_0}{h} & \frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top K_h(T_i - t_0)(\frac{T_i - t_0}{h})^2 \end{pmatrix},$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top K_h(T_i - t_0) &= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0) & \frac{1}{n}\sum_{i=1}^n [\frac{X_i^2}{E(X^2|T_i)} - 1]K_h(T_i - t_0) \\ \frac{1}{n}\sum_{i=1}^n [\frac{X_i^2}{E(X^2|T_i)} - 1]K_h(T_i - t_0) & \frac{1}{n}\sum_{i=1}^n [\frac{X_i^2}{E(X^2|T_i)} - 1]^2 K_h(T_i - t_0)(\frac{T_i - t_0}{h})^2 \end{pmatrix} \\ &\xrightarrow{P} \begin{pmatrix} f(t_0) & 0 \\ 0 & E\{[\frac{X^2}{E(X^2|T)} - 1]^2|T = t_0\}f(t_0) \end{pmatrix}.\end{aligned}$$

Condisering $\frac{1}{n}\tilde{D}^\top K_w\tilde{\mathbf{X}}\Delta_{ga}$ and $\frac{1}{n}\tilde{D}^\top K_w\tilde{\varepsilon}$, let $\frac{1}{n}\tilde{D}^\top K_w\tilde{\mathbf{X}}\Delta_{ga} = \tilde{A}_1^{(1)}$, $\frac{1}{n}\tilde{D}^\top K_w\tilde{\varepsilon} = \tilde{A}_2^{(2)}$, $\tilde{A}_1 = e_1^\top\tilde{A}_1^{(1)}$, $\tilde{A}_2 = e_1^\top\tilde{A}_2^{(2)}$. We have

$$\tilde{D}^\top K_w\tilde{\mathbf{X}}\Delta_{ga} = \begin{pmatrix} \sum_{i=1}^n K_h(T_i - t_0)\tilde{X}_i\tilde{X}_i^\top\Delta_{ga} \\ \sum_{i=1}^n K_h(T_i - t_0)T_{h_i}\tilde{X}_i\tilde{X}_i^\top\Delta_{ga} \end{pmatrix},$$

and

$$\tilde{A}_1^{(1)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\tilde{X}_i\tilde{X}_i^\top\Delta_{ga} \\ \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\frac{T_i - t}{h}\tilde{X}_i\tilde{X}_i^\top\Delta_{ga} \end{pmatrix}.$$

Therefore,

$$\begin{aligned}\tilde{A}_1 &= \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\tilde{X}_{01i}^2[g(T_i) - g(t_0) - g'(t_0)(T_i - t_0)] \\ &\quad + \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\tilde{X}_{01i}\tilde{X}_{02i}[a(T_i) - a(t_0) - a'(t_0)(T_i - t_0)] \\ &= \tilde{A}_{11} + \tilde{A}_{12}.\end{aligned}$$

It is easy to obtain $\tilde{A}_1 = \tilde{A}_{11} + \tilde{A}_{12} = \frac{1}{2}g''(t_0)h^2\mu_2 f(t_0) + o(h^4)$. On the other hand,

$$\begin{aligned}\tilde{A}_2^{(2)} &= \frac{1}{n}\tilde{D}^\top K_w\tilde{\varepsilon} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \tilde{X}_i K_h(T_i - t_0)\tilde{\varepsilon}_i \\ \frac{1}{n}\sum_{i=1}^n \tilde{X}_i K_h(T_i - t_0)\frac{T_i - t_0}{h}\tilde{\varepsilon}_i \end{pmatrix} \\ \tilde{A}_2 &= e_1^\top\tilde{A}_2^{(2)} = \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\tilde{\varepsilon}_i.\end{aligned}$$

Since $\hat{\tilde{\beta}} - \tilde{\beta} = (\frac{1}{n}\tilde{D}^\top K_w\tilde{D})^{-1}(\frac{1}{n}\tilde{D}^\top K_w\tilde{\mathbf{X}}\Delta_{ga} + \frac{1}{n}\tilde{D}^\top K_w\tilde{\varepsilon})$, so we have

$$\hat{\tilde{g}}(t_0) - g(t_0) - \frac{1}{2}h^2\mu_2 g''(t_0) = f^{-1}(t_0)\frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\tilde{\varepsilon}_i + o(h^4). \tag{2.18}$$

