LIMITED DEPENDENT VARIABLE CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS

A Dissertation

by

ZHONGWEN LIANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Economics

LIMITED DEPENDENT VARIABLE CORRELATED RANDOM COEFFICIENT PANEL DATA MODELS

A Dissertation

by

ZHONGWEN LIANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Qi Li
	Joel Zinn
Committee Members,	Dennis W. Jansen
	Ke-Li Xu
Head of Department,	Timothy Gronberg

August 2012

Major Subject: Economics

ABSTRACT

Limited Dependent Variable Correlated Random Coefficient Panel Data Models.
(August 2012)

Zhongwen Liang, B.S., Wuhan University; M.S., Wuhan University
Co-Chairs of Advisory Committee: Dr. Qi Li Dr. Joel Zinn

In this dissertation, I consider linear, binary response correlated random coefficient (CRC) panel data models and a truncated CRC panel data model which are frequently used in economic analysis. I 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 linear CRC models, I give the identification conditions for the average slopes of a linear CRC model with a general nonparametric correlation between regressors and random coefficients. I 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. I 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, I can transfer a binary response CRC model to a linear moment relation. I also construct a semiparametric estimator for the average slopes and derive the $\sqrt{n}$-normality result.

For the truncated CRC panel data model, I obtain the identification and estimation results based on the special regressor method which is used in Khan and Lewbel

(2007). I construct a $\sqrt{n}$ consistent estimator for the population mean of the random coefficient. I also derive the asymptotic distribution of my estimator.

Simulations are given to show the finite sample advantage of my estimators. Further, I use a linear CRC panel data model to reexamine the return from job training. The results show that my estimation method really makes a difference, and the estimated return of training by my method is 7 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 3.16% per 60 hours training.

DEDICATION

To my mother and father

ACKNOWLEDGMENTS

This dissertation was written under the supervision of my chief advisor, Professor Qi Li. In the past five years, I have learned a lot from Professor Li, especially the nonparametric econometric methods. I really admire his deep and broad knowledge, his diligence and excellence in research, and his fastness in thinking. Without his continuous guidance and encouragement and the extensive discussions with him, I could not achieve these research results. I would like to thank Professor Li for leading me to this fruitful field, and all of valuable advice he gives to me.

I am also very grateful to Professor Joel Zinn for serving as the co-chair of my dissertation committee, and to Professor Dennis W. Jansen and Professor Ke-Li Xu for serving as my dissertation committee members. Their knowledge and valuable suggestions broaden my understanding of different aspects of my research.

Thanks also go to all my friends, the department faculty and staff for their help along the way of the pursue of my Ph.D. degree.

Finally, I thank my mother and father for their persistent encouragement and their love.

TABLE OF CONTENTS

Page

LIST OF TABLES

1. INTRODUCTION

Recently, the correlated random coefficient model has drawn much attention. As stated in Heckman et al. (2010), "The correlated random coefficient model is the new centerpiece of a large literature in microeconometrics". In this dissertation, first I consider linear CRC panel data models in the form of

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

where x_{it} denotes regressors with random coefficient β_i, and u_{it} is the error term. Also, I consider binary response CRC panel data models in the form of

$$y_{it} = \mathbf{1}(v_{it}^{\top}\gamma + x_{it}^{\top}\beta_i + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{1.2}$$

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, and u_{it} is the error term. Finally, I consider a truncated CRC panel data model

$$\begin{aligned} y_{it}^{*} &= v_{it}\gamma + x_{it}^{\top}\beta_i + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T) \\ y_{it} &= y_{it}^{*} | y_{it}^{*} \geq 0, \end{aligned} \tag{1.3}$$

where v_{it} denotes regressors with constant coefficient γ, x_{it} denotes regressors with random coefficient β_i, and u_{it} is the error term. Here, x_{it} can include 1 as a component. Thus, the panel data models with fixed effects corresponding to each model are special cases of these models. I allow the general correlation between the random coefficient β_i and the regressor x_{it}. I focus on the nonparametric identification and estimation of the mean of random slope β_i in these models and related transformed models, which will be more specific in later sections.

This dissertation follows the style of *Journal of Econometrics*.

1.1 Linear Models

Linear models are among the mostly used models. The reason is its simplicity and direct economic interpretability. However, for most empirical applications the plain linear models suffer from the lack of flexibility, e.g., the traditional estimators will not be consistent under endogeneity and heterogeneity problem. Recently, correlated random coefficient models are proposed to deal with unobserved heterogeneity problem. Further, with panel data available, we can capture the endogeneity and heterogeneity more easily. In this dissertation, I consider the linear CRC panel data models first. This will also serve as the foundation for the methods I will use for the binary and truncated models.

We can motivate the usefulness of the linear CRC panel data models by an empirical application. In labor economics, we are interested in the return from the job training. We regress the logarithm of wage on a job training variable which is the accumulated hours spent on the job training. Then its coefficient is the rate of return from the training. We know that other things being the same, still different people will get different payoffs even they took same amount of training. This means that there exists unobserved heterogeneity. One way to capture it is to use a random coefficient model. So we will have the coefficient of the job training variable to be random. From the theory of human capital, we know that the marginal return from the job training is diminishing as the level of job training increases. So there is a negative correlation between the job training variable and its coefficient which is the rate of return from the job training. Moreover, there is a selection problem. Individuals with lower marginal return may receive less training, which means there is also a positive correlation. So there must exist correlation between the job training variable and its random coefficient. Also the panel data model gives us the advantage to capture the correlation between regressors and other unobserved heterogeneity by the fixed effects term. A linear CRC panel data model is a good candidate for this type of question.

There is a large literature about the CRC model. Heckman and Vytlacil (1998) is among the very first papers. Motivated by the diminishing return of schooling, they discussed the instrumental variable methods for the cross-sectional setting of CRC model. Wooldridge (2003) gave weaker conditions for the two-stage plug-in estimator proposed by Heckman and Vytlacil (1998). Wooldridge (2005) gave a sufficient condition for the fixed effects estimator to be consistent. Murtazashvili and Wooldridge (2008) investigates the fixed effects instrumental variables estimation for the linear CRC panel data model.

Recently, there is a growing literature on CRC models. Graham and Powell (2012) 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 et al. (2010) and Heckman and Schmierer (2010) investigate the tests of the CRC model.

I discuss the nonparametric identification and estimation of the population mean of the random coefficient β_i for the linear CRC panel data models in Chapter 2. I construct a $\sqrt{n}$ consistent estimator and derive its asymptotic normality.

1.2 Binary Response Models

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 (2012). 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 2.1 and 2.2, 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 et al. (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.

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 (2012), Fox and Gandhi (2010), Hoderlein et al. (2010).

Among the recent developments of panel data models, the nonseparable panel data models is an indispensable part. Chernozhukov et al. (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 (2012) 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 I 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. Dong and Lewbel (2011) give a good survey for this method.

In Chapter 3, I base the identification for binary CRC panel data models on the special regressor method. I construct a $\sqrt{n}$ consistent estimator for the population mean of the random coefficient based on my identification result. Also, the asymptotic normality result is derived.

1.3 Truncated Models

Censored and truncated models are commonly used in economics when we don't have complete observation of the population. Due to the heterogeneity of the population, it is desirable to have models that can take account of the unobserved heterogeneity. One way is to consider a censored or truncated panel data model with additive unobserved individual-specific random variable, i.e. fixed-effects. This was studied by Honoré (1992), who proposed a trimming strategy that can get rid of the unobserved variable via difference. However, the nonadditive heterogeneity arises naturally in economic analysis. In this dissertation, I consider a truncated panel data model which has multiplicative heterogeneity.

The model I consider is as in (1.3). The underlying model is a linear panel data model, and we can observe the dependent variables only when they are strictly positive. I allow the general correlation between the random coefficient β_i and the regressor x_{it}, and I do not assume the distribution function of u_{it} to be known. I focus on the nonparametric identification and estimation of the population mean β of random slope β_i in this model. I assume that $(y_{it}^*, v_{it}, x_{it}, \beta_i)$ are drawn from the underlying untruncated distribution. I use E^* to denote the expectation with respect to this distribution and assume $E^*(u_{it}|x_{i1}, \ldots, x_{iT}, \beta_i) = 0$.

I will use the special regressor method proposed by Lewbel (1998, 2000) for the identification and estimation of our model. Due to the nonadditivity of the unobserved heterogeneity, the idea from Honoré (1992) cannot be generalized to this case. I base the identification on similar idea from Khan and Lewbel (2007) which uses the special regressor method to study a cross-sectional truncated regression model. In Chapter 4, I extend their method to a truncated CRC panel data model. For simplicity, I assume that v_{it} is a scalar regressor and is the special regressor which satisfies three conditions. Further, although the observation of the dependent variable y_{it} can only be partially observed, in order to achieve the identification, I assume that we can estimate the untruncated population distribution of the regressors (v_{it}, x_{it}). Once I get the identification result, I construct a $\sqrt{n}$ consistent estimator from the identification.

2. LINEAR CRC PANEL DATA MODELS

2.1 Identification of Linear CRC Models

In this section I 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.

I consider the linear random coefficient models with general correlation between random coefficients and regressors in sections 2.1 and 2.2. For simplicity, I assume there is no regressor with constant coefficient in model (1.2) in sections 2.1 and 2.2. In section 2.1.1 I first consider a CRC model with cross sectional data. I 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. I 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 2.1.2.

2.1.1 The Cross Sectional Data Case

I 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., $E(u_i|x_i, \alpha_i) = 0$. I 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, I allow for $g(x_i)$ to have any other unknown functional form.

Replacing β_i by $\beta + \alpha_i$, I 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 I 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, I 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. I 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, I have $E(u_i|x_{1i}, x_{2i}) = 0$. Consequently, I have

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

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.

2.1.2 The Panel Data Case

Panel data will provide us more information and help us to identify the unknown functions. For heuristics let us 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

$$\begin{aligned}
\theta_1(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) &= \beta_1 + E(\alpha_{1i}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) \\
\theta_2(x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}) &= \beta_2 + E(\alpha_{2i}|x_{1i1}, x_{2i1}, \ldots, x_{1iT}, x_{2iT}).
\end{aligned}$$

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

$$\begin{pmatrix} \theta_1(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}) \\ \theta_2(x_{11}, x_{21}, \ldots, x_{1T}, x_{2T}) \end{pmatrix} = \left(\begin{pmatrix} x_{11} & x_{21} \\ \vdots & \vdots \\ x_{1T} & x_{2T} \end{pmatrix}^\top \begin{pmatrix} x_{11} & x_{21} \\ \vdots & \vdots \\ x_{1T} & x_{2T} \end{pmatrix} \right)^{-1}$$

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

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, I 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 $z \in \Omega_z$, I can consistently estimate $\theta(z)$ for $z \in \Omega_z$ by

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

where $K_{h,z_j z} = K((z_j - z)/h)$, $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,...,q\}} |z_l - z_{0,l}| \geq \varepsilon_n \text{ for some } z_0 \in \partial\mathcal{S}_z\}$, 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 2.2. Furthermore, I 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.

2.2 A Correlated Random Coefficient Panel Data Model

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

$$y_{it} = x_{it}^{\top}\beta_i + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{2.12}$$

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}.

I can rewrite (2.12) as

$$y_{it} = x_{it}^{\top}\beta + x_{it}^{\top}\alpha_i + u_{it}, \tag{2.13}$$

$E(u_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$. 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 I 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 I have $E(\epsilon_{it}|x_{it}, z_i) = 0$.

Then I have

$$y_{it} = x_{it}^{\top}\beta + x_{it}^{\top}g(z_i) + \epsilon_{it} = x_{it}^{\top}\theta(z_i) + \epsilon_{it}, \tag{2.14}$$

where $\theta(z) = \beta + g(z)$. Note that equation (2.14) is a semiparametric varying coefficient model. Hence, I can estimate $\theta(z)$ by some standard semiparametric estimator, say, kernel-based local constant or local polynomial estimation methods. From $E(g(z_i)) = 0$ I obtain $\beta = E(\theta(z_i))$. Let $\hat{\theta}(z)$ denote a generic semiparametric estimator of $\theta(z)$, I 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 $\Omega_z = \{z \in \mathcal{S}_z : \min_{l\in\{1,\dots,q\}} |z_l - z_{0,l}| \geq \varepsilon_n$ 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 $z_i = \bar{x}_{i\cdot}$, I 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 I take $z_i = x_i = (x_{i1}^{\top}, ..., x_{iT}^{\top})^{\top}$, I can pool the data together and estimate β by

$$\hat{\beta}_{Semi,2} = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{VC,2}(x_i), \tag{2.15}$$

where

$$\hat{\theta}_{VC,2}(x_i) = \left[\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top} K_{h,x_jx_i}\mathbf{1}_{\varepsilon_n}(x_i)\right]^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,x_jx_i}\mathbf{1}_{\varepsilon_n}(x_i), \tag{2.16}$$

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, I will provide detailed proofs without specifying z_i. I 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

$$\hat{\theta}_{LC}(z) = \left(\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}x_{js}^{\top} K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z)\right)^{-1}\sum_{j=1}^{n}\sum_{s=1}^{T} x_{js}y_{js}K_{h,z_jz}\mathbf{1}_{\varepsilon_n}(z), \tag{2.17}$$

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

I introduce some notations and assumptions before I present the asymptotic theories. I 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. I 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. x_{it} and u_{it} have 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 2.2.1. *Under assumptions A1 to A4, I 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

$$
\begin{aligned}
B_{l,LC} &= \mu_\nu \sum_{k_1+k_2=\nu, k_2\neq 0} \frac{1}{k_1!k_2!} E\left[m_i^{-1} (\frac{\partial^{k_1} m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})\right], \\
m_i &= m(z_i) = T^{-1}\sum_{t=1}^{T} E[x_{it}x_{it}^\top | z_i] f(z_i), \\
\frac{\partial^{k_1} m_i}{\partial z_l^{k_1}} &= \frac{\partial^{k_1} m(z)}{\partial z_l^{k_1}}|_{z=z_i}, \ \frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}} = \frac{\partial^{k_2}\theta(z)}{\partial z_l^{k_2}}|_{z=z_i}, \\
V_{LC} &= Var(\theta(z_i)) + T^{-2} Var\left(\sum_{s=1}^{T} (m_i^{-1} f(z_i) x_{is} x_{is}^\top (\alpha_i - E(\alpha_i|z_i)))\right) \\
& \quad + T^{-2} Var\left(\sum_{s=1}^{T} u_{is} m_i^{-1} x_{is} f(z_i)\right).
\end{aligned}
$$

We can see that the semiparametric estimator I give has a $\sqrt{n}$ convergence rate. The reason is well known that taking average can reduce the variance of nonparametric estimators. I also use the high order kernel to reduce the bias. The proof of Theorem 2.2.1 is given in the Appendix A.

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

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

Then I minimize the kernel weighted sum of squared errors

$$
\sum_{j=1}^{n}\sum_{s=1}^{T}\left[y_{js} - \sum_{0\le|k|\le p} x_{js}^\top b_k(z)(z_j - z)^k \right]^2 K_{h,z_j z}, \tag{2.18}
$$

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 $\theta(z)$. I define $\hat{\beta}_{LP} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{LP}(z_i)$.

Now I need $\theta(z)$ to be $p+1$ times differentiable, and the local polynomial estimation cannot be used together with the high order kernel. So I 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. x_{it} and u_{it} have 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 $0 < i \leq p+2$ is an odd integer and $\mu_i = \int k(v)v^i dv \neq 0$, if $0 < i \leq p+2$ 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 2.2.2. *Under assumptions B1 to B4, I have that*

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

where $B_{LP} = P_1 S^{-1} M \sum_{|k|=p+1} \frac{\mu_k h^k}{k!} E[\Theta_i]$*, if* p *is an odd positive integer, or* $B_{LP} = P_1 S^{-1} M \sum_{|k|=p+2} \frac{\mu_k h^k}{k!} E[\Theta_i]$*, if* p *is an even positive integer,* P_1*,* S*,* M *and* Θ_i *are matrices defined in the Appendix A, and*

$$\begin{aligned} V_{LP} &= Var(\theta(z_i)) + T^{-2} Var\left(\sum_{s=1}^{T} (P_1 S(z_i)^{-1} \Gamma_{is} x_{is}^{\top} (\alpha_i - E(\alpha_i|z_i)) f(z_i))\right) \\ &\quad + T^{-2} Var\left(\sum_{s=1}^{T} P_1 S(z_i)^{-1} u_{is} f(z_i) \Gamma_{is}\right), \end{aligned}$$

where Γ_{is} *is also defined in the Appendix A.*

The proof of Theorem 2.2.2 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 I 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 $\hat{\beta}_{LP}$.

