Job Market Paper

Binary Response Correlated Random Coefficient Panel Data Models

Zhongwen Liang*

This version: January 18, 2012

Abstract

In this paper, we consider binary response correlated random coefficient (CRC) panel data models which are frequently used in the analysis of treatment effects and demand of products. We focus on the nonparametric identification and estimation of panel data models under unobserved heterogeneity which is captured by random coefficients and when these random coefficients are correlated with regressors. For the analysis of treatment effects, under some circumstances, the average treatment effect can be estimated via a linear CRC model. We give the identification conditions for the average slopes of a linear CRC model with a general nonparametric correlation between regressors and random coefficients. We construct a $\sqrt{n}$ consistent estimator for the average slopes via varying coefficient regression. The identification of binary response panel data models with unobserved heterogeneity is difficult. We base identification conditions and estimation on the framework of the model with a special regressor, which is a major approach proposed by Lewbel (1998, 2000) to solve the heterogeneity and endogeneity problem in the binary response models. With the help of the additional information on the special regressor, we can transfer a binary response CRC model to a linear moment relation. We also construct a semiparametric estimator for the average slopes and derive the $\sqrt{n}$-normality result. Simulations are given to show the finite sample advantage of our estimators. Further, we use a linear CRC panel data model to reexamine the return from job training. The results show that our estimation method really makes a difference, and the estimated return of training by our method is 10 times as much as the one estimated without considering the correlation between the covariates and random coefficients. It shows that on average the rate of return of job training is 6.1% per 100 hours training, which is reasonable.

Keywords: random coefficient; binary choice; nonparametric; panel data.

JEL Classification: C14, C33.

*Department of Economics, Texas A&M University, College Station, TX 77843-4228.
Email: zhwliang@neo.tamu.edu

1 Introduction

Recently, the correlated random coefficient model has drawn much attention. As stated in Heckman, Schmierer, and Urzua (2010), "The correlated random coefficient model is the new centerpiece of a large literature in microeconometrics". In this paper, we consider binary response CRC panel data models in the form of

$$y_{it} = \mathbf{1}(v_{it}^\top \gamma + x_{it}^\top \beta_i + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{1.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, v_{it} denotes regressors with constant coefficient γ, x_{it} denotes regressors with random coefficient β_i, u_{it} is the error term. x_{it} can include 1 as a component. Thus, the binary choice panel data models with fixed effects are special cases of this model. We allow the general correlation between the random coefficient β_i and the regressor x_{it}. We focus on the nonparametric identification and estimation of the mean of random slope β_i in this model and related transformed models, which will be more specific in later sections.

Binary choice panel data models are widely used by applied researchers. One reason is its direct economic interpretability. Another reason is that given the advantage of panel data with multiple observations of the same individual over several time periods, it is possible to take into account unobserved heterogeneity. The common approach is to include an individual-specific heterogenous effect variable additively, which leads to a correlated random effects model or a fixed effects model. The advantage of this approach is that we can eliminate the unobservable variable by taking the difference between different time periods and get the fixed effects estimator for linear models easily, see e.g. Arellano (2003), Hsiao (2003). This also resolves the incidental parameter problem in linear panel data models. The method of taking difference can also be extended to nonlinear panel data models in certain extent, see Bonhonmme (2011). Though it is convenient to deal with unobserved heterogeneity additively, economic models imply many different non-additive forms, see Browning and Carro (2007), Imbens (2007). Among them, one class is the random coefficient model which arises from the demand analysis with the consideration of the individual heterogeneity.

Random coefficient models have the multiplicative individual heterogeneity. They are popular in empirical analysis of treatment effects and the demand of products. In the analysis of treatment effect, under certain circumstances, the binary choice fixed-effects model can be transferred to a linear random coefficient model with the average treatment effect being the mean of a random coefficient. For instance, in one of the commenting papers for Angrist (2001), Hahn (2001) gives an example on this transformation and discusses the consistency of the fixed effects estimator. Wooldridge (2005) further allows the correlation between regressors and random coefficients and gives the conditions that assure the consistency of the fixed effects estimator. Motivated by the usefulness of linear CRC panel data models from this transformation, we discuss the identification and estimation of the linea CRC panel data models in sections 2 and 3, which will also serve as an important piece towards the semiparametric estimation of the binary response CRC panel data model.

In the literature of demand analysis, Berry, Levinsohn and Pakes (1995) propose to use the random coefficients logit multinomial choice model to study the demand of automobiles which has become the major vehicle of the demand analysis. However, they leave the correlation between the random coefficients and the regressors unconsidered, and have assumptions on the functional form of the distributions of the unobservable variables. In this paper, we study random coefficient binary choice models without specifying the functional form of the distribution of unobservable variables. Also, we allow for non-zero correlation between regressors and random coefficients. For simplicity, we only consider binary choice models.

Recently, there is a growing literature on CRC models. Graham and Powell (2010) discuss the identification and estimation of average partial effects in a class of "irregular" correlated random coefficient panel data models using different information of agents from subpopulations, so called "stayers" and "movers". Due to the irregularity, they get an estimator with slower than $\sqrt{n}$ convergence rate and the normal limiting distribution. Heckman, Schmierer, and Urzua (2009) and Heckman and Schmierer (2010) investigate the tests of the CRC model.

Other related literature includes three aspects: random coefficient models, panel data models with unobserved heterogeneity, and models with a special regressor. Both of these literatures have been developed considerably in the last two decades. Random coefficient models have a long history. Swamy and Tavlas (2007) and Hsiao and Pesaran (2008) are good surveys for these models. For binary random coefficient models, Hoderlein (2009) consider a binary choice model with endogenous regressors under a weak median exclusion restriction. He uses a control function IV approach to identify the local average structural effect of the regressors on the latent variable, and derives $\sqrt{n}$ consistency and the asymptotic distribution of the estimator he proposed. He also proposes tests for heteroscedasticity, overidentification and endogeneity. Some parts of the literature concern distributions of the random coefficients. Recent ones include Arellano and Bonhomme (2010), Fox and Gandhi (2010), Hoderlein, Klemelä and Mammen (2010).

Among the recent developments of panel data models, the nonseparable panel data models is an indispensable part. Chernozhukov, Fernandez-Val and Newey (2009) investigate quantile and average effects in nonseparable panel models. Evdokimov (2010) discusses the identification and estimation of a nonparametric panel data model with nonseparable unobserved heterogeneity. He obtains point identification and estimation via conditional deconvolution. Hoderlein and White (2010) give nonparametric identification in nonseparable panel data models with generalized fixed effects.

The identification of discrete choice model is different from linear models. The framework we adopt in this paper for the identification of the average slope in binary response CRC panel data models is the special regressor method, which assumes the existence of a special regressor with additional information. Proposed by Lewbel (1998, 2000), this method has been exploited extensively in different settings. It is an effective way for identification and estimation of heterogeneity and endogeneity. Honoré and Lewbel (2002) use this method to study a binary choice fixed effects model which allows for general predetermined explanatory variables and give a $\sqrt{n}$ consistent semiparametric estimator. Khan and Lewbel (2007) investigate a truncated regression model using this method and propose a $\sqrt{n}$ consistent and asymptotically normal estimator. Dong and Lewbel (2011) give a good survey for this method.

The rest of the paper is organized as follows. In section 2, we discuss the identification for the linear CRC models with nonparametric correlation between regressors and random coefficients. In section 3, we construct semiparametric estimators for average slopes by kernel methods. We give the asymptotics for both local constant method and local polynomial method. Section 4 gives the identification conditions for the binary response CRC panel data models. The estimator and the $\sqrt{n}$-normality result are given in section 5. We conduct extensive simulations in section 6 to show the finite sample advantage of our estimators. In Section 7, we give an empirical application using a linear CRC panel data model to reexamine the return from job training. It turns out that our method can really make a difference compared with the first difference method for panel data models. Section 8 concludes the paper and discusses further extensions of the paper. All of proofs are relegated to two appendices.

2 Identification of Linear CRC Models

In this section we consider the identification conditions for linear CRC panel data models. The linear CRC panel data models can be motivated as follows, which is given in Hahn (2001).

Suppose we have an unobserved fixed effects panel probit model with two periods, $P(y_{it} = 1|c_i, x_{i1}, x_{i2}) = \Phi(c_i + \theta x_{it})$, $i = 1, \ldots, n$, $t = 1, 2$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function, c_i is the unobserved heterogenous effect, and x_{it} denotes a binary treatment variable. It is difficult to identify the slope coefficient θ without additional assumptions on the conditional distribution of c_i conditioning on (x_{i1}, x_{i2}). However, the average treatment effect $\beta = E[\Phi(c_i + \theta) - \Phi(c_i)]$ can be analyzed by a transformation, i.e., we can transfer the probit model to a linear random coefficient model, $y_{it} = a_i + b_i x_{it} + u_{it}$, $i = 1, \ldots, n$, $t = 1, 2$, where $a_i \equiv \Phi(c_i)$, $b_i \equiv \Phi(c_i + \theta) - \Phi(c_i)$, and $u_{it} \equiv y_{it} - E(y_{it}|x_{i1}, x_{i2}, c_i)$. Hahn assumes the independence of y_{i1} and y_{i2} conditional on (x_{i1}, x_{i2}, c_i). He also assumes $(x_{i1}, x_{i2}) = (0, 1)$ which means no individual is treated in the first period and all are treated in the second period, and which also implies the independence of treatment variables (x_{i1}, x_{i2}) and the unobserved heterogeneity c_i. In general, x_{it} could be correlated with c_i.

We consider the linear random coefficient models with general correlation between random coefficients and regressors in sections 2 and 3. For simplicity, we assume there is no regressor with constant coefficient in model (1.1) in sections 2 and 3. In section 2.1 we first consider a CRC model with cross sectional data. We discuss how to obtain consistent estimate for the mean slope coefficient. In this case, the condition for the identification of the average effect is quite stringent, and may even be unrealistic for many applications. We then show that panel data can provide more information and help to identify the mean slopes. The identification conditions when panel data is available are given in section 2.2.

2.1 The Cross Sectional Data Case

We consider the following CRC model with cross sectional data.

$$y_i = x_i^\top \beta_i + u_i, \qquad (i = 1, ..., n) \tag{2.1}$$

where x_i is a $d \times 1$ vector, $\beta_i = \beta + \alpha_i$ is of dimension $d \times 1$, β is a $d \times 1$ constant vector, α_i is i.i.d. with $(0, \Sigma_\alpha)$, Σ_α is a $d \times d$ positive definite matrix, the superscript $^\top$ denotes the transpose, and u_i is i.i.d. with $(0, \sigma_u^2)$ and is orthogonal to (x_i, α_i), i.e., $E(u_i|x_i, \alpha_i) = 0$. We allow for α_i to be arbitrarily correlated with x_i. Let $E(\alpha_i|x_i) = g(x_i)$, where $g(\cdot)$ is a smooth function but its specific functional form is not specified. For example we could have $g(x_i) = \Gamma(x_i - E(x_i))$, where Γ is $d \times d$ matrix of constants. However, we allow for $g(x_i)$ to have any other unknown functional form.

Replacing β_i by $\beta + \alpha_i$, we can rewrite (2.1) as

$$\begin{aligned} y_i &= x_i^\top \beta + x_i^\top \alpha_i + u_i \\ &= x_i^\top \beta + v_i, \end{aligned} \tag{2.2}$$

where $v_i = x_i^\top \alpha_i + u_i$. Note that $E(v_i|x_i) = x_i^\top E(\alpha_i|x_i) = x_i^\top g(x_i) \neq 0$, so the OLS estimator of β based on (2.2) is biased and inconsistent in general. Indeed it is easy to see that the OLS estimator of β based on (2.2) is given by

$$\begin{aligned} \hat{\beta}_{OLS} &= \beta + \left[n^{-1} \sum_i x_i x_i^\top \right]^{-1} n^{-1} \sum_i [x_i x_i^\top \alpha_i + x_i u_i] \\ &\xrightarrow{p} \beta + [E(x_i x_i^\top)]^{-1} E[x_i x_i^\top \alpha_i], \end{aligned} \tag{2.3}$$

because $E[x_i u_i] = 0$. Hence, whether $\hat{\beta}_{OLS}$ consistently estimates β depends on whether $E[x_i x_i^\top \alpha_i] = 0$ or not.

For expositional simplicity let us consider a simple case that $x_i^\top = (1, \tilde{x}_i)$, where $\tilde{x}_i$ is a scalar. In this case we have ($\alpha_i = (\alpha_{1i}, \alpha_{2i})^\top$)

$$\begin{aligned} E[x_i x_i^\top \alpha_i] &= E\left[\begin{pmatrix} 1 & \tilde{x}_i \\ \tilde{x}_i & \tilde{x}_i^2 \end{pmatrix} \begin{pmatrix} \alpha_{1i} \\ \alpha_{2i} \end{pmatrix} \right] \\ &= \begin{pmatrix} E(\tilde{x}_i \alpha_{2i}) \\ E(\tilde{x}_i \alpha_{1i} + \tilde{x}_i^2 \alpha_{2i}) \end{pmatrix} \end{aligned} \tag{2.4}$$

where we use $E(\alpha_{1i}) = 0$. For $E[x_i x_i^\top \alpha_i]$ to be zero, from (2.4) we know that it requires α_{1i} to be orthogonal to $\tilde{x}_i$, and α_{2i} to be orthogonal to $\tilde{x}_i^2$, which are unlikely to be true in practice. Hence, $\hat{\beta}_{OLS}$ is biased and inconsistent for β in general.

Below we show that a semiparametric estimation method can consistently estimate β in a univariate CRC model. For a general multivariate regression model, additional assumptions are required for identification. For a univariate CRC model

$$y_i = x_i \beta_i + u_i,$$

where x_i is a scalar, $\beta_i = \beta + \alpha_i$, $E(\alpha_i) = 0$ and $E(u_i|x_i, \alpha_i) = 0$. Thus, $E(u_i|x_i) = 0$. Let $g(x_i) = E(\alpha_i|x_i)$, we have

$$E(y_i|x_i = x) = x(\beta + g(x)) \equiv x\theta(x),$$

where $\theta(x) = \beta + g(x)$. If $\theta(x)$ is identified, since $E(g(x_i)) = 0$ by $E(\alpha_i) = 0$, we have $\beta = E(\theta(x_i))$. For the univariate case, it is easy to identify $\theta(x)$ by $\theta(x) = E(y_i|x_i = x)/x$ (for $x \neq 0$). Hence, we can use the standard nonparametric estimation method to estimate $\theta(x)$. Say, by the local constant kernel method:

$$\hat{\theta}(x_i) = \left[\sum_j x_j^2 K_{h,ji}\right]^{-1} \sum_{j=1}^n x_j y_j K_{h,ji},$$

where $K_{h,ji} = K((x_j - x_i)/h)$, $K(\cdot)$ is the kernel density function, and h is the smoothing parameter. Then β can be consistently estimated by $n^{-1}\sum_{i=1}^n \hat{\theta}(x_i)$.

However, for a general multivariate regression model, β is not identified in general if only cross section data is available. We use a bivariate regression model to illustrate the difficulty of identification. Let $x_i = (x_{1i}, x_{2i})^\top$, and we consider a CRC model as

$$y_i = x_{1i}\beta_{1i} + x_{2i}\beta_{2i} + u_i, \tag{2.5}$$

with $\beta_{1i} = \beta_1 + \alpha_{1i}$, $\beta_{2i} = \beta_2 + \alpha_{2i}$, $E(\alpha_{1i}) = 0$, $E(\alpha_{2i}) = 0$, and $E(u_i|x_{1i}, x_{2i}, \alpha_{1i}, \alpha_{2i}) = 0$. Hence, we have $E(u_i|x_{1i}, x_{2i}) = 0$. Consequently, we have

$$E(y_i|x_{1i} = x_1, x_{2i} = x_2) \;\; = \;\; x_1\theta_1(x_1, x_2) + x_2\theta_2(x_1, x_2),$$

where $\theta_1(x_1, x_2) = \beta_1 + E(\alpha_{1i}|x_{1i} = x_1, x_{2i} = x_2)$ and $\theta_2(x_1, x_2) = \beta_2 + E(\alpha_{2i}|x_{1i} = x_1, x_{2i} = x_2)$. However, if we only have cross sectional data, $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are not identified, since $x_1\theta_1(x_1, x_2)+x_2\theta_2(x_1, x_2) = x_1\theta_3(x_1, x_2)+x_2(\frac{x_1}{x_2}\theta_1(x_1, x_2)-\frac{x_1}{x_2}\theta_3(x_1, x_2)+\theta_2(x_1, x_2)) \equiv x_1\theta_3(x_1, x_2)+ x_2\theta_4(x_1, x_2)$, where $\theta_4(x_1, x_2) = \frac{x_1}{x_2}\theta_1(x_1, x_2) - \frac{x_1}{x_2}\theta_3(x_1, x_2) + \theta_2(x_1, x_2)$, if $x_2 \neq 0$.

Put it in another view, from

$$E(y_i|x_{1i} = x_1, x_{2i} = x_2) = x_1\theta_1(x_1, x_2) + x_2\theta_2(x_1, x_2),$$

we have only one equation, and we cannot uniquely identify two unknown functions $\theta_1(\cdot)$ and $\theta_2(\cdot)$. It has infinitely many solutions.

Even though for $d \geq 2$ the cross section data model cannot identify β in general, it is possible to identify β under additional assumptions. Suppose there exists another random variable z_i such that

$$E(\alpha_i|x_{1i}, x_{2i}, z_i) = E(\alpha_i|z_i) = g(z_i), \tag{2.6}$$

for example, we may have $z_i = x_{1i} + x_{2i}$. (2.6) states that α_i is correlated with (x_{i1}, x_{i2}) only through z_i. Then model (2.5) can be rewritten as

$$\begin{aligned} y_i &= x_{1i}(\beta_1 + g_1(z_i)) + x_{2i}(\beta_2 + g_2(z_i)) + \epsilon_i \\ &= x_{i1}\theta_1(z_i) + x_{2i}\theta_2(z_i) + \epsilon_i \\ &= x_i^\top\theta(z_i) + \epsilon_i, \end{aligned} \tag{2.7}$$

where $g_1(z_i) = E(\alpha_{1i}|z_i)$, $g_2(z_i) = E(\alpha_{2i}|z_i)$, $\epsilon_i = x_{1i}(\alpha_{1i} - g_1(z_i)) + x_{2i}(\alpha_{2i} - g_2(z_i)) + u_i$, $x_i = (x_{1i}, x_{2i})^\top$, and $\theta(z_i) = (\theta_1(z_i), \theta_2(z_i))^\top$. By construction, $E(\epsilon_i|x_{1i}, x_{2i}, z_i) = 0$.