Since

$$E(\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i)=E(K_h(T_i-t_0)\tilde{\varepsilon}_i)=E\left(K_h(T_i-t_0)E(\tilde{\varepsilon}_i|X_i,T_i)\right)=0,$$

$$\begin{aligned}E(\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i)^2 &= \frac{1}{n^2}\sum_{i=1}^{n}E\left(K_h^2(T_i-t_0)\tilde{\varepsilon}_i^2\right)+\frac{1}{n^2}\sum_{i\neq j}E\left(K_h(T_i-t_0)\tilde{\varepsilon}_iK_h(T_j-t_0)\tilde{\varepsilon}_j\right)\\ &= \frac{1}{nh}f(t_0)E(\tilde{\varepsilon}_i^2|t_0)\nu_0+o(\frac{1}{nh}),\end{aligned}$$

letting $\hat{\hat{g}}(t_0)=\hat{g}_{TI}(t_0)$, we can get

$$\sqrt{nh}(\hat{\hat{g}}_{TI}(t_0)-g(t_0)-\frac{1}{2}\mu_2h^2g''(t_0))\sim N(0,\nu_0f^{-1}(t_0)E(\tilde{\varepsilon}^2|t_0)),$$

where $\sigma_{TI}^2=\nu_0f^{-1}(t_0)E(\tilde{\varepsilon}^2|t_0)$.

We have

$$\begin{aligned}\frac{1}{nh}\sigma_L^2 &= E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\frac{\varepsilon_iX_i}{E(X^2|T_i)}\right)^2\\ &= E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i\right)^2+E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)(\frac{\varepsilon_iX_i}{E(X^2|T_i)}-\tilde{\varepsilon}_i)\right)^2\\ &\quad +2E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i\left[\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)(\frac{\varepsilon_iX_i}{E(X^2|T_i)}-\tilde{\varepsilon}_i)\right]\right).\end{aligned}$$

Note that

$$\begin{aligned}V=\frac{\varepsilon X}{E(X^2|T)}-\tilde{\varepsilon} &= \frac{(Y-g(T)X)X}{E(X^2|T)}-\frac{YX}{E(X^2|T)}+g(T)+a(T)\left(\frac{X^2}{E(X^2|T)}-1\right)\\ &= g(T)\left(1-\frac{X^2}{E(X^2|T)}\right)+a(T)\left(-1+\frac{X^2}{E(X^2|T)}\right)\\ &= (a(T)-g(T))\,\tilde{X}_{02},\end{aligned}$$

we have

$$\begin{aligned}&E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i\left[\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)(\frac{\varepsilon_iX_i}{E(X^2|t_0)}-\tilde{\varepsilon}_i)\right]\right)\\ &=E\left(\frac{1}{f^2(t_0)}(\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0))^2\tilde{\varepsilon}_iV_i\right)\\ &=E\left(\frac{1}{f^2(t_0)}(\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0))^2E(\tilde{\varepsilon}_i|X_i,T_i)V_i\right)=0.\end{aligned}$$

We can determine that

$$E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\frac{\varepsilon_i X_i}{E(X^2|T_i)}\right)^2$$
$$=E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)\tilde{\varepsilon}_i\right)^2+E\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)(\frac{\varepsilon_i X_i}{E(X^2|T_i)}-\tilde{\varepsilon}_i)\right)^2.$$

That is, $\sigma_L^2 \geq \sigma_{TI}^2$, $\sigma_L^2 = \sigma_{TI}^2 + \sigma_{diff1}^2$, where

$$\begin{aligned}
\sigma_{diff1}^2 &= nhE\left(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n}K_h(T_i-t_0)(\frac{\varepsilon_i X_i}{E(X^2|T_i)}-\tilde{\varepsilon}_i)\right)^2\\
&= h\frac{1}{f^2(t_0)}\int\frac{1}{h^2}K^2(\frac{T_i-t_0}{h})E\left[(\frac{\varepsilon X}{E(X^2|T_i)}-\tilde{\varepsilon})^2|T_i\right]f(T_i)d_{T_i}\\
&= \frac{1}{f(t_0)}\int K^2(u)duE\left[(\frac{\varepsilon X}{E(X^2|t_0)}-\tilde{\varepsilon})^2|t_0\right]+o(h^2)\\
&= \frac{\int K^2(u)du}{f(t_0)}E\left[(a(t_0)-g(t_0))(\frac{X^2}{E(X^2|t_0)}-1)\right]^2.
\end{aligned}$$

□

Theorem 3. *Considering the partial information estimator of the intercept model in a semi-supervised setting, when* $lim\frac{n}{n+m}=v$*, we have*

$$\sqrt{nh}(\hat{g}_{PI}(t_0)-g(t_0)-\tfrac{1}{2}\mu_2h^2g''(t_0))\sim N(0,\sigma_{PI}^2),$$

where $\sigma_{PI}^2 = v_0 f^{-1}(t_0)E(\check{\varepsilon}^2|t_0)$*,* $\sigma_{PI}^2 = \sigma_{TI}^2 + v\sigma_{diff1}^2$*.*

Proof. Considering that in a practical problem the overall information is not completely known, so we can only extract partial information from the observed unlabeled data, we next consider the estimation under partial information (PI). We study the intercept model (2.7) under the semi-supervised setting.