3. BINARY RESPONSE CRC PANEL MODELS

3.1 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 et al. (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. I use this method to identify a binary response CRC panel data model in this paper.

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

$$y_{it} = \mathbf{1}(v_{it} + x_{it}^{\top}\beta_i + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{3.1}$$

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, I assume there exists only one regressor which has constant coefficient and this regressor is the special regressor in model (1.2) to get the model (3.1). The analysis remains similar if I 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, I normalize the coefficient of v_{it} to be 1. If it is negative, I 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, I assume that $E(u_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$, which is the strict exougeneity condition. Also, I 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, I can transfer (3.1) 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. I can identify the average slope if I have enough time period or additional information on z_i as I did in the linear case. I assume the data are independent across i. I 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), I have the following identification proposition.

Proposition 3.1.1. *Under assumptions C1, C2, and C3, let*

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

we have

$$E(y_{it}^*|x_{it}, z_i) = x_{it}^\top\beta + x_{it}^\top g(z_i). \tag{3.2}$$

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

3.2 Estimation of the Binary Response CRC Panel Model

Based on the identification analysis in section 3.1, I can construct the semi-parametric 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 I have an estimator of $\theta(\cdot)$, I can estimate β using $\hat{\beta} = n^{-1}\sum_{i=1}^{n}\hat{\theta}(z_i)$.

From (3.2), I 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, I have to estimate $f_t(v_{it}|x_{it}, z_i)$ and I 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 I estimate $E[x_{it}y_{it}^{*}|z_i]$ by

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

where $\mathbf{1}_{\tau_n,j} = \mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j) = \mathbf{1}\{(v_{jt}, x_{jt}, z_j) \in \Omega_{vxz}\}$, $\Omega_{vxz} = \{a \in \mathcal{S}_{vxz} : \min_{l\in\{1,\ldots,d+q+1\}} |a_l - b_l| \geq \tau_n$, 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 $n \to \infty$. I 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.

I 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, I 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

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

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

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

I list some conditions before I 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, x_{it} and u_{it} have 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 3.2.1. *Under assumptions C1-C7, I have that*

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

where

$$\begin{aligned} V_{LC} &= Var(g(z_i)) + T^{-2} Var\Big(\sum_{t=1}^{T}(m_i^{-1} f_z(z_i) x_{it}\xi_{it} + m_i^{-1} f_z(z_i) x_{it}(E[y_{it}^*|v_i, x_{it}, z_i] \\ &\quad - E[y_{it}^*|x_{it}, z_i]))\Big), \end{aligned}$$

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 3.2.1 is given in the Appendix B.

4. A TRUNCATED CRC PANEL DATA MODEL

4.1 Identification of the Truncated CRC Panel Model

In this section, I discuss the identification of the truncated model (1.3) I discussed in section 1.3 of Chapter 1. My identification result is based on the special regressor method which is similar as the one used in Khan and Lewbel (2007). The idea is to assume the existence of a special regressor which satisfies three conditions, i.e. continuity, conditional independence and relatively large support, which will be more specific below.

Let β be the population mean of β_i, then I have the decomposition $\beta_i = \beta + \alpha_i$, where $E^*(\alpha_i) = 0$. Since β_i and x_{it} are correlated, I introduce z_i to capture this correlation, which satisfies that $E^*(u_{it}|x_{it}, z_i) = 0$ and $E^*(\alpha_i|x_{it}, z_i) = E^*(\alpha_i|z_i) \equiv g(z_i)$, where $g(\cdot)$ is a smooth function. For example I can have $z_i = \bar{x}_i. \equiv T^{-1}\sum_{t=1}^{T} x_{it}$ or $z_i = x_i = (x_{i1}^\top, ..., x_{iT}^\top)^\top$. Define $\epsilon_{it} = x_{it}^\top(\alpha_i - E^*(\alpha_i|z_i)) + u_{it}$. By construction I have $E^*(\epsilon_{it}|x_{it}, z_i) = 0$. Let $\theta(z_i) = \beta + g(z_i)$. Therefore, I have that

$$y_{it}^* = v_{it}\gamma + x_{it}^\top\theta(z_i) + \epsilon_{it}.$$

Since $E^*(\alpha_i) = 0$, I have $E^*(g(z_i)) = E^*(\alpha_i) = 0$ by the law of iterated expectations. Hence, I have $\beta = E^*(\theta(z_i))$. The identification of β depends on the identification of $\theta(\cdot)$.

Recall that I use E^* to denote the expectation under the underlying untruncated population distribution, and I use E to denote the expectation under the truncated distribution. Since I can only partially observe y_{it}^* when $y_{it}^* \geq 0$, I have the following relationship

$$E[h(y_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y_{it} \leq k)|z_i] = \frac{E^*[h(y_{it}^*, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)|z_i]}{P^*(y_{it}^* \geq 0|z_i)},$$

where $h(\cdot)$ is any function of $(y_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})$, $k > 0$ is a constant, and $P^*(y^*_{it} \geq 0|z_i)$ is the conditional probability of the event $\{y^*_{it} \geq 0\}$ under the underlying untruncated probability.

I give some assumptions before I give the identification result.

Assumption D1: Assume $(y_{it}, x_{it}, v_{it}, z_i)$ $(i = 1, \ldots, n, t = 1, \ldots, T)$ are drawn from the model (1.3) with $\gamma \neq 0$, which are independent across the individual index i, and strictly stationary across the time t. The untruncated conditional distribution of v_{it} conditioning on z_i is absolutely continuous with respect to a Lebesgue measure with conditional density function $f^*(v_{it}|z_i)$, which has support $[L, K]$ for some constants L and K, $-\infty \leq L < K \leq \infty$ and for any fixed z_i.

Assumption D2: Assume that conditional on x_{it} and z_i, v_{it} is independent of α_i and u_{it}. Let $F^*_\epsilon(\epsilon_{it}|v_{it}, x_{it}, z_i)$ to denote the underlying untruncated conditional distribution of $\epsilon_{it} = x^\top_{it}(\alpha_i - g(z_i)) + u_{it}$ conditioning on (v_{it}, x_{it}, z_i). This assumption implies that $F^*_\epsilon(\epsilon_{it}|v_{it}, x_{it}, z_i) = F^*_\epsilon(\epsilon_{it}|x_{it}, z_i)$.

Assumption D3: For any $(x_{it}, z_i, \epsilon_{it})$ on the underlying untruncated support of $(x_{it}, z_i, \epsilon_{it})$, we have $[\mathbf{1}(\gamma > 0)L + \mathbf{1}(\gamma < 0)K]\gamma + x^\top_{it}\theta(z_i) + \epsilon_{it} < 0$, and there exists a constant $\tilde{k} > 0$ such that $\tilde{k} \leq [\mathbf{1}(\gamma > 0)K + \mathbf{1}(\gamma < 0)L]\gamma + x^\top_{it}\theta(z_i) + \epsilon_{it}$.

Assumption D4: $E^*(u_{it}|x_{it}, z_i) = 0$, and $\sum_{t=1}^T E^*[x_{it}x^\top_{it}|z_i]$ is invertible.

Assumption D1 to D4 give us the conditions for the identification. Assumption D1 requires the special regressor to be a continuous variable. Assumption D2 means the special regressor is independent of unobserved heterogeneity conditional on the rest of regressors and the random variable z_i we introduce. Assumption D3 requires the support of the special regressor is relatively large. Assumption D4 is the identification condition similar to the linear panel data model which implies that $T \geq d$, where d is the dimension of the regressor x_{it}.

Under the assumptions above, I give the identification result for β. I divide my identification results into three steps. First, given γ I give the theorem on the identification of $\theta(\cdot)$. Second, I discuss how to identify γ. In the end, since the law of iterated expectations imply that $\beta = E^*[\theta(z_i)]$, I can identify β once I have the identification of $\theta(\cdot)$. Let

$$\tilde{y}_{it} = \frac{(y_{it} - v_{it}\gamma)\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)}{E[\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)|z_i]}.$$

Theorem 4.1.1. *Let Assumptions D1 to D4 hold. Let k be any constant satisfying $0 < k \leq \tilde{k}$. Then*

$$\theta(z_i) = \left(\sum_{t=1}^{T} E^*[x_{it}x_{it}^{\top}|z_i]\right)^{-1} \sum_{t=1}^{T} E[x_{it}\tilde{y}_{it}|z_i]. \tag{4.1}$$

Denote

$$\zeta(k) = \frac{1}{T}\sum_{t=1}^{T} \frac{E[2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)]}{E[\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)]}.$$

I have the following identification theorem for γ.

Theorem 4.1.2. *Under Assumptions D1 to D4, and let k and k' be any constants satisfying $0 < k' < k \leq \tilde{k}$. I have*

$$\gamma = \frac{k - k'}{\zeta(k) - \zeta(k')}. \tag{4.2}$$

Once I have the identification result of γ and $\theta(\cdot)$, I can identify β by the equality $\beta = E^*(\theta(z_i))$. In this section, though the observations of y_{it} are not complete, I assume that I can get the full information on the underlying untruncated population distribution of (x_{it}, v_{it}, z_i). In practice, this can be accomplished by the same data set which includes complete observations of the covariates other than just the truncated sample or by an auxiliary data set. This means that $f_t^*(v_{it}|x_{it}, z_i)$ and $E^*(\theta(z_i))$ can be estimated from the data.

4.2 Estimation of the Truncated CRC Panel Model

In this section, I construct our estimator based on the identification results in section 4.1. Recall that $\theta(z_i) = \beta + g(z_i)$. Since $0 = E^*[\alpha_i] = E^*[g(z_i)]$, we have $\beta = E^*[\theta(z_i)]$. Once I have an estimator of $\theta(\cdot)$, I can estimate β using $\hat{\beta} = (n^*)^{-1}\sum_{i=1}^{n^*}\hat{\theta}(z_i)$.

First, I construct the estimator for γ. Denote

$$\begin{aligned}\mu_t(k, z_i) &= E[\mathbf{1}(0 \le y_{it} \le k)/f^*(v_{it}|x_{it}, z_i)|z_i],\\ \mu_t(k) &= E[\mathbf{1}(0 \le y_{it} \le k)/f^*(v_{it}|x_{it}, z_i)].\end{aligned}$$

From (4.1.2), I have to give the estimator for $\mu_t(k, z_i)$. Since $f_t^*(v_{it}|x_{it}, z_i)$ is unknown, I have to estimate $f_t^*(v_{it}|x_{it}, z_i)$ and I estimate it by

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

where $\hat{f}_t^*(v_{it}, x_{it}, z_i) = (n^*H)^{-1}\sum_{k=1}^{n^*} K_h(v_{kt}^* - v_{it}, x_{kt}^* - x_{it}, z_k^* - z_i)$, $\hat{f}_t^*(x_{it}, z_i) = (n^*\tilde{H})^{-1}\sum_{k=1}^{n^*} K_{\tilde{h}}(x_{kt}^* - x_{it}, z_k^* - z_i)$, $H = h_1 \cdots h_{d+q+1}$, $\tilde{H} = h_2 \cdots h_{d+q+1}$, $K_h(u) = \prod_{l=1}^{d+q+1} k\left(\frac{u_l}{h_l}\right)$, $h = (h_1, \ldots, h_{d+q+1})^\top$, and $\tilde{h} = (h_2, \ldots, h_{d+q+1})^\top$. Then I give the estimator for $\mu_t(k, z_i)$ and $\mu_t(k)$ as

$$\begin{aligned}\hat{\mu}_t(k, z_i) &= \frac{(nH')^{-1}\sum_{j=1}^{n}\mathbf{1}(0 \le y_{jt} \le k)K_{h'}(z_j - z_i)/\hat{f}_t^*(v_{jt}|x_{jt}, z_j)}{(nH')^{-1}\sum_{j=1}^{n} K_{h'}(z_j - z_i)},\\ \hat{\mu}_t(k) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0 \le y_{it} \le k)}{\hat{f}_t^*(v_{it}|x_{it}, z_i)}\mathbf{1}_{\tau_n, i},\end{aligned}$$

and the estimator for $\zeta(k)$ can be constructed as

$$\hat{\zeta}(k) = \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{\hat{f}_t^*(v_{it}|x_{it}, z_i)}\mathbf{1}_{\tau_n, i},$$

where $\mathbf{1}_{\tau_n,i} = \mathbf{1}_{\tau_n}(v_{it}, x_{it}, z_i) = \mathbf{1}\{(v_{it}, x_{it}, z_i) \in \Omega_{vxz}\}$, $\Omega_{vxz} = \{a \in \mathcal{S}_{vxz} : \min_{l\in\{1,\dots,d+q+1\}} |a_l - b_l| \geq \tau_n,$ 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_{it}, x_{it}, z_i), $H' = h'_1 \cdots h'_q$, $h' = (h'_1, \dots, h'_q)$, $\|h\|/\tau_n \to 0$, and $\tau_n \to 0$, as $n \to \infty$. I use $\mathbf{1}_{\tau_n}(v_{it}, x_{it}, z_i)$ to truncate the data at the boundary to avoid the singularity problem and the boundary bias. Hence, our estimator of γ is

$$\hat{\gamma} = \frac{k - k'}{\hat{\zeta}(k) - \hat{\zeta}(k')}. \tag{4.3}$$

From (4.1), I have $\theta(z_i) = \left(\sum_{t=1}^{T} E^*[x_{it}x_{it}^\top|z_i]\right)^{-1} \sum_{t=1}^{T} E[x_{it}\tilde{y}_{it}|z_i]$. Since

$$E[x_{it}\tilde{y}_{it}|z_i] = E[x_{it}(y_{it} - v_{it}\gamma)\mathbf{1}(0 \leq y_{it} \leq k)/\mu_t(k, z_i)f_t^*(v_{it}|x_{it}, z_i)|z_i],$$

I estimate $E[x_{it}\tilde{y}_{it}|z_i]$ by

$$\hat{E}[x_{it}\tilde{y}_{it}|z_i] = \frac{(nH')^{-1}\sum_{j=1}^{n} x_{jt}(y_{jt} - v_{jt}\hat{\gamma})\mathbf{1}(0 \leq y_{jt} \leq k)K_{h',ji}\mathbf{1}_{\tau_n,j}/\hat{\mu}_t(k, z_j)\hat{f}^*_{t,v|xz,j}}{(nH')^{-1}\sum_{j=1}^{n} K_{h',ji}},$$

where $f^*_{t,v|xz,j} = f_t^*(v_{jt}|x_{jt}, z_j)$, and $\mathbf{1}_{\tau_n,j} = \mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j) = \mathbf{1}\{(v_{jt}, x_{jt}, z_j) \in \Omega_{vxz}\}$. I use the trimming function $\mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j)$ to trim the data at the boundary to avoid the singularity problem and the boundary bias.

I 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, I will not discuss it here. However, based on the analysis in the linear case, I 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

$$\begin{aligned}\hat{\theta}_{LC}(z_i) &= [\frac{1}{n^*}\sum_{j=1}^{n^*}\sum_{t=1}^{T} x^*_{jt}(x^*_{jt})^\top K_{h',ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i}]^{-1}\frac{1}{n}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{(y_{jt} - v_{jt}\hat{\gamma})}{\hat{f}_t^*(v_{jt}|x_{jt}, z_j)} \\ &\quad \times \frac{\mathbf{1}(0 \leq y_{jt} \leq k))}{\hat{\mu}_t(k, z_j)}K_{h',ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i},\end{aligned}$$

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

$$\hat{\beta}_{LC} = (n^*)^{-1} \sum_{i=1}^{n^*} \hat{\theta}_{LC}(z_i). \tag{4.4}$$

I list some conditions before I present the asymptotic distribution.

Assumption D5: $(y_i^\top, v_i^\top, x_i^\top, z_i^\top)$ 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 D6: $\theta(z)$, $f_t(v, x, z)$, $f_t(v, x)$ and $f_z(z)$ are $\nu + 1$ times continuously differentiable, where ν is an integer defined in the next assumption.

Assumption D7: $K(z) = \prod_{l=1}^q k(z_l)$, 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 D8: As $n \to \infty$, $n/n^* \to c$, $0 \leq c < \infty$, $n^* H'^2/\ln n^* \to \infty$, $\sqrt{n^*}H/\ln n^* \to \infty$, $\|h'\|^{2\nu} \ln n^*/H' \to 0$, $\|h'\|^\nu/H' \to 0$, $n^*||h'||^{2\nu} \to 0$, $n^*||h||^{2\nu} \to 0$, $n^*||\tilde{h}||^{2\nu} \to 0$, $\varepsilon_n \to 0$, $\tau_n \to 0$, $\|h'\|/\varepsilon_n \to 0$, $\|h\|/\tau_n \to 0$, $\varepsilon_n > \tau_n$, $\|h'\|/(\varepsilon_n - \tau_n) \to 0$, $h_l \to 0$ for all $l = 1, ..., d + q + 1$, $h'_l \to 0$ for all $l = 1, ..., q$.

