DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 14-10

Estimation of Semiparametric Regression in Triangular Systems

Xin Geng
University of Colorado Boulder

November 2014

Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

ESTIMATION OF SEMIPARAMETRIC REGRESSION IN TRIANGULAR SYSTEMS*

XIN GENG†
University of Colorado, Boulder

November 2014

Abstract. We propose a kernel based estimator for a partially linear model in triangular systems where endogenous variables appear both in the nonparametric and linear component functions. This model has a wide range of applications in many fields of economics. Compared with the two alternative estimators currently available in the literature for such model, this estimator has an explicit functional form, is much easier to implement, and may significantly outperform theirs in finite sample simulation. Our estimator is inspired by the control function approach of Newey et al. (1999) and was initially proposed by Martins-Filho and Yao (2012). It builds on the additive regression estimation by Kim et al. (1999). We establish: (i) $\sqrt{n}$ asymptotic normality of the estimator for the parametric component, and (ii) consistency and the uniform convergence rate of the estimator for the nonparametric component. In addition, for statistical inference, a consistent estimator for the covariance of the limiting distribution of the parametric estimator is also provided. Various intermediate results will also be of use to theorists.

Keywords. Additive semiparametric regression; Instrumental variables; $\sqrt{n}$-consistent estimation; Nadaraya-Watson kernel estimation; Structural models.

JEL Classifications. C13; C14.

*I am greatly indebted to Carlos Martins-Filho for his guidance, inspiration and encouragement. I would like to thank Xiaodong Liu, Donald Waldman, Kairat Mynbayev for helpful comments and suggestions. All errors are my own.

†Department of Economics, University of Colorado, Boulder, UCB 256, Boulder, CO 80309 (e-mail: Xin.Geng-1@colorado.edu; website: https://sites.google.com/site/xingeng2015)

1 Introduction

Recently there has been a growing interest in estimation of nonparametric regression models with endogenous regressors (Newey et al. (1999); Blundell and Powell (2003); Ai and Chen (2003); Su and Ullah (2008); Otsu (2011)). The problem of endogeneity is widely encountered in empirical models in economics, due to measurement error or simultaneity that arises from individual choices or market equilibrium. Thus, the development of estimation procedures that account for endogeneity has permeated research in Econometrics. Doing so in the context of tightly specified functional forms can be misleading due to the high probability of misspecification. Alternatively, accounting for endogeneity in fully nonparametric models may be undesirable due to reduced precision that results from the well known "curse of dimensionality". Thus, a useful alternative is to consider semiparametric structural models to take advantage of any known functional form information while retaining some nonparametric features.

Semiparametric models that account for endogeneity have been considered by a number of authors (see Li and Racine (2007) Chapter 16 for an introduction). Prominent among these are Ai and Chen (2003) and Otsu (2011) that propose two different sieve estimators for a partially linear model with endogenous regressors in the nonparametric part. In this paper we consider a model that allows for endogeneity on both the parametric and nonparametric components of a regression. Martins-Filho and Yao (2012) proposed a kernel-based semiparametric estimator for such model. Compared with the two natural alternatives in the current available literature (Ai and Chen (2003); Otsu (2011)), this estimator has an explicit functional form, much easier to implement, and a Monte Carlo study suggests that our estimator has a better finite sample performance. However, a full asymptotic characterization of their estimator was not provided. Such characterization is critical for hypothesis testing and inference. In this paper, we establish: (i) $\sqrt{n}$ asymptotic normality of the estimator for the parametric component, and (ii) consistency and the uniform convergence rate of the estimator for the nonparametric component. In addition, we provide a consistent estimator for the covariance of the limiting distribution of the parametric estimator.

We consider the following triangular semiparametric structural model:

$$Y_i = \beta_0 + X_{2i}\beta + m(X_{1i}, Z_{1i}) + \varepsilon_i, \quad \text{for } i = 1, \cdots, n \tag{1}$$

$$X_i = \Pi(Z_i) + U_i \tag{2}$$

$$\mathrm{E}(U_i|Z_i) = 0, \quad \mathrm{E}(\varepsilon_i|Z_i, U_i) = \mathrm{E}(\varepsilon_i|U_i) \tag{3}$$

In (1), the regressand Y_i is a scalar, $Z_{1i} \in \mathbb{R}^{D_{11}}$ is a subvector of $Z_i = (Z_{1i}', Z_{2i}')' \in \mathbb{R}^{D_1}$ with $D_1 = D_{11} + D_{12}$, X_{1i}, X_{2i} are non-overlapping subvectors of $X_i \in \mathbb{R}^{D_2}$ of dimensions D_{21} and D_{22} with $D_2 = D_{21} + D_{22}$, and ε_i is an unobserved scalar random error. $m(\cdot)$ is an unknown real function, $\beta_0 \in \mathbb{R}$ and $\beta \in \mathbb{R}^{D_{22}}$ are unknown coefficients of the linear part. In (2), U_i is a vector of unobserved random errors and $\Pi \colon \mathbb{R}^{D_1} \to \mathbb{R}^{D_2}$ is an unknown function. Let $\mathrm{E}(\cdot)$ denote expectation. Variables X_i are taken as endogenous in that $\mathrm{E}(\varepsilon_i|X_i) \neq 0$, and the variables Z_i are exogenous due to (3). We are interested in estimating β and $m(\cdot)$ consistently up to an additive constant.

Structural models can be viewed as simultaneous equations models, where economic theory is used to guide the construction of a system of equations that describe the relationship among endogenous, exogenous and unobservable variables (Hoyle (1995), Reiss and Wolak (2007)). The triangular system described by (1)-(3) is a special case of a structural model, since all the endogenous variables X_i in (1) can be suitably modeled by exogenous variables Z_i in (2). Triangular models have appeared frequently in economics and other social sciences. For example, the method of "path analysis", which is widely used in sociology, provides a more effective and direct way of modeling mediation, indirect effects; for more, see Lahiri and Schmidt (1978) and Lei and Wu (2007). Partially linear models like (1) have also been studied extensively by Stock (1989), Engle et al. (1986), Heckman (1986), Robinson (1988), Li (1996), Hasan (2012), Lessmann (2014), and among others. However, even though the statistical objectives in these papers may vary, none of them confront the potential endogeneity. For example, Robinson (1988) provided a $\sqrt{n}$-consistent kernel estimator for β under regularity conditions, and based on this, Lessmann (2014) on one hand, tested and verified the inverted-U relationship between spatial inequality and economic development, but on the other hand, to take endogeneity into account, two methods are employed: one is the standard OLS estimation with lagged endogenous variables as part of the regressors, and the other uses a difference GMM estimator. Thus, it would be more convenient and convincing to

employ an estimator that accounts for endogeneity appearing both in the parametric and nonparametric parts of the semiparametric model.

Given (2) and (3), we have $\mathrm{E}(\varepsilon_i|X_{1i},Z_i,U_i) = \mathrm{E}(\varepsilon_i|Z_i,U_i) = \mathrm{E}(\varepsilon_i|U_i)$, and $\mathrm{E}(X_{2i}|X_{1i},Z_i,U_i) = \mathrm{E}(X_{2i}|Z_i,U_i) = X_{2i}$. Note that $\mathrm{E}(\varepsilon_i|U_i)$ is an unknown function of U_i, thus we can denote it by $h(U_i)\colon \mathbb{R}^{D_2} \to \mathbb{R}$, and using(1), we have:

$$\mathrm{E}(Y_i|X_{1i},Z_i,U_i) = \beta_0 + X_{2i}\beta + m(X_{1i},Z_{1i}) + h(U_i) \tag{4}$$

Newey et al. (1999) and Su and Ullah (2008) consider a purely nonparametric structural model with the same conditional mean restriction given in (3). As Newey et al. (1999) put it, (3) is a more general assumption than requiring that (ε_i,U_i) be independent of Z_i and $\mathrm{E}(U_i) = 0$. The added generality may be important in that it allows for conditional heteroskedasticity of the disturbances. Different from the previous literature, this paper allows endogenous X_i to enter the regression not only nonparametrically through $m(\cdot)$ but also linearly. Newey et al. (1999) employ series approximation to exploit the additive structure of the model (as we can see from (4) but without the linear components) and establish the consistency and asymptotic normality for their second-stage estimator of $m(\cdot)$. Su and Ullah (2008) also exploits the additive structure but their estimation is based on local polynomial regression and marginal integration techniques. As discussed in Kim et al. (1999) and Martins-Filho and Yang (2007), the marginal integration estimator (Linton and Hardle (1996)) is not oracle efficient. Thus, Kim et al. (1999) proposed a two-step oracle efficient estimator for the additive nonparametric model. Note that if β were known and realizations of U were observed, (4) is just an additive nonparametric conditional expectation that could be estimated using the pilot or two-step estimator of Kim et al. (1999). We adopt a similar method as their first step pilot estimator does, employing some particular "instrument" function, to derive the identification of our estimator for β. Here, since U is not observed, like Su and Ullah (2008), we replace them by the residuals obtained by regressing X on Z nonparametrically. It can be shown that such a replacement does not impact the asymptotic properties of the resulting estimator.

There are two natural alternative estimators to ours in the current literature, i.e., the sieve minimum distance estimator of Ai and Chen (2003) and the sieve conditional empirical likelihood estimator of Otsu (2011). This paper is different from them in that the object of our estimation is the structural model and not just a conditional expectation,

so that we are able to give identifications and explicit expressions of estimators for each component in the model. Besides, they have a different moment restriction, i.e., $\mathrm{E}(\varepsilon_i|Z_i) = 0$. Strictly speaking, neither restriction is stronger than the other; see Newey et al. (1999). Under additional restrictions: (i) U_i is independent of Z_i, and (ii) $\mathrm{E}(\varepsilon_i) = 0$, the moment restrictions in (3) imply that $\mathrm{E}(\varepsilon_i|Z_i) = 0$. This makes the estimators developed in these two papers and our estimator suitable for the same model, and it turns out in Martins-Filho and Yao (2012) that the latter outperforms the previous two in terms of finite sample performance and ease of their implementation from a computational perspective.

The rest of the paper is organized as follows. Section 2 considers identification, moment conditions, and describes the estimator. Section 3 provides the asymptotic characterization of our proposed estimators and the assumptions we used in our results. Section 4 contains a Monte Carlo study that gives the finite sample performance of our estimators. Section 5 provides a conclusion and gives potential directions for further study. All proofs are given in the Appendix.

2 Estimation

Suppose there are n observations and write $Y = (Y_1, \cdots, Y_n)'$, $X = (X_1, \cdots, X_n)'$, $Z = (Z_1, \cdots, Z_n)'$ for observations on the regressand and regressors for the model (1)-(3). The objective is to estimate the coefficients for the linear component, $\beta_0 \in \mathbb{R}$, $\beta \in \mathbb{R}^{D_{22}}$. Let $v_i = Y_i - \mathrm{E}(Y_i|X_{1i}, Z_i, U_i)$, and rewrite (4) as:

$$Y_i - X_{2i}\beta - \beta_0 = m(X_{1i}, Z_{1i}) + h(U_i) + v_i, \text{ for } i = 1, \cdots, n, \tag{5}$$

where, by construction, $\mathrm{E}(v_i|X_{1i}, Z_i, U_i) = 0$.

Note that if β, β_0 were known and if realizations of U_i were given, (5) could be viewed as an additive nonparametric regression model. With the help of an appropriate choice of "instrument" function, we derive the moment conditions that motivate our estimator for β.

2.1 Moment Conditions

For identification, it is standard to assume that $\mathrm{E}(m(X_{1i}, Z_{1i})) = \mathrm{E}(h(U_i)) = 0$, since each component in an additive nonparametric model can only be estimated up to an constant. For simplicity, let $M = (X_1', Z_1')'$. Like in Kim et al.

(1999), define our "instrument" function as $\eta(M_i,U_i)=\dfrac{f_M(M_i)f_U(U_i)}{\phi(M_i,U_i)}\equiv\eta_i$, where f_M is the joint marginal density of $M_i=(X_{1i}',Z_{1i}')'$, f_U the marginal density of U_i, and ϕ the joint density of M_i and U_i. The essential reason for choosing such "instrument" function lies in that

$$\mathrm{E}\big(\eta(M_i,U_i)|M_i\big)=1;\qquad \mathrm{E}\big(\eta(M_i,U_i)h(U_i)|M_i\big)=0.$$

The equations still hold if we replace the conditioning variable M_i by U_i and $h(U_i)$ by $m(M_i)$. Thus, by pre-multiplying η_i on both sides of (5), and taking conditional expectations given M_i and U_i separately, we have

$$\mathrm{E}\left(\eta_i(Y_i-X_{2i}\beta-\beta_0)\,|\,M_i\right)\;=\;m(M_i);\qquad \mathrm{E}\left(\eta_i(Y_i-X_{2i}\beta-\beta_0)\,|\,U_i\right)\;=\;h(U_i) \tag{6}$$

If β, β_0 were known, we could estimate $m(M_i)$ and $h(U_i)$ based on moment conditions (6) using estimated residuals $\{\hat{U}_i\}_{i=1}^n$ and estimated $\{\hat{\eta}_i\}_{i=1}^n$. Thus, we need to consider estimation of β and β_0. Since $m(M_i)$ and $h(U_i)$ can be expressed as conditional expectations containing β, β_0 in (6), we can plug them into (5), rearranging, with $\beta_0=\mathrm{E}(\eta_i(Y_i-X_{2i}\beta))$, we have

$$Y_i^*\;=\;X_{2i}^*\,\beta\;+\;v_i,\qquad \text{for}\quad i=1,\cdots,n, \tag{7}$$

where $\quad Y_i^*\;\equiv\;Y_i-\mathrm{E}(\eta_iY_i|M_i)-\mathrm{E}(\eta_iY_i|U_i)+\mathrm{E}(\eta_iY_i),\quad$ and $\quad X_{2i}^*\;\equiv\;X_{2i}-\mathrm{E}(\eta_iX_{2i}|M_i)-\mathrm{E}(\eta_iX_{2i}|U_i)+\mathrm{E}(\eta_iX_{2i})$. Note that equation (7) provides infinitely many moment conditions to estimate β, since by pre-multiplying an arbitrary measurable function $L(X_{1i},Z_i,U_i)$, we still have $\mathrm{E}\Big(L(X_{1i},Z_i,U_i)\;v_i\Big|X_{1i},Z_i,U_i\Big)=0$. Here $L(X_{1i},Z_i,U_i)$ can be treated as a normalizing factor that can be chosen conveniently to derive the asymptotic properties of an estimator for β.