In the same way as (ii), the local linear regression by varying coefficients shows that $\hat{\check{\beta}}(t_0) = (\check{D}^\top K_w\check{D})^{-1}(\check{D}^\top K_w\check{Y})$, $\hat{\check{g}}(t_0) = e_1^\top\hat{\check{\beta}}(t_0)$. Where the line i elements of $\check{D}$ are $(\check{X}_i^\top, \check{X}_i^\top(\frac{T_i-t_0}{h}))$ and $\check{X}_i = (\check{X}_{01i}, \check{X}_{02i})^\top$.

First, let us say

$$\check{E}(X^2|T) = \frac{\sum_{i=1}^{n+m}X_i^2K_h(T_i-T)}{\sum_{i=1}^{n+m}K_h(T_i-T)} \triangleq \check{m}(T),$$

and let $n+m=N$. When $T=t$, we say $m(t)=E(X^2|t)$.

From the properties of nonparametric kernel estimation, we can get

$$\sup_t|\check{m}(t)-m(t)| = O_{a.s.}(\frac{\sqrt{\ln N}}{\sqrt{Nh}}+h^2),$$

and from that, we can get

$$\sup|\check{E}(X^2|T=t)-E(X^2|T=t)| = \sup|\check{E}(X^2|t)-E(X^2|t)| = O_p((\frac{\ln N}{Nh})^{\frac{1}{2}}+h^2).$$

So we have

$$\frac{X_i^2}{\check{E}(X^2|T)} - \frac{X_i^2}{E(X^2|T)} = X_i^2 \frac{1}{\check{E}(X^2|T)E(X^2|T)} O_p((\frac{\ln N}{Nh})^{\frac{1}{2}} + h^2),$$

and we let $O_p((\frac{\ln N}{Nh})^{\frac{1}{2}} + h^2) = \triangle_{h_N}$. It is easy to get that

$$\frac{X_i^2}{\check{E}(X^2|T)} = \frac{X_i^2}{E(X^2|T)} + \triangle_{h_N}.$$

From the above description, we have

$$\check{X}_i\check{X}_i^\top = \begin{pmatrix} \check{X}_{01i} \\ \check{X}_{02i} \end{pmatrix}\begin{pmatrix} \check{X}_{01i} \\ \check{X}_{02i} \end{pmatrix}^\top = \begin{pmatrix} 1 & \frac{X_i^2}{\check{E}(X^2|T_i)} - 1 \\ \frac{X_i^2}{\check{E}(X^2|T_i)} - 1 & [\frac{X_i^2}{\check{E}(X^2|T_i)} - 1]^2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{X_i^2}{E(X^2|T_i)} - 1 + \triangle_{h_N} \\ \frac{X_i^2}{E(X^2|T_i)} - 1 + \triangle_{h_N} & [\frac{X_i^2}{E(X^2|T_i)} - 1]^2 + \triangle_{h_N}^2 \end{pmatrix}.$$

We have the following form:

$$\frac{1}{n}\check{D}^\top K_w \check{D} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \check{X}_i\check{X}_i^\top K_h(T_i - t_0) & \frac{1}{n}\sum_{i=1}^n \check{X}_i\check{X}_i^\top K_h(T_i - t_0)\frac{T_i - t_0}{h} \\ \frac{1}{n}\sum_{i=1}^n \check{X}_i\check{X}_i^\top K_h(T_i - t_0)\frac{T_i - t_0}{h} & \frac{1}{n}\sum_{i=1}^n \check{X}_i\check{X}_i^\top K_h(T_i - t_0)(\frac{T_i - t_0}{h})^2 \end{pmatrix}.$$

Similar to the previous steps,

$$\frac{1}{n}\sum_{i=1}^n \check{X}_i\check{X}_i^\top K_h(T_i - t_0) \xrightarrow{P} \begin{pmatrix} f(t_0) & 0 \\ 0 & E\{[\frac{X^2}{E(X^2|T)} - 1]^2|T = t_0\}f(t_0) \end{pmatrix}.$$