Model (2.7) is a varying coefficient model, hence, one can consistently estimate $\theta(z)$ provided that $E(x_i x_i^\top|z_i = z)$ is a nonsingular matrix for almost all $z \in \mathcal{S}_z$, where $\mathcal{S}_z$ is the support of z_i. Then a kernel estimator

$$\hat{\theta}(z) = \left[\sum_{j=1}^{n} x_j x_j^\top K_{h,z_j z}\right]^{-1} \sum_{j=1}^{n} x_j y_j K_{h,z_j z}$$

will consistently estimate $\theta(z)$ under quite general conditions, where $K_{h,z_j z} = K((z_j - z)/h)$. A consistent estimator of β is given by $n^{-1}\sum_{i=1}^{n}\hat{\theta}(z_i)$, and the consistency follows from $E(\theta(z_i)) = \beta$ (because $E(\alpha_i) = 0$ implies $E(g(z_i)) = 0$). However, the existence of such a variable z_i may not be easily justified in practice. Below we show that even without this additional assumption, it is possible to identify β with the help of panel data.

2.2 The Panel Data Case

Panel data will provide us more information and help us to identify the unknown functions. For heuristics we consider an example with a bivariate variable x_{it}, i.e.,

$$y_{it} = x_{1it}\beta_{1i} + x_{2it}\beta_{2i} + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T)$$

with $\beta_{1i} = \beta_1 + \alpha_{1i}$, $\beta_{2i} = \beta_2 + \alpha_{2i}$, $E(\alpha_{1i}) = 0$, $E(\alpha_{2i}) = 0$, and $E(u_{it}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}, \alpha_{1i}, \alpha_{2i}) = 0$.

Then we have $E(u_{it}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) = 0$. Hence, we have

$$\begin{aligned} & E(y_{i1}|x_{1i1} = x_{11}, x_{2i1} = x_{21}, \ldots, x_{1iT} = x_{1T}, x_{2iT} = x_{2T}) \\ = \ & x_{11}\theta_1(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}) + x_{21}\theta_2(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}), \\ & \vdots \\ & E(y_{iT}|x_{1i1} = x_{11}, x_{2i1} = x_{21}, \ldots, x_{1iT} = x_{1T}, x_{2iT} = x_{2T}) \\ = \ & x_{1T}\theta_1(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}) + x_{2T}\theta_2(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}). \end{aligned}$$

where $\theta_1(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) = \beta_1 + E(\alpha_{1i}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT})$ and $\theta_2(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) = \beta_2 + E(\alpha_{2i}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT})$. Once $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are identified, β_1 and β_2 are identified through relations $\beta_1 = E[\theta_1(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT})]$ and $\beta_2 = E[\theta_2(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT})]$, since $E(\alpha_{1i}) = 0$ and $E(\alpha_{2i}) = 0$.

We face a system of linear equations. If $T \geq 2$ and

$$L = \begin{pmatrix} x_{11} & x_{21} \\ \vdots & \vdots \\ x_{1T} & x_{2T} \end{pmatrix}^\top \begin{pmatrix} x_{11} & x_{21} \\ \vdots & \vdots \\ x_{1T} & x_{2T} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^{T} x_{1t}^2 & \sum_{t=1}^{T} x_{1t}x_{2t} \\ \sum_{t=1}^{T} x_{1t}x_{2t} & \sum_{t=1}^{T} x_{2t}^2 \end{pmatrix} \tag{2.8}$$

is nonsingular (i.e., when $(\sum_{t=1}^T x_{1t}^2)(\sum_{t=1}^T x_{2t}^2) > (\sum_{t=1}^T x_{1t}x_{2t})^2$), then we can solve $\theta_1(\cdot)$ and $\theta_2(\cdot)$ uniquely. Specifically, we have

$$\begin{aligned}&\begin{pmatrix}\theta_1(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T})\\ \theta_2(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T})\end{pmatrix}\\ =&\left(\begin{pmatrix}x_{11} & x_{21}\\ \vdots & \vdots\\ x_{1T} & x_{2T}\end{pmatrix}^\top \begin{pmatrix}x_{11} & x_{21}\\ \vdots & \vdots\\ x_{1T} & x_{2T}\end{pmatrix}\right)^{-1} \begin{pmatrix}x_{11} & x_{21}\\ \vdots & \vdots\\ x_{1T} & x_{2T}\end{pmatrix}^\top \begin{pmatrix}E(y_{i1}|x_{1i1} = x_{11}, x_{2i1} = x_{21}, \ldots, x_{1iT} = x_{1T}, x_{2iT} = x_{2T})\\ \cdots\\ E(y_{iT}|x_{1i1} = x_{11}, x_{2i1} = x_{21}, \ldots, x_{1iT} = x_{1T}, x_{2iT} = x_{2T})\end{pmatrix}.\end{aligned}$$

In general, for a panel CRC model with $d \times 1$ vector x_{it}, it requires $T \geq d$. In order the matrix M defined in (2.8) to be invertible, we also need enough variation of x_{it} across t. Once $\theta(\cdot)$ is identified, from $E(\alpha_i) = 0$ we obtain $E(\theta(x_i)) = \beta$. Hence, we can consistently estimate β by

$$\hat{\beta}_{Semi} = \frac{1}{n}\sum_{i=1}^n \hat{\theta}(x_i), \tag{2.9}$$

where $\hat{\theta}(x_i)$ is some standard semiparametric estimator.

In fact when $T \geq d$, one can also first estimate β_i based on individual i's T observations: $\hat{\beta}_{i,OLS} = [\sum_{t=1}^T x_{it}x_{it}^\top]^{-1}\sum_{t=1}^T x_{it}y_{it}$, then average it over i from 1 to n to obtain a group mean (GM) estimator for β given by

$$\hat{\beta}_{GM} = \frac{1}{n}\sum_{i=1}^n \hat{\beta}_{i,OLS}. \tag{2.10}$$

It is easy to show that $\sqrt{n}(\hat{\beta}_{GM} - \beta) \xrightarrow{d} N(0, V_{GM})$, where $V_{GM} = \Sigma_\alpha + V_2$ with $V_2 = E[(\sum_{t=1}^T x_{it}x_{it}^\top)^{-1}(\sum_{t=1}^T\sum_{s=1}^T u_{it}u_{is}x_{it}x_{is}^\top)(\sum_{t=1}^T x_{it}x_{it}^\top)^{-1}]$. If u_{it} is serially uncorrelated and conditionally homoscedastic, then V_2 simplifies to $V_2 = \sigma_u^2 E[(\sum_{t=1}^T x_{it}x_{it}^\top)^{-1}]$, where $\sigma_u^2 = E(u_{it}^2|x_{i1}, ..., x_{iT})$. However, we expect large bias in the finite sample estimation when T is small.

The condition that $T \geq d$ can be relaxed under additional assumptions. Suppose there exists a random variable z_i (z_i can be a vector) such that $E(\alpha_i|x_{it}, z_i) = E(\alpha_i|z_i) \equiv g(z_i)$, for example, we may have $z_i = \bar{x}_{i\cdot} \equiv T^{-1}\sum_{t=1}^T x_{it}$, so that α_i is correlated with $(x_{i1}, ..., x_{iT})$ only through $\bar{x}_{i\cdot}$. In this case we may have $\sum_{t=1}^T E(x_{it}x_{it}^\top|z_i = z)$ to be a nonsingular matrix even when $T < d$. As long as $\sum_{t=1}^T E(x_{it}x_{it}^\top|z_i = z)$ is invertible for almost all $z \in \Omega_z$, we can consistently estimate $\theta(z)$ for $z \in \Omega_z$ by

$$\hat{\theta}(z) = \left[\sum_{j=1}^n\sum_{s=1}^T x_{js}x_{js}^\top K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z)\right]^{-1}\sum_{j=1}^n\sum_{s=1}^T y_{js}x_{js}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z), \tag{2.11}$$

where $K_{h,z_jz} = K((z_j - z)/h)$, $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,\ldots,q\}}|z_l - z_{0,l}| \geq \varepsilon_n \text{ for some } z_0 \in \partial\mathcal{S}_z\}$, and $\mathbf{1}_{\varepsilon_n}(z)$ is a trimming function which ensures to avoid singularity problem and boundary bias and will be more explicit in section 3. Furthermore, we can consistently estimate β by

$$\hat{\beta}_{Semi} = \frac{1}{n}\sum_{i=1}^n \hat{\theta}(z_i),$$

where $\hat{\theta}(z_i)$ is obtained from (2.11) with z being replaced by z_i.

It can be shown that, under some standard regularity conditions, $\sqrt{n}(\hat{\beta}_{Semi} - \beta) \xrightarrow{d} N(0, V)$ for some positive definite matrix V, we discuss the estimation and the asymptotic analysis of $\hat{\beta}_{Semi}$ in the next section.

3 A Correlated Random Coefficient Panel Data Model

In this section we consider a CRC panel data model as follows

$$y_{it} = x_{it}^{\top}\beta_i + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{3.1}$$

where x_{it} is a $d \times 1$ vector, $\beta_i = \beta + \alpha_i$ is of dimension $d \times 1$, β is a $d \times 1$ constant vector, α_i is i.i.d. with $(0, \Sigma_\alpha)$, Σ_α is a $d \times d$ positive definite matrix, and u_{it} is i.i.d. with $(0, \sigma_u^2)$ and is orthogonal to (x_i, α_i). We allow α_i to be correlated with x_{it}.

We can rewrite (3.1) as

$$y_{it} = x_{it}^{\top}\beta + x_{it}^{\top}\alpha_i + u_{it}, \tag{3.2}$$

$E(u_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$. Let z_i satisfy the condition that $E(u_{it}|x_{it}, z_i) = 0$ and $E(\alpha_i|x_{it}, z_i) = E(\alpha_i|z_i) \equiv g(z_i)$. For example we can have $z_i = \bar{x}_{i\cdot} \equiv T^{-1}\sum_{t=1}^{T} x_{it}$ or $z_i = x_i = (x_{i1}^{\top}, ..., x_{iT}^{\top})^{\top}$. Define $\eta_i = \alpha_i - E(\alpha_i|z_i)$ and $\epsilon_{it} = x_{it}^{\top}\eta_i + u_{it}$. By construction we have $E(\epsilon_{it}|z_i) = 0$.

Then we have

$$y_{it} = x_{it}^{\top}\beta + x_{it}^{\top}g(z_i) + \epsilon_{it} = x_{it}^{\top}\theta(z_i) + \epsilon_{it}, \tag{3.3}$$

where $\theta(z) = \beta + g(z)$. Note that equation (3.3) is a semiparametric varying coefficient model. Hence, we can estimate $\theta(z)$ by some standard semiparametric estimator, say, kernel-based local constant or local polynomial estimation methods. From $E(g(z_i)) = 0$ we obtain $\beta = E(\theta(z_i))$. Let $\hat{\theta}(z)$ denote a generic semiparametric estimator of $\theta(z)$, we estimate β by

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}(z_i).$$

Let $\mathbf{1}_{\varepsilon_n}(z_i) = \mathbf{1}\{z_i \in \Omega_z\}$, and $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,...,q\}} |z_l - z_{0,l}| \geq \varepsilon_n \text{ for some } z_0 \in \partial\mathcal{S}_z\}$, where $\partial\mathcal{S}_z$ is the boundary of the compact set $\mathcal{S}_z$ which is the support of z_i, $\|h\|/\varepsilon_n \to 0$ and $\varepsilon_n \to 0$, as $n \to \infty$. If we take $z_i = \bar{x}_{i\cdot}$, we can get a semiparametric estimator using local constant kernel estimation

$$\hat{\beta}_{Semi,1} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{VC,1}(\bar{x}_i),$$

where

$$\hat{\theta}_{VC,1}(\bar{x}_i) = \left[\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top}K_{h,\bar{x}_j\bar{x}_i}\mathbf{1}_{\varepsilon_n}(\bar{x}_{i\cdot})\right]^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,\bar{x}_j\bar{x}_i}\mathbf{1}_{\varepsilon_n}(\bar{x}_{i\cdot}),$$

with $K_{h,\bar{x}_j\bar{x}_i} = \prod_{m=1}^{d} k((\bar{x}_{j\cdot,m} - \bar{x}_{i\cdot,m})/h_m)$.

If we take $z_i = x_i = (x_{i1}^\top, ..., x_{iT}^\top)^\top$. We can pool the data together and estimate β by

$$\hat{\beta}_{Semi,2} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{VC,2}(x_i), \tag{3.4}$$

where

$$\hat{\theta}_{VC,2}(x_i) = \left[\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^\top K_{h,x_jx_i}\mathbf{1}_{\varepsilon_n}(x_i)\right]^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,x_jx_i}\mathbf{1}_{\varepsilon_n}(x_i), \tag{3.5}$$

with $K_{h,x_jx_i} = \prod_{m=1}^{d}\prod_{t=1}^{T} k((x_{jt,m} - x_{it,m})/h_{tm})$.

Since the derivations of asymptotic distributions of $\hat{\beta}_{Semi,1}$ and $\hat{\beta}_{Semi,2}$ are special cases of using different z_i, we will provide detailed proofs without specifying z_i. We consider two types of semiparametric estimators for $\theta(z)$, local constant and local polynomial estimation methods. The local constant estimator of $\theta(z)$ for $z \in \Omega_z$ is given by

$$\hat{\theta}_{LC}(z) = \left(\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^\top K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z)\right)^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z), \tag{3.6}$$

where $K_{h,z_jz} = K((z_j - z)/h) = \prod_{l=1}^{q} k\left(\frac{z_{jl}-z_l}{h_l}\right)$ is the product kernel, $k(\cdot)$ is the univariate kernel function, z_{jl} and z_l are the l^{th}-component of z_j and z, respectively. Then, we define $\hat{\beta}_{LC} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{LC}(z_i)$.

We introduce some notations and assumptions before we present the asymptotic theories. We write $f_i = f(z_i)$. For the $d \times 1$ vector $\theta_i = \theta(z_i)$, we use $\theta_{il} = \theta_l(z_i)$ to denote the l^{th} component of $\theta(z_i)$ and use $||h|| = \sqrt{\sum_{l=1}^{q} h_l^2}$ to denote the usual Euclidean norm. We make following assumptions.

Assumption A1: $(y_i^\top, x_i^\top, z_i^\top)$ are i.i.d. as $(y_1^\top, x_1^\top, z_1^\top)$, where $y_i^\top = (y_{i1}, ..., y_{iT})$, $x_i^\top = (x_{i1}^\top, ..., x_{iT}^\top)$, $x_{it}^\top = (x_{it,1}, ..., x_{it,d})$, $z_i^\top = (z_{i,1}, ..., z_{i,q})$. $z_i^\top$ admits a Lebesgue density function $f(z_1, ..., z_q)$ with $\inf_{z\in\mathcal{S}_z} f(z) > 0$, where $\mathcal{S}_z$ is the support of $z_i^\top$ and is compact. x_{it} is strictly stationary across time t. u_{it} has finite fourth moment.

Assumption A2: $\theta(z)$ and $f(z)$ are $\nu + 1$ times continuously differentiable, where ν is an integer defined in the next assumption.

Assumption A3: $K(z) = \prod_{l=1}^{q} k(z_l)$, where $k(\cdot)$ is a univariate symmetric (around zero) bounded ν^{th} order kernel function with a compact support, i.e., $\int k(v)dv = 1$, $\int k(v)v^j dv = 0$ for $j = 1, ..., \nu - 1$ and $\mu_\nu = \int k(v)v^\nu dv \neq 0$, where ν is a positive even integer, with $\int |k(v)|v^{\nu+2}dv$ being a finite constant.

Assumption A4: As $n \to \infty$, $nh_1 \cdots h_q/\ln n \to \infty$, $\|h\|^{2\nu}\ln n/H \to 0$, $n||h||^{2\nu+2} \to 0$, $\varepsilon_n \to 0$, $\|h\|/\varepsilon_n \to 0$, $h_l \to 0$ for all $l = 1, ..., q$.

Theorem 3.1 *Under assumptions A1 to A4, we have that*

$$\sqrt{n}\left(\hat{\beta}_{LC} - \beta - \sum_{l=1}^{q} h_l^\nu B_{l,LC}\right) \xrightarrow{d} N(0, V_{LC}),$$

where $B_{l,LC} = \mu_\nu \sum_{k_1+k_2=\nu, k_2\neq 0} \frac{1}{k_1!k_2!} E\left[m_i^{-1} (\frac{\partial^{k_1} m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})\right]$, $m_i = m(z_i) = T^{-1}\sum_{t=1}^T E[x_{it}x_{it}^\top|z_i]f(z_i)$, $\frac{\partial^{k_1} m_i}{\partial z_l^{k_1}} = \frac{\partial^{k_1} m(z)}{\partial z_l^{k_1}}|_{z=z_i}$, $\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}} = \frac{\partial^{k_2}\theta(z)}{\partial z_l^{k_2}}|_{z=z_i}$, *and* $V_{LC} = Var(\theta(z_i)) + T^{-1}Var(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i))) + T^{-1}Var(u_{is}m_i^{-1}x_{is}f(z_i))$.

We can see that the semiparametric estimator we give has a $\sqrt{n}$ convergence rate. The reason is well known that taking average can reduce the variance of nonparametric estimators. We also use the high order kernel to reduce the bias. The proof of Theorem 3.1 is given in the Appendix A.

In order to reduce the bias, we also consider the local polynomial estimation. We introduce some notations first. Let

$$k = (k_1, \ldots, k_q), \quad k! = k_1! \times \cdots \times k_q!, \quad |k| = \sum_{i=1}^q k_i,$$
$$z^k = z_1^{k_1} \times \cdots \times z_q^{k_q}, \quad h^k = h_1^{k_1} \cdots h_q^{k_q},$$
$$\sum_{0\le|k|\le p} = \sum_{j=0}^p \sum_{\substack{k_1=0 \\ k_1+\cdots+k_q=j}}^j \cdots \sum_{k_q=0}^j,$$
$$D^k\theta(z) = \frac{\partial^{|k|}\theta(z)}{\partial z_1^{k_1}\cdots\partial z_q^{k_q}}.$$

Then we minimize the kernel weighted sum of squared errors

$$\sum_{j=1}^n \sum_{s=1}^T \left[y_{js} - \sum_{0\le|k|\le p} x_{js}^\top b_k(z)(z_j - z)^k \right]^2 K_{h,z_j z}, \tag{3.7}$$

with respect to each $b_k(z)$ which gives an estimate of $\hat{b}_k(z)$, and $k!\hat{b}_k(z)$ estimates $D^k\theta(z)$. Thus, $\hat{\theta}_{LP} = \hat{b}_0(z)$ is the p^{th} order local polynomial estimator of $\theta(z)$. We define $\hat{\beta}_{LP} = \frac{1}{n}\sum_{i=1}^n \hat{\theta}_{LP}(z_i)$.