In our case, we choose $L(X_{1i},Z_i,U_i)=\sqrt{\eta_i}$. Then, we consider:

$$\sqrt{\eta_i}\,Y_i^*\;=\;\sqrt{\eta_i}\,X_{2i}^*\beta\;+\;\sqrt{\eta_i}\,v_i,\qquad \text{for}\quad i=1,\cdots,n. \tag{8}$$

In matrix form we write $\sqrt{\eta}\,Y^*=\sqrt{\eta}\,X_2^*\beta+\sqrt{\eta}\,v$, where $Y^*=(Y_1^*,\cdots,Y_n^*)'$, $X_2^*=(X_{21}^*,\cdots,X_{2n}^*)'$, $v=(v_1,\cdots,v_n)'$, $\sqrt{\eta}=\mathrm{diag}\{\sqrt{\eta_i}\}_{i=1}^n$, and $\mathrm{E}\big(\sqrt{\eta_i}\,v_i\big|X_{1i},Z_i,U_i\big)=0$. Note that by choosing $\beta_0=\mathrm{E}\big(\eta_i(Y_i-X_{2i}\beta)\big)$ and $L(X_{1i},Z_i,U_i)=$

$\sqrt{\eta_i}$, we have $\mathrm{E}(\eta_i Y_i^*|M_i) = \mathrm{E}(\eta_i Y_i^*|U_i) = \mathrm{E}(\eta_i X_{2i}^*|M_i) = \mathrm{E}(\eta_i X_{2i}^*|U_i) = 0$. These conditions are crucial in establishing the asymptotic properties of our estimator of β, as we will see in later sections. However, a more intuitive reason for choosing such normalizing function is still open to investigation.

Denote the additive components in Y_i^*, X_{2i}^* and corresponding error terms by $\gamma_1(M_i) \equiv \mathrm{E}(\eta_i Y_i|M_i)$, $\gamma_2(U_i) \equiv \mathrm{E}(\eta_i Y_i|U_i)$, $\gamma_3 \equiv \mathrm{E}(\eta_i Y_i)$, $g_1(M_i) \equiv \mathrm{E}(\eta_i X_{2i}|M_i)$, $g_2(U_i) \equiv \mathrm{E}(\eta_i X_{2i}|U_i)$, $g_3 \equiv \mathrm{E}(\eta_i X_{2i})$, $v_{Y1i} \equiv \eta_i Y_i - \gamma_1(M_i)$, $v_{Y2i} \equiv \eta_i Y_i - \gamma_2(U_i)$, $v_{X1i} \equiv \eta_i X_{2i} - g_1(M_i)$, and $v_{X2i} \equiv \eta_i X_{2i} - g_2(U_i)$. Now we have $\sqrt{\eta_i} X_{2i}^*$ as our regressors, and $\mathrm{E}\big(\sqrt{\eta_i} X_{2i}^* v_i\big) = 0$. Equation (8) suggests an estimator of β by inserting estimators of $\sqrt{\eta_i} Y_i^*$ and $\sqrt{\eta_i} X_{2i}^*$ prior to application of a standard rule, such as no-intercept ordinary least square (OLS) method. Note that by (6), we have $m(M_i) = \gamma_1(M_i) - g_1(M_i)\beta - \beta_0$, and $h(U_i) = \gamma_2(U_i) - g_2(U_i)\beta - \beta_0$. Thus to estimate Y_i^*, X_{2i}^*, $m(M_i)$ and $h(U_i)$, we need only to estimate each of their additive components separately. Kernel-based nonparametric estimators are employed throughout this paper. For identification purpose, we need to assume existence and nonsingularity of $\Phi_0 \equiv \mathrm{E}\big(\eta_i X_{2i}^* X_{2i}^{*\prime}\big)$.

2.2 Estimation Procedure

Based on the moment conditions given in Section 2.1, we now describe specific estimation procedure.

1. Obtain a Nadaraya-Watson (NW) estimator for $\Pi(Z_i)$ from (2), with the j^{th} element denoted as

$$\hat{\Pi}_j(Z_i) = \underset{\theta}{\operatorname{argmin}} \frac{1}{n h_1^{D_1}} \sum_{t=1}^{n} (X_{t,j} - \theta)^2 K_1\left(\frac{Z_t - Z_i}{h_1}\right) \text{ for } j = 1, \cdots, D_2,$$

   where $X_{t,j}$ is the j^{th} element of X_t, $h_1 > 0$ is the associated bandwidth, and $K_1 : \mathbb{R}^{D_1} \to \mathbb{R}$ is a multivariate kernel function. Denote the estimates by $\hat{\Pi}(Z_i) \equiv \big(\hat{\Pi}_1(Z_i), \cdots, \hat{\Pi}_{D_2}(Z_i)\big)'$ and calculate the nonparametric residuals $\hat{U}_i \equiv \big(\hat{U}_{i1}, \cdots, \hat{U}_{iD_2}\big)'$, where $\hat{U}_{ij} \equiv X_{i,j} - \hat{\Pi}_j(Z_i)$, for $j = 1, \cdots, D_2$ and $i = 1, \cdots, n$.

2. Obtain Rosenblatt density estimators for f_U, f_M and ϕ:

$$\begin{aligned}
\hat{f}_U(u) &= \frac{1}{n h_2^{D_2}} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - u}{h_2}\right), \qquad \hat{f}_M(m) = \frac{1}{n h_3^{D_3}} \sum_{t=1}^{n} K_3\left(\frac{M_t - m}{h_3}\right), \\
\hat{\phi}(m,u) &= \frac{1}{n h_4^{D_4}} \sum_{t=1}^{n} K_4\left(\frac{(M_t' \quad \hat{U}_t')' - (m' \quad u')'}{h_4}\right),
\end{aligned}$$

where $K_2\colon \mathbb{R}^{D_2} \to \mathbb{R}$, $K_3\colon \mathbb{R}^{D_3} \to \mathbb{R}$, and $K_4\colon \mathbb{R}^{D_4} \to \mathbb{R}$ are multivariate kernel functions, $D_3 \equiv D_{11} + D_{21}$, $D_4 \equiv D_2 + D_{11} + D_{21}$, and $h_i > 0$ are associated bandwidths, $i = 2,3,4$. Thus, a natural estimator for η_i would be $\hat{\eta}(M_i, \hat{U}_i) = \dfrac{\hat{f}_M(M_i)\hat{f}_U(\hat{U}_i)}{\hat{\phi}(M_i, \hat{U}_i)} \equiv \hat{\eta}_i$.

3. Obtain NW estimators for the conditional expectations in Y_i^*, X_{2i}^* as follows:

$$\begin{aligned}
\hat{\gamma}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \sum_{t=1}^{n} K_3\left(\frac{M_t - M_i}{h_3}\right) \hat{\eta}_t Y_t, & \hat{g}_1(M_i) &= \frac{1}{nh_3^{D_3}} \frac{1}{\hat{f}_M(M_i)} \sum_{t=1}^{n} K_3\left(\frac{M_t - M_i}{h_3}\right) \hat{\eta}_t X_{2t}, \\
\hat{\gamma}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - \hat{U}_i}{h_2}\right) \hat{\eta}_t Y_t, & \hat{g}_2(\hat{U}_i) &= \frac{1}{nh_2^{D_2}} \frac{1}{\hat{f}_U(\hat{U}_i)} \sum_{t=1}^{n} K_2\left(\frac{\hat{U}_t - \hat{U}_i}{h_2}\right) \hat{\eta}_t X_{2t}.
\end{aligned} \tag{9}$$

Estimation for expectations γ_3 and g_3 is trivial, as we can just use the population average with $\hat{\eta}_i$ replacing η_i, i.e., $\hat{\gamma}_3 = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t Y_t$, and $\hat{g}_3 = \frac{1}{n}\sum_{t=1}^{n} \hat{\eta}_t X_{2t}$. Thus, estimators for Y_i^* and X_{2i}^* are given as $\hat{Y}_i = Y_i - \hat{\gamma}_1(M_i) - \hat{\gamma}_2(\hat{U}_i) + \hat{\gamma}_3$, $\hat{X}_{2i} = X_{2i} - \hat{g}_1(M_i) - \hat{g}_2(\hat{U}_i) + \hat{g}_3$, for $i = 1, \cdots, n$.

4. Using the estimators $\hat{\eta}_i$, $\hat{Y}_i$ and $\hat{X}_{2i}$ derived in Steps 2 and 3 instead of η_i, Y_i^* and X_{2i}^* in (8), we have the no-intercept OLS estimator for β:

$$\hat{\beta} = \left(\hat{X}_2' \hat{\eta} \hat{X}_2\right)^{-1} \hat{X}_2' \hat{\eta} \hat{Y}, \tag{10}$$

where $\hat{Y} = (\hat{Y}_1, \cdots, \hat{Y}_n)'$, $\hat{X}_2 = (\hat{X}_{21}, \cdots, \hat{X}_{2n})'$, and $\hat{\eta} = \operatorname{diag}\{\hat{\eta}_i\}_{i=1}^n$.

Given $\beta_0 = \mathrm{E}(Y_i - X_{2i}\beta)$ and $\hat{\beta}$, an estimator for β_0 is $\hat{\beta}_0 = \bar{Y} - \bar{X}_2\hat{\beta}$, where $\bar{Y} \equiv \frac{1}{n}\sum_{t=1}^{n} Y_t$, and $\bar{X}_2 \equiv \frac{1}{n}\sum_{t=1}^{n} X_{2t}$. For $m(M_i)$ and $h(U_i)$, we have

$$\hat{m}(M_i) = \hat{\gamma}_1(M_i) - \hat{g}_1(M_i)\hat{\beta} - \hat{\beta}_0, \qquad \hat{h}(\hat{U}_i) = \hat{\gamma}_2(\hat{U}_i) - \hat{g}_2(\hat{U}_i)\hat{\beta} - \hat{\beta}_0. \tag{11}$$

3 Asymptotic Characterization of $\hat{\beta}$

In this section, we study the asymptotic properties of the estimators described in the previous section. We first establish the uniform convergence in probability rate of the Rosenblatt density estimator using estimated residuals $\{\hat{U}_i\}_{i=1}^n$. Second, we give the uniform convergence in probability rate of the NW estimator constructed using estimated residuals $\{\hat{U}_i\}_{i=1}^n$, and third we establish the asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta)$.

3.1 Assumptions

We now provide a list of general assumptions that will be selectively adopted in our theorems and introduce notation. In what follows, C always denotes a generic constant in $\mathbb{R}$ that may vary from case to case. $k^{(j)}(x)$ denotes the j^{th}-order derivative of $k(x)$ evaluated at x.

Assumption A1. The kernels K_i, $i = 1,2,3,4$, satisfy:

$$K_i(x) = \prod_{j=1}^{D_i} k_i(x_j)$$

where D_i is the corresponding dimension of K_i. Assume k_i is symmetric about zero, 4-times partially continuously differentiable and satisfies a) $\int k_i(x)dx = 1$; b) $|k_i^{(j)}(x)||x|^{5+a} \to 0$ as $|x| \to \infty$, $j = 0,\cdots,4$; c) k_i is a kernel of order s_i, i.e., $\int k_i(x)x^j dx = 0$ for $j = 1,\cdots,s_i - 1$, and $\int |k_i(x)||x|^{s_i} dx < C$. Denote $s = \max\{s_i\}_{i=1}^4$.

We adopt the "higher-order" kernel approach to reduce bias. Since global differentiability of the kernel functions is used in order to employ Taylor Theorem in following Theorems 2 and 3, kernels that have compact support are excluded. The ideal candidates have to decay exponentially, and it turns out kernels constructed below based on Hermite polynomial and Gaussian densities are one such class of kernels. Construct the kernel that is of even order $s \geq 2$ by:

$$k_s(x) = \sum_{j=0}^{\frac{1}{2}(s-2)} c_j x^{2j} \phi(x) \tag{12}$$

where $\phi(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$. Given that we can evaluate the moments $m_{2j} = \int x^2 \psi(x)dx$, $0 \leq j \leq \frac{1}{2}(s-2)$, $\{c_j\}_{j=0}^{\frac{1}{2}(s-2)}$ that satisfy the linear system of $\frac{1}{2}(s-2)$ simultaneous equations $\sum_{j=0}^{\frac{1}{2}(s-2)} c_j m_{2(i+j)} = \delta_{i0}$, $0 \leq i \leq \frac{1}{2}(s-2)$ where δ_{ij} Kronecker's delta, will give us the desired kernel. For example, $k_4(x) = \left(\frac{3}{2} - \frac{1}{2}x^2\right)(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$, $k_6(x) = \left(\frac{15}{8} - \frac{5}{4}x^2 + \frac{1}{8}x^4\right)(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$. As discussed in Pagan and Ullah (1999), when higher order kernels with large s are needed, it will be helpful to express them in terms of a recurrence relationship. Rewrite (13) as (for $r \geq 1$) $k_{2r}(x) = P_{2r-2}\phi(x)$, where $P_{2r} = P_{2r-2} + (-1)H_{2r}(2^r r!)^{-1}$ and $H_r(x) = xH_{r-1} - (r-1)H_{r-2}$ is the rth Hermite

polynomial with $H_0 = 1$. Or recursively, with $k_2(x) = \phi(x)$,

$$k_{2r}(x) = k_{2(r-1)}(x) + (-1)^{r-1} H_{2(r-1)}(x)(2^{r-1}(r-1)!)^{-1}\phi(x)$$

Kernels constructed like (12) will satisfy Assumption A1, since they are continuously differentiable of any order everywhere, and when multiplied by any polynomial functions they are all uniformly bounded and absolutely integrable as the tails decay exponentially. We show in Lemma 2 that product kernels satisfying A1 are locally Lipschitz continuous, which is necessary in Lemma 4.

Assumption A2. $\{(X_i', Z_i', Y_i)\}_{i=1}^n$ is a sequence of independent and identically distributed (IID) random vectors that are described by (1), (2) and (3). The density functions $f_M(M_i)$, $f_Z(Z_i)$, $\phi(M_i, U_i)$, $f_{UZ}(U_i, Z_i)$, $f_U(U_i)$ are uniformly bounded away from zero and infinity.