Let $\check{\mathbf{X}} = \begin{pmatrix} \check{X}_1^\top \\ \vdots \\ \check{X}_n^\top \end{pmatrix}$, and considering $\frac{1}{n}\check{D}^\top K_w \check{\mathbf{X}}\Delta_{ga}$ and $\frac{1}{n}\check{D}^\top K_w \check{\varepsilon}$, let $\frac{1}{n}\check{D}^\top K_w \check{\mathbf{X}}\Delta_{ga} = \check{A}_1^{(1)}$, $\frac{1}{n}\check{D}^\top K_w \check{\varepsilon} = \check{A}_2^{(2)}$, $\check{A}_1 = e_1^\top \check{A}_1^{(1)}$, $\check{A}_2 = e_1^\top \check{A}_2^{(2)}$, where the first element of e_1 is 1 and the others are 0. We have

$$\check{D}^\top K_w \check{\mathbf{X}}\Delta_{ga} = \begin{pmatrix} \sum_{i=1}^n K_h(T_i - t_0)\check{X}_i\check{X}_i^\top \Delta_{ga}(T_i) \\ \sum_{i=1}^n K_h(T_i - t_0)\frac{T_i - t_0}{h}\check{X}_i\check{X}_i^\top \Delta_{ga}(T_i) \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \check{A}_1 &= \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\check{X}_{01i}^2[g(T_i) - g(t_0) - g'(t_0)(T_i - t_0)] \\ &\quad + \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)\check{X}_{01i}\check{X}_{02i}[a(T_i) - a(t_0) - a'(t_0)(T_i - t_0)] \\ &= \check{A}_{11} + \check{A}_{12}, \end{aligned}$$

where

$$\check{A}_{11} = \frac{1}{n}\sum_{i=1}^n K_h(T_i - t_0)[g(T_i) - g(t_0) - g'(t_0)(T_i - t_0)],$$

and

$$\begin{aligned}\check{A}_{12} &= \frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\left(\frac{X_i^2}{\check{E}(X^2|T_i)} - 1\right)[a(T_i) - a(t_0) - a'(t_0)(T_i - t_0)] \\ &= \frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\left(\frac{X_i^2}{E(X^2|T_i)} - 1\right)[a(T_i) - a(t_0) - a'(t_0)(T_i - t_0)] + \triangle_{h_N}.\end{aligned}$$

Similarly for $\tilde{A}_1$, $\check{A}_1 = \check{A}_{11} + \check{A}_{12} = \frac{1}{2}g''(t_0)h^2\mu_2 f(t_0) + o(h^4) + \triangle_{h_N}$. On the other hand,

$$\check{A}_2^{(2)} = \frac{1}{n}\check{D}^\top K_w \check{\varepsilon} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i \\ \frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{X}_{02i}\check{\varepsilon}_i \end{pmatrix},$$

$$\check{A}_2 = e_1^\top \check{A}_2^{(2)} = \frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i.$$

Note that $\hat{\check{\beta}} - \check{\beta} = (\frac{1}{n}\check{D}^\top K_w \check{D})^{-1}(\frac{1}{n}\check{D}^\top K_w \check{\mathbf{X}}\Delta_{ga} + \frac{1}{n}\check{D}^\top K_w \check{\varepsilon})$, so we have

$$\hat{\check{g}}(t_0) - g(t_0) - \frac{1}{2}h^2\mu_2 g''(t_0) = f^{-1}(t_0)\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i + o(h^4).$$

Since

$$E(\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i) = E(K_h(T_i - t_0)\check{\varepsilon}_i) = E\left(K_h(T_i - t_0)E(\check{\varepsilon}_i|X, T)\right) = 0,$$

$$E(\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i)^2 = \frac{1}{nh}f(t_0)E(\check{\varepsilon}^2|t_0)\nu_0 + o(\frac{1}{nh}),$$

letting $\hat{\check{g}}(t_0) = \hat{g}_{PI}(t_0)$, we can get

$$\sqrt{nh}(\hat{\check{g}}_{PI}(t_0) - g(t_0) - \frac{1}{2}h^2\mu_2 g''(t_0)) \sim N(0, \nu_0 f^{-1}(t_0)E(\check{\varepsilon}^2|t_0)), \tag{2.19}$$

where $\sigma_{PI}^2 = \nu_0 f^{-1}(t_0)E(\check{\varepsilon}^2|t_0)$. We have

$$\begin{aligned}\frac{1}{nh}\sigma_{PI}^2 &= E(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\check{\varepsilon}_i)^2 \\ &= E(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)\tilde{\varepsilon}_i)^2 + E(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)(\check{\varepsilon}_i - \tilde{\varepsilon}_i))^2 \\ &+ 2E[(\frac{1}{f(t_0)})^2(\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0))(\frac{1}{n}\sum_{j=1}^{n} K_h(T_j - t_0))\tilde{\varepsilon}_i(\check{\varepsilon}_j - \tilde{\varepsilon}_j)],\end{aligned}$$

where

$$E[(\frac{1}{f(t_0)})^2(\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0))(\frac{1}{n}\sum_{j=1}^{n} K_h(T_j - t_0))\tilde{\varepsilon}_i(\check{\varepsilon}_j - \tilde{\varepsilon}_j)] = 0 + o(\frac{1}{nh}).$$