Now we need $\theta(z)$ to be $p+1$ times differentiable, and the local polynomial estimation cannot be used together with the high order kernel. So we give the following assumptions.

Assumption B1: $(y_i^\top, x_i^\top, z_i^\top)$ are i.i.d. as $(y_1^\top, x_1^\top, z_1^\top)$, where $y_i^\top = (y_{i1}, ..., y_{iT})$, $x_i^\top = (x_{i1}^\top, ..., x_{iT}^\top)$, $x_{it}^\top = (x_{it,1}, ..., x_{it,d})$, $z_i^\top = (z_{i,1}, ..., z_{i,q})$. $z_i^\top$ admits a Lebesgue density function $f(z_1, ..., z_q)$ with $\inf_{z\in\mathcal{S}_z} f(z) > 0$, where $\mathcal{S}_z$ is the support of $z_i^\top$ and is compact. x_{it} is strictly stationary across time t. u_{it} has finite fourth moment.
Assumption B2: $\theta(z)$ is $p+1$ times continuously differentiable, and $f(z)$ is three times continuously differentiable.
Assumption B3: $K(z) = \prod_{l=1}^q k(z_l)$, where $k(\cdot)$ is a univariate symmetric (around zero) bounded kernel function with a compact support, i.e., $\int k(v)dv = 1$, $\int k(v)v^i dv = 0$, if $0 < i \le p+2$ is an odd integer and $\mu_i = \int k(v)v^i dv \neq 0$, if $0 < i \le p+2$ is an even integer. We define $\mu_k = \int v_1^{k_1}\cdots v_q^{k_q}\prod_{l=1}^q k(v_l)dv_1\ldots dv_q$ if k is a q-tuple.

Assumption B4: As $n \to \infty$, $nh_1 \cdots h_q / \ln n \to \infty$, $\varepsilon_n \to 0$, $\|h\|/\varepsilon_n \to 0$; if $p > 0$ is an odd integer, $\|h\|^{2p+2} \ln n/H \to 0$, $n||h||^{2p+4} \to 0$; if $p > 0$ is an even integer, $\|h\|^{2p+4} \ln n/H \to 0$, $n||h||^{2p+6} \to 0$; $h_l \to 0$ for all $l = 1, ..., q$.

Theorem 3.2 *Under assumptions B1 to B4, we have that*

$$\sqrt{n}\left(\hat{\beta}_{LP} - \beta - B_{LP}\right) \xrightarrow{d} N(0, V_{LP}),$$

where $B_{LP} = P_1 S^{-1} M \sum_{|k|=p+1} \frac{\mu_k h^k}{k!} E\left[\Theta_i\right]$, *if p is an odd positive integer, or* $B_{LP} = P_1 S^{-1} M \sum_{|k|=p+2} \frac{\mu_k h^k}{k!} E\left[\Theta_i\right]$, *if p is an even positive integer, P_1, S, M and Θ_i are matrices defined in the Appendix A, and* $V_{LP} = Var(\theta(z_i)) + T^{-1} Var(P_1 S(z_i)^{-1} \Gamma_{is} x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) f(z_i)) + T^{-1} Var(P_1 S(z_i)^{-1} u_{is} f(z_i) \Gamma_{is})$, *where Γ_{is} is also defined in the Appendix A.*

The proof of Theorem 3.2 is given in the Appendix A.

Note that if one imposes an additional condition that $n||h||^{2\nu} \to 0$ or $n||h||^{2p+2} \to 0$ as $n \to \infty$ for $\hat{\beta}_{LC}$ or $\hat{\beta}_{LP}$, respectively, then the center term is asymptotically negligible, and we have the following result:

$$\sqrt{n}(\hat{\beta}_{Semi} - \beta) \xrightarrow{d} N(0, V),$$

where $\hat{\beta}_{Semi}$ can be $\hat{\beta}_{LC}$ or $\hat{\beta}_{LP}$.

4 Identification of a Binary Response CRC Panel Model

The identification of the binary response model is different from the linear models. We can identify the coefficients if we assume that the unobserved random terms have known distributions, and this will allow us to estimate the model by conditional maximum likelihood method. However, if we do not assume the distribution of the unobserved terms, the identification becomes problematic. We need to impose additional restrictions on the dependence structure between the regressors and the unobservables. One way to identify the model is transferring the model to a single-index model, which can be estimated nonparametrically. However, the single-index model only admits limited heterogeneity, see Powell, Stock, and Stoker (1989), Ichimura (1993), Klein and Spady (1993), Härdle and Horowitz (1996), Newey and Ruud (2005). Another way of identification is based on the conditional quantile restrictions. Manski (1985, 1988) give the identification conditions in this type for the binary response models. A sufficient condition for the identification of the coefficients is the median independence between the error and the regressors. He also suggests the conditional maximum score estimator to estimate the model. However, the limiting distribution is not standard which is derived by Kim and Pollard (1990). Horowitz (1992) modifies the maximum score estimator to a smoothed maximum score estimator and gets the asymptotic normal distribution. The convergence rates of maximum score estimators are less than $\sqrt{n}$. Chamberlain (2010) shows that the consistent estimation at the

$\sqrt{n}$ convergence rate is possible only when the errors have logistic distributions without other additional assumptions.

The third way of identification and achieving the $\sqrt{n}$ convergence rate is via the special regressor method, which is proposed by Lewbel (1998, 2000). With additional assumptions on the joint distribution of the observables and unobservables based on one special regressor, we can get the identification and the usual parametric estimation rate. We use this method to identify a binary response CRC panel data model in this paper.

We consider a binary response correlated random coefficient panel data model as follows.

$$y_{it} = \mathbf{1}(v_{it} + x_{it}^{\top}\beta_i + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{4.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function, β_i is the individual specific random coefficient, and the superscript $^{\top}$ denotes the transpose. For simplicity, we assume there exists only one regressor which has constant coefficient and this regressor is the special regressor in model (1.1) to get the model (4.1). The analysis remains similar if we assume more regressors with constant coefficients. Let $\beta_i = \beta + \alpha_i$, where $E(\alpha_i) = 0$, then β is the average slope we are interested in. We assume v_{it} is a special regressor, which satisfies three conditions that v_{it} is a continuous random variable, independent of α_i and u_{it} conditional on x_{it}, and has a relatively large support, which will be made more specific below. Here, we normalize the coefficient of v_{it} to be 1. If it is negative, we can use $-v_{it}$ instead of v_{it}. The advantage of including such a special regressor is to allow us to transfer the binary response model into a linear moment condition. Further, we assume that $E(u_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$, which is the strict exougeneity condition. Also, we assume there exists a random vector z_i satisfying the condition that $E(u_{it}|x_{it}, z_i) = 0$ and $E(\alpha_i|x_{it}, z_i) = E(\alpha_i|z_i) \equiv g(z_i)$, for instance $z_i = \bar{x}_{i\cdot} = T^{-1}\sum_{t=T}^{n} x_{it}$ or $z_i = (x_{i1}^{\top}, \ldots, x_{iT}^{\top})^{\top}$. We already saw the identification and estimation in the linear case. With the help of the special regressor, we can transfer (4.1) to a linear moment condition, i.e., $E[(y_{it} - \mathbf{1}(v_{it} > 0))/f_t(v_{it}|x_{it}, z_i)|x_{it}, z_i] = x_{it}^{\top}\beta + x_{it}^{\top}E(\alpha_i|x_{it}, z_i) = x_{it}^{\top}\beta + x_{it}^{\top}g(z_i)$, which is given in the identification proposition below.

Panel data give us more observations for the same individual over different time periods. This brings us the advantage of taking consideration of the heterogenous effects. We can identify the average slope if we have enough time period or additional information on z_i as we did in the linear case. We assume the data are independent across i. We give the assumptions on the special regressor.

Assumption C1: The conditional distribution of v_{it} given x_{it} and z_i has a continuous conditional density function $f_t(v_{it}|x_{it}, z_i)$ with respect to the Lebesgue measure on the real line. Moreover, $f_t(v_{it}|x_{it}, z_i) > 0$, if $f_t(v_{it}|x_{it}, z_i)$ has the real line as the support, and $\inf_{v_{it}\in[L_t,K_t]} f_t(v_{it}|x_{it}, z_i) > 0$, if $[L_t, K_t]$ is compact, where $[L_t, K_t]$ is the support of v_{it} conditional on x_{it} and z_i.

Assumption C2: Assume α_i and u_{it} are independent of v_{it} conditional on x_{it} and z_i. Let $e_{it} = x_{it}^{\top}(\alpha_i - g(z_i)) + u_{it}$ and denote the conditional distribution of e_{it} conditioning on (x_{it}, z_i) as $F_{e_{it}}(e_{it}|x_{it}, z_i)$ with the support Ω_{e_t}.

Assumption C3: The conditional distribution of v_{it} conditional on x_{it} and z_i has support $[L_t, K_t]$ for $-\infty \le L_t < 0 < K_t \le +\infty$, and the support of $-x_{it}^\top\beta - x_{it}^\top g(z_i) - e_{it}$ is a subset of $[L_t, K_t]$.

In the empirical analysis, the existence of the special regressor depends on the context. For instance, the age or date of birth can be chosen as the special regressor. In some situations, it may not be easy to find such a regressor. For more discussions, see Honoré and Lewbel (2002).

Based on these assumptions, similar as Theorem 1 in Honoré and Lewbel (2002), we have the following identification proposition.

Proposition 4.1 *Under assumptions C1, C2, and C3, let*

$$y_{it}^* = \begin{cases} [y_{it} - \mathbf{1}(v_{it} > 0)]/f_t(v_{it}|x_{it}, z_i) & \text{if } v_{it} \in [L_t, K_t], \\ 0 & \text{otherwise.} \end{cases}$$

we have

$$E(y_{it}^*|x_{it}, z_i) = x_{it}^\top\beta + x_{it}^\top g(z_i). \tag{4.2}$$

The proof of this proposition is given in the Appendix B.

5 Estimation of the Binary Response CRC Panel Model

Based on the identification analysis in section 4, we can construct the semiparametric estimator of β using kernel methods. Let $\theta(z_i) = \beta + g(z_i)$. Since $0 = E[\alpha_i] = E[g(z_i)]$, we have $\beta = E[\theta(z_i)]$. Once we have an estimator of $\theta(\cdot)$, we can estimate β using $\hat{\beta} = n^{-1}\sum_{i=1}^n \hat{\theta}(z_i)$.

From (4.2), we have $\theta(z_i) = \left(\sum_{t=1}^T E[x_{it}x_{it}^\top|z_i]\right)^{-1} \sum_{t=1}^T E[x_{it}y_{it}^*|z_i]$. Since $E[x_{it}y_{it}^*|z_i] = E[x_{it}(y_{it} - \mathbf{1}(v_{it} > 0))/f_t(v_{it}|x_{it}, z_i)|z_i]$ and $f_t(v_{it}|x_{it}, z_i)$ is unknown, we have to estimate $f_t(v_{it}|x_{it}, z_i)$ and we estimate it by

$$\hat{f}_t(v_{it}|x_{it}, z_i) = \frac{\hat{f}_t(v_{it}, x_{it}, z_i)}{\hat{f}_t(x_{it}, z_i)} \equiv \frac{(nH)^{-1}\sum_{k=1}^n K_h(v_{kt} - v_{it}, x_{kt} - x_{it}, z_k - z_i)}{(n\tilde{H})^{-1}\sum_{k=1}^n K_{\tilde{h}}(x_{kt} - x_{it}, z_k - z_i)},$$

where $\hat{f}_t(v_{it}, x_{it}, z_i) = (nH)^{-1}\sum_{k=1}^n K_h(v_{kt}-v_{it}, x_{kt}-x_{it}, z_k-z_i)$, $\hat{f}_t(x_{it}, z_i) = (n\tilde{H})^{-1}\sum_{k=1}^n K_{\tilde{h}}(x_{kt}-x_{it}, z_k-z_i)$, $H = h_1 \cdots h_{d+q+1}$, $\tilde{H} = h_2 \cdots h_{d+q+1}$, $K_h(u) = \prod_{l=1}^{d+q+1} k\left(\frac{u_l}{h_l}\right)$, $h = (h_1, \ldots, h_{d+q+1})^\top$ and $\tilde{h} = (h_2, \ldots, h_{d+q+1})^\top$. Then we estimate $E[x_{it}y_{it}^*|z_i]$ by

$$\hat{E}[x_{it}y_{it}^*|z_i] = \frac{(nTH')^{-1}\sum_{j=1}^n x_{jt}(y_{jt} - \mathbf{1}(v_{jt} > 0))K_{h'}(z_j - z_i)\mathbf{1}_{\tau_n,j}/\hat{f}_t(v_{jt}|x_{jt}, z_j)}{(nTH')^{-1}\sum_{j=1}^n K_{h'}(z_j - z_i)},$$

where $\mathbf{1}_{\tau_n,j} = \mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j) = \mathbf{1}\{(v_{jt}, x_{jt}, z_j) \in \Omega_{vxz}\}$, $\Omega_{vxz} = \{a \in \mathcal{S}_{vxz} : \min_{l\in\{1,\ldots,d+q+1\}} |a_l - b_l| \ge \tau_n,$ for some $b \in \partial\mathcal{S}_{vxz}\}$, $\partial\mathcal{S}_{vxz}$ denotes the boundary of the compact set $\mathcal{S}_{vxz}$ which is the support of (v_{jt}, x_{jt}, z_j), $H' = h_1' \cdots h_q'$, $h' = (h_1', \ldots, h_q')$, $\|h\|/\tau_n \to 0$, and $\tau_n \to 0$, as $n \to \infty$.

We use $\mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j)$ to truncate the data at the boundary to avoid the singularity problem and the boundary bias.

We can get an estimator of $\theta(z_i)$ by the local constant kernel method or the local polynomial method. Due to the complexity of the local polynomial kernel estimator, we will not discuss it here. However, based on the analysis in the linear case, we know the derivation will be similar. The local constant kernel estimator $\hat{\theta}_{LC}(z_i)$ for $z_i \in \Omega_z$ is given by

$$\hat{\theta}_{LC}(z_i) = [\sum_{j=1}^{n}\sum_{t=1}^{T} x_{jt}x_{jt}^{\top} K_{h'}(z_j - z_i)\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i}]^{-1} \sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{(y_{jt} - \mathbf{1}(v_{jt} > 0))}{\hat{f}_t(v_{jt}|x_{jt}, z_j)} K_{h'}(z_j - z_i)\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i},$$

where $\mathbf{1}_{\varepsilon_n,i} = \mathbf{1}_{\varepsilon_n}(z_i) = \mathbf{1}\{z_i \in \Omega_z\}$, $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,...,q\}} |z_l - z_{0,l}| \geq \varepsilon_n$ for some $z_0 \in \partial\mathcal{S}_z\}$, $\partial\mathcal{S}_z$ is the boundary of the compact set $\mathcal{S}_z$ which is the support of z_i, $\|h'\|/\varepsilon_n \to 0$ and $\varepsilon_n \to 0$, as $n \to \infty$. Then the local constant kernel estimator of β is given by

$$\hat{\beta}_{LC} = n^{-1}\sum_{i=1}^{n}\hat{\theta}_{LC}(z_i).$$

We list some conditions before we present the asymptotic distribution.

Assumption C4: $(y_i^{\top}, v_i^{\top}, x_i^{\top}, z_i^{\top})$ are i.i.d. as $(y_1^{\top}, v_1^{\top}, x_1^{\top}, z_1^{\top})$, where $y_i^{\top} = (y_{i1}, ..., y_{iT})$, $v_i^{\top} = (v_{i1}, ..., v_{iT})$, $x_i^{\top} = (x_{i1}^{\top}, ..., x_{iT}^{\top})$, $x_{it}^{\top} = (x_{it,1}, ..., x_{it,d})$, $z_i^{\top} = (z_{i,1}, ..., z_{i,q})$. $z_i^{\top}$ admits a Lebesgue density function $f_z(z_1, ..., z_q)$ with $\inf_{z\in\mathcal{S}_z} f_z(z) > 0$, where $\mathcal{S}_z$ is the support of $z_i^{\top}$ and is compact. v_{it} is a continuous scalar random variable with the support $[L_t, K_t]$ on the real line $\mathbf{R}$. (v_{it}, x_{it}, z_i) has a compact support $\mathcal{S}_{vxz}$. v_{it} and x_{it} are strictly stationary across time t and u_{it} has finite fourth moment.

Assumption C5: $\theta(z)$, $f_t(v, x, z)$, $f_t(v, x)$ and $f_z(z)$ are $\nu+1$ times continuously differentiable, where ν is an integer defined in the next assumption.

Assumption C6: $K(z) = \prod_{l=1}^{q} k(z_l)$, where $k(\cdot)$ is a univariate symmetric (around zero) bounded ν^{th} order kernel function with a compact support, i.e., $\int k(v)dv = 1$, $\int k(v)v^j dv = 0$ for $j = 1, ..., \nu - 1$ and $\mu_\nu = \int k(v)v^{\nu}dv \neq 0$, ν is a positive even integer, with $\int |k(v)|v^{\nu+2}dv$ being a finite constant.

Assumption C7: As $n \to \infty$, $nH'^2/\ln n \to \infty$, $\sqrt{n}H/\ln n \to \infty$, $\|h'\|^{2\nu}\ln n/H' \to 0$, $\|h'\|^{\nu}/H' \to 0$, $n||h'||^{2\nu} \to 0$, $n||h||^{2\nu} \to 0$, $n||\tilde{h}||^{2\nu} \to 0$, $\varepsilon_n \to 0$, $\tau_n \to 0$, $\|h'\|/\varepsilon_n \to 0$, $\|h\|/\tau_n \to 0$, $\varepsilon_n > \tau_n$, $\|h'\|/(\varepsilon_n - \tau_n) \to 0$, $h_l \to 0$ for all $l = 1, ..., d + q + 1$, $h'_l \to 0$ for all $l = 1, ..., q$.

Theorem 5.1 *Under assumptions C4-C7, we have that*

$$\sqrt{n}(\hat{\beta}_{LC} - \beta) \xrightarrow{d} N(0, V_{LC}),$$

where $V_{LC} = Var(g(z_i)) + T^{-1}Var\left(m_i^{-1}f_z(z_i)x_{it}\xi_{it}\right)$, $\xi_{it} = y_{it}^{*} - E(y_{it}^{*}|x_{it}, z_i)$, *and* $y_{it}^{*} = [y_{it} - \mathbf{1}(v_{it} > 0)]/f_t(v_{it}|x_{it}, z_i)$, *if* $v_{it} \in [L_t, K_t]$, *and* $y_{it}^{*} = 0$, *otherwise.*

The proof of Theorem 5.1 is given in the Appendix B.