Assumption A3. $\mathrm{E}\big(m(M_i)\big) = \mathrm{E}\big(h(U_i)\big) = 0$, $\mathrm{E}\big(m^2(M_i)\big), \mathrm{E}\big(h^2(U_i)\big) < \infty$, $\mathrm{E}\big(v_i^2 | X_{1i}, Z_i, U_i\big) = \sigma_v^2 < \infty$, $\mathrm{E}\big(U_{ij}^2 | Z_i\big) = \sigma_U^2 < \infty$, $\mathrm{E}\big(v_{X1i,j}^2 | M_i\big) = \sigma_{v_{X1,j}}^2 < \infty$, $\mathrm{E}\big(v_{X2i,j}^2 | U_i\big) = \sigma_{v_{X2,j}}^2 < \infty$, $\mathrm{E}\big(v_{Y1i}^2 | M_i\big) = \sigma_{v_{Y1}}^2 < \infty$, $\mathrm{E}\big(v_{Y2i}^2 | U_i\big) = \sigma_{v_{Y2}}^2 < \infty$, and Cramer's conditions: $\mathrm{E}\big|X_{2i,j}\big|^p \le C^{p-2} p!\, \mathrm{E}\big|X_{2,j}\big|^2 < \infty$, $\mathrm{E}\big(|U_{ij}|^p | Z_i\big) \le C^{p-2} p! \sigma_U^2$, for some $C > 0$, all i, $p = 3, 4, \cdots$, and $j = 1, \cdots, D_2$.

In A3, it is not essential to assume the second conditional moment of those error terms are independent of the conditioning variables. However, the boundedness of the second moment is crucial. Cramer's condition is imposed for some variables due to the use in Lemma 3 of Bernstein's Inequality to establish the uniform order of some specific averages in probability. Thus, each of their higher moments is bounded by the second moment.

Assumption A4. Let C^k denote the class of functions that: (i) is k-times partially continuously differentiable, and (ii) all their partial derivatives up to order k are uniformly bounded. For $d = 1, \cdots, D_2$, $\Pi_d(Z_i), \phi(M_i, U_i), f_Z(Z_i), m(M_i)$, $h(U_i) \in C^s$.

A4 assumes smoothness of the regression functions and uniform bounds of their partial derivatives. This assumption, together with a "higher-order" kernel, gives a desired bias order.

Assumption A5. Denote $L_{in} \equiv \left(\frac{\log n}{n h_i^{D_i}}\right)^{\frac{1}{2}} + h_i^{s_i}$, for $i = 1, \cdots, 4$, and $L_n = \sum_{i=2}^4 L_{in}$, where $h_i \to 0$ as $n \to \infty$ and satisfy:

(i) $h_1 = n^{-\delta}$, where $\frac{1}{2s_1} < \delta < \min_{\{i=2,4\}} \frac{D_i}{D_1(2s_i+D_i)}$, and $\frac{s_1}{D_1} > \max_{\{i=2,4\}} \frac{s_i}{D_i} + \frac{1}{2}$;

(ii) $h_i = n^{-\frac{1}{2s_i+D_i}}$, $s_i \geq D_i/2+2$, for $i = 2,4$;

(iii) $h_3 = n^{-\frac{1}{2s_3+D_3}}$, $s_3 \geq D_3/2$.

Assumption A5 provides the order of all the bandwidths used in the paper. The fact that, using residual estimates $\{\hat{U}_i\}_{i=1}^n$ instead of $\{U_i\}_{i=1}^n$ has no impact on the first-order asymptotic property of our estimator, relies on under-smoothing in the first stage when regressing X on Z nonparametrically and $\Pi(z)$ has to be sufficiently smooth. For h_2, h_3, h_4, the orders are chosen optimally by minimizing the mean squared error of traditional NW kernel estimators.

By Theorem 2.6 in Li and Racine (2007), under A1-A5, for a compact subset $\mathscr{G}_Z \subset \mathbb{R}^{D_1}$, we have

$$\sup_{Z \in \mathscr{G}_Z} \left| \hat{\Pi}(Z_i) - \Pi(Z_i) \right| = O_p(L_{1n}) \tag{13}$$

where $L_{1n} = \left(\frac{\log n}{nh_1^{D_1}}\right)^{1/2} + h_1^{s_1}$. This uniform convergence rate of NW estimator in probability is used throughout this paper. Note that $\hat{f}_U(\hat{U}_i)$ and $\hat{\phi}(M_i, \hat{U}_i)$ are used to approximate $f_U(U_i)$ and $\phi(M_i, U_i)$ in η_i. In Theorem 1, we show that the uniform convergence rate of $\hat{f}_U(\hat{U}_i)$ to $f_U(U_i)$ using $\{\hat{U}_i\}_{i=1}^n$ is no different from that of the traditional Rosenblatt density estimator based on the unobserved $\{U_i\}_{i=1}^n$. A similar result hold for $\hat{\phi}(M_i, \hat{U}_i)$. All proofs of the theorems are provided in Appendix.

Theorem 1. *Under* A1-A5*, for arbitrary convex and compact subsets $\mathscr{G}_Z \subset \mathbb{R}^{D_1}$, $\mathscr{G}_U \subset \mathbb{R}^{D_2}$ and $\mathscr{G}_M \subset \mathbb{R}^{D_3}$, we have*

$$\begin{aligned} &\sup_{\{Z,U\} \in \mathscr{G}_Z \times \mathscr{G}_U} \left| \hat{f}_U(\hat{U}_i) - f_U(U_i) \right| = O_p(L_{2n}), \qquad \sup_{M \in \mathscr{G}_M} \left| \hat{f}_M(M_i) - f_M(M_i) \right| = O_p(L_{3n}), \\ &\sup_{\{Z,U,M\} \in \mathscr{G}_Z \times \mathscr{G}_U \times \mathscr{G}_M} \left| \hat{\phi}(M_i, \hat{U}_i) - \phi(M_i, U_i) \right| = O_p(L_{4n}), \end{aligned} \tag{14}$$

where $\mathscr{G}_Z \times \mathscr{G}_U$ denotes the Cartesian product of sets $\mathscr{G}_Z$ and $\mathscr{G}_U$, $L_{in} = \left(\frac{\log n}{nh_i^{D_i}}\right)^{1/2} + h_i^{s_i}$, for $i = 2,3,4$.

Note that in Theorem 1 we establish the uniform convergence rate of $\hat{f}_U(\hat{U}_i)$ and $\hat{\phi}(M_i, \hat{U}_i)$ over $\mathscr{G}_Z \times \mathscr{G}_U$ and $\mathscr{G}_Z \times \mathscr{G}_U \times \mathscr{G}_M$ separately. This is due to the fact that $\hat{U}_i$ is an estimated residual given by $\hat{U}_i = X_i - \hat{\Pi}(Z_i)$ and the uniform convergence rate of $\hat{\Pi}(Z_i)$ given in (13) is taken over a compact set $\mathscr{G}_Z$. Theorem 1 and A2 together imply

that $|\hat{\eta}_i - \eta_i| = O_p(L_n)$ uniformly, where $L_n = \sum_{i=2}^4 L_{in}$, and consequently we have $|\hat{g}_{3j} - g_{3j}| = O_p(L_n)$. With this result, we are ready to provide the uniform convergence rate of the estimators given in (9).

Theorem 2. *Under* A1-A5, *for arbitrary convex and compact subsets $\mathscr{G}_Z$, $\mathscr{G}_U$ and $\mathscr{G}_M$, we have*

$$\sup_{\{Z,U,M\}\in\mathscr{G}_Z\times\mathscr{G}_U\times\mathscr{G}_M}\left|\hat{g}_2(\hat{U}_i) - g_2(U_i)\right| \;=\; O_p\left(L_n + \frac{L_{1n}}{h_2}\right), \tag{15}$$

Similarly, we have the same uniform convergence rate of $\hat{g}_1(M_i)$, $\hat{\gamma}_1(M_i)$ and $\hat{\gamma}_2(\hat{U}_i)$, as $\hat{g}_2(\hat{U}_i)$ above.

Note that the first term in the order of (15) is not new to us, it is just a sum of uniform orders for different NW estimators. The h_2 in the denominator of the second term comes from a Taylor expansion of the kernel evaluated at the estimated residuals $\{\hat{U}_i\}_{i=1}^n$. With well chosen bandwidths in A5, it is essential to have that L_n^2, $\left(\frac{L_{1n}}{h_2}\right)^2 = o(n^{-1/2})$. This result will help establish the order of elements in $\hat{\beta} - \beta$. Note that

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2\right)^{-1}\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}(\hat{Y} - \hat{X}_2\beta), \tag{16}$$

where
$$\hat{Y} \;=\; Y - \hat{\gamma} = (Y-\gamma) - (\hat{\gamma}-\gamma) \equiv Y^* - V_Y, \qquad \hat{X}_2 \;=\; X_2 - \hat{g} = (X_2 - g) - (\hat{g} - g) \equiv X_2^* - V_X;$$
$$\gamma_i \;\equiv\; \gamma_1(M_i) + \gamma_2(U_i) - \gamma_3, \qquad g_i \;\equiv\; g_1(M_i) + g_2(U_i) - g_3;$$
$$\hat{\gamma}_i \;\equiv\; \hat{\gamma}_1(M_i) + \hat{\gamma}_2(\hat{U}_i) - \hat{\gamma}_3, \qquad \hat{g}_i \;\equiv\; \hat{g}_1(M_i) + \hat{g}_2(\hat{U}_i) - \hat{g}_3;$$
$$V_{Yi} \;=\; \hat{\gamma}_i - \gamma_i \;\equiv\; V_{Y1i} + V_{Y2i} + V_{Y3i}, \qquad V_{Xi} \;=\; \hat{g}_i - g_i \;\equiv\; V_{X1i} + V_{X2i} + V_{X3i}.$$

As we can see in (16), there are basically two parts to deal with. We need to: (i) find the asymptotic behavior of the matrix $\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2$, and (ii) establish asymptotic normality of the second term $\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}(\hat{Y} - \hat{X}_2\beta)$. By Theorem 2, we already have the uniform order of V_{Yi} and V_{Xi}, which are defined above. This result will help take care of (i). However, to establish $\sqrt{n}$ asymptotic normality for the second term, we need to employ a U-statistics of degree 3. Yao and Martins-Filho (2013) provides a direct and convenient method to characterize the asymptotic magnitude of each component in the H-decomposition of a U-statistics, and many places in our proof are built on their results.

In Theorem 3, we derive the $\sqrt{n}$ asymptotic normality of $\hat{\beta}$ by showing that $\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2 \xrightarrow{p} \Phi_0$ and $\frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}(\hat{Y} - \hat{X}_2\beta) \xrightarrow{d} \mathcal{N}(\Phi_1 + \Phi_2)$, where Φ_0, Φ_1 and Φ_2 are given in Theorem 3.

Theorem 3. *Under* A1-A5, *assuming that matrix* Φ_0 *exists and is nonsingular, we have*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \Phi_0^{-1}(\Phi_1 + \Phi_2)\Phi_0^{-1}\right) \tag{17}$$

where

$$\begin{aligned}
\Phi_{0_{(j,k)}} &= \mathrm{E}\left[\eta_t\left(X_{2t,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right)\left(X_{2t,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right)\right];\\
\Phi_{1_{(j,k)}} &= \mathrm{E}\left[\eta_t^2\left(X_{2t,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right)\left(X_{2t,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right)\right]\sigma_v^2;\\
\Phi_{2_{(j,k)}} &= \mathrm{E}\left[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2}\mathrm{E}\Big(\left(\Pi_{2j}(Z_i) - U_{2tj} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}\right)D_d h(U_t)\eta_t\Big|Z_i\Big)\right.\\
&\quad \left.\times\mathrm{E}\Big(\left(\Pi_{2k}(Z_i) - U_{2tk} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}\right)D_\delta h(U_t)\eta_t\Big|Z_i\Big)\mathrm{E}(U_{id}U_{i\delta}\big|Z_i)\right],\\
&\quad \textit{for} \quad j,k = 1,\cdots,D_{22}.
\end{aligned}$$

By Theorem 3, $\hat{\beta}$ is asymptotically unbiased, and has an explicit covariance for the limiting distribution. For statistical inference, we provide consistent estimators for Φ_i, $i = 1,2,3$. By proof of Theorem 3, we have that

$$\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2 \xrightarrow{p} \Phi_0, \qquad \frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}v \xrightarrow{d} \mathcal{N}(0,\Phi_1), \qquad \frac{1}{\sqrt{n}}\hat{X}_2'\hat{\eta}(V_{Y2} - V_{X2}\beta) \xrightarrow{d} \mathcal{N}(0,\Phi_2).$$

Hence, it's easy to show that

$$\hat{\Phi}_0 = \frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2, \quad \hat{\Phi}_1 = \frac{1}{n}\hat{X}_2'\hat{\eta}\hat{v}\hat{v}'\hat{\eta}\hat{X}_2, \quad \hat{\Phi}_2 = \frac{1}{n}\hat{X}_2'\hat{\eta}(V_{Y2} - V_{X2}\hat{\beta})(V_{Y2} - V_{X2}\hat{\beta})'\hat{\eta}\hat{X}_2 \tag{18}$$

are consistent estimators for Φ_0, Φ_1 and Φ_2 separately, where $\hat{v} \equiv Y - X_2\hat{\beta} - \hat{\beta}_0 - \hat{m} - \hat{h}$.

Given Theorems 2, 3 and (11), we have the uniform convergence rate of $\hat{m}(M_i)$ and $\hat{h}(\hat{U}_i)$ at $O_p\left(L_n + \frac{L_{1n}}{h_2}\right)$, which generally worse than that of the traditional NW estimator due to the presence of h_2 in second term. However, it is possible to gain a better rate by implementing a second stage estimator for $m(M_i)$ and $h(U_i)$, or even possibly for β.

With $\hat{\beta}$, $\hat{\beta}_0$, $\hat{m}(M_i)$ and $\hat{h}(\hat{U}_i)$, we can estimate $m(M_i)$ and $h(U_i)$ by $\tilde{m}(M_i)$ and $\tilde{h}(\hat{U}_i)$ using local linear regression:

$$\begin{aligned}\left(\tilde{m}(M_i), \tilde{\delta}(M_i)\right) &= \underset{m,\delta}{\operatorname{argmin}} \frac{1}{n}\sum_{t=1}^{n} \left(Y_{t1} - m - (M_t - M_i)'\delta\right)^2 K_3\left(\frac{M_t - M_i}{h_3}\right), \\ \left(\tilde{h}(\hat{U}_i), \tilde{\eta}(\hat{U}_i)\right) &= \underset{h,\eta}{\operatorname{argmin}} \frac{1}{n}\sum_{t=1}^{n} \left(Y_{t2} - h - (\hat{U}_t - \hat{U}_i)'\eta\right)^2 K_2\left(\frac{\hat{U}_t - \hat{U}_i}{h_2}\right),\end{aligned} \tag{19}$$

where $Y_{t1} = Y_t - X_{2t}\hat{\beta} - \hat{\beta}_0 - \hat{h}(\hat{U}_t)$, $Y_{t2} = Y_t - X_{2t}\hat{\beta} - \hat{\beta}_0 - \hat{m}(M_t)$.