Let $nhE(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)(\check{\varepsilon}_i - \tilde{\varepsilon}_i))^2 = \sigma_{diff2}^2$, so we have

$$\sigma_{PI}^2 = \sigma_{TI}^2 + \sigma_{diff2}^2.$$

Since

$$\begin{aligned}\check{\varepsilon}_i - \tilde{\varepsilon}_i &= (\check{Y}_i - g(T_i)\check{X}_{01i} - a(T_i)\check{X}_{02i}) - (\tilde{Y}_i - g(T_i)\tilde{X}_{01i} - a(T_i)\tilde{X}_{02i}) \\ &= (Y_iX_i - a(T_i)X_i^2)(\frac{1}{\check{E}(X^2|T_i)} - \frac{1}{E(X^2|T_i)}) \\ &= (g(T_i) - a(T_i))(\frac{X_i^2}{\check{E}(X^2|T_i)} - \frac{X_i^2}{E(X^2|T_i)}) + o(\frac{1}{nh}),\end{aligned} \tag{2.20}$$

and

$$\begin{aligned}\frac{X_i^2}{\check{E}(X^2|T)} - \frac{X_i^2}{E(X^2|T)} &= (X_i^2 - \check{E}(X^2|T))(\frac{1}{\check{E}(X^2|T)} - \frac{1}{E(X^2|T)}) + 1 - \frac{\check{E}(X^2|T)}{E(X^2|T)} \\ &= -\frac{\check{E}(X^2|T)}{E(X^2|T)} + 1 + \triangle_{h_N},\end{aligned}$$

we can get

$$\begin{aligned}\check{\varepsilon}_i - \tilde{\varepsilon}_i &= (g(T_i) - a(T_i))(\frac{X_i^2}{\check{E}(X^2|T)} - \frac{X_i^2}{E(X^2|T)}) + o(\frac{1}{nh}) \\ &= (a(T_i) - g(T_i))(\frac{\check{E}(X^2|T)}{E(X^2|T)} - 1) + \triangle_{h_{N_n}} \\ &= (a(T_i) - g(T_i))\left[\frac{\frac{1}{n+m}\sum_{j=1}^{n+m} K_h(T_j - T_i)X_j^2}{\frac{1}{n+m}\sum_{j=1}^{n+m} K_h(T_j - T_i)E(X^2|T)} - 1\right] + \triangle_{h_{N_n}},\end{aligned}$$

where $\triangle_{h_{N_n}} = \triangle_{h_N} + o(\frac{1}{nh})$, and

$$\begin{aligned}&\frac{1}{nh}\sigma_{diff2}^2 \\ =\ & E(\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)(\check{\varepsilon}_i - \tilde{\varepsilon}_i))^2 \\ =\ & E\left[\frac{1}{f(t_0)}\frac{1}{n}\sum_{i=1}^{n} K_h(T_i - t_0)(a(T_i) - g(T_i))\left[\frac{\frac{1}{n+m}\sum_{j=1}^{n+m} K_h(T_j - T_i)X_j^2}{\frac{1}{n+m}\sum_{j=1}^{n+m} K_h(T_j - T_i)E(X^2|T_i)} - 1\right]\right]^2 + O(\triangle_{h_{N_n}})^2 \\ =\ & E\left[\frac{1}{f(t_0)}\frac{1}{n+m}\sum_{j=1}^{n+m} K_h(T_j - T_i)\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(T_i)}K_h(T_i - t_0)(a(T_i) - g(T_i))\left(\frac{X_j^2}{E(X^2|T_i)} - 1\right)\right]\right]^2 + O(\triangle_{h_{N_n}})^2 \\ \triangleq\ & E\left[\frac{1}{f(t_0)}\left[\frac{1}{n+m}\sum_{j=1}^{n+m}\left[X_j^2 W_1\right] - \frac{1}{n+m}\sum_{j=1}^{n+m}[W_2]\right]\right]^2 + O(\triangle_{h_{N_n}})^2,\end{aligned}$$

where

$$\begin{aligned}W_1 &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(T_i)}K_h(T_j - T_i)K_h(T_i - t_0)(a(T_i) - g(T_i))\frac{1}{E(X^2|T_i)} \\ &= K_h(T_j - t_0)(a(t_0) - g(t_0))\frac{1}{E(X^2|t_0)} + o(h^2),\end{aligned}$$

and

$$W_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{f(T_i)}K_h(T_j - T_i)K_h(T_i - t_0)(a(T_i) - g(T_i)) = K_h(T_j - t_0)(a(t_0) - g(t_0)) + o(h^2).$$