Then I have the following asymptotic theorem.

Theorem 4.2.1. *Under assumptions D1-D8, I have that* *(i)* $\sqrt{n}(\hat{\gamma}-\gamma) \xrightarrow{d} N(0, V_\gamma)$*, where* $V_\gamma = E[\psi_t(k)^2]$*,*

$$\begin{aligned}
\psi_t(k) &= \frac{\gamma^2}{k-k'}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\mu_t(k)^{-1}\varphi_k(k) - \phi_t(k)\mu_t(k)^{-2}\eta_t(k) + \mu_t(k')^{-1}\varphi_t(k') \\
&\quad -\phi_t(k')\mu_t(k')^{-2}\eta_t(k')\Big)\Big], \\
\varphi_t(k) &= \frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \eta_t(k) \\
&\quad -cE[\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad +cE[\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*], \\
\phi_t(k) &= \frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \mu_t(k) \\
&\quad -cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad +cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*], \\
\eta_t(k) &= E[2v_{it}\mathbf{1}(0 \le y_{it} \le k)/f_t^*(v_{it}|x_{it}, z_i)];
\end{aligned}$$

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

$$\begin{aligned}
V_{LC} &= E^*(g(z_i^*))^2 + E^*\Bigg(T^{-1}\sum_{t=1}^{T}\Big[m_i^{-1}f_z(z_i^*)x_{it}\xi_{it} \\
&\quad +m_i^{-1}f_z^*(z_i^*)x_{it}^*\Big(E[\tilde{y}_{it}|v_i = v_i^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad -E[\tilde{y}_{it}|x_{it} = x_{it}^*, z_i = z_i^*]\Big) \\
&\quad -m_i^{-1}E^*[x_{it}x_{it}^\top|z_i = z_i^*]\theta_i f_z(z_i^*)\phi_t(k, z_i^*) \\
&\quad -m_i^{-1}f_z(z_i^*)\left(\frac{1}{2\gamma^2}(k^2E[x_{it}|z_i = z_i^*] - kE[x_{it}x_{it}^\top|z_i = z_i^*]\theta(z_i^*))\right) \\
&\quad \times\frac{\psi_t(k)}{\mu_t(k, z_i^*)}\Big]\Bigg)^2, \\
\phi_t(k, z_i^*) &= \frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \mu_t(k, z_i^*)
\end{aligned}$$

$$-cE[\frac{\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*]$$
$$+cE[\frac{\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*],$$

$\xi_{it} = \tilde{y}_{it} - E(\tilde{y}_{it}|x_{it}, z_i)$, *and* $\tilde{y}_{it} = [(y_{it} - v_{it}\gamma)\mathbf{1}(0 \leq y_{it} \leq k)]/f_t(v_{it}|x_{it}, z_i)$, *if* $v_{it} \in [L_t, K_t]$, *and* $y_{it}^* = 0$, *otherwise.*

The proof of Theorem 4.2.1 is given in the Appendix C.

5. MONTE CARLO SIMULATIONS AND EMPIRICAL APPLICATION

5.1 Monte Carlo Simulation Results

In this section, I conduct extensive simulations to examine the finite sample performance of different estimators including semiparametric estimators I proposed in sections 2.2 and 3.2.

5.1.1 Linear CRC Panel Data Models

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

$$y_{it} = \beta_{0i} + x_{it}\beta_{1i} + u_{it}, \qquad (i = 1, ..., n;\ t = 1, ..., T) \tag{5.1}$$

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 $T = 3$. I 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.

I 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 $\hat{\beta}_{FE}$,

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

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

(iii) I 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 I average $\hat{\beta}_{i,OLS}$ to obtain the group mean estimator $\hat{\beta}_{GM}$ as defined in (2.10).

(iv) If we let $z_i = \bar{x}_{i}.$, where $\bar{x}_{i}. = \frac{1}{T}\sum_{t=1}^{T}x_{it}$, then I 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

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

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

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

where $K_{h,x_j.x_i.} = K_h(\bar{x}_{j}. - \bar{x}_{i}.)$, $K(\cdot)$ is a kernel function and h is the smoothing parameter.

(v) If I let $z_i = x_i = (x_{i1}^\top, ..., x_{iT}^\top)^\top$, then I 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.

Below I report the result of a small simulation study. I 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 $\alpha_{1i} = v_{1i} - E(v_{1i})$.

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

where $\bar{x}_i. = T^{-1}\sum_{t=1}^T x_{it}$, 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}.

The simulation results are reported in Table 5.1, Table 5.2, Table 5.3 and Table 5.4, and the results confirm our theoretical analysis in the paper. I can see that in all of these tables, $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FE}$ are not consistent.

Table 5.1

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

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

Table 5.2

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

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

From Table 5.1 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.

The simulation results for DGP2 is given in Table 5.2. 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 5.3

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

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

From Table 5.3 we observe that $\hat{\beta}_{Semi,1}$ has the smallest estimation MSE, followed by $\hat{\beta}_{Semi,2}$ and $\hat{\beta}_{GM}$.

Table 5.4

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

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

Table 5.4 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.

5.1.2 Binary Response CRC Models

In this section, I conduct simulations for binary response CRC models. I compare the estimators as in section 5.1.1 with y_{it} substituted by $\frac{(y_{jt}-\mathbf{1}(v_{jt}>0))}{\hat{f}_t(v_{jt}|x_{jt},z_j)}$. I 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 5.5

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

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

Table 5.6

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

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

Table 5.7

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

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

Table 5.8

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

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

The simulation results are reported in Table 5.5, Table 5.6, Table 5.7 and Table 5.8. We can see that the semiparametric estimators we proposed perform well.

5.1.3 A Truncated CRC Panel Data Model

In this section, I conduct simulations for the truncated CRC panel data model. I generate y_{it} by

$$\begin{aligned} y_{it}^* &= \mathbf{1}(\gamma v_{it} + \beta_{0i} + x_{it}\beta_{1i} + u_{it} > 0), \qquad (i = 1, ..., n;\ t = 1, ..., T;\ T = 3) \\ y_{it} &= y_{it}^* | y_{it}^* \geq 0, \end{aligned}$$

where $\beta_{0i} = \beta_0 + \alpha_{0i}$, $\beta_{1i} = \beta_1 + \alpha_{1i}$, $\beta_0 = 0.5$, $\beta_1 = 1$, $\gamma = 0.5$, 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 $\alpha_{1i} = w_{1i} - E(w_{1i})$.

$$\begin{aligned} DGP9: &\quad v_{it} \text{ is independent of } \alpha_{0i}, \alpha_{1i} \text{ and } u_{it}, \text{ and distributed as } Uniform[-4, 4], \\ &\quad w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = (\bar{x}_{i\cdot} - 1)^2 + \ln(\bar{x}_{i\cdot} + 1) + \eta_{1i}, \\ DGP10: &\quad v_{it} \text{ is independent of } \alpha_{0i}, \alpha_{1i} \text{ and } u_{it}, \text{ and distributed as } Uniform[-4, 4], \\ &\quad w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = \sin(3\bar{x}_{i\cdot}) + \eta_{1i}, \\ DGP11: &\quad v_{it} = \bar{x}_{i\cdot}^2 + w_{it}, \text{ where } w_{it} \sim Uniform[-4, 4], \\ &\quad w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = (\bar{x}_{i\cdot} - 1)^2 + \ln(\bar{x}_{i\cdot} + 1) + \eta_{1i}, \\ DGP12: &\quad v_{it} = \bar{x}_{i\cdot}^2 + w_{it}, \text{ where } w_{it} \sim Uniform[-4, 4], \\ &\quad w_{0i} = (\bar{x}_{i\cdot} - 1)^4 + \eta_{0i}, \text{ and } w_{1i} = \sin(3\bar{x}_{i\cdot}) + \eta_{1i}, \end{aligned}$$

where $\bar{x}_{i\cdot} = T^{-1}\sum_{t=1}^{T} x_{it}$, η_{0i} and η_{1i} are i.i.d. with $Uniform[-0.5, 0.5]$. I use $z_i = \bar{x}_{i\cdot}$, $k = 0.5$ and $k' = 2$ for estimators in (4.3) and (4.4).

The simulation results are reported in Table 5.9, Table 5.10, Table 5.11 and Table 5.12. We can see that the semiparametric estimators we proposed perform well.

Table 5.9
MSE of $\hat{\gamma}$, $\hat{\beta}_0$, $\hat{\beta}_1$ for DGP9

n	$MSE(\hat{\gamma})$	$MSE(\hat{\beta}_0)$	$MSE(\hat{\beta}_1)$
100	0.0029	0.0330	0.8655
200	0.0013	0.0164	0.6551
400	0.0006	0.0099	0.5321

Table 5.10
MSE of $\hat{\gamma}$, $\hat{\beta}_0$, $\hat{\beta}_1$ for DGP10

n	$MSE(\hat{\gamma})$	$MSE(\hat{\beta}_0)$	$MSE(\hat{\beta}_1)$
100	0.0030	0.0334	0.8101
200	0.0014	0.0191	0.5537
400	0.0007	0.0110	0.3952

Table 5.11
MSE of $\hat{\gamma}$,, $\hat{\beta}_0$, $\hat{\beta}_1$ for DGP11

n	$MSE(\hat{\gamma})$	$MSE(\hat{\beta}_0)$	$MSE(\hat{\beta}_1)$
100	0.0029	0.0307	0.8698
200	0.0013	0.0162	0.6612
400	0.0006	0.0097	0.5236

Table 5.12
MSE of $\hat{\gamma}$, $\hat{\beta}_0$, $\hat{\beta}_1$ for DGP12

n	$MSE(\hat{\gamma})$	$MSE(\hat{\beta}_0)$	$MSE(\hat{\beta}_1)$
100	0.0031	0.0335	0.8373
200	0.0014	0.0182	0.5735
400	0.0007	0.0101	0.4084

5.2 An Empirical Application

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

$$\log(wage_{it}) = \beta_{0i}+\beta_{1i}t+\beta_{2i}tenure_{it}+\beta_{3i}educ_{it}+\beta_{4i}union_{it}+\beta_{5i}training_{it}+u_{it}. \quad (5.2)$$

Here, β_{0i} is the fixed effects term which captures the time invariant characteristics of individuals, for instance, gender. I 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. I use edu_{it} to denote years of schooling, $union_{it}$ to denote the union status of the individual, which is also an important factor 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 I 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 a correlation between $training_{it}$ and β_{6i}. These make (5.2) a linear CRC panel data model. Also, random coefficients are used to capture unobserved heterogeneity.

I 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. I 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.

For the estimation of (5.2), I take first difference and get that

$$\log(wage_{it}) - \log(wage_{i,t-1}) = \beta_{1i} + \beta_{2i}\Delta tenure_{it} + \beta_{3i}\Delta educ_{it} + \beta_{4i}\Delta union_{it} + \beta_{5i}\Delta training_{it} + \Delta u_{it}, \quad (5.3)$$

where $\Delta A_{it} = A_{it} - A_{i,t-1}$. The reason I do the first difference is that I can only observe the increment of hours spent on the job training since the last period, not the accumulated hours. Also, it helps me to get rid of the fixed effects term β_{0i}. Then I 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 (5.3), which is equivalent to the first difference estimators for (5.2). I also use the nonparametric method I proposed in (2.15) to estimate (5.3). I report the result in the following table.

Table 5.13

Estimation results of (5.3) by OLS and nonparametric methods

Variables	First difference estimates	Nonparametric estimates
Time trend	5.37%	5.38%
Tenure (weeks)	0.025%	0.017%
Education (years)	2.66%	4.46%
Union	11.47%	16.24%
Job training (per 60 hours)	0.42%	3.16%
Time range: 1988 - 2008 (14 interviews)		
Sample size: 3287		

I use the data of 3287 individuals who took job training during 1988 to 2008. From table 5.13, 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 my nonparametric method for correcting the correlations, I get the return of joining the union is 1.4 times as much as the one estimated by the first difference method. Also, the estimate of the return from job training based on my method is 7 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 5.38%. The increase rate of tenure is 0.017% 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 4.46% for one year more education. Also, I find that the return from joining the union is 16.24%, and the rate of return from job training is 3.16% per 60 hours training. The result for the rate of return from job training is close to the result in Frazis and Loewenstein (2005) which is 3-4 percent for 60 hours of formal training, the median positive amount of training.

Table 5.14
Estimation results of (5.2) with nonlinear functional form in training

Variables	First difference estimates
Time trend	5.52%
Tenure (weeks)	0.025%
Education (years)	2.69%
Union	11.41%
Job training (per 60 hours)	2.79%

Frazis and Loewenstein (2005) proposed to use an optimal functional form which is $(T^{0.35} - 1)/0.35$ for NLSY 79 data for the training variable and use the fixed effects estimators. I use the functional form they proposed and the first difference estimation to estimate the data I gathered, and the results are reported in Table 5.14. We can see that the estimation result is similar as the one from the nonparametric estimation I proposed.

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

6. CONCLUSION

In this dissertation, I discuss the identification and estimation of linear CRC panel data models, binary response CRC panel data models, and a truncated CRC panel data model. I use the linear CRC panel data model to show how I 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, I extend the idea to the binary choice CRC panel data model. The identification of the binary choice model is different from the linear model. I base my identification result on the special regressor method. Moreover, I construct the $\sqrt{n}$ consistent asymptotically normal semiparametric estimators for both models. Further, I did simulations and an empirical application to show the advantage of our estimators.

There are some extensions I am considering. In the example given in section 2.1, the regressor is a discrete variable but I 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), I leave the rigorous derivations for future research. In addition, it is desirable to construct tests for CRC panel data models. I also leave this for further research.

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. Oxford University Press, New York.

Arellano, M., Bonhomme, S., 2012. Identifying distributional characteristics in random coefficient panel data models. Review of Economic Studies (forthcoming).

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

Bonhomme, S., 2012. Functional differencing. Econometrica (forthcoming).

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

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

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

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

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

Fox, J.T., Gandhi, A., 2010. Nonparametric identification and estimation of random coefficients in nonlinear economic models. Working Paper. University of Chicago, 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., 2012. Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models. Econometrica (forthcoming).

Hahn, J., 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.

Heckman, J.J., Vytlacil, E., 1998. Instrumental variables methods for the correlated random coefficient model. Journal of Human Resources 33, 974-987.

Hoderlein, S., 2009. Endogeneity in semiparametric binary random coefficient models. Working Paper. Boston College, Boston.

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

Hoderlein, S., White, H., 2012. Nonparametric identification in nonseparable panel data models with generalized fixed effects. Journal of Econometrics 168, 300-314.

Honoré, B.E., 1992. Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. Econometrica 60, 533-565.

Honoré, B.E., 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 University Press, Cambridge.

Hsiao, C., Pesaran, M.H., 2008. Random coefficient models. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. In: Advanced Studies in Theoretical and Applied Econometrics, vol. 46. Springer-Verlag, Berlin, pp. 185-213.

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: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics: Theory and Applications III. Cambridge University Press, Cambridge, pp. 17-46.

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 University Press, Princeton.

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.

Murtazashvili, I., Wooldridge, J.M., 2008. Fixed effects instrumental variables estimation in correlated random coefficient panel data models. Journal of Econometrics 142, 539-552.

Newey, W.K., Ruud, P.A., 2005. Density weighted linear least squares. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference in Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp. 554-573.

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: Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Blackwell Publishing Ltd, Malden, pp. 410-428.

Wooldridge, J.M., 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. Economics Letters 79, 185-191.

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.

APPENDIX A

Proof of Theorem 2.2.1: I first consider the local constant estimation method. For any $z \in \Omega_z$, we have

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

where

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

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

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.1.1 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

$$\begin{aligned}B_{n1} &= n^{-1}\sum_{i=1}^{n}m(z_i)^{-1}A_{n2}(z_i),\\ B_{n2} &= n^{-1}\sum_{i=1}^{n}m(z_i)^{-1}A_{n3}(z_i),\\ \eta_n &= O_p(||h||^{\nu} + (\ln n/(nH))^{1/2})O_p(\|A_{n2}(z_i)\| + \|A_{n3}(z_i)\|).\end{aligned}$$

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

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

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

$$\begin{aligned} E[H_{n1,ij}] &= (TH)^{-1}\sum_{s=1}^{T}E[m_i^{-1}x_{js}x_{js}^{\top}(\theta_j-\theta_i)K_{h,ij}] \\ &= (TH)^{-1}\sum_{s=1}^{T}E[m_i^{-1}E(x_{js}x_{js}^{\top}|z_j)(\theta_j-\theta_i)K_{h,ij}] \end{aligned}$$

$$
\begin{aligned}
&= H^{-1}E[m_i^{-1}m_jf_j^{-1}(\theta_j-\theta_i)K_{h,ij}] \\
&= H^{-1}\int\int m_i^{-1}f_im_j(\theta_j-\theta_i)K_{h,ij}dz_idz_j \\
&= \int\int m_i^{-1}f_im(z_i+hv)(\theta(z_i+hv)-\theta_i)K(v)dvdz_i \\
&= \mu_\nu\sum_{l=1}^{q}\sum_{k_1+k_2=\nu,k_2\neq0}\frac{h_l^\nu}{k_1!k_2!}\int m_i^{-1}f_i(\frac{\partial^{k_1}m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})dz_i+O(||h||^{\nu+1}) \\
&= \sum_{l=1}^{q}h_l^\nu B_{l,LC}+O_p(||h||^{\nu+1}),
\end{aligned}
$$

where $B_{l,LC}=\mu_\nu\sum_{k_1+k_2=\nu,k_2\neq0}\frac{1}{k_1!k_2!}E\left[m_i^{-1}(\frac{\partial^{k_1}m_i}{\partial z_l^{k_1}})(\frac{\partial^{k_2}\theta_i}{\partial z_l^{k_2}})\right]$.