And a second stage estimator for β is given as

$$\tilde{\beta} = (X_2'X_2)^{-1}X_2'\tilde{Y} \tag{20}$$

where $\tilde{Y}$ is $n \times 1$ with ith element $\tilde{Y}_i = Y_i - \tilde{m}(M_i) - \tilde{h}(\hat{U}_i) - \hat{\beta}_0$, and $X_2 = (X_{21}', \cdots, X_{2n}')'$.

In this paper, we will not provide asymptotic properties for these second stage estimators and we will leave them for future study. However, we will provide a simple Monte Carlo study for both estimators in the two stages in the next section.

4 Monte Carlo Study

In this section, we investigate the finite sample performance of the proposed estimators $\hat{\beta}$, $\hat{m}(\cdot)$, and $\tilde{\beta}$, $\tilde{m}(\cdot)$ for future reference. Consider the following data generating processes (DGPs):

$$\text{DGP}_1: \quad Y_i = \text{Ln}(|X_{1i} - 1| + 1)\,\text{sgn}(X_{1i} - 1) + X_{2i}\beta + \beta_0 + \varepsilon_i$$

$$\text{DGP}_2: \quad Y_i = \frac{\exp(X_{1i})}{1 + c\exp(X_{1i})} + X_{2i}\beta + \beta_0 + \varepsilon_i$$

for $i = 1, \cdots, n$. The sample size n is set at 100 and 400. In both DGPs, Z_{1i} and Z_{2i} are generated independently from a $N(0,1)$, and construct $X_{1i} = Z_{1i} + Z_{2i} + U_{1i}$ and $X_{2i} = Z_{1i}^2 + Z_{2i}^2 + U_{2i}$. ε_i and $U_i = (U_{1i}, U_{2i})$ are generated as $\begin{pmatrix} \varepsilon_i \\ U_i \end{pmatrix} \sim NID\left(0, \begin{pmatrix} 1 & \theta & \theta \\ \theta & 1 & \theta^2 \\ \theta & \theta^2 & 1 \end{pmatrix}\right)$, where the values $\theta = 0.3, 0.6$, and 0.9 indicate weak, moderate, and strong endogeneity. It is easy to verify that $E(\varepsilon_i|Z_i) = 0$, $E(U_i|Z_i) = 0$, and thus $E(\varepsilon_i|U_i, Z_i) = E(\varepsilon_i|U_i) = \dfrac{\theta}{1+\theta^2}(U_{1i} + U_{2i})$.

We set the parameters $\beta = 1, \beta_0 = 1$ and $c = 3$, and perform 1000 repetitions for each experiment design.

The implementation of the estimator requires a choice of kernel function $K_i(\cdot)$ for $i = 1, \cdots, 4$ and bandwidth sequences. For all kernels, products of an univariate Epanechnikov kernel were used: $k(x) = \frac{3}{4\sqrt{5}}(1 - x^2)I(|x| < \sqrt{5})$, where $I(\cdot)$ is an indicator function. Note that even though Epanechnikov kernel is not continuously differentiable at the boundaries of its support, it does satisfy all other assumptions given in A1. We are using it instead of the kernel constructed by Gaussian distribution since in finite sample it performs better. Bandwidths were selected with the simple rule-of-thumb bandwidth $1.25\hat{\sigma}(W_i)h_i$, for $i = 1, 2, 3, 4$, where $\hat{\sigma}(W_i)$ is the sample standard deviation of the variable W_i, $W_1 = Z_i$, $W_2 = \hat{U}_i$, $W_3 = (X_{1i}, Z_{1i})$, and $W = (X_{1i}, Z_{1i}, \hat{U}_i)$. In our two DPGs, we have $D_1 = 2$, $D_2 = 2$, $D_3 = 1$, $D_4 = 3$. Thus we choose $\delta = \frac{1}{11}$, $s_1 = 6$, $s_2 = 4$, $s_3 = 4$, $s_4 = 4$.

In Table 1, we list the finite sample performances in terms of bias (B), standard deviation (S), and root mean squared error (R) for the estimation of β, and the mean of root mean squared error (M) for estimating $m(\cdot)$ obtained by averaging across the realized values of (X_{1i}, Z_{1i}). Results for both of the two stages estimators $\left(\hat{\beta}, \hat{m}(\cdot)\right)$ and $\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$ are listed. To avoid any extreme estimates, results are only shown for the $10-90\%$ quantile range of sample estimates. As it is shown in the table, the estimator's performance, in terms of the above measures, improves significantly with the sample size. For example, for DGP_1, when $\theta = 0.3$, root mean squared error of $\hat{\beta}$ drops nearly 40% from 0.09 to 0.055 when we increase the sample size from 100 to 400. Besides, it turns out that our estimators have correctly accounted for the endogeneity problem as controlling for the DGP and the sample size, the root mean squared error of $\hat{\beta}$ does not change much as the degree of endogeneity (θ) increases. As we predicted, the second stage estimators $\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$ outperform the first stage $\left(\hat{\beta}, \hat{m}(\cdot)\right)$ in all aspects, suggesting a significant improvement in asymptotic efficiency for both parametric and nonparametric estimation. Superiority of our estimator compared with those in Ai and Chen (2003) and Otsu (2011) in terms of finite sample performance will not be provided here, as it is already established in Martins-Filho and Yao (2012).

Table 1
Finite sample performances.

	$\theta = 0.3$				$\theta = 0.6$				$\theta = 0.9$			
	B	S	R	M	B	S	R	M	B	S	R	M
DGP_1	$n = 100$											
$\left(\hat{\beta}, \hat{m}(\cdot)\right)$	0.065	0.062	0.09	0.66	0.069	0.056	0.089	0.644	0.069	0.057	0.09	0.625
$\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$	0.004	0.08	0.08	0.427	0.006	0.074	0.074	0.417	0.0001	0.076	0.076	0.417
	$n = 400$											
$\left(\hat{\beta}, \hat{m}(\cdot)\right)$	0.045	0.032	0.055	0.677	0.042	0.032	0.053	0.658	0.048	0.031	0.057	0.634
$\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$	-0.029	0.044	0.052	0.397	-0.037	0.044	0.057	0.388	-0.034	0.04	0.053	0.388
DGP_2	$n = 100$											
$\left(\hat{\beta}, \hat{m}(\cdot)\right)$	0.078	0.06	0.098	1.38	0.089	0.064	0.109	1.369	0.105	0.064	0.123	1.353
$\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$	-0.013	0.081	0.082	1.07	-0.001	0.087	0.087	1.082	0.017	0.087	0.089	1.098
	$n = 400$											
$\left(\hat{\beta}, \hat{m}(\cdot)\right)$	0.072	0.032	0.079	1.417	0.069	0.034	0.077	1.41	0.086	0.034	0.092	1.387
$\left(\tilde{\beta}, \tilde{m}(\cdot)\right)$	-0.047	0.043	0.064	1.027	-0.051	0.047	0.07	1.034	-0.03	0.049	0.057	1.052

5 Conclusion and extensions

In this paper we study a partially linear model in triangular systems where endogenous variables appear both in nonparametric and linear components. The estimation is based upon the control function approach of Newey et al. (1999) and an additive regression estimation method of Kim et al. (1999). NW kernel estimator is used for the nonparametric estimation. We establish the $\sqrt{n}$ asymptotic normality of our estimator for the linear component and uniform convergence rate of estimator for the nonparametric component. Estimators for the covariance of the limiting distribution of the parametric estimator are provided. Our simple Monte Carlo study suggests good finite sample properties, and may significantly outperform the estimators of (Ai and Chen, 2003) and Otsu (2011) as Martins-Filho and Yao (2012) implies.

In the future, there are still some aspects to be investigated, for example, the asymptotic normality of the nonparametric component, optimal bandwidths selection. And our theoretical results can be extended in three directions. First, the Monte Carlo results reveal that, one can pursue one step further to obtain a potentially asymptotically more efficient estimator for both the nonparametric and linear component functions, as we discussed in Remark 8. Second,

like Newey et al. (1999), Kim et al. (1999), Ai and Chen (2003) and Otsu (2011), we study an IID process. A potential extension would be allowing some weak dependence like Su and Ullah (2008), and investigate whether theorems exhibited in our paper still hold. Third, we will provide some empirical applications of our estimator. For example, we can apply our estimators to the empirical model of Lessmann (2014), and test the inverted-U relationship between spatial development and economic development directly, with the endogeneity problem being taken care of.

Appendix

This appendix presents the proof of the main theorems and lemmas. We give all the notation used in the proof and a basic introduction to the U-statistics.

Throughout the proofs, C will represent an inconsequential and arbitrary constant that may take different values in different context. For a scalar variable x, $f'(x)$ denotes the derivative of $f(x)$ evaluated at x. For $D\times 1$ vectors γ,β, define $\gamma^{\beta}=\prod_{d=1}^{D}\gamma_d^{\beta_d}$, $|\beta|=\sum_{d=1}^{D}\beta_d$, $D_d f(\gamma)=\frac{\partial}{\partial d}f(\gamma)$, $D_{dk}^2 f(\gamma)=\frac{\partial^2}{\partial d\partial k}f(\gamma)$, $D^{\beta}f(\gamma)=\frac{\partial^{|\beta|}}{\partial_1^{\beta_1}\cdots\partial_D^{\beta_D}}f(\gamma)$. $\mathbf{J}f(\gamma)$ and $\mathbf{H}f(\gamma)$ denote the Jacobian and Hessian matrix of $f(\gamma)$. Note that for a scalar function $f(\gamma)$, $\mathbf{J}f(\gamma)$ is exactly the transpose of the gradient vector of $f(\gamma)$. $A\times B$ denotes the Cartesian product of two sets A and B. χ_A denotes the indicator function for the set A, $P(A)$ denotes the probability of event A in the probability space $(\Omega,\mathscr{F},P)$, $\mathrm{V}(\cdot)$ denotes variance.

U-statistics will be repeatedly used in the proofs. Let $\{P_i\}_{i=1}^{n}$ be a sequence of IID random variables and $\phi_n(P_{i_1},\cdots,P_{i_k})$ be a symmetric kernel function that depends on n. Then a U-statistic U_n of degree k is defined as

$$U_n=\binom{n}{2}^{-1}\sum_{(n,k)}\phi_n(P_{i_1},\cdots,P_{i_k}),$$

where $\sum_{(n,k)}$ denotes the sum over all subsets $1\le i_1<\cdots<i_k\le n$ of $\{1,\cdots,n\}$. Now let $\phi_{cn}(z_1,\cdots,z_c)=\mathrm{E}(\phi_n(P_1,\cdots,P_c,P_{c+1},\cdots,P_k)|P_1=p_1,\cdots,P_c=p_c)$, $\sigma_{cn}^2=\mathrm{V}(\phi_{cn}(P_1,\cdots,P_c))$ and $\theta_n=\mathrm{E}(\phi_n(P_{i_1},\cdots,P_{i_k})$. In addition, recursively define $h_n^{(1)}(p_1)=\phi_{1n}(p_1)-\theta_n,\cdots,h_n^{(c)}(p_1,\cdots,p_c)=\phi_{cn}(p_1,\cdots,p_c)-\sum_{j=1}^{c-1}\sum_{(c,j)}h_n^{(j)}(p_{i_1},\cdots,p_{i_j})-\theta_n$ for $c=2,\cdots,$

k. By Hoeffding's H-decomposition in Hoeffding (1961) we have

$$U_n = \theta_n + \sum_{j=1}^{k} \binom{k}{j} H_n^{(j)}(P_{i_1}, \cdots, P_{i_j}),$$

where $H_n^{(j)}(P_{i_1}, \cdots, P_{i_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(P_{i_1}, \cdots, P_{i_j})$. The order of U_n can be determined by studying each $H_n^{(j)}$ and θ_n in the finite sum. By Theorem 1 in Yao and Martins-Filho (2013), the order of $H_n^{(j)}$ is determined by n and the leading variance σ_{jn}^2. Throughout the proofs, we will use $\{P_i\}_{i=1}^n$ and the above notation to characterize the U-statistics of interest, denoted by U_n .

Theorem 1 *Proof.* By uniform convergence rate of Rosenblatt density estimator given in Theorem 1.4 of Li and Racine (2007), we have $\sup_{M \in \mathscr{G}_M} \left| \hat{f}_M(M_i) - f_M(M_i) \right| = O_p(L_{3n})$. Similarly, for the first equation in (14), we only need to focus on $|\hat{f}_U(\hat{U}_i) - \hat{f}_U(U_i)|$.

Denote $\hat{K}_{2ti} = K_2\left(\frac{\hat{U}_t - \hat{U}_i}{h_2}\right)$, $K_{2ti} = K_2\left(\frac{U_t - U_i}{h_2}\right)$, and other kernels similarly. Since K_2 is 4-times partially continuously differentiable, by Taylor Theorem,

$$\begin{aligned}
\left| \hat{f}_U(\hat{U}_i) - \hat{f}_U(U_i) \right| &= \left| \frac{1}{n h_2^{D_2}} \sum_{t=1}^{n} \left(\hat{K}_{2ti} - K_{2ti} \right) \right| \\
&= \left| \frac{1}{n h_2^{D_2}} \sum_{t=1}^{n} \left(\sum_{|\beta|=1}^{3} \frac{H^\beta}{|\beta|!} D^\beta K_{2ti} + \sum_{|\beta|=4} \frac{H^\beta}{|\beta|!} D^\beta K_2 \left(\frac{U_t - U_i}{h_2} + \lambda H \right) \right) \right| \\
&\leq \sum_{i=1}^{4} |T_i|
\end{aligned}$$

where $H \equiv \frac{1}{h_2}(\hat{U}_t - U_t) - \frac{1}{h_2}(\hat{U}_i - U_i)$, $\lambda \in (0,1)$.

Next, examine the uniform order of $|T_i|$ over $\{Z,U\} \in \mathscr{G}_Z \times \mathscr{G}_U$ for $i = 1, \cdots, 4$.