6 Monte Carlo Simulation Results

In this section, we conduct extensive simulations to examine the finite sample performance of different estimators including semiparametric estimators we proposed in sections 3 and 5.

6.1 Linear CRC panel data models

In this subsection, we consider a simple linear panel data model

$$y_{it} = \beta_{0i} + x_{it}\beta_{1i} + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{6.1}$$

where x_{it} is a scalar random variable, $\beta_{0i} = \beta_0 + \alpha_{0i}$, $\beta_{1i} = \beta_1 + \alpha_{1i}$, α_{0i} is i.i.d. with $(0, \sigma_0^2)$, α_{1i} is i.i.d. with $(0, \sigma_1^2)$, and u_{it} is i.i.d. with $(0, \sigma_u^2)$ and is independent with (x_{it}, α_i). $n = 100, 200, 400$ and $T = 3$. We report the estimated mean squared error (MSE) computed by

$$MSE(\hat{\beta}_s) = \frac{1}{n_r}\sum_{j=1}^{n_r}\left[\hat{\beta}_{s,j} - \beta_s\right]^2, \qquad \text{for } s = 0, 1,$$

where $\hat{\beta}$ is one of five estimators, $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$, which are defined below, $\hat{\beta}_{s,j}$ is the value of $\hat{\beta}_s$ in the j^{th} simulation replication, $n_r = 1,000$ is the number of replications.

6.1.1 Estimators Considered

We will compare the following five estimators:

(i) The OLS estimator of regressing y_{it} on $(1, x_{it})$, i.e., $\hat{\beta}_{OLS}$ is from the linear regression

$$y_{it} = \beta_0 + x_{it}\beta_1 + u_{it}.$$

Let $\tilde{x}_{it} = (1, x_{it})^\top$, then

$$\hat{\beta}_{OLS} = (\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}^\top)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{x}_{it}y_{it}.$$

(ii) The fixed-effects estimator $\hat{\beta}_{FE}$,

$$\hat{\beta}_{FE} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i\cdot})^2},$$

where $\bar{x}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ and $\bar{y}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$. We can see that the fixed-effects estimator cannot estimate β_0. We only report its estimation results for β_1.

(iii) We estimate β_i using each individual's data, i.e., $\hat{\beta}_{i,OLS} = [\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}^\top]^{-1}\sum_{t=1}^{T}\tilde{x}_{it}y_{it}$. Then we average $\hat{\beta}_{i,OLS}$ to obtain the group mean estimator $\hat{\beta}_{GM}$ as defined in (2.10).

(iv) If we let $z_i = \bar{x}_{i\cdot}$, where $\bar{x}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, then we can get the semiparametric estimator $\hat{\beta}_{Semi,1}$. That is, $\hat{\beta}_{Semi,1}$ is the average of the varying coefficient estimator $\hat{\theta}_{VC,1}$ of the following varying coefficient model

$$y_{it} = \theta_0(z_i) + x_{it}\theta_1(z_i) + u_{it}.$$

$\hat{\beta}_{Semi,1} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{VC,1}(\bar{x}_{i\cdot})$, where

$$\hat{\theta}_{VC,1}(\bar{x}_{i\cdot}) = (\sum_{j=1}^{n}\sum_{t=1}^{T} \tilde{x}_{jt}\tilde{x}_{jt}^{\top} K_h(\bar{x}_{j\cdot} - \bar{x}_{i\cdot})\mathbf{1}_{\varepsilon_n}(\bar{x}_{i\cdot}))^{-1} \sum_{j=1}^{n}\sum_{t=1}^{T} \tilde{x}_{jt} y_{jt} K_h(\bar{x}_{j\cdot} - \bar{x}_{i\cdot})\mathbf{1}_{\varepsilon_n}(\bar{x}_{i\cdot}),$$

where $K(\cdot)$ is a kernel function and h is the smoothing parameter.

(v) If we let $z_i = x_i = (x_{i1}^{\top}, ..., x_{iT}^{\top})^{\top}$, then we can get the semiparametric estimator $\hat{\beta}_{Semi,2}$. That is, $\hat{\beta}_{Semi,2}$ is the average of the varying coefficient estimator $\hat{\theta}_{VC,2}$ of the following varying coefficient model

$$y_{it} = \theta_0(z_i) + x_{it}\theta_1(z_i) + u_{it}.$$

$\hat{\beta}_{Semi,2} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{VC,2}(x_i)$, where

$$\hat{\theta}_{VC,2}(z_i) = (\sum_{j=1}^{n}\sum_{t=1}^{T} \tilde{x}_{jt}\tilde{x}_{jt}^{\top} K_h(z_j - z_i)\mathbf{1}_{\varepsilon_n}(x_i))^{-1} \sum_{j=1}^{n}\sum_{t=1}^{T} \tilde{x}_{jt} y_{jt} K_h(z_j - z_i)\mathbf{1}_{\varepsilon_n}(x_i),$$

where $K(\cdot)$ is a multivariate kernel function and h is a vector of smoothing parameters.

6.1.2 Data Generating Processes

Below we report the result of a small simulation study. We generate y_{it} by

$$y_{it} = \beta_{0i} + x_{it}\beta_{1i} + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T;\ T = 3)$$

where $\beta_{0i} = \beta_0 + \alpha_{0i}$, $\beta_{1i} = \beta_1 + \alpha_{1i}$, $\beta_0 = 1$, $\beta_1 = 1$, x_{it} is i.i.d. with $Gamma(1, 1)$, and u_{it} is i.i.d. with $N(0, 1)$. α_{0i} and α_{1i} are generated in the following ways, where $\alpha_{0i} = v_{0i} - E(v_{0i})$ and $\alpha_{1i} = v_{1i} - E(v_{1i})$.

$$
\begin{aligned}
DGP1: &\quad v_{0i} = \bar{x}_{i\cdot} + \eta_{0i}, \text{ and } v_{1i} = \bar{x}_{i\cdot} + \eta_{1i},\\
DGP2: &\quad v_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } v_{1i} = (\bar{x}_{i\cdot} - 1)^2 + \ln(\bar{x}_{i\cdot} + 1) + \eta_{1i},\\
DGP3: &\quad v_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } v_{1i} = \sin(3\bar{x}_{i\cdot}) + \eta_{1i},\\
DGP4: &\quad v_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } v_{1i} = (x_{i1}^2 + x_{i2}^2 + x_{i3}^2)/9 + \eta_{1i},
\end{aligned}
$$

where $\bar{x}_{i\cdot} = T^{-1}\sum_{t=1}^{T} x_{it}$, η_{0i} and η_{1i} are i.i.d. with $Uniform[-1, 1]$.

In both DGP1 to DGP4 above, α_{0i} and α_{1i} are correlated with x_{it}.

Table 1: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP1

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	0.1727	n/a	0.0511	0.0193	0.0239	1.7706	0.1100	2.2231	0.1739	0.2532
200	0.1695	n/a	0.0252	0.0103	0.0131	1.7876	0.0788	0.6199	0.1052	0.1596
400	0.1691	n/a	0.0170	0.0056	0.0079	1.7740	0.0619	0.6050	0.0602	0.0981

6.1.3 Simulation Results

The simulation results are reported in Table 1, Table 2, Table 3 and Table 4, and the results confirm our theoretical analysis in the paper. We can see that in all of these tables, $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FE}$ are not consistent.

From Table 1 we observe the followings: $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ have the smaller estimation MSE than $\hat{\beta}_{GM}$. The GM estimator has the large estimation MSE because of the short panel of $T = 3$ so that each individual estimator has large variance. Though averaging over individuals makes it a consistent estimator, its finite sample MSE is still large.

Table 2: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP2

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	2.6718	n/a	0.2425	0.2012	0.2120	34.9186	1.1697	2.2223	0.0973	0.1843
200	2.5887	n/a	0.1229	0.1049	0.1102	32.0093	1.0391	0.6196	0.0603	0.1166
400	2.4841	n/a	0.0768	0.0632	0.0664	29.3801	1.0430	0.6048	0.0348	0.0692

The simulation results for DGP2 is given in Table 2. Note that for DGP2, $\hat{\beta}_{Semi,1}$ performs the best, followed by $\hat{\beta}_{Semi,2}$, and with $\hat{\beta}_{GM}$ far behind.

Table 3: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP3

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	1.3804	n/a	0.2425	0.2032	0.2142	17.3218	0.2184	2.2223	0.1251	0.2007
200	1.3286	n/a	0.1229	0.1057	0.1116	14.9118	0.1826	0.6196	0.0790	0.1281
400	1.2416	n/a	0.0768	0.0635	0.0673	12.7015	0.1630	0.6048	0.0453	0.0768

From Table 3 we observe that $\hat{\beta}_{Semi,1}$ has the smallest estimation MSE, followed by $\hat{\beta}_{Semi,2}$ and $\hat{\beta}_{GM}$.

Table 4 reports simulation results for DGP4, we can see that $\hat{\beta}_{Semi,1}$ and $\hat{\beta}_{Semi,2}$ are consistent.

The simulation results reported in this section show that our proposed semiparametric estimators $\hat{\beta}_{Semi,1}$ and $\hat{\beta}_{Semi,2}$ perform well.

Table 4: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP4

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	2.7451	n/a	0.2425	0.2105	0.2115	36.0380	2.0803	2.2334	0.1287	0.1834
200	2.6751	n/a	0.1229	0.1125	0.1098	33.2559	1.8795	0.6224	0.0691	0.1080
400	2.6186	n/a	0.0768	0.0701	0.0662	31.2719	1.9394	0.6077	0.0394	0.0631

6.2 Binary Response CRC Models

In this section, we conduct simulations for binary response CRC models. We compare the estimators as in section 6.1.1 with y_{it} substituted by $\frac{(y_{jt}-\mathbf{1}(v_{jt}>0))}{\hat{f}_t(v_{jt}|x_{jt},z_j)}$. We generate y_{it} by

$$y_{it} = \mathbf{1}(v_{it} + \beta_{0i} + x_{it}\beta_{1i} + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T;\ T = 3)$$

where $\beta_{0i} = \beta_0 + \alpha_{0i}$, $\beta_{1i} = \beta_1 + \alpha_{1i}$, $\beta_0 = 0.5$, $\beta_1 = 1$, x_{it} is i.i.d. with $Gamma(1, 1/3)$, and u_{it} is i.i.d. with $Uniform[-0.5, 0.5]$. α_{0i} and α_{1i} are generated in the following ways, where $\alpha_{0i} = w_{0i} - E(w_{0i})$ and $\alpha_{1i} = w_{1i} - E(w_{1i})$.

$$\begin{aligned}
DGP5: \quad & v_{it} \text{ is independent of } \alpha_{0i},\ \alpha_{1i} \text{ and } u_{it}, \text{ and distributed as } Uniform[-4, 4],\\
& w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = (\bar{x}_{i\cdot} - 1)^2 + \ln(\bar{x}_{i\cdot} + 1) + \eta_{1i},\\
DGP6: \quad & v_{it} \text{ is independent of } \alpha_{0i},\ \alpha_{1i} \text{ and } u_{it}, \text{ and distributed as } Uniform[-4, 4],\\
& w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = \sin(3\bar{x}_{i\cdot}) + \eta_{1i},\\
DGP7: \quad & v_{it} = \bar{x}_{i\cdot}^2 + w_{it}, \text{ where } w_{it} \sim Uniform[-4, 4],\\
& w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = (\bar{x}_{i\cdot} - 1)^2 + \ln(\bar{x}_{i\cdot} + 1) + \eta_{1i},\\
DGP8: \quad & v_{it} = \bar{x}_{i\cdot}^2 + w_{it}, \text{ where } w_{it} \sim Uniform[-4, 4],\\
& w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = \sin(3\bar{x}_{i\cdot}) + \eta_{1i},
\end{aligned}$$

where $\bar{x}_{i\cdot} = T^{-1}\sum_{t=1}^{T} x_{it}$, η_{0i} and η_{1i} are i.i.d. with $Uniform[-0.5, 0.5]$.

Table 5: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP5

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	0.0231	n/a	0.7049	0.0288	0.0474	0.6119	0.4586	15.2617	0.4788	0.6449
200	0.0133	n/a	0.1123	0.0134	0.0298	0.5513	0.3767	3.5648	0.2706	0.3271
400	0.0105	n/a	0.0528	0.0070	0.0197	0.5156	0.3262	1.7518	0.1812	0.2069

The simulation results are reported in Table 5, Table 6, Table 7 and Table 8. We can see that the semiparametric estimators we proposed perform well.

Table 6: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP6

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	0.0225	n/a	0.7078	0.0294	0.0489	0.4491	0.3688	14.2614	0.4306	0.6242
200	0.0114	n/a	0.1019	0.0135	0.0302	0.3794	0.2820	3.0915	0.2419	0.3166
400	0.0086	n/a	0.0539	0.0072	0.0195	0.3413	0.2341	1.6976	0.1602	0.2064

Table 7: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP7

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	0.0230	n/a	0.7132	0.0289	0.0461	0.6083	0.4543	15.7561	0.4661	0.6270
200	0.0144	n/a	0.1083	0.0139	0.0294	0.5699	0.3879	3.7572	0.2681	0.3204
400	0.0112	n/a	0.0496	0.0072	0.0192	0.5269	0.3356	1.7287	0.1821	0.2013

Table 8: MSE of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{GM}$, $\hat{\beta}_{Semi,1}$, $\hat{\beta}_{Semi,2}$ for DGP8

	$MSE(\hat{\beta}_0)$					$MSE(\hat{\beta}_1)$				
n	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FE}$	$\hat{\beta}_{GM}$	$\hat{\beta}_{Semi,1}$	$\hat{\beta}_{Semi,2}$
100	0.0220	n/a	0.7349	0.0292	0.0477	0.4434	0.3668	14.7899	0.4226	0.5975
200	0.0125	n/a	0.0970	0.0144	0.0306	0.3898	0.2946	3.2012	0.2448	0.3160
400	0.0088	n/a	0.0524	0.0073	0.0193	0.3385	0.2319	1.6847	0.1571	0.1958

7 An Empirical Application

In this section, we use the linear CRC panel data model to reexamine the return of on-the-job training. We consider the following simple wage equation

$$\log(wage_{it}) = \beta_{0i} + \beta_{1i}t + \beta_{2i}tenure_{it} + \beta_{3i}tenure^2_{it} + \beta_{4i}educ_{it} + \beta_{5i}union_{it} + \beta_{6i}training_{it} + u_{it}. \quad (7.1)$$

Here, β_{0i} is the fixed effects term which captures the time invariant characteristics of individuals, for instance, gender. We include a time trend to capture the individual wage growth. $tenure_{it}$ denotes weeks an individual has worked for the current employer, which describes the working experience. $tenure^2_{it}$ is used to capture the nonlinear effect of working experience as Mincer's wage equation. We use edu_{it} to denote years of schooling, $union_{it}$ to denote the union status of the individual, which is also an important ingredient for the wage, and $training_{it}$ to denote accumulated hours spent on the job training until time t. Then β_{5i} is the return from joining the union, and β_{6i} is the rate of return from the job training. Though some people took the job after finished the education, the years of schooling occasionally change for some other people, so we include an education term in the equation.

We know that people make decisions on whether to join the union depending on how much benefit they can get from this activity. Thus, there exists a correlation between $union_{it}$ and

β_{5i}. From the theory of human capital, we know that the marginal return of the job training is diminishing as the level of the training increases. Therefore, there is also a correlation between $training_{it}$ and β_{6i}. These make (7.1) a linear CRC panel data model. Also, random coefficients are used to capture unobserved heterogeneity.

We use 1979 cohort data from the National Longitudinal Survey of Youth (NLSY). The 1979 cohort data in NLSY is a data set of 12,686 individuals who were aged 14 to 21 in 1979, and interviewed every year from 1979 to 1994, and every two years after 1994. In 1988 and after, individuals were asked about the spell of their job training, i.e., weeks they spent on the training since last interview and hours per week spent on the training. We use the product of the weeks and hours to calculate the increment of hours spent on the job training since the last interview. The data also include other information about individuals, such as hourly wage, tenure, union status, years of schooling, etc. The descriptive statistics are given in Tables 10 and 11 at the end of the paper.

For the estimation of (7.1), we take first difference and get that

$$\begin{aligned}\log(wage_{it}) - \log(wage_{i,t-1}) &= \beta_{1i} + \beta_{2i}\Delta tenure_{it} + \beta_{3i}\Delta tenure_{it}^2 + \beta_{4i}\Delta educ_{it} + \beta_{5i}\Delta union_{it} \\ &\quad + \beta_{6i}\Delta training_{it} + \Delta u_{it}, \qquad (7.2)\end{aligned}$$

where $\Delta A_{it} = A_{it} - A_{i,t-1}$. The reason we do the first difference is that we can only observe the increment of hours spent on the job training since the last period, not the accumulated hours. Also, it helps us to get rid of the fixed effects term β_{0i}. Then we can use the OLS approach to estimate $\bar{\beta}_{1i}$, $\bar{\beta}_{2i}$, $\bar{\beta}_{3i}$, $\bar{\beta}_{4i}$, $\bar{\beta}_{5i}$ and $\bar{\beta}_{6i}$ which are population means of the random coefficients in (7.2), which is equivalent to fixed effects estimators for (7.1). We also use the nonparametric method we proposed in (3.4) to estimate (7.2). We report the result in the following table.

Table 9: Estimation results of (7.2) by OLS and nonparametric methods

Variables	First difference estimates	Nonparametric estimates
Time trend	5.67%	6.80%
Tenure (weeks)	0.053%	0.067%
Tenure^2	-0	-0
Education (years)	2.57%	4.29%
Union	11.33%	16.5%
Job training (per 100 hours)	0.6%	6.1%
Time range: 1988 - 2008 (14 interviews)		
Sample size: 3287		

We use the data of 3287 individuals who took job training during 1988 to 2008. From table 9, we can see that the first difference estimators underestimate the rate of return from the job training and joining the union. This is consistent with the discussions in the literature, e.g.

Frazis and Loewenstein (2005). Using our nonparametric method for correcting the correlations, we get the return of joining the union is 1.6 times as much as the one estimated by the first difference method. Also, the estimate of the return from job training based on our method is 10 times as much as the one estimated by the first difference method.

From the estimation results, we can see that the yearly increase rate of wage is 6.8%. The increase rate of tenure is 0.067% per week. The reason this is small is that for most people who continuously work for a same employer, the tenure is proportional to the difference of time. So part of the increase from tenure is absorbed in the yearly increment. Moreover, we can see that there is no obvious nonlinear effect of the tenure due to the similar reason as tenure. The rate of return of education is 4.29% for one year more education. Also, we find that the return from joining the union is 16.5%, and the rate of return from job training is 6.1% per 100 hours training.