Also, we have

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

and

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

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

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}.

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

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

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

Also, by Cauchy-Schwarz inequality we have that

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

where $\mathbf{1}_{\varepsilon_n,i} = \mathbf{1}_{\varepsilon_n}(z_i)$, and $A^{\otimes 2}$ 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,

$$\begin{aligned}&Var\Big(\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T}\left(m_i^{-1}f(z_i)x_{is}x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\mathbf{1}_{\varepsilon_n}(z_i)\Big) \\ =\ &Var\left(\frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T}\left(m_i^{-1}f(z_i)x_{is}x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i)\right)\right) + o(1).\end{aligned}$$

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

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

by the Lindeberg central limit theorm, where

$$\begin{aligned} V_{LC} &= Var(v_i) + T^{-2}Var\left(\sum_{s=1}^{T}(m_i^{-1}f(z_i)x_{is}x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) + u_{is}m_i^{-1}x_{is}f(z_i))\right) \\ &= Var(\theta_i) + T^{-2}Var\left(\sum_{s=1}^{T}m_i^{-1}f(z_i)x_{is}x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i))\right) \\ &\quad + T^{-2}Var\left(\sum_{s=1}^{T}u_{is}m_i^{-1}x_{is}f(z_i)\right) \end{aligned}$$

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.1.1. *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,*

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

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

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

which completes the proof of Lemma A.1.1.

Proof of Theorem 2.2.2: Now, we consider the local polynomial estimation method.

The minimization of (2.18) leads to the set of equations

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

where

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

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., $(0,\dots,0,r)$ is the first element and $(r,0,\dots,0)$ is the last one. Denote this vector by $\tau_{n,r}(z)$. Let $\tau_n = (\tau_{n,0}(z)^\top, \tau_{n,1}(z)^\top, \dots, \tau_{n,p}(z)^\top)^\top$. 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 $\hat{\delta}(z) = (\hat{\delta}_{n,0}(z)^\top, \hat{\delta}_{n,1}(z)^\top, \dots, \hat{\delta}_{n,p}(z)^\top)^\top$. 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 $t^*_{n,i} = (nTH)^{-1}\sum_{j=1}^{n}$ $\sum_{s=1}^{T} x_{js}x_{js}^\top(\theta(z_j) - \sum_{0\le|k|\le p}\frac{1}{k!}(D^k\theta(z))(z_j - z)^k)\left(\frac{z_i-z}{h}\right)^i K_{h,z_j z}$ and $(nTH)^{-1}\sum_{j=1}^{n}$ $\sum_{s=1}^{T} x_{js}\epsilon_{js}\left(\frac{z_i-z}{h}\right)^i K_{h,z_j z}$, respectively.

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

$$\begin{aligned} \hat{\beta}_{LP} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{LP}(z_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}\theta(z_i) + \frac{1}{n}\sum_{i=1}^{n} P_1 S_n(z_i)^{-1}\left[C_{n1}(z_i) + C_{n2}(z_i)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\theta(z_i) + \frac{1}{n}\sum_{i=1}^{n} P_1 S(z_i)^{-1}\left[C_{n1}(z_i) + C_{n2}(z_i)\right] + (s.o.), \end{aligned}$$

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

Similar as in the proof of Theorem2.2.1, we have that if $p > 0$ is an odd integer,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} P_1 S(z_i)^{-1} C_{n1}(z_i) &= \sum_{|k|=p+1} \frac{\mu_k h^k}{k!} P_1 E\left[S_i^{-1} M_i \Theta_i\right] \\
&\quad + O_p(\|h\|^{p+2} + n^{-1}H^{-1/2}\|h\|^{p+1}) \\
&= B_{LP} + O_p(\|h\|^{p+2} + n^{-1}H^{-1/2}\|h\|^{p+1}), \\
\frac{1}{n}\sum_{i=1}^{n} P_1 S(z_i)^{-1} C_{n2}(z_i) &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{s=1}^{T} P_1 S(z_i)^{-1}\Gamma_{is} x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) f(z_i) \\
&\quad + \frac{1}{nT}\sum_{i=1}^{n}\sum_{s=1}^{T} P_1 S(z_i)^{-1} u_{is} f(z_i)\Gamma_{is} \\
&\quad + O_p(\|h\|^2/\sqrt{n} + (nH^{1/2})^{-1}),
\end{aligned}$$

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 $N \times 1$ 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 $\mu_{\alpha+\gamma}$, $0 \le |\alpha| \le p$, $0 \le |\gamma| \le p$, and M for the $N \times 1$ 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 $B_{LP} = P_1 S^{-1} M \sum_{|k|=p+1} \frac{\mu_k h^k}{k!} E\left[\Theta_i\right]$.

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

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} P_1 S(z_i)^{-1} C_{n1}(z_i) &= \sum_{|k|=p+2} \frac{\mu_k h^k}{k!} P_1 E\left[S_i^{-1} M_i \Theta_i\right] \\ &\quad + O_p(\|h\|^{p+4} + n^{-1} H^{-1/2} \|h\|^{p+2}) \\ &= P_1 S^{-1} M \sum_{|k|=p+2} \frac{\mu_k h^k}{k!} E\left[\Theta_i\right] \\ &\quad + O_p(\|h\|^{p+4} + n^{-1} H^{-1/2} \|h\|^{p+2}) \\ &= B_{LP} + O_p(\|h\|^{p+4} + n^{-1} H^{-1/2} \|h\|^{p+2}).\end{aligned}$$

Therefore, we have that

$$\begin{aligned}&\sqrt{n}\left(\hat{\beta}_{LP} - \beta - B_{LP}\right) \\ = \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\theta_i - \beta) + \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T} P_1 S(z_i)^{-1} \Gamma_{is} x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) f(z_i) \\ & + \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\sum_{s=1}^{T} P_1 S(z_i)^{-1} u_{is} f(z_i) \Gamma_{is} + O_p(\zeta_n) \\ \xrightarrow{d} \;& N(0, V_{LP})\end{aligned} \tag{A.9}$$

by the Lindeberg central limit theorem, where

$$\begin{aligned}V_{LP} &= Var(\theta_i) + T^{-2} Var\left(\sum_{s=1}^{T} P_1 S(z_i)^{-1} \Gamma_{is} x_{is}^{\top}(\alpha_i - E(\alpha_i|z_i)) f(z_i)\right) \\ &\quad + T^{-2} Var\left(\sum_{s=1}^{T} P_1 S(z_i)^{-1} u_{is} f(z_i) \Gamma_{is}\right)\end{aligned}$$

$$\zeta_n = (nH)^{-1/2} + (n||h||^{2p+4})^{1/2} + (nH)^{-1/2}\|h\|^{p+1} + \sqrt{n\|h\|^{2p+2}}(\ln n/(nH))^{1/2}$$
$$+\|h\|(nH)^{-1/2} + (nH)^{-1/2}(\ln n/(nH))^{1/2} = o_p(1)$$

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 3.1.1: 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

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

$$
\begin{aligned}
&= \int_{L_t}^{K_t}\int_{\Omega_{e_t}}[\mathbf{1}(v_{it}+x_{it}^\top\theta(z_i)+e_{it}>0)-\mathbf{1}(v_{it}>0)]dF_{e_{it}}(e_{it}|v_{it},x_{it},z_i)dv_{it}\\
&= \int_{\Omega_{e_t}}\int_{L_t}^{K_t}[\mathbf{1}(v_{it}>s_{it})-\mathbf{1}(v_{it}>0)]dv_{it}dF_{e_{it}}(e_{it}|x_{it},z_i) \qquad \text{Let } (s_{it}=-x_{it}^\top\theta(z_i)-e_{it})\\
&= \int_{\Omega_{e_t}}\int_{L_t}^{K_t}[(\mathbf{1}(v_{it}>s_{it})-\mathbf{1}(v_{it}>0))\mathbf{1}(s_{it}\le 0)+(\mathbf{1}(v_{it}>s_{it})-\mathbf{1}(v_{it}>0))\mathbf{1}(s_{it}>0)]\\
&\quad dv_{it}dF_{e_{it}}(e_{it}|x_{it},z_i)\\
&= \int_{\Omega_{e_t}}\int_{L_t}^{K_t}[\mathbf{1}(s_{it}<v_{it}\le 0)\mathbf{1}(s_{it}\le 0)-\mathbf{1}(0<v_{it}\le s_{it})\mathbf{1}(s_{it}>0)]dv_{it}dF_{e_{it}}(e_{it}|x_{it},z_i)\\
&= \int_{\Omega_{e_t}}[\mathbf{1}(s_{it}\le 0)\int_{s_{it}}^{0}1dv_{it}-\mathbf{1}(s_{it}>0)\int_0^{s_{it}}1dv_{it}]dF_{e_{it}}(e_{it}|x_{it},z_i)\\
&= \int_{\Omega_{e_t}}-s_{it}dF_{e_{it}}(e_{it}|x_{it},z_i)\\
&= \int_{\Omega_{e_t}}(x_{it}^\top\theta(z_i)+e_{it})dF_{e_{it}}(e_{it}|x_{it},z_i)\\
&= x_{it}^\top\theta(z_i)+E(e_{it}|x_{it},z_i)\\
&= x_{it}^\top\theta(z_i).
\end{aligned}
$$

This completes the proof.

We give some shorthand notations first. These notations will be used throughout the proof of Theorem 3.2.1. Let

$$
\begin{aligned}
K_{h',z,jz} &= K_{h'}(z_j-z),\ K_{h',z,ji}=K_{h'}(z_j-z_i),\ K_{h',z,ij}=K_{h'}(z_i-z_j),\\
K_{h',z,jk} &= K_{h'}(z_j-z_k),\ K_{h',z,kj}=K_{h'}(z_k-z_j),\ K_{h',z,ki}=K_{h'}(z_k-z_i),\\
K_{h',z,ik} &= K_{h'}(z_i-z_k),\ K_{h,vxz,kj}=K_h(v_{kt}-v_{jt},x_{kt}-x_{jt},z_k-z_j),\\
K_{h,vxz,ki} &= K_h(v_{kt}-v_{it},x_{kt}-x_{it},z_k-z_i),
\end{aligned}
$$

$$
\begin{aligned}
K_{h,vxz,ij} &= K_h(v_{it}-v_{jt}, x_{it}-x_{jt}, z_i-z_j),\\
K_{h,vxz,jk} &= K_h(v_{jt}-v_{kt}, x_{jt}-x_{kt}, z_j-z_k),\\
K_{h,vxz,ik} &= K_h(v_{it}-v_{kt}, x_{it}-x_{kt}, z_i-z_k),\\
K_{h,vxz,ji} &= K_h(v_{jt}-v_{it}, x_{jt}-x_{it}, z_j-z_i),\\
K_{h,vxz,mj} &= K_h(v_{mt}-v_{jt}, x_{mt}-x_{jt}, z_m-z_j),\ K_{\tilde{h},xz,mj} = K_{\tilde{h}}(x_{mt}-x_{jt}, z_m-z_j),\\
\hat{f}_{t,v|xz,j} &= \hat{f}_t(v_{jt}|x_{jt}, z_j),\ f_{t,v|xz,j} = f_t(v_{jt}|x_{jt}, z_j),\ \hat{f}_{t,vxz,j} = \hat{f}_t(v_{jt}, x_{jt}, z_j),\\
f_{t,vxz,j} &= f_t(v_{jt}, x_{jt}, z_j),\ f_{t,vxz,i} = f_t(v_{it}, x_{it}, z_i),\ f_{t,vxz,k} = f_t(v_{kt}, x_{kt}, z_k),\\
f^{-1}_{t,vxz,j} &= f^{-1}_t(v_{jt}, x_{jt}, z_j),\ f^{-1}_{t,vxz,i} = f^{-1}_t(v_{it}, x_{it}, z_i),\ f^{-1}_{t,vxz,k} = f^{-1}_t(v_{kt}, x_{kt}, z_k),\\
\hat{f}_{t,xz,j} &= \hat{f}_t(x_{jt}, z_j),\ f_{t,xz,j} = f_t(x_{jt}, z_j),\\
\mathbf{1}_{\tau_n,j} &= \mathbf{1}_{\tau_n}(v_{jt}, x_{jt}, z_j),\ \mathbf{1}_{\tau_n,i} = \mathbf{1}_{\tau_n}(v_{it}, x_{it}, z_i),\ \mathbf{1}_{\tau_n,k} = \mathbf{1}_{\tau_n}(v_{kt}, x_{kt}, z_k),\\
\theta_j &= \theta(z_j),\ \theta_i = \theta(z_i),\ \theta_k = \theta(z_k),\\
m_i &= m(z_i) = T^{-1}\sum_{s=1}^{T} E[x_{is}x_{is}^\top|z_i]f_z(z_i),\ m_j = m(z_j),\ m_k = m(z_k).
\end{aligned}
$$

Proof of Theorem 3.2.1: 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

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

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.1.2, 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

$$\hat{f}_{t,vxz,j} = (nH)^{-1}\sum_{m=1}^{n}K_{h,vxz,mj} \quad \text{and} \quad \hat{f}_{t,xz,j} = (n\tilde{H})^{-1}\sum_{m=1}^{n}K_{\tilde{h},xz,mj},$$

we have

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

Then, we have that

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

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

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

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

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

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

$$\begin{aligned} B_{n1,2} + B_{n1,3} = & \; O_p\Big(\|h'\|^\nu\|h\|^\nu + \|h'\|^\nu\|\tilde{h}\|^\nu + n^{-1/2}\|h'\|^\nu + (n^{3/2}H'^{1/2}H^{1/2})^{-1} \\ & \times\|h'\|\|h\| + (n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|\tilde{h}\|\Big). \end{aligned}$$

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

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

From Hansen (2008), we have

$$
\begin{aligned}
\sup_{(v,x,z)\in\Omega_{vxz}}|\hat{f}_t(v,x,z)-f_t(v,x,z)| &= O_p(\|h\|^{\nu}+(\ln n)^{1/2}(nH)^{-1/2}), \\
\sup_{(x,z)\in P_{xz}(\Omega_{vxz})}|\hat{f}_t(x,z)-f_t(x,z)| &= O_p(\|\tilde{h}\|^{\nu}+(\ln n)^{1/2}(n\tilde{H})^{-1/2}),
\end{aligned}
$$

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

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

Hence,

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

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

$$\begin{aligned}&B_{n2,1} \\ &= \frac{1}{n}\sum_{i=1}^{n}m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n}x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{\hat{f}_{t,vxz,j}-f_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i} \\ &= \frac{1}{n}\sum_{i=1}^{n}m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n}x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\frac{(nH)^{-1}\sum_{k=1}^{n}K_{h,vxz,kj}-f_{t,vxz,j}}{f_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i} \\ &= (n^3TH'H)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{t=1}^{T}m_i^{-1}x_{jt}x_{jt}^{\top}\theta_i K_{h',z,ji}\left(K_{h,vxz,kj}-Hf_{t,vxz,j}\right)f_{t,vxz,j}^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n,i}\end{aligned}$$

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

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

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

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

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

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

By standard calculations, we have

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

Also, it is easy to see that

$$B_{n2,1,5,3} = O_p(n^{-1}), \text{ and } B_{n2,1,5,4} = O_p((n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|).$$

Hence, we have that

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

Similarly, we can show that

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

Similar as the derivation of $B_{n1,4}$, we have $B_{n2,3} = O_p(\|h\|^{2\nu} + (\ln n)(nH)^{-1} + \|h\|^\nu\|\tilde{h}\|^\nu + (\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Denote

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

By (B.1), we have

$$B_{n3} = \frac{1}{n}\sum_{i=1}^{n} m_i^{-1}A_{n5}(z_i)$$

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

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

$$
B_{n3,1,1} = (n^2TH')^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1}x_{it}\xi_{it}K_{h'}(0)\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_i),
$$

and

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

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

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

$$= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it}\xi_{it}\mathbf{1}_{\tau_n,i} + O_p(\|h'\|^{\nu+1}/\sqrt{n}). \tag{B.6}$$

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

$$\begin{aligned}
B_{n3,2} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it} E[\xi_{it}|x_{it}, z_i]\mathbf{1}_{\tau_n,i} + O_p\Big((n^2 H'\tilde{H})^{-1} + \|\tilde{h}\|^\nu (nH')^{-1} \\
&\quad + (n\tilde{H})^{-1} + \|\tilde{h}\|^\nu + (n^{3/2} H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|\Big) \\
&= O_p((n^2 H'\tilde{H})^{-1} + \|\tilde{h}\|^\nu (nH')^{-1} + (n\tilde{H})^{-1} + \|\tilde{h}\|^\nu + (n^{3/2} H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|), \\
B_{n3,3} &= -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it} E[\xi_{it}|v_i, x_{it}, z_i]\mathbf{1}_{\tau_n,i} + O_p\Big((n^2 H' H)^{-1} \\
&\quad + \|h\|^\nu (nH')^{-1} + (nH)^{-1} + \|h\|^\nu + n^{-1} + (n^{3/2} H'^{1/2} H^{1/2})^{-1}\|h\|\Big) \\
&= -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it}(E[y_{it}^*|v_i, x_{it}, z_i] - E[y_{it}^*|x_{it}, z_i])\mathbf{1}_{\tau_n,i} \\
&\quad + O_p\Big((n^2 H' H)^{-1} + \|h\|^\nu (nH')^{-1} + (nH)^{-1} + \|h\|^\nu + n^{-1} \\
&\quad + (n^{3/2} H'^{1/2} H^{1/2})^{-1}\|h\|\Big),
\end{aligned}$$

since $E[\xi_{it}|x_{it}, z_i] = 0$.