1.
$$\begin{aligned}
|T_1| \leq & \sum_{|\beta|=1} \left| \frac{1}{n h_2^{D_2}} \sum_{t=1}^{n} H^\beta D^\beta K_{2ti} \right| \\
\leq & \sum_{d=1}^{D_2} \left(\left| \frac{1}{n h_2^{D_2+1}} \sum_{t=1}^{n} (\hat{U}_{id} - U_{id}) D_d K_{2ti} + \frac{1}{n h_2^{D_2+1}} \sum_{t=1}^{n} (\hat{U}_{td} - U_{td}) D_d K_{2ti} \right| \right) \\
\equiv & \sum_{d=1}^{D_2} (|T_{11} + T_{12}|).
\end{aligned}$$

1.1. $|T_{11}| \leq \left|\hat{U}_{id} - U_{id}\right| \left|(nh_2^{D_2+1})^{-1} \sum_{t=1}^n D_d K_{2ti}\right| \equiv \left|\hat{U}_{id} - U_{id}\right| \; C_1(U_i)$. It can be shown that by Lemma 4,

$$\mathrm{E}\left(C_1(U_i)\right) \to D_d f_U(U_i), \qquad \sup_{U \in \mathscr{G}_U} |C_1(U_i) - \mathrm{E}\left(C_1(U_i)\right)| = O_p\left(\left(\frac{\log n}{nh_2^{D_2+2}}\right)^{\frac{1}{2}}\right) = o_p(1)$$

Thus $\sup_{U \in \mathscr{G}_U} |C_1(U_i)| = O_p(1)$. Note that $\left|\hat{U}_{id} - U_{id}\right| = \left|\hat{\Pi}_d(Z_i) - \Pi_d(Z_i)\right|$, and by uniform convergence rate of Nadaraya-Watson estimator, we have $\sup_{Z \in \mathscr{G}_Z} \left|\hat{U}_{id} - U_{id}\right| = O_p(L_{1n})$. Consequently, $|T_{11}| = O_p(L_{1n})$ uniformly.

1.2. Given $\hat{\Pi}_d(Z_t) = \frac{1}{nh_1^{D_1} \hat{f}_Z(Z_t)} \sum_{l=1}^n K_{1lt} X_{l,d}$, and $\hat{f}_Z(Z_t) = \frac{1}{nh_1^{D_1}} \sum_{l=1}^n K_{1lt}$, we have

$$\begin{aligned}
-(\hat{U}_{td} - U_{td}) = & \quad \hat{\Pi}_d(Z_t) - \Pi_d(Z_t) \\
= & \quad \frac{1}{nh_1^{D_1} \hat{f}_Z(Z_t)} \sum_{l=1}^n K_{1lt} \Big(U_{ld} + \Pi_d(Z_l) - \Pi_d(Z_t)\Big) \\
= & \quad \left\{\frac{1}{nh_1^{D_1} f_Z(Z_t)} \sum_{l=1}^n K_{1lt} \Big(U_{ld} + \Pi_d(Z_l) - \Pi_d(Z_t)\Big)\right\} (1 + O_p(L_{1n})) \qquad \text{(A.1)}
\end{aligned}$$

by the uniform order of Rosenblatt density estimator $\hat{f}_Z(Z_t)$. Thus

$$\begin{aligned}
|T_{12}| \leq & \quad \left| \frac{1}{n^2} \sum_{t=1}^n \sum_{l=1}^n \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} U_{ld} \right. \\
& \quad \left. + \frac{1}{n^2} \sum_{t=1}^n \sum_{l=1}^n \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} \Big(\Pi_d(Z_l) - \Pi_d(Z_t)\Big) \right| \Big(1 + O_p(L_{1n})\Big) \\
\equiv & \quad \Big|T_{121} + T_{122}\Big| \Big(1 + O_p(L_{1n})\Big) \\
|T_{121}| = & \quad \left| \frac{1}{n^2} \sum_{t=1}^n \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_1(0) D_d K_{2ti} U_{td} \right. \\
& \quad \left. + \frac{1}{2} \left(O\left(n^{-3} + \binom{n}{2}^{-1}\right)\right) \sum_{\substack{t=1 \\ t \neq l}}^n \sum_{l=1}^n \frac{1}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} K_{1lt} D_d K_{2ti} U_{ld} \right| \\
\equiv & \quad \Big|E_1 + E_2\Big|.
\end{aligned}$$

We can show that $|E_1| = O_p\Big((nh_1^{D_1} h_2)^{-1}\Big)$ by Markov's Inequality, and $|E_2| \leq C|U_n|$, where $U_n = \binom{n}{2}^{-1} \sum_{t=1}^n \sum_{\substack{l=1 \\ t \neq l}}^n \frac{K_{1lt} D_d K_{2ti}}{h_1^{D_1} h_2^{D_2+1} f_Z(Z_t)} U_{ld} \equiv \binom{n}{2}^{-1} \sum_{t=1}^n \sum_{\substack{l=1 \\ t \neq l}}^n \psi_{nlt} \equiv \binom{n}{2}^{-1} \sum_{t=1}^n \sum_{\substack{l=1 \\ t<l}}^n \phi_{nlt} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ is a U-statistic. $\theta_n = \mathrm{E}(\phi_{nlt}) = 0$ in this case. $H_n^{(1)} = \frac{1}{n} \sum_{l=1}^n h_n^{(1)}(U_i, P_l) = \frac{1}{n} \sum_{l=1}^n \phi_{1n}(U_i, P_l) = \frac{1}{n} \sum_{l=1}^n \mathrm{E}(\phi_{nlt} | U_i, P_l) = \frac{1}{n} \sum_{l=1}^n U_{ld} c(U_i, Z_l)$, where $c(U_i, Z_l) \equiv \int K_1(\gamma) K_2(\psi) D_d f_{U|Z}(U_i + h_2\psi | Z_l - h_1\gamma) d\gamma d\psi$. Given Cramer's

condition in A3, by Lemma 3, we have $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|H_n^{(1)}| = O_p\left((\log n/n)^{1/2}\right)$. For $H_n^{(2)}$, by Theorem 1 in Yao and Martins-Filho (2013), $H_n^{(2)} = (\sigma_{2n}^2/n^2)^{1/2}O_p(1)$. And $\sigma_{2n}^2 \equiv \mathrm{V}(\phi_{nlt}) = \mathrm{E}(\phi_{nlt}^2) \leq 4\mathrm{E}(\psi_{nlt}^2) = O\left((h_1^{D_1}h_2^{D_2+2})^{-1}\right)$. Thus $H_n^{(2)} = (n^2h_1^{D_1}h_2^{D_2+2})^{-1/2}O_p(1)$ uniformly. In sum, $|T_{121}| = O_p\Big((nh_1^{D_1}h_2)^{-1} + (\log n/n)^{1/2} + (n^2h_1^{D_1}h_2^{D_2+2})^{-1/2}\Big) = O_p(L_{1n})$ uniformly by A5.

The order of $|D_{122}|$ could be analyzed in the same way, given that Π and f_Z are s_1 times partially continuously differentiable, and K_1 is a multivariate kernel of order s_1, we have

$|T_{122}| = O_p\left(h_1^{s_1} + (\log n/n)^{1/2} + (n^2h_1^{D_1-2}h_2^{D_2+2})^{-1/2}\right) = O_p(L_{1n})$ uniformly by A5.

In sum, $\sup\limits_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|T_1| = O_p(L_{1n})$.

2. $|T_2| \leq \sum_{|\beta|=2}\left|\frac{1}{nh_2^{D_2}}\sum_{t=1}^n H^\beta D^\beta K_{2ti}\right|$, when 1 appears in the d^{th} and k^{th} position of β, we have:
$$\left|\frac{1}{nh_2^{D_2}}\sum_{t=1}^n H^\beta D^\beta K_{2ti}\right| \leq \left|\frac{1}{2nh_2^{D_2+2}}\sum_{t=1}^n\left[(\hat{U}_{td}-U_{td})-(\hat{U}_{id}-U_{id})\right]\left[(\hat{U}_{tk}-U_{tk})-(\hat{U}_{ik}-U_{ik})\right]D_{dk}^2K_{2ti}\right|.$$

Since $\sup_{Z\in\mathscr{G}_Z}\left|\hat{U}_{ab}-U_{ab}\right| = O_p(L_{1n})$, for $a=i,j$ and $b=d,k$, we have $|T_2| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right)\frac{1}{nh_2^{D_2}}\sum_{t=1}^n\left|D_{dk}^2K_{2ti}\right| \equiv O_p\left(\frac{L_{1n}^2}{h_2^2}\right)C_2(U_i)$ uniformly. As $\mathrm{E}|C_2(U_i)| = O(1)$ uniformly for $U_i\in\mathscr{G}_U$, we have $\sup_{U\in\mathscr{G}_U}|C_2(U_i)| = O_p(1)$ by Markov's Inequality. Thus, $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|T_2| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right)$.

3. Similarly, $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|T_3| = O_p\left(\frac{L_{1n}^3}{h_2^3}\right)$.

4. $|T_4|$ is different from $|T_2|$ and $|T_3|$ in that $\sup\limits_{U\in\mathscr{G}_U}|C_4(U_i)| = O_p(1/h_2^{D_2})$, where $C_4(U_i) \equiv \frac{1}{nh_2^{D_2}}\sum_{t=1}^n\left|D^\beta K_{2ti}\right|$, for any $|\beta|=4$, thus $\sup\limits_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|T_4| = O_p\left(\frac{L_{1n}^4}{h_2^{D_2+4}}\right)$.

By A5, it can be shown that $|T_2|,|T_3|,|T_4| = o_p(n^{-1/2})$, and $L_{1n} = O(L_{2n})$, which gives us

$$\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|\hat{f}_U(\hat{U}_i) - f_U(U_i)| = O_p(L_{2n})$$

Uniform order of $\left|\hat{\phi}(M_i,\hat{U}_i) - \phi(M_i,U_i)\right|$ is derived in the similar way under A5.

□

Theorem 3 *Proof.* We start with the j^{th} element of $\hat{g}_2(\hat{U}_i) - g_2(U_i)$. Note that

$$\begin{aligned}
\hat{g}_2(\hat{U}_i) - g_2(U_i) &= \frac{1}{nh_2^{D_2}\hat{f}_U(\hat{U}_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\hat{\eta}_t X_{2t,j} - g_{2j}(U_i) \\
&= \frac{1}{nh_2^{D_2}\hat{f}_U(\hat{U}_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\underbrace{\Big\{(\hat{\eta}_t-\eta_t)X_{2t,j} + v_{X2t,j} + \big((g_{2j}(U_t) - g_{2j}(U_i)\big)\Big\}}_{C_{X2ti}} \\
&= \Bigg\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}K_{2ti}C_{X2ti} + \frac{1}{nh_2^{D_2+1}f_U(U_i)}\sum_{t=1}^{n}\mathbf{J}K_{2ti}\big(\hat{U}_t - U_t - (\hat{U}_i - U_i)\big)C_{X2ti} \\
&\quad + \frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}R_{ti}C_{X2ti}\Bigg\}\big(1 + O_p(L_{2n})\big) \qquad \text{(A.2)} \\
&\equiv \left(\sum_{k=1}^{3}T_k\right)\big(1 + O_p(L_{2n})\big),
\end{aligned}$$

where R_{ti} is the remainder term of a Taylor expansion of $\hat{K}_{2ti}$ at $\left(\frac{U_t - U_i}{h_2}\right)$.

We will show that $T_1 = O_p(L_n)$, $T_2 = O_p\left(\frac{L_{1n}}{h_2}\right)$, and $T_3 = o_p(n^{-1/2})$, which completes the proof.

1. Let $T_1 \equiv \sum_{k=1}^{3} T_{1k}$, according to the three components in C_{X2ti}. By Theorem 2 and A2, we have that

$$\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|\hat{\eta}_t - \eta_t| = O_p(L_{2n} + L_{3n} + L_{4n}) \equiv O_p(L_n).$$

By Markov's Inequality, $|T_{11}| \le O_p(L_n)\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}|K_{2ti}X_{2t,j}| = O_p(L_n)$, since by Lemma 1 and A3 ,

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{nh_2^{D_2}}\sum_{t=1}^{n}|K_{2ti}X_{2t,j}|\right) &= \frac{1}{h_2^{D_2}}\mathrm{E}\Big(|K_{2ti}||\Pi_{2j}(Z_t) + U_t|\Big) \\
&= \int |K_2(\gamma)||\Pi_{2j}(Z_t) + U_i + h_2\gamma|f_{UZ}(U_i + h_2\gamma, Z_t)d\gamma dZ_t \\
&\to \int |K_2(\gamma)|d\gamma\Big(\int |\Pi_{2j}(Z_t)|f_{UZ}(U_i, Z_t)dZ_t + U_i f_U(U_i)\Big) < \infty.
\end{aligned}$$

By Chebyshev's Inequality, we have $|T_{12}| = \frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}\left|K_{2ti}v_{X2t,j}\right| = O_p\left((nh_2^{D_2})^{-1/2}\right) = O_p(L_n)$, since $\mathrm{E}(T_{12}) = 0$, and $\mathrm{V}(T_{12}) = \mathrm{E}(T_{12}^2) \le \frac{C}{nh_2^{2D_2}}\mathrm{E}\big(K_{2ti}^2 v_{X2t,j}^2\big) = O\Big((nh_2^{D_2})^{-1}\Big)$.

For T_{13}, note that by Taylor Theorem,

$$\begin{aligned}
\mathrm{E}(T_{13}) &= \frac{1}{h_2^{D_2} f_U(U_i)} \mathrm{E}\Big(K_{2ti}\big(g_{2j}(U_t) - g_{2j}(U_i)\big)\Big) \\
&= \frac{1}{f_U(U_i)} \int K_2(\gamma)\big(g_{2j}(U_i + h_2\gamma) - g_{2j}(U_i)\big) f_U(U_i + h_2\gamma) d\gamma \\
&= O(h_2^{s_2}),
\end{aligned}$$

since K_2 is of order s_2, $g_{2j}(U_t), f_U(U_t) \in C^{s_2}$ and all the partial derivatives of $g_{2j}(U_t)$ up to order s_2 are uniformly bounded by A4. $\mathrm{V}(T_{13}) \leq \mathrm{E}(T_{13}^2) \leq \frac{C}{nh_2^{2D_2}} \mathrm{E}\Big(K_{2ti}^2\big(g_{2j}(U_t) - g_{2j}(U_i)\big)^2\Big) = O\big((nh_2^{D_2-2})^{-1}\big) = o(1)$. Thus, $|T_{13}| = O_p(h_2^{s_2}) = O_p(L_n)$.