Overall, the estimator we proposed can make a difference compared with the usual first difference estimation. The magnitude of these values are very reasonable.

8 Conclusion

In this paper, we discuss the identification and estimation of a linear CRC panel data model and a binary response CRC panel data model. We use the linear CRC panel data model to show how we deal with the general correlation between random coefficients and regressors in the CRC model. Also, the linear CRC panel data model has usefulness in its own for the analysis of the average treatment effect. Further, we extend the idea to the binary choice CRC panel data model. The identification of the binary choice model is different from the linear model. We base our identification result on the special regressor method. Moreover, we construct the $\sqrt{n}$ consistent asymptotically normal semiparametric estimators for both models. Further, we did simulations and an empirical application to show the advantage of our estimators.

There are some extensions we are considering. In the example given in section 2, the regressor is a discrete variable but we mainly discuss the identification and estimation results for continuous variables in this paper. Though, similar discussions can be made by using kernel smoothing method for discrete variables as in Li and Racine (2007), we leave the rigorous derivations for future research. In addition, it is desirable to construct tests for CRC panel data models. We also leave this for further research.

Appendix A

Proof of Theorem 3.1: We first consider the local constant estimation method. For any $z \in \Omega_z$, we have

$$\begin{aligned}
\hat{\theta}_{LC}(z) &= \left[\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z)\right]^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z) \\
&= \theta(z) + \left[\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z)\right]^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}[x_{js}^{\top}(\theta(z_j)-\theta(z))+\epsilon_{js}]K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z) \\
&= \theta(z) + A_{n1}(z)^{-1}\left[A_{n2}(z)+A_{n3}(z)\right], \qquad \text{(A.1)}
\end{aligned}$$

where

$$\begin{aligned}
A_{n1}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z), \\
A_{n2}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top}(\theta(z_j)-\theta(z))K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z), \\
A_{n3}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}\epsilon_{js}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z),
\end{aligned}$$

with $H = h_1 \cdots h_q$ and $K_{h,z_jz} = K((z_j - z)/h) = \prod_{s=1}^{q} k((z_{js}-z_s)/h_s)$.

Using (A.1) we have

$$\begin{aligned}
\hat{\beta}_{LC} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{LC}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\theta(z_i) + \frac{1}{n}\sum_{i=1}^{n}A_{n1}(z_i)^{-1}\left[A_{n2}(z_i)+A_{n3}(z_i)\right].
\end{aligned}$$

By Lemma A.1 we have uniformly in $z \in \Omega_z$,

$$A_{n1}(z)^{-1} = m(z)^{-1} + O_p(||h||^{\nu} + (\ln n/(nH))^{1/2}),$$

where $m(z) = T^{-1}\sum_{s=1}^{T} E[x_{js}x_{js}^{\top}|z_j = z]f(z)$.

So we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}A_{n1}(z_i)^{-1}\left[A_{n2}(z_i)+A_{n3}(z_i)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}m(z_i)^{-1}\left[A_{n2}(z_i)+A_{n3}(z_i)\right] + \eta_n \\
&\equiv B_{n1} + B_{n2} + \eta_n,
\end{aligned}$$

where $B_{n1} = n^{-1}\sum_{i=1}^n m(z_i)^{-1}A_{n2}(z_i)$, $B_{n2} = n^{-1}\sum_{i=1}^n m(z_i)^{-1}A_{n3}(z_i)$, $\eta_n = O_p(||h||^\nu + (\ln n/(nH))^{1/2})O_p(\|A_{n2}(z_i)\| + \|A_{n3}(z_i)\|)$. B_{n1} and B_{n2} correspond to 'bias' and 'variance' terms, respectively.

We first consider B_{n1}. Note that B_{n1} can be written as a second order U-statistic.

$$B_{n1} = n^{-2}\frac{n(n-1)}{2}\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n H_{n1,ij} \equiv n^{-2}\frac{n(n-1)}{2}U_{n1},$$

where $H_{n1,ij} = (TH)^{-1}\sum_{s=1}^T [m(z_i)^{-1}x_{js}x_{js}^\top(\theta_j-\theta_i)\mathbf{1}_{\varepsilon_n}(z_i) + m(z_j)^{-1}x_{is}x_{is}^\top(\theta_i-\theta_j)\mathbf{1}_{\varepsilon_n}(z_j)]K_{h,ji}$, $K_{h,ji} = K_h((z_j-z_i)/h)$. Using the U-statistic H-decomposition we have

$$U_{n1} = E[H_{n1,ij}] + \frac{2}{n}\sum_{i=1}^n [H_{n1,i} - E(H_{n1,i})] + \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n [H_{n1,ij} - H_{n1,i} - H_{n1,j} + E(H_{n1,ij})],$$

where $H_{n1,i} = E[H_{n1,ij}|w_i]$, $w_i = (x_i, z_i) = (x_{i1},\ldots,x_{iT}, z_i)$.

Since $\|h\|/\varepsilon_n \to 0$ and the kernel function $K(\cdot)$ has a compact support, the trimming function $\mathbf{1}_{\varepsilon_n}(z_i)$ will ensure that all of the points which have boundary effects are excluded from our estimated locations. We have that

$$\begin{aligned}
E[H_{n1,ij}] &= (TH)^{-1}\sum_{s=1}^T E[m_i^{-1}x_{js}x_{js}^\top(\theta_j-\theta_i)K_{h,ij}] \\
&= (TH)^{-1}\sum_{s=1}^T E[m_i^{-1}E(x_{js}x_{js}^\top|z_j)(\theta_j-\theta_i)K_{h,ij}] \\
&= H^{-1}E[m_i^{-1}m_jf_j^{-1}(\theta_j-\theta_i)K_{h,ij}] \\
&= H^{-1}\int\int m_i^{-1}f_im_j(\theta_j-\theta_i)K_{h,ij}dz_idz_j \\
&= \int\int m_i^{-1}f_im(z_i+hv)(\theta(z_i+hv)-\theta_i)K(v)dvdz_i \\
&= \mu_\nu\sum_{l=1}^q\sum_{k_1+k_2=\nu,k_2\neq 0}\frac{h_l^\nu}{k_1!k_2!}\int m_i^{-1}f_i(\frac{\partial^{k_1}m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})dz_i + O(||h||^{\nu+1}) \\
&= \sum_{l=1}^q h_l^\nu B_{l,LC} + O_p(||h||^{\nu+1}),
\end{aligned}$$

where $B_{l,LC} = \mu_\nu\sum_{k_1+k_2=\nu,k_2\neq 0}\frac{1}{k_1!k_2!}E\left[m_i^{-1}(\frac{\partial^{k_1}m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})\right]$.

Also, we have

$$\begin{aligned}
&E\left[\left(\frac{2}{n}\sum_{i=1}^n[H_{n1,i}-E(H_{n1,i})]\right)\left(\frac{2}{n}\sum_{i=1}^n[H_{n1,i}-E(H_{n1,i})]\right)^\top\right] \\
=\ & Var\left[\frac{2}{n}\sum_{i=1}^n[H_{n1,i}-E(H_{n1,i})]\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{4}{n^2}\sum_{i=1}^{n} Var[H_{n1,i} - E(H_{n1,i})] \\
&= \frac{4}{n^2}\sum_{i=1}^{n} E\left[[H_{n1,i} - E(H_{n1,i})][H_{n1,i} - E(H_{n1,i})]^\top\right] \\
&= O(n^{-1}\|h\|^{2\nu}),
\end{aligned}
$$

and

$$
\begin{aligned}
& Var\left[\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}[H_{n1,ij} - H_{n1,i} - H_{n1,j} + E(H_{n1,ij})]\right] \\
= & \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}^{n} Var\left[H_{n1,ij} - H_{n1,i} - H_{n1,j} + E(H_{n1,ij})\right] \\
= & \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}^{n} E\left[\left[H_{n1,ij} - H_{n1,i} - H_{n1,j} + E(H_{n1,ij})\right]\left[H_{n1,ij} - H_{n1,i} - H_{n1,j} + E(H_{n1,ij})\right]^\top\right] \\
= & O(n^{-2}H^{-1}\|h\|^2).
\end{aligned}
$$

Hence, $B_{n1} = \sum_{l=1}^{q} h_l^\nu B_{l,LC} + O_p(||h||^{\nu+1} + n^{-1}H^{-1/2}\|h\|)$.

We decompose B_{n2} into two terms

$$
B_{n2} = B_{n2,1} + B_{n2,2},
$$

where

$$
\begin{aligned}
B_{n2,1} &= (n^2TH)^{-1}\sum_{i=1}^{n}\sum_{s=1}^{T} m(z_i)^{-1}x_{is}\epsilon_{is}K(0)\mathbf{1}_{\varepsilon_n}(z_i), \\
B_{n2,2} &= (n^2TH)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{s=1}^{T} m(z_i)^{-1}x_{js}\epsilon_{js}K_{h,ji}\mathbf{1}_{\varepsilon_n}(z_i).
\end{aligned}
$$

It is easy to see that $E[B_{n2,1}] = 0$ and $E[||B_{n2,1}||^2] = (n^4H^2)^{-1}O(n) = O((n^3H^2)^{-1})$. Hence, $B_{n2,1} = O_p((n^{3/2}H)^{-1})$.

$B_{n2,2}$ can be written as a second order U-statistic.

$$
B_{n2,2} = n^{-2}\frac{n(n-1)}{2}U_{n2},
$$

where $U_{n2} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} H_{n2,ij}$, $H_{n2,ij} = (TH)^{-1}\sum_{s=1}^{T}(m_i^{-1}x_{js}\epsilon_{js}\mathbf{1}_{\varepsilon_n}(z_i) + m_j^{-1}x_{is}\epsilon_{is}\mathbf{1}_{\varepsilon_n}(z_j))K_{h,ij}$.

Since U_{n2} has zero mean, its H-decomposition is given by

$$
U_{n2} = U_{n2,1} + U_{n2,2},
$$

where $U_{n2,1} = \frac{2}{n}\sum_{i=1}^n H_{n2,i}$ and $U_{n2,2} = \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j>i}^n [H_{n2,ij} - H_{n2,i} - H_{n2,j}]$, $H_{n2,i} = E[H_{n2,ij}|w_i]$, $w_i = (x_i, \alpha_i, z_i, u_i) = (x_{i1}, \ldots, x_{iT}, \alpha_i, z_i, u_{i1}, \ldots, u_{iT})$. It is easy to show that $U_{n2,1}$ is the leading term of U_{n2}.

$$\begin{aligned}
U_{n2,1} &= \frac{1}{nTH}\sum_{i=1}^n\sum_{s=1}^T E\left[(m_i^{-1}x_{js}\epsilon_{js}\mathbf{1}_{\varepsilon_n}(z_i) + m_j^{-1}x_{is}\epsilon_{is}\mathbf{1}_{\varepsilon_n}(z_j))K_{h,ij}|w_i\right] \\
&= \frac{1}{nTH}\sum_{i=1}^n\sum_{s=1}^T E\Big[\Big(m_i^{-1}x_{js}x_{js}^\top(\alpha_j - E(\alpha_j|z_j))\mathbf{1}_{\varepsilon_n}(z_i) + m_i^{-1}x_{js}u_{js}\mathbf{1}_{\varepsilon_n}(z_i) + m_j^{-1}x_{is}x_{is}^\top \\
&\quad (\alpha_i - E(\alpha_i|z_i))\mathbf{1}_{\varepsilon_n}(z_j) + m_j^{-1}x_{is}u_{is}\mathbf{1}_{\varepsilon_n}(z_j)\Big)K_{h,ij}|w_i\Big] \\
&= \frac{1}{nTH}\sum_{i=1}^n\sum_{s=1}^T \left(E[m_j^{-1}K_{h,ij}|w_i]x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i))\mathbf{1}_{\varepsilon_n}(z_i) + u_{is}\mathbf{1}_{\varepsilon_n}(z_i)E[m_j^{-1}x_{is}K_{h,ij}|w_i]\right) \\
&= \frac{1}{nT}\sum_{i=1}^n\sum_{s=1}^T \left(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\mathbf{1}_{\varepsilon_n}(z_i) + O_p(\|h\|^{\nu+1}/\sqrt{n}).
\end{aligned} \tag{A.2}$$

It is easy to evaluate its second moment $E[||U_{n2,2}||^2] = (n^4H^2)^{-1}n^2O(H) = O((n^2H)^{-1})$. Hence, $U_{n2,2} = O_p((nH^{1/2})^{-1})$.

Summarizing the above, we have shown that

$$\begin{aligned}
\hat{\beta}_{LC} &= \frac{1}{n}\sum_{i=1}^n\theta(z_i) + \frac{1}{nT}\sum_{i=1}^n\sum_{s=1}^T\Big(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\Big)\mathbf{1}_{\varepsilon_n}(z_i) \\
&\quad + B_{LC} + O_p\Big((nH^{1/2})^{-1} + ||h||^{\nu+1} + ((nH^{1/2})^{-1} + ||h||^\nu)(||h||^\nu + (\ln n/(nH))^{1/2})\Big).
\end{aligned} \tag{A.3}$$

Also, by Cauchy-Schwarz inequality we have that

$$\begin{aligned}
& E\|\frac{1}{nT}\sum_{i=1}^n\sum_{s=1}^T\left(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)^{\otimes 2}(1 - \mathbf{1}_{\varepsilon_n}(z_i))\| \\
\leq\ & \{E(\|m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\|^2)P(z_i \in \mathcal{S}_z\backslash\Omega_z)\}^{1/2},
\end{aligned}$$

where $A^{\otimes 2}$ denotes $AA^\top$ for any matrix A. Since the density function $f_z(z_i)$ of z_i is bounded and the volume of the set that is within a distance ε_n of $\partial\mathcal{S}_z$ is proportional to ε_n, we have that $P(z_i \in \mathcal{S}_z\backslash\Omega_z) = O(\varepsilon_n)$. Hence, $Var\Big(\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{s=1}^T\left(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\mathbf{1}_{\varepsilon_n}(z_i)\Big) = Var\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{s=1}^T\left(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\right) + o(1)$.

Hence, by noting that $\beta = E[\theta(z_i)]$ and letting $v_i = \theta(z_i) - \beta$, we have

$$\begin{aligned}
& \sqrt{n}\left(\hat{\beta}_{LC} - \beta - B_{LC}\right) \\
=\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n v_i + \frac{1}{\sqrt{n}T}\sum_{i=1}^n\sum_{s=1}^T\left(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\mathbf{1}_{\varepsilon_n}(z_i) + O_p(\zeta_n) \\
\xrightarrow{d}\ & N(0, V_{LC})
\end{aligned} \tag{A.4}$$

by the Lindeberg central limit theorm, where $V_{LC} = Var(v_i) + T^{-1}Var(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)) = Var(\theta_i) + T^{-1}Var(m_i^{-1}f(z_i)x_{is}x_{is}^\top(\alpha_i - E(\alpha_i|z_i))) + T^{-1}Var(u_{is}m_i^{-1} x_{is}f(z_i))$ and $\zeta_n = (nH)^{-1/2} + (n||h||^{2\nu+2})^{1/2} + (nH)^{-1/2}\|h\| + \sqrt{n}\|h\|^{2\nu} + \sqrt{n\|h\|^{2\nu}}(\ln n/(nH))^{1/2} + \|h\|^{\nu}(nH)^{-1/2} + (nH)^{-1/2}(\ln n/(nH))^{1/2} = o_p(1)$.

Lemma A.1 *Define* $A_{n1}(z) = \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^\top K_{h,z_jz}$, *and* $m(z) = T^{-1}\sum_{s=1}^{T} E[x_{js}x_{js}^\top | z_j = z]f(z)$, *where* $K_{h,z_jz} = \prod_{l=1}^{q} k\left(\frac{z_{jl}-z_l}{h_l}\right)$, *then under Assumptions A1-A4,*

$$A_{n1}(z)^{-1} \quad = \quad m(z)^{-1} + O_p\left(||h||^{\nu} + (\ln n)^{1/2}(nH)^{-1/2}\right),$$

uniformly in $z \in \Omega_z$, *where* $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,\dots,q\}} |z_l - z_{0,l}| \geq \varepsilon_n$ *for some* $z_0 \in \partial\mathcal{S}_z\}$, $\partial\mathcal{S}_z$ *is the boundary of the compact set* $\mathcal{S}_z$, $\varepsilon_n \to 0$ *and* $\|h\|/\varepsilon_n \to 0$, *as* $n \to \infty$.

Proof: First, we have

$$E[A_{n1}(z)] = m(z) + O\left(\|h\|^{\nu}\right), \tag{A.5}$$

uniformly in $z \in \Omega_z$. Following similar arguments used in Masry (1996) when deriving uniform convergence rates for nonparametric kernel estimators, we know that

$$A_{n1}(z) - E[A_{n1}(z)] = O_p\left(\frac{(\ln n)^{1/2}}{(nH)^{1/2}}\right), \tag{A.6}$$

uniformly in $z \in \Omega_z$.

Combining (A.5) and (A.6) we have

$$A_{n1}(z) - m(z) = O_p\left(||h||^{\nu} + (\ln n)^{1/2}(nH)^{-1/2}\right), \tag{A.7}$$

uniformly in $z \in \Omega_z$.

Using (A.7) we obtain

$$\begin{aligned} A_{n1}(z)^{-1} &= [m(z) + A_{n1}(z) - m(z)]^{-1} \\ &= m(z)^{-1} - m(z)^{-1}[A_{n1}(z) - m(z)]\, m(z)^{-1} + O_p\left(\|A_{n1}(z) - m(z))\|^2\right) \\ &= m(z)^{-1} + O_p\left(||h||^{\nu} + (\ln n)^{1/2}(nh_1\cdots h_q)^{-1/2}\right), \end{aligned}$$

which completes the proof of Lemma A.1.

Proof of Theorem 3.2: Now, we consider the local polynomial estimation method.