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

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

$P((v_{it},x_{it},z_i)\in\Omega_{vxz})$ 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 $Var(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}m_i^{-1}f_z(z_i)x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}) = Var(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}m_i^{-1}f_z(z_i)x_{it}\xi_{it}) + o(1)$.

Therefore, we have that

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

by the Lindeberg central limit theorem, where

$$V_{LC} = Var(g(z_i)) + T^{-2}Var\Big(\sum_{t=1}^{T}(m_i^{-1}f_z(z_i)x_{it}\xi_{it}$$

$$+m_i^{-1} f_z(z_i) x_{it}(E[y_{it}^*|v_i, x_{it}, z_i] - E[y_{it}^*|x_{it}, z_i]))\Big),$$

$$\begin{aligned}
\delta_n &= \sqrt{n}\|h'\|^{\nu} + \sqrt{n}(nH'^{1/2})^{-1}\|h'\| + \sqrt{n}n^{-1/2}\|h'\|^{\nu} \\
&+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|h\| + \sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|\tilde{h}\| \\
&+\sqrt{n}(\ln n)(nH)^{-1} + \sqrt{n}\|h\|^{\nu}\|\tilde{h}\|^{\nu} + \sqrt{n}(\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2} \\
&+\sqrt{n}(n^2H'H)^{-1} + \sqrt{n}\|h\|^{\nu}(nH')^{-1} + \sqrt{n}(nH)^{-1} + \sqrt{n}\|h\|^{\nu} + \sqrt{n}n^{-1} \\
&+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\| + \sqrt{n}(n^2H'\tilde{H})^{-1} + \sqrt{n}\|\tilde{h}\|^{\nu}(nH')^{-1} + \sqrt{n}(n\tilde{H})^{-1} \\
&+\sqrt{n}\|\tilde{h}\|^{\nu} + \sqrt{n}(n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\| + \sqrt{n}\|h'\|^{\nu+1}/\sqrt{n} + \sqrt{n}(nH'^{1/2})^{-1} \\
&+\sqrt{n}\eta_n = o_p(1),
\end{aligned}$$

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.1.2. *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* $m(z) = T^{-1}\sum_{s=1}^{T} E[x_{js}x_{js}^{\top}|z_j = z]f_z(z)$, *where* $K_{h',z_jz} = \prod_{l=1}^{q} k\left(\frac{z_{jl}-z_l}{h_l}\right)$, *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

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

which completes the proof of Lemma B.1.2.

APPENDIX C

Similar as Theorem 2.1 in Khan and Lewbel (2007), we can prove the following useful lemmas.

Lemma C.1.3. *Let $h(v_{it}, x_{it}, z_i, \epsilon_{it})$ be any function. If*

$$F_\epsilon^*(\epsilon_{it}|v_{it}, x_{it}, z_i) = F_\epsilon^*(\epsilon_{it}|x_{it}, z_i),$$

and the support of the random variable v_{it} is the interval $[L, K]$, then

$$E^*\left[\frac{h(v_{it}, x_{it}, z_i, \epsilon_{it})}{f_t^*(v_{it}|x_{it}, z_i)}\middle|x_{it}, z_i\right] = E^*\left[\int_L^K h(v_{it}, x_{it}, z_i, \epsilon_{it})dv_{it}\middle|x_{it}, z_i\right]. \qquad \text{(C.1)}$$

Proof of Lemma C.1.3: It is easy to see that

$$\begin{aligned}
E^*\left[\frac{h(v_{it}, x_{it}, z_i, \epsilon_{it})}{f_t^*(v_{it}|x_{it}, z_i)}\middle|x_{it}, z_i\right] &= E^*\left[\frac{E^*[h(v_{it}, x_{it}, z_i, \epsilon_{it})|v_{it}, x_{it}, z_i]}{f_t^*(v_{it}|x_{it}, z_i)}\middle|x_{it}, z_i\right] \\
&= \int_L^K \frac{E^*[h(v_{it}, x_{it}, z_i, \epsilon_{it})|v_{it}, x_{it}, z_i]}{f_t^*(v_{it}|x_{it}, z_i)} f_t^*(v_{it}|x_{it}, z_i)dv_{it} \\
&= \int_L^K E^*[h(v_{it}, x_{it}, z_i, \epsilon_{it})|v_{it}, x_{it}, z_i]dv_{it} \\
&= \int_L^K \int h(v_{it}, x_{it}, z_i, \epsilon_{it})dF_\epsilon^*(\epsilon_{it}|v_{it}, x_{it}, z_i)dv_{it} \\
&= \int_L^K \int h(v_{it}, x_{it}, z_i, \epsilon_{it})dF_\epsilon^*(\epsilon_{it}|x_{it}, z_i)dv_{it} \\
&= E^*\left[\int_L^K h(v_{it}, x_{it}, z_i, \epsilon_{it})dv_{it}\middle|x_{it}, z_i\right],
\end{aligned}$$

which completes the proof.

Lemma C.1.4. *Let Assumptions D1 to D4 hold. Let $H(y_{it}^*, x_{it}, z_i, \epsilon_{it})$ be any function that is differentiable in y_{it}^*. Let k be any constant that satisfies $0 \leq k \leq \tilde{k}$. Then*

$$\begin{aligned} & E^*\left[\frac{\partial H(y_{it}^*, x_{it}, z_i, \epsilon_{it})}{\partial y_{it}^*}\frac{\mathbf{1}(0 \leq y_{it}^* \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\Big|x_{it}, z_i\right] \\ = \; & E^*\left[\frac{H(k, x_{it}, z_i, \epsilon_{it}) - H(0, x_{it}, z_i, \epsilon_{it})}{|\gamma|}\Big|x_{it}, z_i\right]. \end{aligned} \tag{C.2}$$

Proof of Lemma C.1.4: By (C.1), we have that

$$\begin{aligned} & E^*\left[\frac{\partial H(y_{it}^*, x_{it}, z_{it}, \epsilon_{it})}{\partial y_{it}^*}\frac{\mathbf{1}(0 \leq y_{it}^* \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\Big|x_{it}, z_i\right] \\ = \; & E^*\left[\int_L^K \frac{\partial H[y_{it}^*(v_{it}, x_{it}, z_i, \epsilon_{it}), x_{it}, z_i, \epsilon_{it}]}{\partial y_{it}^*(v_{it}, x_{it}, z_i, \epsilon_{it})}\mathbf{1}(0 \leq y_{it}^*(v_{it}, x_{it}, z_i, \epsilon_{it}) \leq k)dv_{it}\Big|x_{it}, z_i\right] \\ = \; & \begin{cases} E^*\left[\int_{L\gamma + x_{it}^\top\theta(z_i)+\epsilon_{it}}^{K\gamma + x_{it}^\top\theta(z_i)+\epsilon_{it}} \frac{\partial H(y_{it}^*, x_{it}, z_i, \epsilon_{it})}{\partial y_{it}^*}\mathbf{1}(0 \leq y_{it}^* \leq k)dy_{it}^*/\gamma\Big|x_{it}, z_i\right] & \text{if } \gamma > 0, \\ -E^*\left[\int_{K\gamma + x_{it}^\top\theta(z_i)+\epsilon_{it}}^{L\gamma + x_{it}^\top\theta(z_i)+\epsilon_{it}} \frac{\partial H(y_{it}^*, x_{it}, z_i, \epsilon_{it})}{\partial y_{it}^*}\mathbf{1}(0 \leq y_{it}^* \leq k)dy_{it}^*/\gamma\Big|x_{it}, z_i\right] & \text{if } \gamma < 0. \end{cases} \end{aligned}$$

By Assumptions D1 and D3 and $0 < k \leq \tilde{k}$, we obtain that

$$\begin{aligned} & E^*\left[\frac{\partial H(y_{it}^*, x_{it}, z_i, \epsilon_{it})}{\partial y_{it}^*}\frac{\mathbf{1}(0 \leq y_{it}^* \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\Big|x_{it}, z_i\right] \\ = \; & E^*\left[\int_0^k \frac{\partial H(y_{it}^*, x_{it}, z_i, \epsilon_{it})}{\partial y_{it}^*}dy_{it}^*/|\gamma|\Big|x_{it}, z_i\right], \end{aligned} \tag{C.3}$$

which completes the proof.

Proof of Theorem 4.1.1: Since for any function $h(y_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})$

$$E[h(y_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y_{it} \leq k)|z_i] = \frac{E^*[h(y^*_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y^*_{it} \leq k)|z_i]}{P^*(y^*_{it} \geq 0|z_i)},$$

we have that

$$E[\mathbf{1}(0 \leq y_{it} \leq k)/f^*_t(v_{it}|x_{it}, z_i)|z_i] = \frac{k}{|\gamma|P^*(y^*_{it} \geq 0|z_i)} \tag{C.4}$$

by (C.3). Also, we have

$$\begin{aligned}
&\sum_{t=1}^{T} E\left[\frac{x_{it}(y_{it} - v_{it}\gamma_0)\mathbf{1}(0 \leq y_{it} \leq k)}{f^*_t(v_{it}|x_{it}, z_i)}\middle| z_i\right] \\
= &\sum_{t=1}^{T} \frac{E^*[x_{it}(y^*_{it} - v_{it}\gamma_0)\mathbf{1}(0 \leq y^*_{it} \leq k)/f^*_t(v_{it}|x_{it}, z_i)|z_i]}{P^*(y^*_{it} \geq 0|z_i)} \\
= &\sum_{t=1}^{T} \frac{E^*[x_{it}(x_{it}^\top\beta_i + \epsilon_{it})\mathbf{1}(0 \leq y^*_{it} \leq k)/f^*_t(v_{it}|x_{it}, z_i)|z_i]}{P^*(y^*_{it} \geq 0|z_i)} \\
= &\sum_{t=1}^{T} \frac{E^*[E^*[x_{it}(x_{it}^\top\beta_i + \epsilon_{it})\mathbf{1}(0 \leq y^*_{it} \leq k)/f^*_t(v_{it}|x_{it}, z_i)|x_{it}, z_i]|z_i]}{P^*(y^*_{it} \geq 0|z_i)} \\
= &\sum_{t=1}^{T} \frac{k(E^*[x_{it}x_{it}^\top|z_i]\theta(z_i) + E^*[x_{it}\epsilon_{it}|z_i])}{|\gamma|P^*(y^*_{it} \geq 0|z_i)}.
\end{aligned}$$

Hence, by Assumption A4 and $E^*(\epsilon_{it}|x_{it}, z_i) = 0$, we have that

$$\sum_{t=1}^{T} E[x_{it}\tilde{y}_{it}|z_i] = (\sum_{t=1}^{T} E^*[x_{it}x_{it}^\top|z_i])\theta(z_i).$$

Therefore, we get that

$$\theta(z_i) = \left(\sum_{t=1}^{T} E^*[x_{it} x_{it}^\top | z_i] \right)^{-1} \sum_{t=1}^{T} E[x_{it} \tilde{y}_{it} | z_i].$$

Proof of Theorem 4.1.2: Since $v_{it}\mathbf{1}(0 \leq y_{it} \leq k) = \gamma^{-1}(y_{it}^* - x_{it}^\top \beta_i - u_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)$ and

$$E[h(y_{it}, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y_{it} \leq k)] = \frac{E^*[h(y_{it}^*, x_{it}, v_{it}, z_i, \epsilon_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)]}{P^*(y_{it}^* \geq 0)},$$

we have that

$$\begin{aligned}
& E[v_{it}\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)] \\
= \; & E[\gamma^{-1}(y_{it}^* - x_{it}^\top \beta_i - u_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)/f_t^*(v_{it}|x_{it}, z_i)] \\
= \; & \frac{E^*[\gamma^{-1}(y_{it}^* - x_{it}^\top \beta_i - u_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)/f_t^*(v_{it}|x_{it}, z_i)]}{P^*(y_{it}^* \geq 0)} \\
= \; & \frac{E^*[\gamma^{-1} y_{it}^* \mathbf{1}(0 \leq y_{it}^* \leq k)/f_t^*(v_{it}|x_{it}, z_i)]}{P^*(y_{it}^* \geq 0)} \\
& + \frac{E^*[\gamma^{-1}(x_{it}^\top \beta_i - u_{it})\mathbf{1}(0 \leq y_{it}^* \leq k)/f_t^*(v_{it}|x_{it}, z_i)]}{P^*(y_{it}^* \geq 0)} \\
= \; & \left(\frac{k^2}{2\gamma|\gamma|} - \frac{kE^*[x_{it}^\top \beta_i - u_{it}]}{|\gamma|} \right) / P^*(\tilde{y}_{it} \geq 0).
\end{aligned}$$

Also, we have that

$$E[\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)] = \frac{k}{|\gamma| P^*(y_{it}^* \geq 0)}.$$

Therefore, we have that

$$\zeta(k) = \frac{k}{\gamma} - \frac{1}{T}\sum_{t=1}^{T} 2E^*[x_{it}^\top \beta_i - u_{it}].$$

Hence, we have that

$$\gamma = \frac{k - k'}{\zeta(k) - \zeta(k')}.$$

We give some shorthand notations first. These notations will be used throughout the proof of Theorem 4.2.1. Define

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

$$\begin{aligned}
K^*_{h,vxz,ij} &= K_h(v^*_{it} - v_{jt}, x^*_{it} - x_{jt}, z^*_i - z_j),\\
K^*_{h,vxz,jk} &= K_h(v^*_{jt} - v_{kt}, x^*_{jt} - x_{kt}, z^*_j - z_k),\\
K^*_{h,vxz,ik} &= K_h(v^*_{it} - v_{kt}, x^*_{it} - x_{kt}, z^*_i - z_k),\\
K^*_{h,vxz,ji} &= K_h(v^*_{jt} - v_{it}, x^*_{jt} - x_{it}, z^*_j - z_i),\\
K^*_{h,vxz,mi} &= K_h(v^*_{mt} - v_{it}, x^*_{mt} - x_{it}, z^*_m - z_i),\ K^*_{\tilde{h},xz,mi} = K_{\tilde{h}}(x^*_{mt} - x_{it}, z^*_m - z_i),\\
K^*_{h,vxz,mj} &= K_h(v^*_{mt} - v_{jt}, x^*_{mt} - x_{jt}, z^*_m - z_j),\ K^*_{\tilde{h},xz,mj} = K_{\tilde{h}}(x^*_{mt} - x_{jt}, z^*_m - z_j).
\end{aligned}$$

Proof of Theorem 4.2.1: Since $\hat{f}^*_{t,v|xz,i} = \frac{\hat{f}^*_{t,vxz,i}}{\hat{f}^*_{t,xz,i}}$, where

$$\begin{aligned}
\hat{f}^*_{t,vxz,i} &= (n^*H)^{-1}\sum_{m=1}^{n^*} K^*_{h,vxz,mi},\\
\hat{f}^*_{t,xz,i} &= (n^*\tilde{H})^{-1}\sum_{m=1}^{n^*} K^*_{\tilde{h},xz,mi},
\end{aligned}$$

we have

$$\begin{aligned}
\frac{f^*_{t,v|xz,i}}{\hat{f}^*_{t,v|xz,i}} &= 1 + f^*_{t,v|xz,i}\left(\frac{1}{\hat{f}^*_{t,v|xz,i}} - \frac{1}{f^*_{t,v|xz,i}}\right)\\
&= 1 + \frac{\hat{f}^*_{t,xz,i} - f^*_{t,xz,i}}{f^*_{t,xz,i}} + \frac{f^*_{t,vxz,i} - \hat{f}^*_{t,vxz,i}}{f^*_{t,vxz,i}}\\
&\quad + \frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i} - \hat{f}^*_{t,vxz,i}f^*_{t,xz,i})(f^*_{t,vxz,i} - \hat{f}^*_{t,vxz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}\hat{f}^*_{t,vxz,i}}. \qquad \text{(C.5)}
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\mu}_t(k) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_t^*(v_{it}|x_{it},z_i)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i}\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} + \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_{t,vxz,i}^*-\hat{f}_{t,vxz,i}^*}{f_{t,vxz,i}^*}\mathbf{1}_{\tau_n,i}\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{\hat{f}_{t,xz,i}^*-f_{t,xz,i}^*}{f_{t,xz,i}^*}\mathbf{1}_{\tau_n,i}\\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{(f_{t,vxz,i}^*\hat{f}_{t,xz,i}^*-\hat{f}_{t,vxz,i}^*f_{t,xz,i}^*)(f_{t,vxz,i}^*-\hat{f}_{t,vxz,i}^*)}{f_{t,xz,i}^*f_{t,vxz,i}^*\hat{f}_{t,vxz,i}^*}\mathbf{1}_{\tau_n,i}\\
&\equiv \mu_{t1}(k)+\mu_{t2}(k)+\mu_{t3}(k)+\mu_{t4}(k).
\end{aligned}$$

Since $\|h'\|/\tau_n \to 0$ and the kernel function $K(\cdot)$ has a compact support, the trimming function $\mathbf{1}_{\tau_n,i}$ will ensure that all of the points which have boundary effects are excluded from our estimated locations. By Lindeberg's central limit theorem, we have $\mu_{t1}(k) - E[\mathbf{1}(0\le y_{it}\le k)/f_t^*(v_{it}|x_{it},z_i)] = O_p(n^{-1/2})$.