2. For T_2, we have

$$\begin{aligned}
T_2 &= \frac{1}{nh_2^{D_2+1} f_U(U_i)} \sum_{t=1}^{n} \mathbf{J} K_{2ti}\big(\hat{U}_t - U_t - (\hat{U}_i - U_i)\big) C_{X2ti} \\
&= O_p\Big(\frac{L_{1n}}{h_2}\Big) \sum_{d=1}^{D_2} \frac{1}{nh_2^{D_2} f_U(U_i)} \sum_{t=1}^{n} \Big| D_d K_{2ti}\Big((\hat{\eta}_t - \eta_t) X_{2t,j} + v_{X2t,j} + \big((g_{2j}(U_t) - g_{2j}(U_i)\big)\Big)\Big| \\
&= O_p\Big(\frac{L_{1n}}{h_2}\Big),
\end{aligned}$$

similarly as finding order of $|T_{11}|$ by Markov's Inequality.

3. R_{ti} is the remainder term of a Taylor expansion of $\hat{K}_{2ti}$ at $\Big(\frac{U_t - U_i}{h_2}\Big)$, thus $R_{ti} = \sum_{|\beta|=2}^{3} \frac{1}{|\beta|!} D^\beta K_{2ti} H^\beta$ $+ \sum_{|\beta|=4} \frac{1}{4!} D^\beta K_2\Big(\frac{\hat{U}_{ti} - U_{ti}}{h_2}\Big) H^\beta$, where $\Big(\frac{\hat{U}_{ti} - U_{ti}}{h_2}\Big) \equiv \Big(\frac{\hat{U}_i - U_i}{h_2}\Big) + \lambda H$, $\lambda \in (0,1)$, and $H = \frac{1}{h_2}\big(\hat{U}_t - U_t - (\hat{U}_i - U_i)\big)$. Thus, let $T_3 \equiv \sum_{k=1}^{3} T_{3k}$, with

$$\begin{aligned}
T_{31} &= \sum_{d=1}^{D_2} \sum_{l=1}^{D_2} \frac{1}{2nh_2^{D_2+2} f_U(U_i)} \sum_{t=1}^{n} D_{dl}^2 K_{2ti}\Big(\hat{U}_{td} - U_{td} - (\hat{U}_{id} - U_{id})\Big)\Big(\hat{U}_{tl} - U_{tl} - (\hat{U}_{il} - U_{il})\Big) C_{X2ti} \\
&\leq O_p\left(\frac{L_{1n}^2}{h_2^2}\right) \frac{1}{nh_2^{D_2}} \sum_{t=1}^{n} \left| D_{dl}^2 K_{2ti} C_{X2ti}\right| = O_p\left(\frac{L_{1n}^2}{h_2^2}\right)
\end{aligned}$$

by Lemma 1 and A3. Similarly, $T_{32} = O_p\Big(\frac{L_{1n}^3}{h_2^3}\Big)$. By A1, $T_{33} \leq O_p\Big(\frac{L_{1n}^4}{h_2^{D_2+4}}\Big) \frac{1}{n} \sum_{t=1}^{n} \big|C_{X2ti}\big| = O_p\Big(\frac{L_{1n}^4}{h_2^{D_2+4}}\Big)$. By A5, we can show that $|T_3| = O_p\Big(\frac{L_{1n}^2}{h_2^2} + \frac{L_{1n}^3}{h_2^3} + \frac{L_{1n}^4}{h_2^{D_2+4}}\Big) = o_p(n^{-1/2})$ uniformly.

Combining 1-3, we have $\sup_{\{Z,U\}\in\mathscr{G}_Z\times\mathscr{G}_U}|\hat{g}_2(\hat{U}_i)-g_2(U_i)| = O_p\left(L_n+\frac{L_{1n}}{h_2}\right)$. For $\hat{g}_{1j}(M_i)-g_{1j}(M_i)$, note that

$$\begin{aligned}\hat{g}_{1j}(M_i)-g_{1j}(M_i) &= \frac{1}{nh_2^{D_2}\hat{f}_M(\hat{M}_i)}\sum_{t=1}^{n}K_{3ti}\hat{\eta}_t X_{2t,j}-g_{1j}(M_i)\\ &= \left\{\frac{1}{nh_2^{D_2}f_M(M_i)}\sum_{t=1}^{n}K_{3ti}\underbrace{\left\{(\hat{\eta}_t-\eta_t)X_{2t,j}+v_{X1t,j}+\big((g_{1j}(M_t)-g_{1j}(M_i)\big)\right\}}_{C_{X1ti}}\right\}\big(1+O_p(L_{3n})\big) \quad \text{(A.3)}\end{aligned}$$

Thus order of $|\hat{g}_2(\hat{U}_i)-g_2(U_i)|$ can be found similarly to $|T_1|$ in part 1. For $|\hat{g}_{3j}-g_{3j}|$, we have

$$\begin{aligned}\hat{g}_{3j}-g_{3j} &= \frac{1}{n}\sum_{t=1}^{n}\hat{\eta}_t X_{2t,j}-\mathrm{E}(\eta_i X_{2i,j}) = \frac{1}{n}\sum_{t=1}^{n}(\hat{\eta}_t-\eta_t)X_{2t,j}+\frac{1}{n}\sum_{t=1}^{n}\eta_t X_{2t,j}-\mathrm{E}(\eta_i X_{2i,j})\\ &= O_p(L_n)+O_p(n^{-1/2}) = O_p(L_n)\end{aligned}$$

As to $|\hat{\gamma}_k|$, proof is similar to $|\hat{g}_k|$ for $k=1,2,3$, thus they will not be provided here.

□

Theorem 4 *Proof.* Note that $m=\gamma_1-g_1\beta-\beta_0$, $h=\gamma_2-g_2\beta-\beta_0$, we have

$$\begin{aligned}\hat{Y}-\hat{X}_2\beta &= Y-\hat{\gamma}-(X_2-\hat{g})\beta = Y-X_2\beta-(\hat{\gamma}-\hat{g}\beta)\\ &= Y-X_2\beta-m-h-\beta_0-\big((\hat{\gamma}-\gamma)-(\hat{g}-g)\beta\big)\\ &= v-\sum_{k=1}^{3}(V_{Yk}-V_{Xk}\beta).\end{aligned}$$

Thus $\hat{\beta}-\beta = \big(\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2\big)^{-1}\frac{1}{n}\hat{X}_2'\hat{\eta}\big(\hat{Y}-\hat{X}_2\beta\big) = \Big(\big(\frac{1}{n}\hat{X}_2'\eta\hat{X}_2\big)^{-1}\frac{1}{n}\hat{X}_2'\eta\big(\hat{Y}-\hat{X}_2\beta\big)\Big)\big(1+O_p(L_n)\big)^2$, where

$$\begin{aligned}\frac{1}{n}\hat{X}_2'\eta\hat{X}_2 &= \frac{1}{n}X_2^{*\prime}\eta X_2^{*}-\frac{1}{n}X_2^{*\prime}\eta V_X-\frac{1}{n}V_X'\eta X_2^{*}+\frac{1}{n}V_X'\eta V_X \equiv \sum_{k=1}^{4}A_k,\\ \frac{1}{n}\hat{X}_2'\eta(\hat{Y}-\hat{X}_2\beta) &= \frac{1}{n}\hat{X}_2'\eta v-\frac{1}{n}\hat{X}_2'\eta(V_{Y1}-V_{X1}\beta)-\frac{1}{n}\hat{X}_2'\eta(V_{Y2}-V_{X2}\beta)-\frac{1}{n}\hat{X}_2'\eta(V_{Y3}-V_{X3}\beta) \equiv \sum_{k=1}^{4}B_k.\end{aligned}$$

The proof has five steps:

(1) We show that $A_1 \xrightarrow{p} \Phi_0$ and $A_2, A_3, A_4 = o_p(1)$.

(2) We show that $\sqrt{n}B_1 \xrightarrow{d} \mathcal{N}(0,\Phi_1)$.

(3) We show that $B_2, B_4 = o_p(n^{-1/2})$.

(4) Let $a_{ni} = \sum_{d=1}^{D_2} \frac{U_{id}}{2h_1^{D_1}h_2^{D_2}} \mathrm{E}\Big(\frac{\eta_l X_{2l}^* D_d K_{2tl} K_{1il}}{f_U(U_l) f_Z(Z_l)} \mathbf{J}h(U_l)\big(\frac{U_t - U_l}{h_2}\big) \big| Z_i \Big)$. We show that $B_3 = \frac{1}{n}\sum_{i=1}^n a_{ni} + o_p(n^{-1/2})$.

(5) Combine (1)-(4), we show that $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}\big(0, \Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}\big)$.

Step 1: By Kolmogorov's LLN and A3, $A_1 = \frac{1}{n}\sum_{i=1}^n \eta_i X_{2i}^* X_{2i}^{*\prime} \xrightarrow{p} \Phi_0$, where

$$\Phi_{0_{(j,k)}} = \mathrm{E}(\eta_t X_{2t,j}^* X_{2t,k}^*) = \mathrm{E}\Big\{ \eta_t (X_{2t,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j})(X_{2t,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}) \Big\} < \infty,$$

since $\{\eta_i X_{2i}^* X_{2i}^{*\prime}\}_{i=1}^n$ is an IID sequence, and $\mathrm{E}|\eta_i X_{2i,k}^* X_{2i,j}^*| < \infty$ due to

(i) η_i is uniformly bounded;

(ii) $\mathrm{E}|X_{2i,j}X_{2i,k}| \le \Big(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(X_{2i,k}^2)\Big)^{1/2} < \infty$ by Cauchy-Schwarz Inequality;

(iii) $\mathrm{E}|X_{2i,j} g_{1k}(M_i)| \le \Big(\mathrm{E}(X_{2i,j}^2)\mathrm{E}(g_{1k}^2(M_i))\Big)^{1/2}$;

(iv) $\mathrm{E}(g_{1k}^2(M_i)) = \mathrm{E}(\mathrm{E}(\eta_i X_{2i,k}|M)^2) \le \mathrm{E}(\mathrm{E}(\eta_i^2 X_{2i,k}^2|M)) = \mathrm{E}(\eta_i^2 X_{2i,k}^2) < \infty$.

By the non-singularity of Φ_0 in A3, we have $A_1^{-1} \xrightarrow{p} \Phi_0^{-1}$. And for $-A_2 = \frac{1}{n}\sum_{i=1}^n \eta_i X_{2i}^* V_{Xi}'$, the $(k,j)^{\text{th}}$ element $-A_{2_{(k,j)}} = \frac{1}{n}\sum_{i=1}^n \eta_i X_{2i,k}^* V_{Xi,j} \le O_p(L_n)\frac{1}{n}\sum_{i=1}^n |\eta_i X_{2i,k}| = O_p(L_n) = o_p(1)$, by Theorem 3. Similarly we have $A_3, A_4 = o_p(1)$. Thus,

$$\left(\frac{1}{n}\hat{X}_2'\eta\hat{X}_2\right)^{-1} \xrightarrow{p} \Phi_0^{-1}.$$

Step 2: Note that $B_1 = \frac{1}{n}\sum_{i=1}^n \hat{X}_{2i}\eta_i v_i = \frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i v_i - \frac{1}{n}\sum_{i=1}^n V_{Xi}^*\eta_i v_i \equiv B_{11} - B_{12}$. By Levy central limit theorem and Cramer-Wold device, we have $\sqrt{n}B_{11} \xrightarrow{d} \mathcal{N}(0,\Phi_1)$, since

(i). $\{X_{2i}^*\eta_i v_i\}_{i=1}^n$ is IID; (ii). $\mathrm{E}(X_{2i}^*\eta_i v_i) = 0$; (iii). $\mathrm{E}(v_i^2|Z_i,U_i) = \sigma_v^2$;

(iv). $\mathrm{V}(X_{2i}^*\eta_i v_i) = \mathrm{E}(X_{2i}^*\eta_i^2 v_i^2 X_{2i}^{*\prime}) = \sigma_v^2 \mathrm{E}(\eta_i^2 X_{2i}^* X_{2i}^{*\prime})) \equiv \Phi_1 < \infty$, where

$$\Phi_{1_{(j,k)}} \;=\; \sigma_v^2\, \mathrm{E}(\eta_t^2 X_{2t,j}^* X_{2t,k}^*) \;=\; \sigma_v^2\, \mathrm{E}\Big\{ \eta_t^2 (X_{2t,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j})(X_{2t,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}) \Big\} < \infty.$$

For B_{12}, the j^{th} element can be written as

$$B_{12,j} = \frac{1}{n}\sum_{i=1}^{n} V_{Xi,j}\eta_i v_i = \frac{1}{n}\sum_{i=1}^{n} V_{X1i,j}\eta_i v_i + \frac{1}{n}\sum_{i=1}^{n} V_{X2i,j}\eta_i v_i - \frac{1}{n}\sum_{i=1}^{n} V_{X3i,j}\eta_i v_i \equiv \sum_{k=1}^{3} B_{12k}.$$

We show that $B_{12k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

Note that $B_{123} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{g}_{3j} - g_{3j}\right)\eta_i v_i = \left(\hat{g}_{3j} - g_{3j}\right)\frac{1}{n}\sum_{i=1}^{n}\eta_i v_i = O_p(L_n)O_p(n^{-1/2}) = O_p(n^{-1/2})$. By A.3 in Theorem 3, we have

$$B_{121} = \left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} C_{X1ti,j}\right\}\left(1 + O_p(L_{3n})\right) \equiv \left(\sum_{k=1}^{3} B_{121k}\right)\left(1 + O_p(L_{3n})\right)$$

where $B_{1211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}(\hat{\eta}_t - \eta_t)X_{2t,j}$, $B_{1212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} v_{X1t,j}$,

$B_{1213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\left(g_{1j}(M_t) - g_{1j}(M_i)\right)$.