The minimization of (3.7) leads to the set of equations

$$t_{n,i}(z) = \sum_{0\leq|k|\leq p} h^k \hat{b}_k(z) s_{n,i+k}(z),\ 0 \leq |i| \leq p \tag{A.8}$$

where

$$\begin{aligned} t_{n,i}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}\left(\frac{z_i - z}{h}\right)^{i} K_{h,z_jz}, \\ s_{n,i+k}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^\top\left(\frac{z_i - z}{h}\right)^{i+k} K_{h,z_jz}. \end{aligned}$$

We put the set of equations (A.8) into a lexicographical order. Let $N_r = \binom{r+q-1}{q-1}$ be the number of distinct q-tuples i with $|i| = r$. Stacking $t_{n,i}(z)$, $|i| = r$ up into a column vector according to these N_r q-tuples by a lexicographical order, i.e., $(0, \ldots, 0, r)$ is the first element and $(r, 0, \ldots, 0)$ is the last one. Denote this vector by $\tau_{n,r}(z)$. Let $\tau_n = (\tau_{n,0}(z)^\top, \tau_{n,1}(z)^\top, \ldots, \tau_{n,p}(z)^\top)^\top$. Note that the column vector $\tau_n(z)$ is of dimension $N = \sum_{i=0}^{p} N_i \times d$. Similarly, we can arrange $h^k \hat{b}_k(z)$, $0 \le |k| \le p$ into a $N \times 1$ column vector according to the lexicographical order of k as $\hat{\delta}(z) = (\hat{\delta}_{n,0}(z)^\top, \hat{\delta}_{n,1}(z)^\top, \ldots, \hat{\delta}_{n,p}(z)^\top)^\top$. Finally, we arrange $s_{n,i+k}(z)$ into a matrix $(S_{n,|i|,|k|}(z))_{N\times N}$, where columns are according the lexicographical order of i and rows are following the lexicographical order of k. Thus, denote the $N \times N$ matrix $S_n(z)$ by

$$S_n(z) = \begin{pmatrix} S_{n,0,0}(z) & S_{n,0,1}(z) & \cdots & S_{n,0,p}(z) \\ S_{n,1,0}(z) & S_{n,1,1}(z) & \cdots & S_{n,1,p}(z) \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p,0}(z) & S_{n,p,1}(z) & \cdots & S_{n,p,p}(z) \end{pmatrix}.$$

Hence, $\hat{\delta}(z) = S_n(z)^{-1}\tau_n(z)$. Let $P_1 = e_1^\top \otimes I_{d\times d}$, where $e_1 = (1, 0, \ldots, 0)^\top$ is a $(\sum_{i=0}^{p} N_i) \times 1$ vector containing the first element as 1 and others as 0, $I_{d\times d}$ is the $d \times d$ identity matrix, and $\otimes$ is the kronecker product. Then $\hat{\theta}_{LP}(z) = P_1\hat{\delta}(z)$.

Using similar arguments in Masry (1996), we can show that

$$S_n(z) \;\; = \;\; S(z) + O_p\left(||h|| + (\ln n)^{1/2}(nH)^{-1/2}\right),$$

uniformly in $z \in \Omega_z$, where $S(z) = (S_{|i|,|k|}(z))_{N\times N}$ has each element corresponding to $S_n(z)$, for the corresponding element $s_{i+k}(z)$ in $S(z)$, $s_{i+k}(z) = T^{-1}\sum_{s=1}^{T} E[x_{js}x_{js}^\top | z_j = z] f(z)\mu_{i+k}$, and $\mu_{i+k} = \int u^{i+k}K(u)du$.

Hence,

$$S_n(z)^{-1} \;\; = \;\; S(z)^{-1} + O_p\left(||h|| + (\ln n)^{1/2}(nH)^{-1/2}\right),$$

uniformly in $z \in \Omega_z$.

We can write $t_{n,i}(z)$ as

$$\begin{aligned} t_{n,i}(z) &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}\left(\frac{z_i - z}{h}\right)^i K_{h,z_j z} \\ &= \frac{1}{nTH}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}(x_{js}^\top\theta(z_j) + \epsilon_{js})\left(\frac{z_i - z}{h}\right)^i K_{h,z_j z}. \end{aligned}$$

Also, we have that

$$\hat{\delta}(z) = \delta(z) + S_n(z)^{-1}(C_{n1}(z) + C_{n2}(z)),$$

where $\delta(z)$ is corresponding to $\hat{\delta}(z)$ with elements from $h^k D^k\theta(z)/k!$ instead of $h^k\hat{b}_k(z)$, $C_{n1}(z)$ and C_{n2} are $N \times 1$ vectors with elements from $t^*_{n,i} = (nTH)^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^\top(\theta(z_j) -$

$\sum_{0\leq|k|\leq p}\frac{1}{k!}(D^k\theta(z))(z_j-z)^k)\left(\frac{z_i-z}{h}\right)^i K_{h,z_jz}$ and $(nTH)^{-1}\sum_{j=1}^n\sum_{s=1}^T x_{js}\epsilon_{js}\left(\frac{z_i-z}{h}\right)^i K_{h,z_jz}$, respectively.

Since $\hat{\theta}_{LP}(z)=P_1\hat{\delta}(z)$, we have

$$
\begin{aligned}
\hat{\beta}_{LP} &= \frac{1}{n}\sum_{i=1}^n\hat{\theta}_{LP}(z_i)\\
&= \frac{1}{n}\sum_{i=1}^n\theta(z_i)+\frac{1}{n}\sum_{i=1}^n P_1S_n(z_i)^{-1}\left[C_{n1}(z_i)+C_{n2}(z_i)\right]\\
&= \frac{1}{n}\sum_{i=1}^n\theta(z_i)+\frac{1}{n}\sum_{i=1}^n P_1S(z_i)^{-1}\left[C_{n1}(z_i)+C_{n2}(z_i)\right]+(s.o.),
\end{aligned}
$$

where $(s.o.)$ denotes terms with smaller orders.

Similar as in the proof of Theorem3.1, we have that if $p>0$ is an odd integer,

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n P_1S(z_i)^{-1}C_{n1}(z_i) &= \sum_{|k|=p+1}\frac{\mu_kh^k}{k!}P_1E\left[S_i^{-1}M_i\Theta_i\right]+O_p(\|h\|^{p+2}+n^{-1}H^{-1/2}\|h\|^{p+1})\\
&= B_{LP}+O_p(\|h\|^{p+2}+n^{-1}H^{-1/2}\|h\|^{p+1}),\\
\frac{1}{n}\sum_{i=1}^n P_1S(z_i)^{-1}C_{n2}(z_i) &= \frac{1}{nT}\sum_{i=1}^n\sum_{s=1}^T P_1S(z_i)^{-1}\Gamma_{is}x_{is}^\top(\alpha_i-E(\alpha_i|z_i))f(z_i)\\
&\quad+\frac{1}{nT}\sum_{i=1}^n\sum_{s=1}^T P_1S(z_i)^{-1}u_{is}f(z_i)\Gamma_{is}+O_p(\|h\|^2/\sqrt{n}+(nH^{1/2})^{-1}),
\end{aligned}
$$

where $M_i=M(z_i)=(M_{0,p+1}(z_i)^\top,M_{1,p+1}(z_i)^\top,\ldots,M_{p,p+1}(z_i)^\top)^\top$, $M_{j,p+1}(z)$ is corresponding to $S_{n,j,p+1}(z)$ which is similar as elements in $S_n(z)$, $\Theta_i=\Theta(z_i)$ which has the elements from $(1/k!)D^k\theta(z)|_{z=z_i}$ using the lexicographical order, and Γ_{is} is a $N\times1$ column vector with elements from $x_{is}\mu_\alpha$ following the lexicographical order. The elements in $M(z)$ are from $s_{\alpha+p+1}=T^{-1}\sum_{s=1}^T E[x_{js}x_{js}^\top|z_j=z]f(z)\mu_{\alpha+p+1}$. If we denote S for the $N\times N$ matrix which has the elements from $\mu_{\alpha+\gamma}$, $0\leq|\alpha|\leq p$, $0\leq|\gamma|\leq p$, and M for the $N\times1$ vector which has the elements from $\mu_{\alpha+p+1}$ following the lexicographical order introduced earlier. We have that $S_i^{-1}M_i=S^{-1}M$. Thus $B_{LP}=P_1S^{-1}M\sum_{|k|=p+1}\frac{\mu_kh^k}{k!}E\left[\Theta_i\right]$.

If $p>0$ is an even integer, we have that

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^n P_1S(z_i)^{-1}C_{n1}(z_i) &= \sum_{|k|=p+2}\frac{\mu_kh^k}{k!}P_1E\left[S_i^{-1}M_i\Theta_i\right]+O_p(\|h\|^{p+4}+n^{-1}H^{-1/2}\|h\|^{p+2})\\
&= P_1S^{-1}M\sum_{|k|=p+2}\frac{\mu_kh^k}{k!}E\left[\Theta_i\right]+O_p(\|h\|^{p+4}+n^{-1}H^{-1/2}\|h\|^{p+2})\\
&= B_{LP}+O_p(\|h\|^{p+4}+n^{-1}H^{-1/2}\|h\|^{p+2}).
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
& \sqrt{n}\left(\hat{\beta}_{LP}-\beta-B_{LP}\right) \\
= \quad & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\theta_i-\beta)+\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T}P_1S(z_i)^{-1}\Gamma_{is}x_{is}^{\top}(\alpha_i-E(\alpha_i|z_i))f(z_i) \\
& +\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T}P_1S(z_i)^{-1}u_{is}f(z_i)\Gamma_{is}+O_p(\zeta_n) \\
\stackrel{d}{\to} \quad & N(0,V_{LP}) \qquad \text{(A.9)}
\end{aligned}
$$

by the Lindeberg central limit theorem, where $V_{LP}=Var(\theta_i)+T^{-1}Var(P_1S(z_i)^{-1}\Gamma_{is}x_{is}^{\top}(\alpha_i-E(\alpha_i|z_i))f(z_i))+T^{-1}Var(P_1S(z_i)^{-1}u_{is}f(z_i)\Gamma_{is})$ and $\zeta_n=(nH)^{-1/2}+(n||h||^{2p+4})^{1/2}+(nH)^{-1/2}\|h\|^{p+1}+\sqrt{n\|h\|^{2p+2}}(\ln n/(nH))^{1/2}+\|h\|(nH)^{-1/2}+(nH)^{-1/2}(\ln n/(nH))^{1/2}=o_p(1)$ if $p>0$ is an odd integer, or $\zeta_n=(nH)^{-1/2}+(n||h||^{2p+8})^{1/2}+(nH)^{-1/2}\|h\|^{p+2}+\sqrt{n}||h||^{p+3}+\sqrt{n\|h\|^{2p+4}}(\ln n/(nH))^{1/2}+\|h\|(nH)^{-1/2}+(nH)^{-1/2}(\ln n/(nH))^{1/2}=o_p(1)$ if $p>0$ is an even integer.

Appendix B

Proof of Proposition 4.1: Since $\beta_i=\beta+\alpha_i$, $g(z_i)=E(\alpha_i|x_{it},z_i)=E(\alpha_i|z_i)$, we have

$$
\begin{aligned}
y_{it} &= \mathbf{1}(v_{it}+x_{it}^{\top}\beta+x_{it}^{\top}\alpha_i+u_{it}>0) \\
&= \mathbf{1}(v_{it}+x_{it}^{\top}\beta+x_{it}^{\top}g(z_i)+x_{it}^{\top}(\alpha_i-g(z_i))+u_{it}>0) \\
&= \mathbf{1}(v_{it}+x_{it}^{\top}\theta(z_i)+e_{it}>0),
\end{aligned}
$$

where $\theta(z_i)=\beta+g(z_i)$, and $e_{it}=x_{it}^{\top}(\alpha_i-g(z_i))+u_{it}$. Since $E(u_{it}|x_{it},z_i)=0$, we have $E(e_{it}|x_{it},z_i)=E[x_{it}^{\top}(\alpha_i-g(z_i))|x_{it},z_i]+E(u_{it}|x_{it},z_i)=x_{it}^{\top}E[(\alpha_i-g(z_i))|x_{it},z_i]+E[u_{it}|x_{it},z_i]=0$.

From Assumption C2, we have the conditional distribution $F_{e_{it}}(e_{it}|v_{it},x_{it},z_i)$ of e_{it} conditioning on (v_{it},x_{it},z_i) satisfies that $F_{e_{it}}(e_{it}|v_{it},x_{it},z_i)=F_{e_{it}}(e_{it}|x_{it},z_i)$. Also,

$$
y_{it}^{*}=\begin{cases}[y_{it}-\mathbf{1}(v_{it}>0)]/f_t(v_{it}|x_{it},z_i) & \text{if } v_{it}\in[L_t,K_t] \\ 0 & \text{otherwise}\end{cases},
$$

then

$$
\begin{aligned}
E(y_{it}^{*}|x_{it},z_i) &= E\left[(y_{it}-\mathbf{1}(v_{it}>0))/f_t(v_{it}|x_{it},z_i)|x_{it},z_i\right] \\
&= E\left[\frac{E[y_{it}-\mathbf{1}(v_{it}>0)|v_{it},x_{it},z_i]}{f_t(v_{it}|x_{it},z_i)}|x_{it},z_i\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{L_t}^{K_t} \frac{E[y_{it} - \mathbf{1}(v_{it} > 0)|v_{it}, x_{it}, z_i]}{f_t(v_{it}|x_{it}, z_i)} f_t(v_{it}|x_{it}, z_i) dv_{it} \\
&= \int_{L_t}^{K_t} \int_{\Omega_{e_t}} [\mathbf{1}(v_{it} + x_{it}^\top \theta(z_i) + e_{it} > 0) - \mathbf{1}(v_{it} > 0)] dF_{e_{it}}(e_{it}|v_{it}, x_{it}, z_i) dv_{it} \\
&= \int_{\Omega_{e_t}} \int_{L_t}^{K_t} [\mathbf{1}(v_{it} > s_{it}) - \mathbf{1}(v_{it} > 0)] dv_{it} dF_{e_{it}}(e_{it}|x_{it}, z_i) \qquad \text{Let } (s_{it} = -x_{it}^\top \theta(z_i) - e_{it}) \\
&= \int_{\Omega_{e_t}} \int_{L_t}^{K_t} [(\mathbf{1}(v_{it} > s_{it}) - \mathbf{1}(v_{it} > 0))\mathbf{1}(s_{it} \le 0) + (\mathbf{1}(v_{it} > s_{it}) - \mathbf{1}(v_{it} > 0))\mathbf{1}(s_{it} > 0)] dv_{it} dF_{e_{it}}(e_{it}|x_{it}, z_i) \\
&= \int_{\Omega_{e_t}} \int_{L_t}^{K_t} [\mathbf{1}(s_{it} < v_{it} \le 0)\mathbf{1}(s_{it} \le 0) - \mathbf{1}(0 < v_{it} \le s_{it})\mathbf{1}(s_{it} > 0)] dv_{it} dF_{e_{it}}(e_{it}|x_{it}, z_i) \\
&= \int_{\Omega_{e_t}} [\mathbf{1}(s_{it} \le 0) \int_{s_{it}}^{0} 1 dv_{it} - \mathbf{1}(s_{it} > 0) \int_{0}^{s_{it}} 1 dv_{it}] dF_{e_{it}}(e_{it}|x_{it}, z_i) \\
&= \int_{\Omega_{e_t}} -s_{it} dF_{e_{it}}(e_{it}|x_{it}, z_i) \\
&= \int_{\Omega_{e_t}} (x_{it}^\top \theta(z_i) + e_{it}) dF_{e_{it}}(e_{it}|x_{it}, z_i) \\
&= x_{it}^\top \theta(z_i) + E(e_{it}|x_{it}, z_i) \\
&= x_{it}^\top \theta(z_i).
\end{aligned}
$$

This completes the proof.

We give some shorthand notations first. These notations will be used throughout the proof of Theorem 5.1. Let

$$
\begin{aligned}
K_{h',z,jz} &= K_{h'}(z_j - z),\ K_{h',z,ji} = K_{h'}(z_j - z_i),\ K_{h',z,ij} = K_{h'}(z_i - z_j),\ K_{h',z,jk} = K_{h'}(z_j - z_k), \\
K_{h',z,kj} &= K_{h'}(z_k - z_j),\ K_{h',z,ki} = K_{h'}(z_k - z_i),\ K_{h',z,ik} = K_{h'}(z_i - z_k), \\
\hat{f}_{t,v|xz,j} &= \hat{f}_t(v_{jt}|x_{jt}, z_j),\ f_{t,v|xz,j} = f_t(v_{jt}|x_{jt}, z_j),\ \hat{f}_{t,vxz,j} = \hat{f}_t(v_{jt}, x_{jt}, z_j),\ f_{t,vxz,j} = f_t(v_{jt}, x_{jt}, z_j), \\
f_{t,vxz,i} &= f_t(v_{it}, x_{it}, z_i),\ f_{t,vxz,k} = f_t(v_{kt}, x_{kt}, z_k), \\
f_{t,vxz,j}^{-1} &= f_t^{-1}(v_{jt}, x_{jt}, z_j),\ f_{t,vxz,i}^{-1} = f_t^{-1}(v_{it}, x_{it}, z_i),\ f_{t,vxz,k}^{-1} = f_t^{-1}(v_{kt}, x_{kt}, z_k), \\
\hat{f}_{t,xz,j} &= \hat{f}_t(x_{jt}, z_j),\ f_{t,xz,j} = f_t(x_{jt}, z_j), \\
\mathbf{1}_{\tau_n,j} &= \mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j),\ \mathbf{1}_{\tau_n,i} = \mathbf{1}_{\tau_n}(v_{it}, x_{it}, z_i),\ \mathbf{1}_{\tau_n,k} = \mathbf{1}_{\tau_n}(v_{kt}, x_{kt}, z_k), \\
\theta_j &= \theta(z_j),\ \theta_i = \theta(z_i),\ \theta_k = \theta(z_k), \\
m_i &= m(z_i) = T^{-1} \sum_{s=1}^{T} E[x_{is} x_{is}^\top | z_i] f_z(z_i),\ m_j = m(z_j),\ m_k = m(z_k), \\
K_{h,vxz,kj} &= K_h(v_{kt} - v_{jt}, x_{kt} - x_{jt}, z_k - z_j),\ K_{h,vxz,ki} = K_h(v_{kt} - v_{it}, x_{kt} - x_{it}, z_k - z_i), \\
K_{h,vxz,ij} &= K_h(v_{it} - v_{jt}, x_{it} - x_{jt}, z_i - z_j),\ K_{h,vxz,jk} = K_h(v_{jt} - v_{kt}, x_{jt} - x_{kt}, z_j - z_k), \\
K_{h,vxz,ik} &= K_h(v_{it} - v_{kt}, x_{it} - x_{kt}, z_i - z_k),\ K_{h,vxz,ji} = K_h(v_{jt} - v_{it}, x_{jt} - x_{it}, z_j - z_i), \\
K_{h,vxz,mj} &= K_h(v_{mt} - v_{jt}, x_{mt} - x_{jt}, z_m - z_j),\ K_{\tilde{h},xz,mj} = K_{\tilde{h}}(x_{mt} - x_{jt}, z_m - z_j).
\end{aligned}
$$

Proof of Theorem 5.1: For $z \in \Omega_z$, let

$$\begin{aligned} A_{n1}(z) &= (nTH')^{-1} \sum_{j=1}^{n} \sum_{t=1}^{T} x_{jt} x_{jt}^\top K_{h',z,jz} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z), \\ A_{n2}(z) &= (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} \left(x_{jt}(y_{jt} - \mathbf{1}(v_{jt} > 0)) K_{h',z,jz} / \hat{f}_{t,v|xz,j} \right) \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z). \end{aligned}$$

We have that

$$\begin{aligned} \hat{\theta}_{LC}(z) &= A_{n1}(z)^{-1} A_{n2}(z) \\ &= A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} E(y_{jt}^* | x_{jt}, z_j) K_{h',z,jz} \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &\quad + A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} (y_{jt}^* - E(y_{jt}^* | x_{jt}, z_j)) K_{h',z,jz} \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &= \theta(z) + A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} x_{jt}^\top \left(\theta_j \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} - \theta(z) \right) K_{h',z,jz} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &\quad + A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} (y_{jt}^* - E(y_{jt}^* | x_{jt}, z_j)) K_{h',z,jz} \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &= \theta(z) + A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} x_{jt}^\top (\theta_j - \theta(z)) \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} K_{h',z,jz} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &\quad - A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} x_{jt}^\top \theta(z) K_{h',z,jz} \left(1 - \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \right) \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &\quad + A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} (y_{jt}^* - E(y_{jt}^* | x_{jt}, z_j)) K_{h',z,jz} \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z) \\ &\equiv \theta(z) + A_{n1}(z)^{-1} A_{n3}(z) + A_{n1}(z)^{-1} A_{n4}(z) + A_{n1}(z)^{-1} A_{n5}(z), \end{aligned}$$

where

$$\begin{aligned} A_{n3}(z) &= (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} x_{jt}^\top (\theta_j - \theta(z)) \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} K_{h',z,jz} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z), \\ A_{n4}(z) &= -(nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} x_{jt}^\top \theta(z) K_{h',z,jz} \left(1 - \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \right) \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z), \\ A_{n5}(z) &= A_{n1}(z)^{-1} (nTH')^{-1} \sum_{t=1}^{T} \sum_{j=1}^{n} x_{jt} (y_{jt}^* - E(y_{jt}^* | x_{jt}, z_j)) K_{h',z,jz} \frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} \mathbf{1}_{\tau_n,j} \mathbf{1}_{\varepsilon_n}(z). \end{aligned}$$

By Lemma B.1 we have uniformly in $z \in \Omega_z$,

$$A_{n1}(z)^{-1} = m(z)^{-1} + O_p\left(||h'||^\nu + (\ln n)^{1/2} (nH')^{-1/2} \right),$$

where $m(z) = T^{-1}\sum_{s=1}^{T} E[x_{js}x_{js}^{\top}|z_j = z]f_z(z)$.