We can see that $\mu_{t2}(k)$ and $\mu_{t3}(k)$ can be written as a second-order U-statistics. By similar argument as in proving (A.32) and (A.33) in Khan and Lewbel (2007), we have that

$$\begin{aligned}
\mu_{t2}(k) &= -\frac{n}{n^*}\frac{1}{n}\sum_{i=1}^{n}E[\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it}=v_{it}^*,x_{it}=x_{it}^*,z_i=z_i^*]\\
&\quad +E[\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}]+o_p(n^{-1/2}),\\
\mu_{t3}(k) &= \frac{n}{n^*}\frac{1}{n}\sum_{i=1}^{n}E[\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it}=x_{it}^*,z_i=z_i^*]-E[\frac{\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}]\\
&\quad +o_p(n^{-1/2}).
\end{aligned}$$

For $\mu_{t4}(k)$, we have

$$\begin{aligned} & E(\|\mu_{t4}(k)\|) \\ \leq\ & E\left(\|\frac{\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i} - \hat{f}^*_{t,vxz,i}f^*_{t,xz,i})(f^*_{t,vxz,i} - \hat{f}^*_{t,vxz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}\hat{f}^*_{t,vxz,i}}\mathbf{1}_{\tau_n,i}\|\right) \\ \leq\ & E\left(\|\frac{\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\| \left|\frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i} - \hat{f}^*_{t,vxz,i}f^*_{t,xz,i})(f^*_{t,vxz,i} - \hat{f}^*_{t,vxz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}\hat{f}^*_{t,vxz,i}}\mathbf{1}_{\tau_n,i}\right|\right). \end{aligned} \tag{C.6}$$

From Hansen (2008), we have

$$\begin{aligned} \sup_{(v,x,z)\in\Omega_{vxz}} |\hat{f}_t^*(v,x,z) - f_t^*(v,x,z)| &= O_p(\|h\|^\nu + (\ln n^*)^{1/2}(n^*H)^{-1/2}), \\ \sup_{(x,z)\in P_{xz}(\Omega_{vxz})} |\hat{f}_t^*(x,z) - f_t^*(x,z)| &= O_p(\|\tilde{h}\|^\nu + (\ln n^*)^{1/2}(n^*\tilde{H})^{-1/2}), \end{aligned}$$

where $P_{xz}(\,\cdot\,)$ is the projection of Cartesian product. Hence, we have that $\mu_{t4}(k) = O_p(\|h\|^{2\nu} + (\ln n^*)(n^*H)^{-1} + \|h\|^\nu\|\tilde{h}\|^\nu + (\ln n^*)(n^*)^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Thus, we have that $\hat{\mu}_t(k) - \mu_t(k) = O_p(n^{-1/2})$.

Since

$$\frac{1}{\hat{\mu}_t(k)} = \frac{1}{\mu_t(k)} - \frac{\hat{\mu}_t(k) - \mu_t(k)}{\mu_t(k)^2} + \frac{(\hat{\mu}_t(k) - \mu_t(k))^2}{\hat{\mu}_t(k)\mu_t(k)^2}, \tag{C.7}$$

we have that

$$\hat{\zeta}(k) = \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{\hat{f}_t^*(v_{it}|x_{it}, z_i)}\mathbf{1}_{\tau_n,i}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{(\hat{\mu}_t(k)-\mu_t(k))^2}{\hat{\mu}_t(k)\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&\equiv \zeta_1(k)+\zeta_2(k)+\zeta_3(k),
\end{aligned}
$$

by (C.5), we have that

$$
\begin{aligned}
&\zeta_1(k) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_t^*(v_{it}|x_{it},z_i)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_{t,vxz,i}^*-\hat{f}_{t,vxz,i}^*}{f_{t,vxz,i}^*}\mathbf{1}_{\tau_n,i} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{\hat{f}_{t,xz,i}^*-f_{t,xz,i}^*}{f_{t,xz,i}^*}\mathbf{1}_{\tau_n,i} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\le y_{it}\le k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{(f_{t,vxz,i}^*\hat{f}_{t,xz,i}^*-\hat{f}_{t,vxz,i}^*f_{t,xz,i}^*)}{f_{t,xz,i}^*f_{t,vxz,i}^*} \\
&\quad \times\frac{(f_{t,vxz,i}^*-\hat{f}_{t,vxz,i}^*)}{\hat{f}_{t,vxz,i}^*}\mathbf{1}_{\tau_n,i} \\
&= \zeta_{1,1}(k)+\zeta_{1,2}(k)+\zeta_{1,3}(k)+\zeta_{1,4}(k).
\end{aligned}
$$

By Lindeberg's central limit theorem and the same argument for the trimming function as in the previous proof, we have $\zeta_{1,1}(k) - T^{-1}\sum_{t=1}^{T}\mu_t(k)^{-1}E[2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)/f_t^*(v_{it}|x_{it}, z_i)] = O_p(n^{-1/2})$.

For $\zeta_{1,2}(k)$, we have that

$$\begin{aligned}
\zeta_{1,2}(k) &= \frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}^* - \hat{f}_{t,vxz,i}^*}{f_{t,vxz,i}^*}\mathbf{1}_{\tau_n,i} \\
&= -\frac{1}{T}\sum_{t=1}^{T}\mu_t(k)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{(n^*H)^{-1}\sum_{j=1}^{n^*}K_{h,vxz,ji}^* - f_{t,vxz,i}^*}{f_{t,vxz,i}^*} \\
&\quad \times\mathbf{1}_{\tau_n,i} \\
&= -(nTn^*H)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n*}\mu_t(k)^{-1}\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}(K_{h,vxz,ji}^* - Hf_{t,vxz,i}^*) \\
&\quad \times(f_{t,vxz,i}^*)^{-1}\mathbf{1}_{\tau_n,i}.
\end{aligned}$$

$\zeta_{1,2}(k)$ can be written as a second-order U-statistics. By the similar argument as in proving (A.32) and (A.33) in Khan and Lewbel (2007), we have that

$$\begin{aligned}
&\zeta_{1,2}(k) \\
= &-\frac{1}{n}\sum_{i=1}^{n}T^{-1}\sum_{t=1}^{T}\mu_t(k)^{-1}\Big(E[\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&-E[\frac{2v_{it}\mathbf{1}(0 \leq y_{it} \leq k)}{f_t^*(v_{it}|x_{it}, z_i)}]\Big) + o_p(n^{-1/2}).
\end{aligned}$$

Similarly, we have that

$$\zeta_{1,3}(k) = \frac{1}{n}\sum_{i=1}^{n} T^{-1}\sum_{t=1}^{T}\mu_t(k)^{-1}\Big(E[\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f_{t,xz,i}}{f^*_{t,xz,i}}|x_{it}=x^*_{it},z_i=z_i^*] - E[\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}]\Big) + o_p(n^{-1/2}).$$

For $\zeta_{1,4}(k)$, we have

$$\begin{aligned} & E(\|\zeta_{1,4}(k)\|) \\ \leq\ & E\left(\|\mu(k)^{-1}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i}-\hat{f}^*_{t,vxz,i}f^*_{t,xz,i})(f^*_{t,vxz,i}-\hat{f}^*_{t,vxz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}\hat{f}^*_{t,vxz,i}}\mathbf{1}_{\tau_n,i}\|\right) \\ \leq\ & E\left(\|\mu(k)^{-1}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\|\left|\frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i}-\hat{f}^*_{t,vxz,i}f^*_{t,xz,i})(f^*_{t,vxz,i}-\hat{f}^*_{t,vxz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}\hat{f}^*_{t,vxz,i}}\mathbf{1}_{\tau_n,i}\right|\right). \end{aligned}$$

Similar as (C.6), we have that $\zeta_{1,4}(k) = O_p(\|h\|^{2\nu} + (\ln n^*)(n^*H)^{-1} + \|h\|^{\nu}\|\tilde{h}\|^{\nu} + (\ln n^*)(n^*)^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

For $\zeta_2(k)$, we have

$$\begin{aligned} \zeta_2(k) &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{\hat{f}_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\ &= -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\mathbf{1}_{\tau_n,i} \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{\hat{f}^*_{t,xz,i}-f^*_{t,xz,i}}{f^*_{t,xz,i}}\mathbf{1}_{\tau_n,i} \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{f^*_{t,vxz,i}-\hat{f}^*_{t,vxz,i}}{f^*_{t,vxz,i}}\mathbf{1}_{\tau_n,i} \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{\mu}_t(k)-\mu_t(k)}{\mu_t(k)^2}\frac{1}{n}\sum_{i=1}^{n}\frac{2v_{it}\mathbf{1}(0\leq y_{it}\leq k)}{f_t^*(v_{it}|x_{it},z_i)}\frac{(f^*_{t,vxz,i}\hat{f}^*_{t,xz,i}-\hat{f}^*_{t,vxz,i}f^*_{t,xz,i})}{f^*_{t,xz,i}f^*_{t,vxz,i}} \end{aligned}$$

$$\times \frac{(f^*_{t,vxz,i} - \hat{f}^*_{t,vxz,i})}{\hat{f}^*_{t,vxz,i}} \mathbf{1}_{\tau_n,i}$$
$$\equiv \zeta_{2,1}(k) + \zeta_{2,2}(k) + \zeta_{2,3}(k) + \zeta_{2,4}(k).$$

Hence, we have

$$\begin{aligned}
& \zeta_{2,1}(k) \\
= & -\frac{1}{T}\sum_{t=1}^{T} \frac{\hat{\mu}_t(k) - \mu_t(k)}{\mu_t(k)^2} \frac{1}{n}\sum_{i=1}^{n} \frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f^*_t(v_{it}|x_{it}, z_i)} \mathbf{1}_{\tau_n,i} \\
= & -\frac{1}{T}\sum_{t=1}^{T} \frac{\hat{\mu}_t(k) - \mu_t(k)}{\mu_t(k)^2} E[2v_{it}\mathbf{1}(0 \le y_{it} \le k)/f^*_t(v_{it}|x_{it}, z_i)] \\
& -\frac{1}{T}\sum_{t=1}^{T} \frac{\hat{\mu}_t(k) - \mu_t(k)}{\mu_t(k)^2} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f^*_t(v_{it}|x_{it}, z_i)} \mathbf{1}_{\tau_n,i} - E[\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f^*_t(v_{it}|x_{it}, z_i)}] \right) \\
= & -\frac{1}{T}\sum_{t=1}^{T} \frac{\hat{\mu}_t(k) - \mu_t(k)}{\mu_t(k)^2} E[2v_{it}\mathbf{1}(0 \le y_{it} \le k)/f^*_t(v_{it}|x_{it}, z_i)] + O_p(n^{-1}) \\
= & -\frac{1}{T}\sum_{t=1}^{T} \frac{\mu_{t,1}(k) - \mu_t(k) + \mu_{t,2}(k) + \mu_{t,3}(k)}{\mu_t(k)^2} E[2v_{it}\mathbf{1}(0 \le y_{it} \le k)/f^*_t(v_{it}|x_{it}, z_i)] \\
& +O_p(n^{-1}).
\end{aligned}$$

Also, we have $\zeta_{2,2}(k) = O_p(n^{-1})$, and $\zeta_{2,3}(k) = O_p(n^{-1})$. Since $\sup_{1\le t\le T} \|\hat{\mu}_t(k) - \mu_t(k)\| = O_p(n^{-1/2})$, similar as (C.6) we have $\zeta_{2,4}(k) = o_p(n^{-1/2})$. It is easy to see that $\zeta_3(k) = o_p(n^{-1/2})$.

Hence, we have $\hat{\zeta}(k) - \zeta(k) = O_p(n^{-1/2})$.

Next, we have that

$$
\begin{aligned}
\hat{\gamma} &= \frac{k-k'}{\hat{\zeta}(k)-\hat{\zeta}(k')} \\
&= \frac{k-k'}{\zeta(k)-\zeta(k')} - (k-k')\frac{\hat{\zeta}(k)-\hat{\zeta}(k')-(\zeta(k)-\zeta(k'))}{(\zeta(k)-\zeta(k'))^2} \\
&\quad +(k-k')\frac{\left(\hat{\zeta}(k)-\hat{\zeta}(k')-(\zeta(k)-\zeta(k'))\right)^2}{(\hat{\zeta}(k)-\hat{\zeta}(k'))(\zeta(k)-\zeta(k'))^2} \\
&= \gamma - (k-k')\frac{\hat{\zeta}(k)-\hat{\zeta}(k')-(\zeta(k)-\zeta(k'))}{(\zeta(k)-\zeta(k'))^2} \\
&\quad +(k-k')\frac{\left(\hat{\zeta}(k)-\hat{\zeta}(k')-(\zeta(k)-\zeta(k'))\right)^2}{(\hat{\zeta}(k)-\hat{\zeta}(k'))(\zeta(k)-\zeta(k'))^2}
\end{aligned}
$$

by Theorem 4.1.2. Hence, by Lindeberg's central limit theorem we obtain that

$$
\begin{aligned}
&\sqrt{n}(\hat{\gamma}-\gamma) \\
&= -\sqrt{n}(k-k')\frac{\hat{\zeta}(k)-\hat{\zeta}(k')-(\zeta(k)-\zeta(k'))}{(\zeta(k)-\zeta(k'))^2} + o_p(1) \\
&= \sqrt{n}\frac{\gamma^2}{k-k'}[(\zeta_{1,1}(k')-\zeta(k')+\zeta_{1,2}(k')+\zeta_{1,3}(k')+\zeta_{2,1}(k')) \\
&\quad -(\zeta_{1,1}(k)-\zeta(k)+\zeta_{1,2}(k)+\zeta_{1,3}(k)+\zeta_{2,1}(k))] + o_p(1) \\
&\xrightarrow{d} N(0, V_\gamma),
\end{aligned}
$$

where $V_\gamma = E[\psi_t(k)^2]$,

$$
\begin{aligned}
\psi_t(k) &= \frac{\gamma^2}{k-k'}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\mu_t(k)^{-1}\varphi_k(k) - \phi_t(k)\mu_t(k)^{-2}\eta_t(k) + \mu_t(k')^{-1}\varphi_t(k') \\
&\quad -\phi_t(k')\mu_t(k')^{-2}\eta_t(k')\Big)\Big],
\end{aligned}
$$

$$
\begin{aligned}
\varphi_t(k) &= \frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \eta_t(k) \\
&\quad -cE[\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad +cE[\frac{2v_{it}\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*], \\
\phi_t(k) &= \frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \mu_t(k) \\
&\quad -cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad +cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} \frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*], \\
\eta_t(k) &= E[2v_{it}\mathbf{1}(0 \le y_{it} \le k)/f_t^*(v_{it}|x_{it}, z_i)].
\end{aligned}
$$

This completes the proof of the first part of Theorem 4.2.1.