We show that $B_{121k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

(1a). Since $\hat{\eta}_t - \eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{1211} = B'_{1211}O_p(L_n)$,

where $B'_{1211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t X_{2t,j} \equiv E_{1n} + E_{2n}$, with

$$E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\eta_i v_i K_3(0)}{h_3^{D_3} f_M(M_i)}\eta_i X_{2i,j}, \qquad E_{2n} = \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t X_{2t,j}$$

By Chebyshev's Inequality and that $\mathrm{E}(E_{1n}) = 0$, $\mathrm{V}(E_{1n}) = \mathrm{E}(E_{1n}^2) = \frac{1}{n^2}\frac{1}{n}\mathrm{E}\left(\frac{\eta_i^2 v_i^2 K_3^2(0)}{h_3^{2D_3} f_M^2(M_i)}\eta_i^2 X_{2i,j}^2\right) = O\left(n^{-3}h_3^{-2D_3}\right)$, we have $E_{1n} = O_p\left(n^{-1/2}(nh_3^{D_3})^{-1}\right) = o_p(n^{-1/2})$. $E_{2n} \leq CU_n$, where U_n is a U-statistic such that $U_n = \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nit} = \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i<t}}^{n}\phi_{nit}$ with $\psi_{nit} = \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t X_{2t,j}$. Since $\mathrm{E}(v_i|M_i) = 0$, we have $\theta_n = 0$, $\phi_{1n} = \mathrm{E}(\phi_{nit}|P_i) = \mathrm{E}(\psi_{nit}|P_i) = \frac{\eta_i v_i}{f_M(M_i)}\int K_3(\gamma)g_{1j}(M_i + h_3\gamma)f_M(M_i + h_3\gamma)d\gamma \equiv \frac{\eta_i v_i}{f_M(M_i)}C_1(M_i)$. Thus $\sigma_{1n}^2 = \mathrm{V}(\phi_{1n}) \leq \mathrm{E}(\phi_{1n}^2) \leq C\sigma_v^2\mathrm{E}\left(\frac{C_1^2(M_i)}{f_M^2(M_i)}\right) \to \mathrm{E}\left(g_{1j}^2(M_i)\right) < \infty$ by Lemma 1 and A3. By Theorem 1 in Yao and Martins-Filho (2013), $H_n^{(1)} = O_p\left(\left(\frac{\sigma_{1n}^2}{n}\right)^{1/2}\right) = O_p(n^{-1/2})$. $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \leq C\mathrm{E}(\psi_{nit}^2) \leq \frac{C\sigma_v^2}{h_3^{2D_3}}\mathrm{E}\left(K_{3ti}^2(\eta_t X_{2t,j})^2\right) = O(h_3^{-D_3})$ by

Lemma 1 and A3. $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$. In sum, $B_{1211} = O_p(n^{-1/2})O_p(L_n)$ $= o_p(n^{-1/2})$.

(1b). $B_{1212} = \frac{1}{n^2}\sum_{i=1}^n\sum_{t=1}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} v_{X1t,j} \equiv E_{1n} + E_{2n}$.

$E_{1n} = o_p(n^{-1/2})$ as $\mathrm{E}(E_{1n}) = 0$, $\quad \mathrm{V}(E_{1n}) = \frac{1}{n^2}\frac{1}{n}\mathrm{E}\Big(\frac{\eta_i^2 v_i^2 K_3^2(0)}{h_3^{2D_3} f_M^2(M_i)} v_{X1i,j}^2\Big) = O\big(n^{-3}h_3^{-2D_3}\big) = o_p(n^{-1})$.

$E_{2n} \le CU_n \equiv C\binom{n}{2}^{-1}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n \psi_{nit}$ with $\psi_{nit} = \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)} v_{X1t,j}$. We analyze each component in $U_n = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ by Hoeffding's decomposition in Hoeffding (1961).

- $\theta_n = \sigma_{1n}^2 = 0$, as $\mathrm{E}(v_i|M_i) = \mathrm{E}(v_{X1t,j}|M_t) = 0$;
- $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi_{nit}^2) \le \frac{C\sigma_v^2\sigma_{X1,j}^2}{h_3^{2D_3}}\mathrm{E}(K_{3ti}^2) = O(h_3^{-D_3})$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$.

We have $B_{1212} = o_p(n^{-1/2})$.

(1c). $B_{1213} = \frac{1}{n^2}\sum_{i=1}^n\sum_{\substack{t=1\\ i\neq t}}^n \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\big(g_{1j}(M_t) - g_{1j}(M_i)\big) \le CU_n$, where $U_n = \binom{n}{2}^{-1}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n \psi_{nit}$ with $\psi_{nit} = \frac{\eta_i v_i K_{3ti}}{h_3^{D_3} f_M(M_i)}\big(g_{1j}(M_t) - g_{1j}(M_i)\big)$ is a U-statistic of degree 2.

- $\theta_n = \mathrm{E}(\phi_{nit}|P_t) = 0$, as $\mathrm{E}(v_i|M_i) = 0$.
- $\phi_{1n} = \mathrm{E}(\phi_{nit}|P_i) = \frac{\eta_i v_i}{h_3^{D_3} f_M(M_i)}\mathrm{E}\Big(K_{3ti}\big(g_{1j}(M_t) - g_{1j}(M_i)\big)\Big|M_i\Big) \le \frac{Ch_3^{s_3}|\eta_i v_i|}{f_M(M_i)}$.
- $\sigma_{1n}^2 \le \mathrm{E}(\phi_{1n}^2) = O(h_3^{2s_3}) = o(1)$.
- $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi_{nit}^2) \le \frac{C\sigma_v^2}{h_3^{2D_3}}\mathrm{E}\Big(K_{3ti}^2\big(g_{1j}(M_t) - g_{1j}(M_i)\big)^2\Big) = O\big(h_3^{-D_3+2}\big)$.
- $H_n^{(1)} = O_p\Big(\big(\frac{\sigma_{1n}^2}{n}\big)^{1/2}\Big) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p(n^{-1/2}(nh_3^{D_3-2})^{-1/2}) = o_p(n^{-1/2})$.

We have $B_{1213} = o_p(n^{-1/2})$.

By (1a)-(1c), we have $B_{121} = o_p(n^{-1/2})$.

For B_{122}, since $\frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^n R_{ti}C_{X2ti} = o_p(n^{-1/2})$ uniformly, by A.2 in Theorem 3, we have

$$B_{122} = \frac{1}{n}\sum_{i=1}^n V_{X2i,j}\eta_i v_i = \left(\sum_{k=1}^3 B_{122k}\right)\big(1 + O_p(L_{2n})\big) + o_p(n^{-1/2}),$$

where $B_{1221} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{X2ti}$, $B_{1222} = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)C_{X2ti}$,

$B_{1223} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_t - U_t)C_{X2ti}$, $C_{X2ti} = (\hat{\eta}_t - \eta_t)X_{2t,j} + v_{X2t,j} + \big((g_{2j}(U_t) - g_{2j}(U_i)\big)$.

Similarly to B_{121} we just analyzed, we have $B_{1221} = o_p(n^{-1/2})$, with U_i replacing M_i. B_{1222} and B_{1223} could be studied similarly, here we only show that $B_{1222} = o_p(n^{-1/2})$. By the three components in C_{X2ti}, let $B_{1222} = \sum_{k=1}^{3} B_{1222k}$. We show that $B_{1222k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

(2a). $B_{12221} = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)(\hat{\eta}_t - \eta_t)X_{2t,j}$

$\leq O_p(L_n)O_p\left(\frac{L_{1n}}{h_2}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{d=1}^{D_2}\frac{\left|\eta_i v_i \eta_t X_{2t,j} D_d K_{2ti}\right|}{h_2^{D_2} f_U(U_i)}$

$= O_p(L_n)O_p\left(\frac{L_{1n}}{h_2}\right) = o_p(n^{-1/2})$

since $L_n^2, \left(\frac{L_{1n}}{h_2}\right)^2 = o_p(n^{-1/2})$ by A5.

(2b). By A.1 in Theorem 2, we have

$B_{12222} = -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}(\hat{U}_{id} - U_{id})$

$= \left\{\sum_{d=1}^{D_2}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\right\}\big(1 + O_p(L_{1n})\big)$

$\equiv \left\{\sum_{d=1}^{D_2}(T_{1d} + T_{2d})\right\}\big(1 + O_p(L_{1n})\big)$.

We show that $T_{1d}, T_{2d} = o_p(n^{-1/2})$.

(i) $T_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \binom{n}{3}^{-1}\sum_{i<t<l}\phi_{nitl}$ be a U-statistic of degree 3. We analyze each component in $U_n = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ by Hoeffding's decomposition in Hoeffding (1961).

- $\theta_n, \mathrm{E}(\phi_{nitl}|P_i), \mathrm{E}(\phi_{nitl}|P_i, P_t) = 0$, as $\mathrm{E}(v_i|M_i, U_i) = \mathrm{E}(v_{X2t,j}|U_t) = \mathrm{E}(U_{ld}|Z_l) = 0$;
- $\sigma_{1n}^2 = \sigma_{2n}^2 = 0$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1} h_2^{D_2+2})^{-1}\big)$;

- $H_n^{(1)} = H_n^{(2)} = 0$, $H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big) = O_p\Big(\big(n^3 h_1^{D_1} h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

$$\text{if} \quad i=t=l, \qquad \frac{1}{n^3}\sum_{i=1}^{n}\psi_{niii}$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\frac{\eta_i v_i v_{X2i,j} D_d K_2(0) K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n^2 h_1^{D_1} h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i=t\neq l, \qquad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\psi_{niil}$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\frac{\eta_i v_i v_{X2i,j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} = O_p\Big(\big(n h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i=l\neq t, \qquad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{niti}$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_1(0)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n h_1^{D_1} h_2\big)^{-1}\Big) = o_p(n^{-1/2});$$

$$\text{if} \quad i\neq t=l, \qquad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nitt}$$
$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{td} = O_p\Big(\big(n h_2\big)^{-1}\Big) = o_p(n^{-1/2}).$$

In sum, we have $T_{1d} = o_p(n^{-1/2})$.

(ii) $T_{2d} = \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\dfrac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \equiv \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$

If $i\neq t\neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3.

- $\theta_n = \mathrm{E}(\phi_{nitl}|P_i) = \mathrm{E}(\psi_{nitl}|P_i,P_l) = \mathrm{E}(\psi_{nitl}|P_t,P_l) = 0$, as $\mathrm{E}(v_i|M_i,U_i) = \mathrm{E}(v_{X2t,j}|U_t) = 0$;
- $\sigma_{1n}^2 = 0$;
- $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \frac{\eta_i v_i v_{X2t,j} D_d K_{2ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\mathrm{E}\Big(K_{1li}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big|Z_i\Big) \leq \frac{C h_1^{s_1}\big|\eta_i v_i v_{X2t,j} D_d K_{2ti}\big|}{h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}$;
- $\sigma_{2n}^2 \leq C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i,P_t)\big) = O\Big(\frac{h_1^{2s_1}}{h_2^{D_2+2}}\Big)$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1-2} h_2^{D_2+2})^{-1}\big)$;

- $H_n^{(1)} = 0, \quad H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\Big(h_1^{s_1}\big(n^2 h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2}), \quad H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big)$
  $= O_p\Big(\big(n^3 h_1^{D_1-2} h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2}).$

  We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $\quad i = t = l, \qquad i = l \neq t, \qquad \psi_{nitl} = 0;$

if $\quad i = t \neq l, \qquad \dfrac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^{n}\sum_{l=1}^{n}\psi_{niil}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq l}}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i v_{X2i,j} D_d K_2(0) K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $\quad i \neq t = l, \qquad \dfrac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt}$

$$= \frac{1}{n^3}\sum_{\substack{i=1 \\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i v_{X2t,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_t) - \Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2\big)^{-1}\Big) = o_p(n^{-1/2}).$$

We have $B_{12222} = o_p(n^{-1/2})$.

(2c). Similar to part (2b), we have

$$\begin{aligned}
B_{12223} &= -\sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i\big(g_{2j}(U_t) - g_{2j}(U_i)\big) D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}(\hat{U}_{id} - U_{id}) \\
&= \left\{\sum_{d=1}^{D_2}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i\big(g_{2j}(U_t) - g_{2j}(U_i)\big) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\right\}\big(1 + O_p(L_{1n})\big) \\
&\equiv \left\{\sum_{d=1}^{D_2}(W_{1d} + W_{2d})\right\}\big(1 + O_p(L_{1n})\big).
\end{aligned}$$

We show that $W_{1d}, W_{2d} = o_p(n^{-1/2})$.

(i) $W_{1d} = \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i v_i\big(g_{2j}(U_t) - g_{2j}(U_i)\big) D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3:

- $\theta_n = \sigma_{1n}^2 = \mathrm{E}(\psi_{nitl}|P_i, P_t) = \mathrm{E}(\psi_{nitl}|P_l, P_t) = 0, \quad \text{as} \quad \mathrm{E}(v_i|M_i, U_i) = \mathrm{E}(U_{ld}|Z_l) = 0;$
- $\mathrm{E}(\psi_{nitl}|P_i, P_l) = \frac{\eta_i v_i U_{ld} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\mathrm{E}\Big(D_d K_{2ti}\big(g_{2j}(U_t) - g_{2j}(U_i)\big)\Big|U_i\Big) \leq \frac{C\big|\eta_i v_i U_{ld} K_{1li}\big|}{h_1^{D_1} f_U(U_i) f_Z(Z_i)};$

- $\sigma_{2n}^2 \le C\mathrm{E}\left(\mathrm{E}^2(\psi_{nitl}|P_i,P_l)\right) = O\big(h_1^{-D_1}\big)$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1}h_2^{D_2+2})^{-1}\big)$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\Big(\big(n^2h_1^{D_1}\big)^{-1/2}\Big) = o_p(n^{-1/2})$, $H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big)$
  $= O_p\Big(\big(n^3h_1^{D_1}h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

  We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $i=t=l$, $i=t\neq l$, $\psi_{nitl}=0$;

if $i=l\neq t$,
$$\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{niti} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i U_{id} D_d K_{2ti} K_1(0)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_{2j}(U_t)-g_{2j}(U_i)\big) = O_p\Big(\big(nh_1^{D_1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $i\neq t=l$,
$$\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt} = \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i U_{td} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_{2j}(U_t)-g_{2j}(U_i)\big) = O_p(n^{-1}) = o_p(n^{-1/2}).$$

In sum, we have $W_{1d} = o_p(n^{-1/2})$.

(ii) $W_{2d} = \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\dfrac{\eta_i v_i D_d K_{2ti} K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big) \equiv \dfrac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}.$

If $i\neq t\neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3.