So, we can get, for any given t_0

$$\begin{aligned}
\sigma^2_{diff2} &= nhE\left[\frac{1}{f(t_0)}\frac{1}{n+m}\sum_{j=1}^{n+m}\left[K_h(T_j - t_0)(a(t_0) - g(t_0))\left[\frac{X_j^2}{E(X^2|t_0)} - 1\right]\right] + o(h^2)\right]^2 + O(nh\triangle^2_{h_{N_n}}) \\
&= nh\left[\frac{1}{f^2(t_0)}\frac{1}{n+m}E\left[K_h(T_j - t_0)(a(t_0) - g(t_0))\left(\frac{X^2}{E(X^2|t_0)} - 1\right)\right]^2 + o(h^4)\right] + O(nh\triangle^2_{h_{N_n}}) \\
&= nh\left[\frac{1}{f(t_0)}\frac{1}{n+m}E\left[(a(t_0) - g(t_0))\left(\frac{X^2}{E(X^2|t_0)} - 1\right)\right]^2\frac{\int K^2(u)du}{h} + o(h^4)\right] + O(nh\triangle^2_{h_{N_n}}) \\
&= \frac{n}{n+m}\frac{\int K^2(u)du}{f(t_0)}E\left[(a(t_0) - g(t_0))\left(\frac{X^2}{E(X^2|t_0)} - 1\right)\right]^2 + o(nh^5) + O(nh\triangle^2_{h_{N_n}})
\end{aligned}$$

Suppose $lim\frac{n}{n+m} = v$, we can get $\sigma^2_{diff2} = v\sigma^2_{diff1}$, then we have $\sigma^2_{PI} = \sigma^2_{TI} + v\sigma^2_{diff1}$. □

Remark 2. *Theorems 1–3 show that if σ^2_{diff1} is not equal to zero, then σ^2_L is larger than σ^2_{TI}, and, further, if $v < 1$, then σ^2_L is also larger than σ^2_{PI}. This means that the proposed semi-supervised estimators could be more efficient than the supervised estimators.*

3. Simulation studies and results

To study the finite sample performance of the proposed semi-supervised estimates, this section uses some numerical simulations and a real data analysis to compare the estimated performances of the locally weighted linear estimation of the varying coefficient model (LWLR), the total information estimator based on the intercept model (TI) and the semi-supervised estimator under the partial information setting (PI).

3.1. Monte Carlo simulations

Example. *Consider*

$$\begin{aligned}
Model\,(I): &\quad Y = sin(\pi T)X + \underbrace{\frac{1}{2}TX^2 + \delta}_{\varepsilon 1}, \\
Model\,(II): &\quad Y = (1-T)^2X + \underbrace{exp(-\frac{1}{2}T)X^2 + \delta}_{\varepsilon 2},
\end{aligned}$$

where $X \sim N(0,1)$, T comes from the uniform distribution $U(0,1)$ and random error $\delta \sim N(0,1)$. At the same time, $\varepsilon 1$ and $\varepsilon 2$ are affected by X and T. The labeled samples (Y_i, X_i, T_i), $i = 1, 2, ..., n$ and the unlabeled samples (X_i, T_i), $i = n+1, n+2, ..., n+m$ are from Model (I) or (II) with sample size n and m, *respectively. Let the kernel function be the Epanechnikov kernel function,* $K(u) = 0.75(1-u^2)I(|u| \le 1)$. *We determine the bandwidth h of the kernel function through the process of cross-validation (CV). By employing CV, we aim to identify an optimal bandwidth value that enhances the performance of the kernel function in the given context. For numerical simulation, the labeled sample size is* n=10, 20, 40, 60 *and the unlabeled sample size is* $m = n, 3n, 10n$ *for each model. To compare the pros and cons of the estimators obtained by different methods, the estimation curves are obtained by 1000 repeated simulations, and the estimation effect is evaluated by the root mean squared error (RMSE) and standard deviation (STD) of the estimations, where* $RMSE = \sqrt{\frac{1}{q}\sum_{j=1}^{q}\frac{1}{n}\sum_{k=1}^{n}[\hat{g}_j(t_k) - g(t_k)]^2}$, *and* $STD = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\frac{1}{q}\sum_{j=1}^{q}[\hat{g}_j(t_k) - \frac{1}{q}\sum_{j=1}^{q}\hat{g}_j(t_k)]^2}$. *The expression "*$\hat{g}_j(t_k)$*" represents the jth estimation of the coefficient function at the discrete time point* t_k, *where* $\{t_k, k = 1, 2, \ldots, n\}$ *refers to the appropriate grid point. The simulation results are shown in Table 1.*

Table 1. RMSE and STD derived in Model (I) and (II).