Next, we prove the second part of Theorem 4.2.1.

For $z \in \Omega_z$, let

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

Recall that

$$
\begin{aligned}
\tilde{y}_{jt} &= \frac{(y_{jt} - v_{jt}\gamma)\mathbf{1}(0 \le y_{jt} \le k)/f_t^*(v_{jt}|x_{jt}, z_j)}{E[\mathbf{1}(0 \le y_{jt} \le k)/f_t^*(v_{jt}|x_{jt}, z_j)|z_j]} \\
&= \frac{(y_{jt} - v_{jt}\gamma)\mathbf{1}(0 \le y_{jt} \le k)/f_t^*(v_{jt}|x_{jt}, z_j)}{\mu_t(k, z_j)}.
\end{aligned}
$$

Using (C.5) and an equality similar as (C.7), we have that

$$
\begin{aligned}
& \hat{\theta}_{LC}(z) \\
= \;& A_{n1}(z)^{-1}A_{n2}(z) \\
= \;& \theta(z) + \Big(A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} \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) - \theta(z)\Big) \\
& + A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}(\tilde{y}_{jt} - E(\tilde{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) \\
& - A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\tilde{y}_{jt}\frac{\hat{\mu}_t(k, z_j) - \mu_t(k, z_j)}{\mu_t(k, 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) \\
& - A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{v_{jt}\mathbf{1}(0 \leq y_{jt} \leq k)(\hat{\gamma} - \gamma)}{\mu_t(k, z_j)f_{t,v|xz,j}^*} \\
& \times 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) \\
& + A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\tilde{y}_{jt}\frac{(\hat{\mu}_t(k, z_j) - \mu_t(k, z_j))^2}{\hat{\mu}_t(k, z_j)\mu_t(k, z_j)} \\
& \times 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) \\
& + A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{v_{jt}\mathbf{1}(0 \leq y_{jt} \leq k)(\hat{\gamma} - \gamma)(\hat{\mu}_t(k, z_j) - \mu_t(k, z_j))}{\mu_t(k, z_j)^2 f_{t,v|xz,j}^*} \\
& \times 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) \\
& - A_{n1}(z)^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{v_{jt}\mathbf{1}(0 \leq y_{jt} \leq k)(\hat{\gamma} - \gamma)(\hat{\mu}_t(k, z_j) - \mu_t(k, z_j))^2}{\hat{\mu}_t(k, z_j)\mu_t(k, z_j)^2 f_{t,v|xz,j}^*} \\
& \times 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_{n3}(z) + A_{n1}(z)^{-1}A_{n4}(z) + A_{n1}(z)^{-1}A_{n5}(z) + A_{n1}(z)^{-1}A_{n6}(z) + A_{n7}(z).
\end{aligned}
$$

By Lemma C.1.5 we have uniformly in $z \in \Omega_z$,

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

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

Let $m_i = m(z_i^*)$. Then, we have that

$$\begin{aligned}\hat{\beta}_{LC} &= \frac{1}{n^*}\sum_{i=1}^{n^*}\hat{\theta}_{LC}(z_i^*) \\ &= \beta + \frac{1}{n^*}\sum_{i=1}^{n^*} g(z_i^*) + \frac{1}{n^*}\sum_{i=1}^{n^*} A_{n3}(z_i^*) + \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}\Big[A_{n4}(z_i^*) + A_{n5}(z_i^*) \\ &\quad + A_{n6}(z_i^*) + A_{n7}(z_i^*)\Big] + \frac{1}{n^*}\sum_{i=1}^{n^*} A_{n8}(z_i^*) + \eta_n,\end{aligned}$$

where $\eta_n = O_p\left(||h'||^{\nu} + (\ln n^*)^{1/2}(n^*H')^{-1/2}\right) O_p(\|A_{n4}(z_i^*)\| + \|A_{n5}(z_i^*)\| + \|A_{n6}(z_i^*)\| + \|A_{n7}(z_i^*)\|)$.

Since $\hat{f}^*_{t,v|xz,j} = \frac{\hat{f}^*_{t,vxz,j}}{\hat{f}^*_{t,xz,j}}$, where

$$\hat{f}^*_{t,vxz,j} = (n^*H)^{-1}\sum_{m=1}^{n^*} K^*_{h,vxz,mj} \quad \text{and} \quad \hat{f}^*_{t,xz,j} = (n^*\tilde{H})^{-1}\sum_{m=1}^{n^*} K^*_{\tilde{h},xz,mj},$$

we have

$$\frac{f^*_{t,v|xz,j}}{\hat{f}^*_{t,v|xz,j}} = 1 + f^*_{t,v|xz,j}\left(\frac{1}{\hat{f}^*_{t,v|xz,j}} - \frac{1}{f^*_{t,v|xz,j}}\right)$$

$$= 1 + \frac{\hat{f}^*_{t,xz,j} - f^*_{t,xz,j}}{f^*_{t,xz,j}} + \frac{f^*_{t,vxz,j} - \hat{f}^*_{t,vxz,j}}{f^*_{t,vxz,j}} + \frac{(f^*_{t,vxz,j}\hat{f}^*_{t,xz,j} - \hat{f}^*_{t,vxz,j}f^*_{t,xz,j})}{f^*_{t,xz,j}f^*_{t,vxz,j}} \times \frac{(f^*_{t,vxz,j} - \hat{f}^*_{t,vxz,j})}{\hat{f}^*_{t,vxz,j}}. \quad \text{(C.8)}$$

Then, we have that

$$\begin{aligned}
B_{n1} &= \frac{1}{n^*}\sum_{i=1}^{n^*} A_{n3}(z_i^*) \\
&= \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}T^{-1}\sum_{t=1}^{T}\left(n^{-1}\sum_{j=1}^{n} x_{jt}x_{jt}^\top\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)}(H')^{-1}K_{h',z,ji}\mathbf{1}_{\tau_n,j}\right. \\
&\quad \left. -(n^*)^{-1}\sum_{j=1}^{n^*} x_{jt}^*(x_{jt}^*)^\top\theta_i(H')^{-1}K_{h',z^*,ji}\mathbf{1}^*_{\tau_n,j}\right)\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&\quad +\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^\top\theta_j\frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)}K_{h',z,ji} \\
&\quad \times\frac{\hat{f}^*_{t,xz,j} - f^*_{t,xz,j}}{f^*_{t,xz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&\quad +\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^\top\theta_j\frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)}K_{h',z,ji} \\
&\quad \times\frac{f^*_{t,vxz,j} - \hat{f}^*_{t,vxz,j}}{f^*_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&\quad +\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^\top\theta_j\frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)}K_{h',z,ji} \\
&\quad \times\frac{(f^*_{t,vxz,j}\hat{f}^*_{t,xz,j} - \hat{f}^*_{t,vxz,j}f^*_{t,xz,j})}{f^*_{t,xz,j}f^*_{t,vxz,j}}\frac{(f^*_{t,vxz,j} - \hat{f}^*_{t,vxz,j})}{\hat{f}^*_{t,vxz,j}}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&\quad +O_p\left((||h'||^\nu + (\ln n^*)^{1/2}(n^*H')^{-1/2})^2\right) \\
&\equiv B_{n1,1} + B_{n1,2} + B_{n1,3} + B_{n1,4}.
\end{aligned}$$

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

$$\begin{aligned}B_{n1,1} &= \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}T^{-1}\sum_{t=1}^{T}\Bigg(n^{-1}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}(H')^{-1}K_{h',z,ji}\mathbf{1}_{\tau_n,j}\\ &\quad -(n^*)^{-1}\sum_{j=1}^{n^*} x_{jt}^*(x_{jt}^*)^{\top}\theta_i(H')^{-1}K_{h',z^*,ji}\mathbf{1}_{\tau_n,j}^*\Bigg)\mathbf{1}_{\varepsilon_n}(z_i^*).\end{aligned}$$

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

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

where

$$\begin{aligned}H_{n1,ij} &= (TH')^{-1}\sum_{t=1}^{T}[m_i^{-1}x_{jt}x_{jt}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}\mathbf{1}(i\leq n)K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)\\ &\quad -\frac{n}{n^*}m_i^{-1}x_{jt}^*(x_{jt}^*)^{\top}\theta_iK_{h',z^*,ji}\mathbf{1}_{\tau_n,j}^*\mathbf{1}_{\varepsilon_n}(z_i^*) + m_j^{-1}x_{it}x_{it}^{\top}\theta_i\frac{P^*(y_{it}^*\geq 0|z_i)}{P^*(y_{it}^*\geq 0|x_{it},z_i)}\\ &\quad \times\mathbf{1}(j\leq n)K_{h',z,ji}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j^*) - \frac{n}{n^*}m_j^{-1}x_{it}^*(x_{it}^*)^{\top}\theta_jK_{h',z^*,ji}\mathbf{1}_{\tau_n,i}^*\mathbf{1}_{\varepsilon_n}(z_j^*).\end{aligned}$$

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

$$E[H_{n1,ij}] = O_p(\|h'\|^{\nu}).$$

Also, we have

$$\begin{aligned} & E\left[\left(\frac{1}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right)\left(\frac{1}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right)^{\top}\right] \\ = \;& Var\left[\frac{1}{n}\sum_{i=1}^{n}[H_{n1,i}-E(H_{n1,i})]\right] \\ = \;& O(n^{-1}\|h'\|^{2\nu}), \end{aligned} \tag{C.9}$$

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

$$E\left[\left(\frac{1}{n^*}\sum_{i=1}^{n^*}[H^*_{n1,i}-E(H^*_{n1,i})]\right)\left(\frac{1}{n^*}\sum_{i=1}^{n^*}[H^*_{n1,i}-E(H^*_{n1,i})]\right)^{\top}\right] = O((n^*)^{-1}\|h'\|^{2\nu}),$$

where $H^*_{n1,i} = E[H_{n1,ij}|w_i^*]$, $w_i^* = ((x_{i1}^*)^{\top}, \ldots, (x_{iT}^*)^{\top}, z_i^*)$, and

$$\begin{aligned} & Var\left[\frac{2}{n^*(n^*-1)}\sum_{i=1}^{n^*}\sum_{j>i}^{n^*}\left[H_{n1,ij}-H_{n1,i}-H_{n1,j}-H^*_{n1,i}-H^*_{n1,j}+E(H_{n1,ij})\right]\right] \\ = \;& \frac{4}{(n^*)^2(n^*-1)^2}\sum_{i=1}^{n^*}\sum_{j>i}^{n^*} Var\left[H_{n1,ij}-H_{n1,i}-H_{n1,j}+E(H_{n1,ij})\right] \\ = \;& O((n^*)^{-2}H'^{-1}\|h'\|^2). \end{aligned} \tag{C.10}$$

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

Let

$$B_{n2} \;=\; B_{n1,2} + B_{n1,3} + B_{n1,4}.$$

Hence,

$$\begin{aligned} B_{n2} &= \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji}\frac{\hat{f}_{t,vxz,j}^* - f_{t,vxz,j}^*}{f_{t,vxz,j}^*} \\ &\quad \times \mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &\quad + \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji}\frac{f_{t,xz,j}^* - \hat{f}_{t,xz,j}^*}{f_{t,xz,j}^*} \\ &\quad \times \mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &\quad + \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji} \\ &\quad \times \frac{(\hat{f}_{t,vxz,j}^* f_{t,xz,j}^* - f_{t,vxz,j}^*\hat{f}_{t,xz,j}^*)}{f_{t,xz,j}^* f_{t,vxz,j}^*}\frac{(f_{t,vxz,j}^* - \hat{f}_{t,vxz,j}^*)}{\hat{f}_{t,vxz,j}^*}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &\equiv B_{n2,1} + B_{n2,2} + B_{n2,3}. \end{aligned}$$

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

$$\begin{aligned} B_{n2,1} &= \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji} \\ &\quad \times \frac{\hat{f}_{t,vxz,j}^* - f_{t,vxz,j}^*}{f_{t,vxz,j}^*}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &= \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji} \\ &\quad \times \frac{(n^*H)^{-1}\sum_{k=1}^{n^*} K_{h,vxz,kj}^* - f_{t,vxz,j}^*}{f_{t,vxz,j}^*}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &= (n(n^*)^2TH'H)^{-1}\sum_{i=1}^{n^*}\sum_{j=1}^{n}\sum_{k=1}^{n^*}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ji} \\ &\quad \times \left(K_{h,vxz,kj}^* - Hf_{t,vxz,j}^*\right)(f_{t,vxz,j}^*)^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\ &= (n(n^*)^2TH'H)^{-1}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1}x_{it}x_{it}^{\top}\theta_j \frac{P^*(y_{jt}^* \geq 0|z_j)}{P^*(y_{jt}^* \geq 0|x_{jt}, z_j)} K_{h',z,ii} \\ &\quad \times \left(K_h^*(0) - Hf_{t,vxz,i}^*\right)(f_{t,vzx,i}^*)^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \end{aligned}$$

$$
\begin{aligned}
&+(n(n^*)^2TH'H)^{-1}\sum_{i=1}^{n^*}\sum_{k\neq i}^{n^*}\sum_{t=1}^{T} m_i^{-1}x_{it}x_{it}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}K_{h',z,ii}\\
&\times\left(K^*_{h,vxz,ki}-Hf^*_{t,vxz,i}\right)(f^*_{t,vzx,i})^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)\\
&+(n(n^*)^2TH'H)^{-1}\sum_{i=1}^{n^*}\sum_{j\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}K_{h',z,ji}\\
&\times\left(K^*_{h}(0)-Hf^*_{t,vxz,j}\right)(f^*_{t,vxz,j})^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)\\
&+(n(n^*)^2TH'H)^{-1}\sum_{i=1}^{n^*}\sum_{j\neq i}^{n}\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}K_{h',z,ji}\\
&\times\left(K^*_{h,vxz,ij}-Hf^*_{t,vxz,j}\right)(f^*_{t,vxz,j})^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)\\
&+(n(n^*)^2TH'H)^{-1}\sum\sum_{i\neq j\neq k}\sum\sum_{t=1}^{T} m_i^{-1}x_{jt}x_{jt}^{\top}\theta_j\frac{P^*(y_{jt}^*\geq 0|z_j)}{P^*(y_{jt}^*\geq 0|x_{jt},z_j)}K_{h',z,ji}\\
&\times\left(K^*_{h,vxz,kj}-Hf^*_{t,vxz,j}\right)(f^*_{t,vxz,j})^{-1}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)\\
\equiv\ & B_{n2,1,1}+B_{n2,1,2}+B_{n2,1,3}+B_{n2,1,4}+B_{n2,1,5}.
\end{aligned}
$$

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

$$B_{n2,1,2} = O_p(\|h\|^{\nu}(n^*H')^{-1}),\ B_{n2,1,3} = O_p((n^*H)^{-1}),\ \text{and} B_{n2,1,4} = O_p(\|h\|^{\nu}(n^*)^{-1}).$$

By the theory of two sample U-statistics, we have that

$$
\begin{aligned}
& B_{n2,1,5}\\
=\ & \frac{1}{n^*}\sum_{i=1}^{n^*}T^{-1}\sum_{t=1}^{T}\left(m_i^{-1}x_{it}x_{it}^{\top}\theta_i\frac{P^*(y_{it}^*\geq 0|z_i)}{P^*(y_{it}^*\geq 0|x_{it},z_i)}f_z^*(z_i^*)-E^*[P^*(y_{it}^*\geq 0|z_i)\theta_i]\right)\\
& +O_p(\|h\|^{\nu}+(n^*)^{-1}+(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|).
\end{aligned}
$$

Hence, we have that

$$
\begin{aligned}
& B_{n2,1} \\
= \ & \frac{1}{n^*}\sum_{i=1}^{n^*} T^{-1}\sum_{t=1}^{T}\left(m_i^{-1}x_{it}x_{it}^\top\theta_i\frac{P^*(y_{it}^*\geq 0|z_i)}{P^*(y_{it}^*\geq 0|x_{it},z_i)}f_z^*(z_i^*) - E^*[P^*(y_{it}^*\geq 0|z_i)\theta_i]\right) \\
& +O_p\Big((n^2H'H)^{-1} + \|h\|^\nu(nH')^{-1} + (nH)^{-1} + \|h\|^\nu + n^{-1} \\
& +(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|\Big). \qquad \text{(C.11)}
\end{aligned}
$$

Similarly, we can show that

$$
\begin{aligned}
& B_{n2,2} \\
= \ & -\frac{1}{n^*}\sum_{i=1}^{n^*} T^{-1}\sum_{t=1}^{T}\left(m_i^{-1}x_{it}x_{it}^\top\theta_i\frac{P^*(y_{it}^*\geq 0|z_i)}{P^*(y_{it}^*\geq 0|x_{it},z_i)}f_z^*(z_i^*) - E^*[P^*(y_{it}^*\geq 0|z_i)\theta_i]\right) \\
& +O_p((n^2H'\tilde{H})^{-1} + \|\tilde{h}\|^\nu(nH')^{-1} + (n\tilde{H})^{-1} + \|\tilde{h}\|^\nu + n^{-1} \\
& +(n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|). \qquad \text{(C.12)}
\end{aligned}
$$

Similar as the derivation of $B_{n1,4}$, we have $B_{n2,3} = O_p(\|h\|^{2\nu} + (\ln n)(nH)^{-1} + \|h\|^\nu\|\tilde{h}\|^\nu + (\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2})$.