Then, we have that

$$\begin{aligned}
\hat{\beta}_{LC} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{LC}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\theta_i + \frac{1}{n}\sum_{i=1}^{n}A_{n1}(z_i)^{-1}\left[A_{n3}(z_i) + A_{n4}(z_i) + A_{n5}(z_i)\right] \\
&= \beta + \frac{1}{n}\sum_{i=1}^{n}g(z_i) + \frac{1}{n}\sum_{i=1}^{n}m_i^{-1}\left[A_{n3}(z_i) + A_{n4}(z_i) + A_{n5}(z_i)\right] + \eta_n,
\end{aligned}$$

where $\eta_n = O_p\left(||h'||^{\nu} + (\ln n)^{1/2}(nH')^{-1/2}\right) O_p(\|A_{n3}(z_i)\| + \|A_{n4}(z_i)\| + \|A_{n5}(z_i)\|)$.

Since $\hat{f}_{t,v|xz,j} = \frac{\hat{f}_{t,vxz,j}}{\hat{f}_{t,xz,j}}$, where $\hat{f}_{t,vxz,j} = (nH)^{-1}\sum_{m=1}^{n} K_{h,vxz,mj}$, $\hat{f}_{t,xz,j} = (n\tilde{H})^{-1}\sum_{m=1}^{n} K_{\tilde{h},xz,mj}$, we have

$$\begin{aligned}
\frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} &= 1 + f_{t,v|xz,j}\left(\frac{1}{\hat{f}_{t,v|xz,j}} - \frac{1}{f_{t,v|xz,j}}\right) \\
&= 1 + \frac{\hat{f}_{t,xz,j} - f_{t,xz,j}}{f_{t,xz,j}} + \frac{f_{t,vxz,j} - \hat{f}_{t,vxz,j}}{f_{t,vxz,j}} + \frac{(f_{t,vxz,j}\hat{f}_{t,xz,j} - \hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j} - \hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}.
\end{aligned} \tag{B.1}$$

Then, we have that

$$\begin{aligned}
& B_{n1} \\
= & \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}A_{n3}(z_i) \\
= & \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}(\theta_j - \theta_i)K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
& + \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}(\theta_j - \theta_i)K_{h',z,ji}\frac{\hat{f}_{t,xz,j} - f_{t,xz,j}}{f_{t,xz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
& + \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}(\theta_j - \theta_i)K_{h',z,ji}\frac{f_{t,vxz,j} - \hat{f}_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
& + \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}(\theta_j - \theta_i)K_{h',z,ji}\frac{(f_{t,vxz,j}\hat{f}_{t,xz,j} - \hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j} - \hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}} \\
& \times\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
\equiv & B_{n1,1} + B_{n1,2} + B_{n1,3} + B_{n1,4}.
\end{aligned}$$

First we consider $B_{n1,1}$. We have

$$B_{n1,1} = \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}(\theta_j - \theta_i)K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i).$$

Further, $B_{n1,1}$ can be written as a second order U-statistic.

$$B_{n1,1} = n^{-2}\frac{n(n-1)}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} H_{n1,ij} \equiv n^{-2}\frac{n(n-1)}{2}U_{n1},$$

where $H_{n1,ij} = (TH')^{-1}\sum_{t=1}^{T}[m_i^{-1}x_{jt}x_{jt}^{\top}(\theta_j-\theta_i)\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)+m_j^{-1}x_{it}x_{it}^{\top}(\theta_i-\theta_j)\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j)]K_{h',z,ji}$. Using the U-statistic H-decomposition we have

$$U_{n1} = E[H_{n1,ij}]+\frac{2}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]+\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left[H_{n1,ij}-H_{n1,i}-H_{n1,j}+E(H_{n1,ij})\right],$$

where $H_{n1,i} = E[H_{n1,ij}|w_i]$, $w_i = (v_i, x_i, z_i) = (v_{i1},\ldots,v_{iT},x_{i1},\ldots,x_{iT},z_i)$.

Since $\varepsilon_n > \tau_n$ and $\|h'\|/(\varepsilon_n-\tau_n)\to 0$ and the kernel function $K(\cdot)$ has a compact support, the trimming functions $\mathbf{1}_{\tau_n,j}$ and $\mathbf{1}_{\varepsilon_n}(z_i)$ will ensure that all of the points which have boundary effects are excluded from our estimated locations. We have

$$\begin{aligned} E[H_{n1,ij}] &= (TH')^{-1}\sum_{t=1}^{T}E[m_i^{-1}x_{jt}x_{jt}^{\top}(\theta_j-\theta_i)K_{h',ij}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)] \\ &= \sum_{l=1}^{q}h_l'^{\nu}B_{l,LC}+O_p(\|h'\|^{\nu+1}), \end{aligned}$$

where $B_{l,LC} = \mu_\nu\sum_{k_1+k_2=\nu,k_2\neq 0}\frac{1}{k_1!k_2!}E\left[m_i^{-1}(\frac{\partial^{k_1}m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})\right]$. Also, we have

$$\begin{aligned} & E\left[\left(\frac{2}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right)\left(\frac{2}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right)^{\top}\right] \\ &= Var\left[\frac{2}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right] \\ &= O(n^{-1}\|h'\|^{2\nu}), \end{aligned} \tag{B.2}$$

and

$$\begin{aligned} & Var\left[\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left[H_{n1,ij}-H_{n1,i}-H_{n1,j}+E(H_{n1,ij})\right]\right] \\ &= \frac{4}{n^2(n-1)^2}\sum_{i=1}^{n}\sum_{j>i}^{n}Var\left[H_{n1,ij}-H_{n1,i}-H_{n1,j}+E(H_{n1,ij})\right] \\ &= O(n^{-2}H'^{-1}\|h'\|^2). \end{aligned} \tag{B.3}$$

Thus, $B_{n1,1} = O_p(\|h'\|^{\nu}+(nH'^{1/2})^{-1}\|h'\|)$.

Then, we evaluate $B_{n1,2}$ and $B_{n1,3}$, and by U-statistics Hoeffding decomposition, we have that $B_{n1,2}+B_{n1,3} = O_p(\|h'\|^{\nu}\|h\|^{\nu}+\|h'\|^{\nu}\|\tilde{h}\|^{\nu}+n^{-1/2}\|h'\|^{\nu}+(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|h\|+$

$(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|\tilde{h}\|)$. We omit the detailed derivation here to save the space. However, the procedure is similar as the derivation of the order of $B_{n2,1,5}$ where the details are provided.

For $B_{n1,4}$, we have

$$\begin{aligned}
& E(\|B_{n1,4}\|) \\
\leq\ & (TH')^{-1}\sum_{t=1}^{T} E\left(\|m_i^{-1}x_{jt}x_{jt}^{\top}(\theta_j-\theta_i)K_{h',z,ji}\frac{(f_{t,vxz,j}\hat{f}_{t,xz,j}-\hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j}-\hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\|\right) \\
\leq\ & (TH')^{-1}\sum_{t=1}^{T} E\left(\|m_i^{-1}x_{jt}x_{jt}^{\top}(\theta_j-\theta_i)K_{h',z,ji}\mathbf{1}_{\varepsilon_n}(z_i)\|\left|\frac{(f_{t,vxz,j}\hat{f}_{t,xz,j}-\hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j}-\hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\right|\right).
\end{aligned}$$

From Hansen (2008), we have $\sup_{(v,x,z)\in\Omega_{vxz}}|\hat{f}_t(v,x,z)-f_t(v,x,z)| = O_p(\|h\|^{\nu}+(\ln n)^{1/2}(nH)^{-1/2})$ and $\sup_{(x,z)\in P_{xz}(\Omega_{vxz})}|\hat{f}_t(x,z)-f_t(x,z)| = O_p(\|\tilde{h}\|^{\nu}+(\ln n)^{1/2}(n\tilde{H})^{-1/2})$, where $P_{xz}(\cdot)$ is the projection of Cartesian product. Hence, we have that $B_{n1,4} = O_p(\|h'\|\|h\|^{2\nu}+\|h'\|(\ln n)(nH)^{-1}+\|h'\|\|h\|^{\nu}\|\tilde{h}\|^{\nu}+\|h'\|(\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Let

$$\begin{aligned}
B_{n2} &= -\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}A_{n4}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\left(1-\frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}}\right)\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i).
\end{aligned}$$

From the equation (B.1), we have

$$1-\frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}} = -\frac{\hat{f}_{t,xz,j}-f_{t,xz,j}}{f_{t,xz,j}}-\frac{f_{t,vxz,j}-\hat{f}_{t,vxz,j}}{f_{t,vxz,j}}-\frac{(f_{t,vxz,j}\hat{f}_{t,xz,j}-\hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j}-\hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}.$$

Hence,

$$\begin{aligned}
& B_{n2} \\
=\ & -\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{\hat{f}_{t,vxz,j}-f_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
& -\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{f_{t,xz,j}-\hat{f}_{t,xz,j}}{f_{t,xz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
& -\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{(\hat{f}_{t,vxz,j}f_{t,xz,j}-f_{t,vxz,j}\hat{f}_{t,xz,j})(f_{t,vxz,j}-\hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
\equiv\ & -B_{n2,1}-B_{n2,2}-B_{n2,3}.
\end{aligned}$$

First, we consider $B_{n2,1}$. We have

$$\begin{aligned}
B_{n2,1} &= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{\hat{f}_{t,vxz,j}-f_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{(nH)^{-1}\sum_{k=1}^{n}K_{h,vxz,kj}-f_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)
\end{aligned}$$

$$
\begin{aligned}
&= (n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^\top\theta_i K_{h',z,ji}\left(K_{h,vxz,kj}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&= (n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}x_{it}x_{it}^\top\theta_i K_{h'}(0)\left(K_{h}(0)-Hf_{t,vxz,i}\right)f_{t,vzx,i}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{k\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{it}x_{it}^\top\theta_i K_{h'}(0)\left(K_{h,vxz,ki}-Hf_{t,vxz,i}\right)f_{t,vzx,i}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^\top\theta_i K_{h',z,ji}\left(K_{h}(0)-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^\top\theta_i K_{h',z,ji}\left(K_{h,vxz,ij}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(n^3TH'H)^{-1}\sum\sum_{i\neq j\neq k}\sum\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^\top\theta_i K_{h',z,ji}\left(K_{h,vxz,kj}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\equiv B_{n2,1,1}+B_{n2,1,2}+B_{n2,1,3}+B_{n2,1,4}+B_{n2,1,5}.
\end{aligned}
$$

It is easy to see that $B_{n2,1,1} = O_p((n^2H'H)^{-1})$, $B_{n2,1,2}$, $B_{n2,1,3}$ and $B_{n2,1,4}$ can be written as second order U-statistics, and $B_{n2,1,5}$ can be written a third order U-statistic. Also, by the Hoeffding decomposition, we have that $B_{n2,1,2} = O_p(\|h\|^\nu(nH')^{-1})$, $B_{n2,1,3} = O_p((nH)^{-1})$, and $B_{n2,1,4} = O_p(\|h\|^\nu n^{-1})$.

We can write $B_{n2,1,5}$ as $B_{n2,1,5} = n^{-3}\sum\sum\sum_{1\le i<j<k\le n}\psi_n(v_i,x_i,z_i,v_j,x_j,z_j,v_k,x_k,z_k)$, where

$$
\begin{aligned}
&\psi_n(v_i,x_i,z_i,v_j,x_j,z_j,v_k,x_k,z_k)\\
&= (TH'H)^{-1}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^\top\theta_i K_{h',z,ji}\left(K_{h,vxz,kj}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(TH'H)^{-1}\sum_{t=1}^{T} m_j^{-1}x_{it}x_{it}^\top\theta_j K_{h',z,ij}\left(K_{h,vxz,ki}-Hf_{t,vxz,i}\right)f_{t,vzx,i}^{-1}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j)\\
&\quad +(TH'H)^{-1}\sum_{t=1}^{T} m_k^{-1}x_{jt}x_{jt}^\top\theta_k K_{h',z,jk}\left(K_{h,vxz,ij}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_k)\\
&\quad +(TH'H)^{-1}\sum_{t=1}^{T} m_i^{-1}x_{kt}x_{kt}^\top\theta_i K_{h',z,ki}\left(K_{h,vxz,jk}-Hf_{t,vzx,k}\right)f_{t,vzx,k}^{-1}\mathbf{1}_{\tau_n,k}\mathbf{1}_{\varepsilon_n}(z_i)\\
&\quad +(TH'H)^{-1}\sum_{t=1}^{T} m_k^{-1}x_{it}x_{it}^\top\theta_k K_{h',z,ik}\left(K_{h,vxz,ji}-Hf_{t,vxz,i}\right)f_{t,vzx,i}^{-1}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_k)\\
&\quad +(TH'H)^{-1}\sum_{t=1}^{T} m_j^{-1}x_{kt}x_{kt}^\top\theta_j K_{h',z,kj}\left(K_{h,vxz,ik}-Hf_{t,vzx,k}\right)f_{t,vzx,k}^{-1}\mathbf{1}_{\tau_n,k}\mathbf{1}_{\varepsilon_n}(z_j).
\end{aligned}
$$

Let $w_i = (v_{i1}, \ldots, v_{iT}, x_{i1}, \ldots, x_{iT}, z_i)$, by the Hoeffding decomposition, we have

$$\begin{aligned} B_{n2,1,5} &= n^{-3}(n(n-1)(n-2)/6)\Big[E(\psi_n) + \frac{3}{n}\sum_{i=1}^{n}\Big(E[\psi_n|w_i] - E(\psi_n)\Big) \\ &\quad + \frac{6}{n(n-1)}\sum_{1\le i<j\le n}\Big(E[\psi_n|w_i,w_j] - E[\psi_n|w_i] - E[\psi_n|w_j] + E[\psi_n]\Big) \\ &\quad + \frac{6}{n(n-1)(n-2)}\sum_{1\le i<j<k\le n}\Big(\psi_n - E[\psi_n|w_i,w_j] - E[\psi_n|w_i,w_k] - E[\psi_n|w_j,w_k] \\ &\quad + E[\psi_n|w_i] + E[\psi_n|w_j] + E[\psi_n|w_k] - E[\psi_n]\Big)\Big] \\ &\equiv B_{n2,1,5,1} + B_{n2,1,5,2} + B_{n2,1,5,3} + B_{n2,1,5,4}. \end{aligned}$$

By standard calculations, we have

$$\begin{aligned} B_{n2,1,5,1} &= (n^{-3}n(n-1)(n-2)/6)E[\psi_n] = O_p(\|h\|^{\nu}), \\ B_{n2,1,5,2} &= (n^{-3}n(n-1)(n-2)/6)\frac{3}{n}\sum_{i=1}^{n}\Big(E[\psi_n|w_i] - E(\psi_n)\Big) \\ &= \frac{1}{n}\sum_{i=1}^{n}T^{-1}\sum_{t=1}^{T}\big(m_i^{-1}x_{it}x_{it}^{\top}\theta_i f_z(z_i) - E[\theta_i]\big) + O_p(\|h\|^{\nu} + n^{-1}). \end{aligned}$$

Also, it is easy to see that $B_{n2,1,5,3} = O_p(n^{-1})$, and $B_{n2,1,5,4} = O_p((n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|)$.