			LWLR	TI	PI		
		n \ m	–	–	n	3n	10n
Model(I)	RMSE	10	6.1491	3.5380	3.8323	3.7895	3.8279
		20	2.2740	1.6919	1.6775	1.7452	1.7575
		40	0.8886	0.7658	0.8036	0.7853	0.7644
		60	0.7847	0.7580	0.7813	0.7781	0.7764
	STD	10	6.1192	3.4779	3.7803	3.7370	3.7762
		20	2.1842	1.5691	1.5536	1.6264	1.6396
		40	0.3583	0.3053	0.3505	0.3512	0.3019
		60	0.2476	0.2394	0.2163	0.2044	0.2340
Model(II)	RMSE	10	6.1408	3.5271	3.8242	3.7821	3.8205
		20	2.2067	1.6047	1.5934	1.6616	1.6692
		40	0.6967	0.5785	0.6292	0.6046	0.5761
		60	0.5003	0.4976	0.5088	0.4972	0.4990
	STD	10	6.1225	3.4927	3.7945	3.7522	3.7909
		20	2.1888	1.5801	1.5685	1.6376	1.6453
		40	0.4704	0.3405	0.4208	0.3831	0.3365
		60	0.2445	0.2390	0.2415	0.2383	0.2420

As can be seen from Table 1, the proposed total information estimator and the semi-supervised partial information estimator based on the intercept model are generally superior to the ordinary locally weighted linear regression estimator, and the RMSE and STD obtained by the former two methods are generally smaller than the latter one. The proposed semi-supervised intercept model can improve the efficiency of the estimator.

3.2. Real data application

In recent years, the introduction of shared bicycles has enriched the travel and transportation options of urban residents and solved the "last kilometer" travel problem of residents. Shared cycling has a significant impact on building a larger riding community, minimizing greenhouse gas emissions and improving public health and transport issues. But, shared bicycle travel is easily affected by the weather environment, where the temperature is an important disturbance factor. We study the data set of the number of shared bicycles rented hourly in Seoul from 2017.12–2018.11 [11]. There are a total of N=8760 samples in the data set. To study the superiority of the estimation method proposed in this paper, the variables selected include count of bikes rented at each hour (vehicles, Y), hour of the day ($0:00-23:00$, T) and temperature (°C, X). At the same time, n=1000 labeled samples and m=7760 unlabeled samples are randomly selected in the original data set to construct a semi-supervised setting, and the semi-supervised intercept model method proposed in this paper is used for analysis. We establish the following model:

$$Y = \beta(T)X + \varepsilon. \tag{3.1}$$

Compare the estimated R^2 obtained by semi-supervised intercept model estimation R^2_{PI}, the local linear regression estimation $R^2_{LWLR_N}$ under all supervised data and $R^2_{LWLR_n}$ under n supervised samples. At the same time, in order to eliminate the influence of randomness, we repeated $K = 50$ times and calculated the mean of the corresponding $R^2(k)$ respectively, that is, $R^2 = \frac{1}{K}\sum_{k=1}^{K} R^2(k)$. The goodness of fit based on N samples is $R^2_{LWLR_N} = 0.8531$, the goodness of fit based on n samples is $R^2_{LWLR_n}$=0.7651 and $R^2_{PI} = 0.8411$, the effect is significantly better than that of the local linear regression estimation without using unlabeled data.

4. Conclusions

Semi-supervised data is becoming more common, and most semi-supervised learning methods focus on classification tasks, or solving linear regression models, with less emphasis on varying coefficient models. Therefore, a good estimate of the coefficient function of the varying coefficient model is given in this work. The key idea is to introduce an intercept model to replace the original varying coefficient regression model and perform the estimation under a semi-supervised setting; that is, the information of unlabeled data is utilized in the estimation process. It is further proved that the new estimates have good asymptotic properties. At the same time, the asymptotic properties of the new estimate is better than that of the conventional locally weighted linear regression estimators.