Denote

$$\xi_{jt} = \tilde{y}_{jt} - E(\tilde{y}_{jt}|x_{jt}, z_j).$$

By (C.8), we have

$$B_{n3} = \frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}A_{n4}(z_i^*)$$

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

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

$$
B_{n3,1,1} = (nn^*TH')^{-1}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1}x_{it}\xi_{it}K_{h',z,ii}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_i^*),
$$

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

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

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

where

$$H_{n3,ij} \;=\; (TH')^{-1}\sum_{t=1}^{T}(m_i^{-1}x_{jt}\xi_{jt}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*)K_{h',ij}\mathbf{1}(j\leq n) + m_j^{-1}x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}\mathbf{1}_{\varepsilon_n}(z_j^*)$$
$$\times K_{h',ji}\mathbf{1}(i\leq n)).$$

Then, by using two sample U-statistics, we have

$$U_{n3} \;=\; \frac{1}{n^*T}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1}f_z(z_i^*)x_{it}\xi_{it}\mathbf{1}_{\tau_n,i} + O_p(\|h'\|^{\nu+1}/\sqrt{n^*} + (n^*H'^{1/2})^{-1}). \tag{C.13}$$

Then we consider $B_{n3,2}$, $B_{n3,3}$, and $B_{n3,4}$. Similar as (C.11) and (C.12), we have that

$$B_{n3,2} \;=\; \frac{1}{n^*T}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1}f_z^*(z_i^*)x_{it}^*E[\xi_{it}|x_{it}=x_{it}^*, z_i=z_i^*]\mathbf{1}_{\tau_n,i}$$

$$+O_p((n^2H'\tilde{H})^{-1} + \|\tilde{h}\|^{\nu}(nH')^{-1} + (n\tilde{H})^{-1} + \|\tilde{h}\|^{\nu} + (n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|)$$
$$= O_p((n^2H'\tilde{H})^{-1} + \|\tilde{h}\|^{\nu}(nH')^{-1} + (n\tilde{H})^{-1} + \|\tilde{h}\|^{\nu} + (n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\|),$$

and

$$\begin{aligned}
B_{n3,3} &= -\frac{1}{n^*T}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1} f_z^*(z_i^*) x_{it}^* E[\xi_{it}|v_i = v_i^*, x_{it} = x_{it}^*, z_i = z_i^*]\mathbf{1}_{\tau_n,i} \\
&\quad +O_p\Big((n^2H'H)^{-1} + \|h\|^{\nu}(nH')^{-1} + (nH)^{-1} + \|h\|^{\nu} + n^{-1} \\
&\quad +(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|\Big) \\
&= -\frac{1}{n^*T}\sum_{i=1}^{n^*}\sum_{t=1}^{T} m_i^{-1} f_z^*(z_i^*) x_{it}^* (E[\tilde{y}_{it}|v_i = v_i^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad -E[\tilde{y}_{it}|x_{it} = x_{it}^*, z_i = z_i^*])\mathbf{1}_{\tau_n,i} + O_p\Big((n^2H'H)^{-1} + \|h\|^{\nu}(nH')^{-1} \\
&\quad +(nH)^{-1} + \|h\|^{\nu} + n^{-1} + (n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\|\Big),
\end{aligned}$$

since $E[\xi_{it}|x_{it}, z_i] = 0$.

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

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

$P((v_{it}, x_{it}, z_i) \in \Omega_{vxz})$ 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 $Var(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it}\xi_{it}\mathbf{1}_{\tau_n,i}) = Var(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} m_i^{-1} f_z(z_i) x_{it}\xi_{it}) + o(1)$.

Further, we have

$$
\begin{aligned}
\hat{\mu}_t(k, z_i) &= \frac{(nH')^{-1}\sum_{j=1}^{n}\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}/\hat{f}_t^*(v_{jt}|x_{jt}, z_j)}{(nH')^{-1}\sum_{j=1}^{n} K_{h',ji}} \\
&= \frac{(nH')^{-1}\sum_{j=1}^{n}\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}/f_t^*(v_{jt}|x_{jt}, z_j)}{(nH')^{-1}\sum_{j=1}^{n} K_{h',ji}} \frac{f_t^*(v_{jt}|x_{jt}, z_j)}{\hat{f}_t^*(v_{jt}|x_{jt}, z_j)} \\
&= (nH')^{-1}\sum_{j=1}^{n} f(z_i)^{-1} \frac{\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}}{f_t^*(v_{jt}|x_{jt}, z_j)} \\
&\quad + (nH')^{-1}\sum_{j=1}^{n} f(z_i)^{-1} \frac{\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}}{f_t^*(v_{jt}|x_{jt}, z_j)} \frac{f_{t,vxz,j}^* - \hat{f}_{t,vxz,j}^*}{f_{t,vxz,j}^*} \\
&\quad + (nH')^{-1}\sum_{j=1}^{n} f(z_i)^{-1} \frac{\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}}{f_t^*(v_{jt}|x_{jt}, z_j)} \frac{\hat{f}_{t,xz,j}^* - f_{t,xz,j}^*}{f_{t,xz,j}^*} \\
&\quad + (nH')^{-1}\sum_{j=1}^{n} f(z_i)^{-1} \frac{\mathbf{1}(0 \le y_{jt} \le k) K_{h',ji}}{f_t^*(v_{jt}|x_{jt}, z_j)} \frac{(f_{t,vxz,j}^* \hat{f}_{t,xz,j}^* - \hat{f}_{t,vxz,j}^* f_{t,xz,j}^*)}{f_{t,xz,j}^* f_{t,vxz,j}^*} \\
&\quad \times \frac{(f_{t,vxz,j}^* - \hat{f}_{t,vxz,j}^*)}{\hat{f}_{t,vxz,j}^*} + o_p(n^{1/2}) \\
&\equiv \mu_{t1}(k, z_i) + \mu_{t2}(k, z_i) + \mu_{t3}(k, z_i) + \mu_{t4}(k, z_i).
\end{aligned}
$$

We can see that $\mu_{t2}(k, z_i)$ and $\mu_{t3}(k, z_i)$ can be written as a second-order U-statistics. By similar argument as in proving (A.32) and (A.33) in Khan and Lewbel (2007), we have that

$$\begin{aligned}
\mu_{t2}(k, z_i) &= -\frac{n}{n^*}\frac{1}{n}\sum_{i=1}^{n} E[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad + E[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}|z_i] + o_p(n^{-1/2}), \\
\mu_{t3}(k, z_i) &= \frac{n}{n^*}\frac{1}{n}\sum_{i=1}^{n} E[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*] \\
&\quad - E[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}|z_i] + o_p(n^{-1/2}).
\end{aligned}$$

Further, we have that $\mu_{t4}(k, z_i) = O_p(\|h\|^{2\nu} + (\ln n^*)(n^* H)^{-1} + \|h\|^{\nu}\|\tilde{h}\|^{\nu} + (\ln n^*)(n^*)^{-1} H^{-1/2}\tilde{H}^{-1/2})$.

We have

$$\begin{aligned}
B_{n4} &= -\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\tilde{y}_{jt}\frac{\hat{\mu}_t(k, z_j) - \mu_t(k, z_j)}{\mu_t(k, z_j)} K_{h',z,jz} \\
&\quad \times \frac{f_{t,v|xz,j}^*}{\hat{f}_{t,v|xz,j}^*}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&= -\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\tilde{y}_{jt} \\
&\quad \times \frac{\mu_{t1}(k, z_j) - \mu_t(k, z_j) + \mu_{t2}(k, z_j) + \mu_{t3}(k, z_j)}{\mu_t(k, z_j)} K_{h',z,ji}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*) \\
&\quad + o_p((n^*)^{-1/2}).
\end{aligned}$$

By U-statistic Hoeffding decomposition, we have that

$$B_{n4} = \frac{1}{n^*}\sum_{i=1}^{n^*} T^{-1}\sum_{t=1}^{T} m_i^{-1} E^*[x_{it}x_{it}^\top | z_i = z_i^*]\theta_i f_z(z_i^*)\phi_t(k, z_i^*)\mathbf{1}_{\tau_n,i} + o_p((n^*)^{-1/2}),$$

where

$$\begin{aligned}\phi_t(k, z_i^*) &= \frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)} - \mu_t(k, z_i^*) \\ &\quad -cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,vxz,i}}{f_{t,vxz,i}^*}|v_{it} = v_{it}^*, x_{it} = x_{it}^*, z_i = z_i^*] \\ &\quad +cE[\frac{\mathbf{1}(0 \le y_{it} \le k)}{f_t^*(v_{it}|x_{it}, z_i)}\frac{f_{t,xz,i}}{f_{t,xz,i}^*}|x_{it} = x_{it}^*, z_i = z_i^*].\end{aligned}$$

Also, let

$$\begin{aligned}B_{n5} &= -\frac{1}{n^*}\sum_{i=1}^{n^*} m_i^{-1}(nTH')^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n} x_{jt}\frac{v_{jt}\mathbf{1}(0 \le y_{jt} \le k)(\hat{\gamma} - \gamma)}{\mu_t(k, z_j)f_{t,v|xz,j}^*}K_{h',z,ji} \\ &\quad \times\frac{f_{t,v|xz,j}^*}{\hat{f}_{t,v|xz,j}^*}\mathbf{1}_{\tau_n,j}\mathbf{1}_{\varepsilon_n}(z_i^*).\end{aligned}$$

By using the projection of U-statistics, we have that

$$\begin{aligned}B_{n5} &= \frac{1}{n^*}\sum_{i=1}^{n^*} T^{-1}\sum_{t=1}^{T} m_i^{-1} f_z(z_i^*)\left(\frac{1}{2\gamma^2}(k^2 E[x_{it}|z_i = z_i^*] - kE[x_{it}x_{it}^\top|z_i = z_i^*]\theta(z_i^*))\right) \\ &\quad \times\frac{\psi_t(k)}{\mu_t(k, z_i^*)}\mathbf{1}_{\tau_n,i} + o_p((n^*)^{-1/2}).\end{aligned}$$

Therefore, we have that

$$
\begin{aligned}
\sqrt{n^*}(\hat{\beta}_{LC} - \beta) &= \frac{1}{\sqrt{n^*}} \sum_{i=1}^{n^*} g(z_i^*) - \frac{1}{\sqrt{n^*}T} \sum_{i=1}^{n^*} \sum_{t=1}^{T} m_i^{-1} f_z(z_i^*) x_{it} \xi_{it} \mathbf{1}_{\tau_n,i} \\
&- \frac{1}{\sqrt{n^*}T} \sum_{i=1}^{n^*} \sum_{t=1}^{T} m_i^{-1} f_z^*(z_i^*) x_{it}^* (E[\tilde{y}_{it}|v_i = v_i^*, x_{it} = x_{it}^*, z_i = z_i^*] \\
&- E[\tilde{y}_{it}|x_{it} = x_{it}^*, z_i = z_i^*]) \mathbf{1}_{\tau_n,i} \\
&+ \frac{1}{\sqrt{n^*}} \sum_{i=1}^{n^*} T^{-1} \sum_{t=1}^{T} m_i^{-1} E^*[x_{it} x_{it}^\top | z_i = z_i^*] \theta_i f_z(z_i^*) \phi_t(k, z_i^*) \mathbf{1}_{\tau_n,i} \\
&+ \frac{1}{\sqrt{n^*}} \sum_{i=1}^{n^*} T^{-1} \sum_{t=1}^{T} m_i^{-1} f_z(z_i^*) \Big(\frac{1}{2\gamma^2} (k^2 E[x_{it}|z_i = z_i^*] \\
&- kE[x_{it} x_{it}^\top | z_i = z_i^*] \theta(z_i^*)) \Big) \frac{\psi_t(k)}{\mu_t(k, z_i^*)} \mathbf{1}_{\tau_n,i} \\
&+ O_p(\delta_n) \\
&\xrightarrow{d} N(0, V_{LC})
\end{aligned}
$$

by the Lindeberg central limit theorem, where

$$
\begin{aligned}
V_{LC} &= E^*(g(z_i^*))^2 \\
&+ E^* \Bigg(T^{-1} \sum_{t=1}^{T} \Bigg[m_i^{-1} f_z(z_i^*) x_{it} \xi_{it} \\
&+ m_i^{-1} f_z^*(z_i^*) x_{it}^* \Big(E[\tilde{y}_{it}|v_i = v_i^*, x_{it} = x_{it}^*, z_i = z_i^*] - E[\tilde{y}_{it}|x_{it} = x_{it}^*, z_i = z_i^*] \Big) \\
&- m_i^{-1} E^*[x_{it} x_{it}^\top | z_i = z_i^*] \theta_i f_z(z_i^*) \phi_t(k, z_i^*) \\
&- m_i^{-1} f_z(z_i^*) \left(\frac{1}{2\gamma^2} (k^2 E[x_{it}|z_i = z_i^*] - kE[x_{it} x_{it}^\top | z_i = z_i^*] \theta(z_i^*)) \right) \frac{\psi_t(k)}{\mu_t(k, z_i^*)} \Bigg] \Bigg)^2,
\end{aligned}
$$

$$\begin{aligned}\delta_n &= \sqrt{n}\|h'\|^{\nu} + \sqrt{n}(nH'^{1/2})^{-1}\|h'\| + \sqrt{n}n^{-1/2}\|h'\|^{\nu} \\ &+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|h\| + \sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h'\|\|\tilde{h}\| \\ &+\sqrt{n}(\ln n)(nH)^{-1} + \sqrt{n}\|h\|^{\nu}\|\tilde{h}\|^{\nu} + \sqrt{n}(\ln n)n^{-1}H^{-1/2}\tilde{H}^{-1/2} + \sqrt{n}(n^2H'H)^{-1} \\ &+\sqrt{n}\|h\|^{\nu}(nH')^{-1} + \sqrt{n}(nH)^{-1} + \sqrt{n}\|h\|^{\nu} + \sqrt{n}n^{-1} \\ &+\sqrt{n}(n^{3/2}H'^{1/2}H^{1/2})^{-1}\|h\| + \sqrt{n}(n^2H'\tilde{H})^{-1} + \sqrt{n}\|\tilde{h}\|^{\nu}(nH')^{-1} + \sqrt{n}(n\tilde{H})^{-1} \\ &+\sqrt{n}\|\tilde{h}\|^{\nu} + \sqrt{n}(n^{3/2}H'^{1/2}\tilde{H}^{1/2})^{-1}\|\tilde{h}\| + \sqrt{n}\|h'\|^{\nu+1}/\sqrt{n} + \sqrt{n}(nH'^{1/2})^{-1} \\ &+\sqrt{n}\eta_n = o_p(1),\end{aligned}$$

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 C.1.5. *Define* $A_{n1}(z) = \frac{1}{n^*TH'}\sum_{j=1}^{n^*}\sum_{s=1}^{T} x^*_{js}(x^*_{js})^\top K_{h',z^*_j z}$*, 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 B5-B8,*

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

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

$$E^*[A_{n1}(z)] = m(z) + O\left(\|h'\|^{\nu}\right), \tag{C.14}$$

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

$$A_{n1}(z) - E^*[A_{n1}(z)] = O_p\left(\frac{(\ln n^*)^{1/2}}{(n^* H')^{1/2}}\right), \qquad \text{(C.15)}$$

uniformly in $z \in \Omega_z$.

Combining (C.14) and (C.15) we obtain

$$A_{n1}(z) - m(z) = O_p\left(||h'||^{\nu} + (\ln n^*)^{1/2} (n^* H')^{-1/2}\right), \qquad \text{(C.16)}$$

uniformly in $z \in \Omega_z$.

Using (C.16) we obtain

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

which completes the proof of Lemma C.1.5.