Hence, we have that

$$\begin{aligned} B_{n2,2} &= \frac{1}{n}\sum_{i=1}^{n}T^{-1}\sum_{t=1}^{T}\big(m_i^{-1}x_{it}x_{it}^{\top}\theta_i f_z(z_i) - E[\theta_i]\big) \\ &\quad + O_p((n^2H'H)^{-1} + \|h\|^{\nu}(nH')^{-1} + (nH)^{-1} + \|h\|^{\nu} + n^{-1} + (n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|). \end{aligned} \tag{B.4}$$

Similarly, we can show that

$$\begin{aligned} B_{n2,3} &= -\frac{1}{n}\sum_{i=1}^{n}T^{-1}\sum_{t=1}^{T}\big(m_i^{-1}x_{it}x_{it}^{\top}\theta_i f_z(z_i) - E[\theta_i]\big) \\ &\quad + O_p((n^2H'\tilde{H})^{-1} + \|\tilde{h}\|^{\nu}(nH')^{-1} + (n\tilde{H})^{-1} + \|\tilde{h}\|^{\nu} + n^{-1} + (n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|). \end{aligned} \tag{B.5}$$

Similar as the derivation of $B_{n1,4}$, we have $B_{n2,4} = O_p(\|h\|^{2\nu} + (\ln n)(nH)^{-1} + \|h\|^{\nu}\|\tilde{h}\|^{\nu} + (\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Denote

$$\xi_{jt} = y_{jt}^{*} - E(y_{jt}^{*}|x_{jt}, z_j).$$

By (B.1), we have

$$
\begin{aligned}
B_{n3} &= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1} A_{n5}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}(y_{jt}^* - E(y_{jt}^*|x_{jt}, z_j))K_{h',z,ji}\frac{f_{t,v|xz,j}}{\hat{f}_{t,v|xz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\xi_{jt}K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\xi_{jt}K_{h',z,ji}\frac{\hat{f}_{t,xz,j} - f_{t,xz,j}}{f_{t,xz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\xi_{jt}K_{h',z,ji}\frac{f_{t,vxz,j} - \hat{f}_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\xi_{jt}K_{h',z,ji}\frac{(f_{t,vxz,j}\hat{f}_{t,xz,j} - \hat{f}_{t,vxz,j}f_{t,xz,j})(f_{t,vxz,j} - \hat{f}_{t,vxz,j})}{f_{t,xz,j}f_{t,vxz,j}\hat{f}_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i) \\
&\equiv B_{n3,1} + B_{n3,2} + B_{n3,3} + B_{n3,4}.
\end{aligned}
$$

Then $E[B_{n3,1}] = 0$. We have

$$
B_{n3,1} = \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\xi_{jt}K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i).
$$

Moreover, we can decompose $B_{n3,1}$ into two terms

$$
B_{n3,1} = B_{n3,1,1} + B_{n3,1,2},
$$

where

$$
\begin{aligned}
B_{n3,1,1} &= (n^2TH')^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}x_{it}\xi_{it}K_{h'}(0)\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_i) \\
B_{n3,1,2} &= (n^2TH')^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}\xi_{jt}K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i).
\end{aligned}
$$

It is easy to see that $E[B_{n3,1,1}] = 0$ and $E[||B_{n3,1,1}||^2] = (n^4H'^2)^{-1}O(n) = O((n^3H'^2)^{-1})$. Hence, $B_{n3,1,1} = O_p((n^{3/2}H')^{-1})$.

Also, $B_{n3,1,2}$ can be written as a second order U-statistic.

$$
B_{n3,1,2} = n^{-2}\frac{n(n-1)}{2}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} H_{n3,ij} \equiv n^{-2}\frac{n(n-1)}{2}U_{n3},
$$

where $H_{n3,ij} = (TH')^{-1}\sum_{t=1}^{T}(m_i^{-1}x_{jt}\xi_{jt}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)+m_j^{-1}x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j))K_{h',ij}$. Since U_{n3} has zero mean, its H-decomposition is given by

$$U_{n3} = U_{n3,1} + U_{n3,2},$$

where $U_{n3,1} = \frac{2}{n}\sum_{i=1}^{n} H_{n3,i}$, $U_{n3,2} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}[H_{n3,ij} - H_{n3,i} - H_{n3,j}]$, $H_{n3,i} = E[H_{n3,ij}|w_i]$, and $w_i = (v_i, x_i, \alpha_i, z_i, u_i) = (v_{i1}, \ldots, v_{iT}, x_{i1}, \ldots, x_{iT}, \alpha_i, z_i, u_{i1}, \ldots, u_{iT})$. Then, we have

$$\begin{aligned} U_{n3,1} &= \frac{1}{nTH'}\sum_{i=1}^{n}\sum_{t=1}^{T} E\left[(m_i^{-1}x_{jt}\xi_{jt}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i)+m_j^{-1}x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j))K_{h',ij}|w_i\right] \\ &= \frac{1}{nTH'}\sum_{i=1}^{n}\sum_{t=1}^{T} E[m_j^{-1}K_{h,ij}\mathbf{1}_{\varepsilon_n}(z_j)|w_i]x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}, \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}f_z(z_i)x_{it}\xi_{it}\mathbf{1}_{\tau_n,i} + O_p(\|h'\|^{\nu+1}/\sqrt{n}). \qquad \text{(B.6)} \end{aligned}$$

Also, we have $E[||U_{n2,2}||^2] = (n^4H'^2)^{-1}n^2O(H') = O((n^2H')^{-1})$. Hence, $U_{n2,2} = O_p((nH'^{1/2})^{-1})$.

Then we consider $B_{n3,2}$, $B_{n3,3}$, and $B_{n3,4}$. Similar as (B.4) and (B.5), we have that $B_{n3,2} + B_{n3,3} = O_p((n^2H'H)^{-1}+\|h\|^{\nu}(nH')^{-1}+(nH)^{-1}+\|h\|^{\nu}+n^{-1}+(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|(n^2H'\tilde{H})^{-1}+\|\tilde{h}\|^{\nu}(nH')^{-1}+(n\tilde{H})^{-1}+\|\tilde{h}\|^{\nu}+(n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|)$. Similar as the derivation of $B_{n1,4}$, we have $B_{n3,4} = O_p(\|h\|^{2\nu}+(\ln n)(nH)^{-1}+\|h\|^{\nu}\|\tilde{h}\|^{\nu}+(\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Moreover, by Cauchy-Schwarz inequality, we have that

$$E\|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(m_i^{-1}f_z(z_i)x_{it}\xi_{it})^{\otimes 2}(1-\mathbf{1}_{\tau_n,i})\| \leq \{E(\|m_i^{-1}f_z(z_i)x_{it}\xi_{it}\|^2)P((v_i,x_i,z_i)\in\Omega_{vxz})\}^{1/2}.$$

$P((v_{it}, x_{it}, z_i) \in \Omega_{vxz})$ is the probability that (v_{it}, x_{it}, z_i) is within a distance τ_n of the boundary $\partial\mathcal{S}_{vxz}$ of $\mathcal{S}_{vxz}$. Since the joint density function $f_{vxz}(v_{it}, x_{it}, z_i)$ of (v_{it}, x_{it}, z_i) is bounded and the volume of the set that is within a distance τ_n of $\partial\mathcal{S}_{vxz}$ is proportional to τ_n, we have that $P((v_{it}, x_{it}, z_i) \in \Omega_{vxz}) = O(\tau_n)$. Hence, we have $Var(\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}f_z(z_i)x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}) = Var(\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}f_z(z_i)x_{it}\xi_{it}) + o(1)$.

Therefore, we have that

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{LC} - \beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(z_i) - \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}f_z(z_i)x_{it}\xi_{it}\mathbf{1}_{\tau_n,i} + O_p(\delta_n) \\ &\xrightarrow{d} N(0, V_{LC}) \end{aligned}$$

by the Lindeberg central limit theorem, where $V_{LC} = Var(g(z_i)) + T^{-1}Var\Big(m_i^{-1}f_z(z_i)x_{it}\xi_{it}\Big)$, $\delta_n = \sqrt{n}\|h'\|^{\nu}+\sqrt{n}(nH'^{1/2})^{-1}\|h'\|+\sqrt{n}n^{-1/2}\|h'\|^{\nu}+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|h\|+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|\tilde{h}\|+\sqrt{n}(\ln n)(nH)^{-1}+\sqrt{n}\|h\|^{\nu}\|\tilde{h}\|^{\nu}+\sqrt{n}(\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2}+\sqrt{n}(n^2H'H)^{-1}+\sqrt{n}\|h\|^{\nu}(nH')^{-1}+\sqrt{n}(nH)^{-1}+\sqrt{n}\|h\|^{\nu}+\sqrt{n}n^{-1}+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|+\sqrt{n}(n^2H'\tilde{H})^{-1}+\sqrt{n}\|\tilde{h}\|^{\nu}(nH')^{-1}+\sqrt{n}(n\tilde{H})^{-1}+\sqrt{n}\|\tilde{h}\|^{\nu}+\sqrt{n}(n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|+\sqrt{n}\|h'\|^{\nu+1}/\sqrt{n}+\sqrt{n}(nH'^{1/2})^{-1}+\sqrt{n}\eta_n = o_p(1)$, and $\sqrt{n}\eta_n = \sqrt{n}O_p\Big(||h'||^{\nu}+(\ln n)^{1/2}(nH')^{-1/2}\Big)O_p(\|h'\|^{\nu}+(nH')^{-1/2}) = o_p(1)$.

Lemma B.1 *Define* $A_{n1}(z) = \frac{1}{nTH'} \sum_{j=1}^{n} \sum_{s=1}^{T} x_{js} x_{js}^{\top} K_{h',z_j z}$, *and* $m(z) = T^{-1} \sum_{s=1}^{T} E[x_{js} x_{js}^{\top} | z_j = z] f_z(z)$, *where* $K_{h',z_j z} = \prod_{l=1}^{q} k\left(\frac{z_{jl}-z_l}{h_l}\right)$, *then under Assumptions B4-B7,*

$$A_{n1}(z)^{-1} \;=\; m(z)^{-1} + O_p\left(||h'||^{\nu} + (\ln n)^{1/2}(nH')^{-1/2}\right),$$

uniformly in $z \in \Omega_z$, *where* $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,\ldots,q\}} |z_l - z_{0,l}| \geq \varepsilon_n$ *for some* $z_0 \in \partial\mathcal{S}_z\}$, $\partial\mathcal{S}_z$ *is the boundary of the compact set* $\mathcal{S}_z$, $\varepsilon_n \to 0$ *and* $\|h'\|/\varepsilon_n \to 0$.

Proof: First, we have

$$E[A_{n1}(z)] = m(z) + O\left(\|h'\|^{\nu}\right), \tag{B.7}$$

uniformly in $z \in \Omega_z$. Following similar arguments used in Masry (1996) when deriving uniform convergence rates for nonparametric kernel estimators, we know that

$$A_{n1}(z) - E[A_{n1}(z)] = O_p\left(\frac{(\ln n)^{1/2}}{(nH')^{1/2}}\right), \tag{B.8}$$

uniformly in $z \in \Omega_z$.

Combining (B.7) and (B.8) we obtain

$$A_{n1}(z) - m(z) = O_p\left(||h'||^{\nu} + (\ln n)^{1/2}(nH')^{-1/2}\right), \tag{B.9}$$

uniformly in $z \in \Omega_z$.

Using (B.9) we obtain

$$\begin{aligned} A_{n1}(z)^{-1} &= [m(z) + A_{n1}(z) - m(z)]^{-1} \\ &= m(z)^{-1} - m(z)^{-1}[A_{n1}(z) - m(z)]\, m(z)^{-1} + O_p\left(\|A_{n1}(z) - m(z))\|^2\right) \\ &= m(z)^{-1} + O_p\left(||h'||^{\nu} + (\ln n)^{1/2}(nH')^{-1/2}\right), \end{aligned}$$

which completes the proof of Lemma B.1.

References

Angrist, J.D., 2001. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice (with discussion). Journal of Business and Economics Statistics 19, 2-28.

Arellano, M., 2003. Panel Data Econometrics. New York: Oxford University Press.

Arellano, M., Bonhomme, S., 2010. Identifying Distributional Characteristics in Random Coefficient Panel Data Models. Working paper, CEMFI.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841-890.

Bonhomme, S., 2011. Functional differencing. Econometrica, *forthcoming.*

Browning, M., Carro, J., 2007. Heterogeneity and microeconometrics modeling. In: R. Blundell, W. Newey and T. Persson, Eds., Advances in Economics and Econometrics: Theory and Applications III, 47-74. Cambridge: Cambridge University Press.

Chamberlain, G., 2010. Binary response models for panel data: Identification and information. Econometrica 78, 159-168.

Chernozhukov, V., Fernandez-Val, I., Newey, W.K., 2009. Quantile and average effects in nonseparable panel models. Working paper, MIT.

Dong, Y., Lewbel, A., 2011. Simple estimators for binary choice models with endogenous regressors. Working paper, Boston College.

Evdokimov, K., 2010. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working paper, Princeton University.

Fox, J.T., Gandhi, A., 2010. Nonparametric identification and estimation of random coefficients in nonlinear economic models. Working paper, University of Chicago.

Frazis, H., Loewenstein, M.A., 2005. Reexamining the returns to training: functional form, magnitude, and interpretation. Journal of Human Resources 40, 453-476.

Graham, B.S., Powell, J.L., 2010. Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models. Working paper, New York University.

Hahn, Jinyong, 2001. Comment: Binary regressors in nonlinear panel-data models with fixed effects. Journal of Business and Economic Statistics 19, 16-17.

Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Härdle, W., Horowitz, J.L., 1996. Direct semiparametric estimation of single-index models with discrete covariates. Journal of the American Statistical Association 91, 1632-1640.

Heckman, J.J., Schmierer, D.A., 2010. Tests of hypotheses arising in the correlated random coefficient model. Economic Modelling 27, 1355-1367.

Heckman, J.J., Schmierer, D.A., Urzua, S.S., 2010. Testing the correlated random coefficient model. Journal of Econometrics 158, 177-203.

Hoderlein, S., 2009. Endogeneity in Semiparametric Binary Random Coefficient Models. Working paper, Boston College.

Hoderlein, S., Klemelä, J., Mammen, E., 2010. Analyzing the random coefficient model nonparametrically. Econometric Theory 26, 804-837.

Honoré, B., Lewbel, A., 2002. Semiparametric binary choice panel data models without strict exogeneity. Econometrica 70, 2053-2063.

Horowitz, J.L., 1992. A smoothed maximum score estimator for the binary response model. Econometrica 60, 505-532.

Hsiao, C., 2003. Analysis of Panel Data. Cambridge: Cambridge University Press.

Hsiao, C., Pesaran, M.H., 2008. Random Coefficient Models. In L. Mátyás and P. Sevestre eds., Advanced Studies in Theoretical and Applied Econometrics, Vol. 46, The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, Springer-Verlag, Berlin.

Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58, 71-120.

Imbens, G.W., 2007. Nonadditive models with endogenous regressors. In: R. Blundell, W. Newey and T. Persson, Eds., Advances in Economics and Econometrics: Theory and Applications III, 17-46. Cambridge: Cambridge University Press.

Khan, S., Lewbel, A., 2007. Weighted and two-stage least squares estimation of semiparametric truncated regression models. Econometric Theory 23, 309-347.

Kim, J., Pollard, D., 1990. Cube root asymptotics. Annals of Statistics 18, 191-219.

Klein, R., Spady, R.H., 1993. An efficient semiparametric estimator for binary response models. Econometrica 61, 387-421.

Lewbel, A., 1998. Semiparametric latent variable model estimation with endogenous or mismeasured regressors. Econometrica 66, 105-121.

Lewbel, A., 2000. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. Journal of Econometrics 97, 145-177.

Li, Q., Racine, J.S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton: Princeton University Press.

Manski, C.F., 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. Journal of Econometrics 27, 313-334.

Manski, C.F., 1988. Identification of binary response models. Journal of the American Statistical Association 83, 729-738.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis 17, 571-599.

Newey, W.K., Ruud, P.A., 2005. Density Weighted Linear Least Squares. In D.W.K Andrews and J.H. Stock eds., Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg, 554-573, Cambridge: Cambridge University Press.

Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403-1430.

Swamy, P., Tavlas, G.S., 2007. Random Coefficient Models. In B.H. Baltagi ed., A Companion to Theoretical Econometrics, 410-428, Malden: Blackwell Publishing Ltd.

Wooldridge, J.M., 2005. Fixed-effects related estimators for correlated random-coefficient and treatment-effect panel data models. The Review of Economics and Statistics 87, 385-390.

Table 10: Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
wage88	833.3858	476.068	1	4333
wage89	905.739	503.6028	1	4807
wage90	995.7764	523.8137	1	4615
wage91	1053.024	549.6688	1	4700
wage92	1108.457	581.8279	1	5500
wage93	1180.624	642.1308	1	7500
wage94	1263.622	722.7841	1	10000
wage96	1445.186	874.871	1	10852
wage98	1606.613	1009.838	1	13898
wage00	1823.811	1149.412	1	11627
wage02	1963.478	1270.139	1	19230
wage04	2110.454	1440.742	64	24038
wage06	2300.247	1615.956	91	19230
wage08	2453.701	1790.974	109	37652
tenure88	140.1427	145.2536	0	781
tenure89	157.0955	157.2331	0	819
tenure90	177.3152	171.4312	0	873
tenure91	195.9717	182.7982	0	922
tenure92	219.6279	196.6144	0	976
tenure93	239.1874	211.0686	0	1023
tenure94	259.6474	226.6629	0	987
tenure96	291.85	253.4582	0	1077
tenure98	321.2476	282.6207	0	1477
tenure00	362.4685	313.0751	0	1588
tenure02	395.1795	336.1148	0	1693
tenure04	435.4588	366.2723	0	1797
tenure06	467.4472	397.1599	0	1887
tenure08	506.1135	426.1333	1	1969
educ88	13.37664	2.11318	3	20
educ89	13.44478	2.15468	3	20
educ90	13.49224	2.186354	3	20
educ91	13.53301	2.212822	3	20
educ92	13.58047	2.221079	3	20
educ93	13.61698	2.238841	3	20
educ94	13.64649	2.249989	3	20
educ96	13.70611	2.284811	3	20
educ98	13.74232	2.284587	3	20
educ00	13.80925	2.320448	3	20
educ02	13.84484	2.334274	3	20
educ04	13.89808	2.362523	3	20
educ06	13.93824	2.377009	3	20
educ08	13.99361	2.364841	3	20

Table 11: Descriptive Statistics Continued

Variable	Mean	Standard Deviation	Minimum	Maximum
union88	.11439	.3183328	0	1
union89	.1432918	.3504234	0	1
union90	.1557651	.3626877	0	1
union91	.1618497	.3683689	0	1
union92	.1664131	.3725078	0	1
union93	.1621539	.368648	0	1
union94	.1545482	.3615285	0	1
union96	.1785823	.3830604	0	1
union98	.1776696	.3822925	0	1
union00	.1953149	.3965032	0	1
union02	.1877092	.3905392	0	1
union04	.1846669	.3880861	0	1
union06	.1895345	.3919923	0	1
union08	.1895345	.3919923	0	1
training88	67.50563	291.9215	0	4160
training89	36.02221	276.4029	0	9248
training90	35.15972	206.4548	0	5760
training91	24.14603	141.7969	0	3456
training92	28.16428	190.6329	0	4608
training93	23.47946	159.2761	0	6240
training94	14.16002	83.9495	0	2304
training96	31.96714	158.5174	0	3044
training98	28.36294	174.2248	0	5824
training00	25.21022	134.9312	0	4320
training02	19.87344	131.2677	0	5280
training04	21.02282	146.5059	0	4704
training06	16.5467	146.6135	0	6120
training08	14.4028	106.4687	0	3840