Finally, the method is applied to study the effect of temperature on the demand for shared bicycle rental. The coefficient function of the shared bicycle rental demand model is well estimated, and the demand is predicted.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Peng Lai's research is supported by National Natural Science Foundation of China (11771215). Yanqiu Zhou's research is supported by Guangxi Science and Technology Base and Talent Project (2020ACI9151), and Guangxi University Young and Middle-aged Teachers' Basic Research Ability Improvement Project (2021KY0343).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. D. Azriel, L. Brown, M. Sklar, R. Berk, A. Buja, L. Zhao, Semi-supervised linear regression, *J. Am. Stat. Assoc.*, **117** (2022), 2238–2251. https://doi.org/10.1080/01621459.2021.1915320
2. T. Cai, Z. Guo, Semi-supervised inference for explained variance in high dimensional linear regression and its applications, *J. R. Stat. Soc. B*, **82** (2020), 391–419. https://doi.org/10.1111/rssb.12357
3. Z. Cai, J. Fan, R. Li, Efficient estimation and inferences for varying-coefficient models, *J. Am. Stat. Assoc.*, **95** (2000), 888–902. https://dx.doi.or/10.1080/01621459.2000.10474280
4. A. Chakrabortty, T. Cai, Efficient and adaptive linear regression in semi-supervised settings, *Ann. Statist.*, **46** (2018), 1541–1572. https://doi.org/10.1214/17-AOS1594
5. J. Fan, R. Gijbels, *Local polynomial modelling and its applications*, New York: Routledge, 1996. https://doi.org/10.1201/9780203748725
6. J. Fan, W. Zhang, Statistical estimation in varying coefficient models, *Ann. Statist.*, **27** (1999), 1491–1518. https://dx.doi.or/10.1214/aos/1017939139
7. Z. Fang, L. Lu, F. Liu, G. Zhang, Semi-supervised heterogeneous domain adaptation: theory and algorithms, *IEEE T. Pattern Anal.*, **45** (2023), 1087–1105. https://dx.doi.or/10.1109/TPAMI.2022.3146234
8. S. Feng, B. Li, H. Yu, Y. Liu, Q. Yang, Semi-supervised federated heterogeneous transfer learning, *Knowl.-Based Syst.*, **252** (2022), 109384. https://dx.doi.or/10.1016/j.knosys.2022.109384
9. P. Lai, Q. Zhang, H. Lian, Q. Wang, Efficient estimation for the heteroscedastic single-index varying coefficient models, *Stat. Probabil. Lett.*, **110** (2016), 84–93. https://dx.doi.or/10.1016/j.spl.2015.12.005
10. C. Merz, D. Clair, W. Bond, SeMi-supervised adaptive resonance theory (SMART2), *Proceedings of International Joint Conference on Neural Networks*, 1992, 851–856. https://dx.doi.org/10.1109/IJCNN.1992.227046
11. V. Sathishkumar, J. Park, Y. Cho, Using data mining techniques for bike sharing demand prediction in metropolitan city, *Comput. Commun.*, **153** (2020), 353–366. https://doi.org/10.1016/j.comcom.2020.02.007
12. L. Tang, Z. Zhou, Weighted local linear CQR for varying-coefficient models with missing covariates, *TEST*, **24** (2015), 583–604. https://doi.org/10.1007/s11749-014-0425-z
13. V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, A. Solin, Y. Bengio, et al., Interpolation consistency training for semi-supervised learning, *Neural Networks*, **145** (2022), 90–106. https://dx.doi.or/10.1016/j.neunet.2021.10.008
14. J. Wang, X. Shen, W. Pan, On efficient large margin semisupervised learning: method and theory, *J. Mach. Learn. Res.*, **10** (2009), 719–742.

15. A. Zhang, L. Brown, T. Cai, Semi-supervised inference: general theory and estimation of means, *Ann. Statist.*, **47** (2019), 2538–2566. https://doi.org/10.1214/18-AOS1756

16. Y. Zhang, J. Bradic, High-dimensional semi-supervised learning: in search of optimal inference of the mean, *Biometrika*, **109** (2022), 387–403. https://doi.org/10.1093/biomet/asab042

17. W. Zhao, J. Li, H. Lian, Adaptive varying-coefficient linear quantile model: a profiled estimating equations approach, *Ann. Inst. Stat. Math.*, **70** (2018), 553–582. https://dx.doi.or/10.1007/s10463-017-0599-8

18. Z. Zhou, M. Li, Semi-supervised regression with co-training, *Proceedings of the 19th international joint conference on Artificial intelligence*, 2005, 908–913.

19. X. Zhu, *Semi-supervised learning literature survey*, Madison: University of Wisconsin-Madison Publisher, 2005.

AIMS Press