- $\theta_n = \mathrm{E}(\psi_{nitl}|P_t) = \mathrm{E}(\psi_{nitl}|P_l) = \mathrm{E}(\psi_{nitl}|P_t,P_l) = 0$, as $\mathrm{E}(v_i|Z_i,U_i,M_i) = 0$;
- $\mathrm{E}(\psi_{nitl}|P_i) = \frac{\eta_i v_i}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(D_d K_{2ti}K_{1li}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|P_i\Big) \le \frac{Ch_1^{s_1}\eta_i v_i}{f_U(U_i)f_Z(Z_i)}$;
- $\sigma_{1n}^2 \le C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i)\big) \le Ch_1^{2s_1} = o(1)$;
- $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \frac{\eta_i v_i\left(g_{2j}(U_t)-g_{2j}(U_i)\right)D_d K_{2ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(K_{1li}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|Z_i\Big) \le \frac{Ch_1^{s_1}\left|\eta_i v_i\left(g_{2j}(U_t)-g_{2j}(U_i)\right)D_d K_{2ti}\right|}{h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}$,
  $\mathrm{E}(\psi_{nitl}|P_i,P_l) = \frac{\eta_i v_i K_{1li}\left(\Pi_d(Z_l)-\Pi_d(Z_i)\right)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(D_d K_{2ti}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\Big|U_i\Big) \le \frac{C\left|\eta_i v_i K_{1li}\left(\Pi_d(Z_l)-\Pi_d(Z_i)\right)\right|}{h_1^{D_1}f_U(U_i)f_Z(Z_i)}$;
- $\sigma_{2n}^2 \le C\mathrm{E}\left(\mathrm{E}^2(\psi_{nitl}|P_i,P_t)+\mathrm{E}^2(\psi_{nitl}|P_i,P_l)\right) = O\Big(\frac{h_1^{2s_1}}{h_2^{D_2}} + \frac{1}{h_1^{D_1-2}}\Big)$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \le C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1-2}h_2^{D_2})^{-1}\big)$;

- $H_n^{(1)} = O_p\Big(\big(\frac{\sigma_{1n}^2}{n}\big)^{1/2}\Big) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\Big(h_1^{s_1}\big(n^2h_2^{D_2}\big)^{-1/2} + \big(n^2h_1^{D_1-2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$, $H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big) = O_p\Big(\big(n^3h_1^{D_1-2}h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

  We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $i = t = l$, $\quad i = l \neq t$, $\quad i = t \neq l$, $\quad \psi_{nitl} = 0$;

if $i \neq t = l$,
$$\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt}$$
$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i v_i D_d K_{2ti}K_{1ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(g_{2j}(U_t)-g_{2j}(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)$$
$$= O_p\Big(\frac{h_1}{n}\Big) = o_p(n^{-1/2}).$$

We have $B_{12223} = o_p(n^{-1/2})$. By (2a)-(2c), we have $B_{122} = o_p(n^{-1/2})$.

Combing all the terms in step 2, we have $B_1 = B_{11} + o_p(n^{-1/2})$, where $\sqrt{n}B_{11} \xrightarrow{d} \mathcal{N}(0,\Phi_1)$.

Step 3: We first show that $B_4 = o_p(n^{-1/2})$. Note that $-B_4 = \frac{1}{n}\hat{X}_2'\eta(V_{Y3} - V_{X3}\beta) = \frac{1}{n}\hat{X}_2'\eta V_{Y3} - \frac{1}{n}\hat{X}_2'\eta V_{X3}\beta \equiv B_{41} + B_{42}$. By Theorem 3, we have $|V_{Xi}|, |V_{Y3}| = O_p(L_n)$. Thus $B_{41} = V_{Y3}\big(\frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i\big) = O_p(L_n)\big(O_p(n^{-1/2}) + O_p(L_n)\big) = o_p(n^{-1/2})$ by A5. Similarly, $-B_2 = \frac{1}{n}\hat{X}_2'\eta(V_{Y2} - V_{X2}\beta) \equiv B_{21} + B_{22}$, and we will show that $B_{21} = o_p(n^{-1/2})$. $B_{22} = o_p(n^{-1/2})$ follows by the same arguments. Note that $B_{21} = \frac{1}{n}\sum_{i=1}^n X_{2i}^*\eta_i V_{Y1i} - \frac{1}{n}\sum_{i=1}^n V_{Xi}\eta_i V_{Y1i} \equiv B_{21}' + o_p(n^{-1/2})$ by Theorem 3. And by expression of V_{Y1i} similar to V_{X1i} given in A.3 of Theorem 3, we have the j^{th} element of B_{21}' as

$$\begin{aligned} B_{21,j}' &= \left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_M(M_i)}C_{Y1ti}\right)\big(1 + O_p(L_{3n})\big) \\ &= \left(\sum_{k=1}^{3}B_{21k}\right)\big(1 + O_p(L_{3n})\big), \end{aligned}$$

where $\quad B_{211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_M(M_i)}(\hat{\eta}_t - \eta_t)Y_t, \qquad B_{212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3}f_M(M_i)}v_{Y1t}$

$$B_{213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3} f_M(M_i)}\big(\gamma_1(M_t)-\gamma_1(M_i)\big).$$

We show that $B_{21k} = o_p(n^{-1/2})$ for $k = 1,2,3$.

(3a). Since $\hat{\eta}_t - \eta_t = \eta_t O_p(L_n)$ uniformly, we have $B_{211} = B'_{211} O_p(L_n)$,

$$\text{where} \qquad B'_{211} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3} f_M(M_i)}\eta_t Y_t \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{1n}+E_{2n}, \quad \text{with}$$
$$E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\eta_i X_{2i,j}^* K_3(0)}{h_3^{D_3} f_M(M_i)}\eta_i Y_i, \qquad E_{2n} = \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nit}$$

By Markov's Inequality and that $\mathrm{E}|E_{1n}| \le \frac{C}{nh_3^{D_3}}\mathrm{E}|X_{2i,j}^* Y_i| = O\big((nh_3^{D_3})^{-1}\big)$, we have $E_{1n} = o_p(n^{-1/2})$. $E_{2n} \le CU_n$, where U_n is a U-statistic of degree 2 such that $U_n = \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nit}$.

- $\theta_n = \mathrm{E}(\psi_{nit}|P_t) = 0$, as $\mathrm{E}(\eta_i X_{2i,j}^*|M_i) = 0$;
- $\phi_{1n} = \mathrm{E}(\psi_{nit}|P_i) = \frac{\eta_i X_{2i,j}^*}{h_3^{D_3} f_M(M_i)}\mathrm{E}\big(K_{3ti}\eta_t Y_t\big|M_i\big) = \frac{\eta_i X_{2i,j}^*}{f_M(M_i)}\int K(\psi)\gamma_1(M_i+h_3\psi)f_M(M_i+h_3\psi)d\psi$;
- $\sigma_{1n}^2 \le C\mathrm{E}\big(\mathrm{E}^2(\psi_{nit}|P_i)\big) = O(1)$ by Lemma 1 and A3.
- $\sigma_{2n}^2 \le C\mathrm{E}(\psi_{nit}^2) = \frac{C}{h_3^{2D_3}}\mathrm{E}\left(\frac{(\eta_i X_{2i,j}^*)^2(\eta_t Y_t)^2 K_{3ti}^2}{f_M^2(M_i)}\right) = O\big(h_3^{-D_3}\big)$;
- $H_n^{(1)} = O_p\Big(\big(\frac{\sigma_{1n}^2}{n}\big)^{1/2}\Big) = O_p(n^{-1/2})$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\Big(\big(n^2 h_3^{D_3}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

Thus $B'_{211} = O_p(n^{-1/2})$, and $B_{211} = O_p(n^{-1/2})O_p(L_n) = o_p(n^{-1/2})$.

(3b). $B_{212} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3} f_M(M_i)}v_{Y1t} \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\psi_{nit} \equiv E_{1n}+E_{2n}$, where
$E_{1n} = \frac{1}{n^2}\sum_{i=1}^{n}\psi_{nii} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\eta_i X_{2i,j}^* K_3(0)}{h_3^{D_3} f_M(M_i)}v_{Y1i} = O_p\big((nh_3^{D_3})^{-1}\big) = o_p(n^{-1/2})$;
$E_{2n} \le CU_n$, and $U_n = \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nit} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ is a U-statistic of degree 2.

- $\theta_n = \sigma_{1n}^2 = 0$, as $\mathrm{E}(\eta_i X_{2i,j}^*|M_i) = \mathrm{E}(v_{Y1t}|M_t) = 0$;
- $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \le C\mathrm{E}(\psi_{nit}^2) = O(h_3^{-D_3})$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p((n^2 h_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$.

We have $B_{212} = o_p(n^{-1/2})$.

(3c). $B_{213} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i X_{2i,j}^* K_{3ti}}{h_3^{D_3} f_M(M_i)}\big(\gamma_1(M_t)-\gamma_1(M_i)\big) \equiv \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nit} \leq CU_n$, where $U_n = \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nit}$ is a U-statistic of degree 2.

- $\theta_n = \mathrm{E}(\phi_{nit}|P_t) = 0$, as $\mathrm{E}(\eta_i X_{2i,j}^*|M_i) = 0$;
- $\phi_{1n} = \mathrm{E}(\phi_{nit}|P_i) = \frac{\eta_i X_{2i,j}^*}{h_3^{D_3} f_M(M_i)}\mathrm{E}\Big(K_{3ti}\big(\gamma_1(M_t)-\gamma_1(M_i)\big)\Big|M_i\Big) \leq \frac{Ch_3^{s_3}|\eta_i X_{2i,j}^*|}{f_M(M_i)}$;
- $\sigma_{1n}^2 \leq \mathrm{E}(\phi_{1n}^2) = O(h_3^{2s_3}) = o(1)$;
- $\sigma_{2n}^2 = \mathrm{V}(\phi_{nit}) \leq C\mathrm{E}(\psi_{nit}^2) = O\big(h_3^{-D_3+2}\big)$;
- $H_n^{(1)} = O_p\Big(\big(\frac{\sigma_{1n}^2}{n}\big)^{1/2}\Big) = o_p(n^{-1/2})$, $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p(n^{-1/2}(nh_3^{D_3-2})^{-1/2}) = o_p(n^{-1/2})$.

We have $B_{213} = o_p(n^{-1/2})$.

By (3a)-(3c), we have $B_{21} = o_p(n^{-1/2})$.

Step 4: For B_3, we have $-B_3 = \frac{1}{n}\hat{X}_2'\eta(V_{Y2} - V_{X2}\beta) \equiv B_{31} + B_{32}$. We will focus on B_{31} here, since B_{32} has a similar structure to B_{31} and could be analyzed accordingly. By Theorem 3, we have $B_{31} = \frac{1}{n}\sum_{i=1}^{n} X_{2i}^*\eta_i V_{Y2i} - \frac{1}{n}\sum_{i=1}^{n} V_{Xi}\eta_i V_{Y2i} \equiv B_{31}' + o_p(n^{-1/2})$. Similar to A.2 given in Theorem 3, by Taylor Theorem, we have

$$
\begin{aligned}
V_{Y2i} &= \hat{\gamma}_2(\hat{U}_i) - \gamma_2(U_i) = \frac{1}{nh_2^{D_2}\hat{f}_U(\hat{U}_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\hat{\eta}_t Y_t - \gamma_2(U_i) \\
&= \left\{\frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\Big(\underbrace{(\hat{\eta}_t-\eta_t)Y_t + v_{Y2t} + \big(\gamma_2(U_t)-\gamma_2(U_i)\big)}_{C_{Y2ti}}\Big)\right\}\Big(1+O_p(L_{2n})\Big) \\
&= \left\{\frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^{n}K_{2ti}C_{Y2ti} + \frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^{n}\mathbf{J}K_{2ti}\Big(\hat{U}_t - U_t - \big(\hat{U}_i - U_i\big)\Big)C_{Y2ti}\right. \\
&\quad \left. + \frac{1}{nh_2^{D_2+1} f_U(U_i)}\sum_{t=1}^{n}R_{ti}C_{Y2ti}\right\}\Big(1+O_p(L_{2n})\Big)
\end{aligned}
$$

where R_{ti} is the remainder term of a Taylor expansion of $\hat{K}_{2ti}$ at $\Big(\frac{U_t-U_i}{h_2}\Big)$.

Similar to the T_3 term in Theorem 3, we have $\frac{1}{nh_2^{D_2} f_U(U_i)}\sum_{t=1}^{n}R_{ti}C_{Y2ti} = o_p(n^{-1/2})$ uniformly. Thus, we have the j^{th} element of B_{31}' as

$$B_{31,j}' = \sum_{k=1}^{3} B_{31k} + o_p(n^{-1/2})$$

where $B_{311} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j} K_{2ti}}{h_2^{D_2} f_U(U_i)} C_{Y2ti}$, $B_{312} = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)C_{Y2ti}$,

$B_{313} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_t - U_t)C_{Y2ti}$, and $C_{Y2ti} = (\hat{\eta}_t - \eta_t)Y_t + v_{Y2t} + \big(\gamma_2(U_t) - \gamma_2(U_i)\big)$.

We will show that $B_{311} = B_{313} = o_p(n^{-1/2})$ and $B_{312} = \frac{1}{n}\sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$, where

$$a_{1ni,j} = \sum_{d=1}^{D_2}\frac{U_{id}}{2h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X^*_{2l,j} D_d K_{2tl} K_{1il}}{f_U(U_l) f_Z(Z_l)}\mathbf{J}\gamma_2(U_l)\Big(\frac{U_t - U_l}{h_2}\Big)\middle| Z_i\right).$$

The components in B_{311} are similar to B_{121} with U_i replacing M_i, $\eta_i X^*_{2i,j}$ replacing $\eta_i v_i$, C_{Y2ti} replacing $C_{X1ti,j}$, and $\mathrm{E}(\eta_i X^*_{2i,j}|U_i) = 0$ replacing $E(\eta_i v_i|M_i) = 0$. By the same arguments in (1a)-(1c), we have $B_{311} = o_p(n^{-1/2})$. By the three components in C_{Y2ti}, we have $-B_{312} \equiv \sum_{k=1}^{3} B_{312k}$, with

$B_{3121} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)(\hat{\eta}_t - \eta_t)Y_t$, $B_{3122} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)v_{Y2t}$,

$B_{3123} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)\big(\gamma_2(U_t) - \gamma_2(U_i)\big)$.

We show that $B_{3121} = B_{3122} = o_p(n^{-1/2})$, and $B_{3123} = \frac{1}{n}\sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$.

(4a). $B_{3121} = \left\{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j}}{h_2^{D_2+1} f_U(U_i)}\mathbf{J}K_{2ti}(\hat{U}_i - U_i)\eta_t Y_t\right\} O_p(L_n)$

$\le O_p(L_n) O_p\left(\frac{L_{1n}}{h_2}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{d=1}^{D_2}\frac{\big|\eta_i X^*_{2i,j}\eta_t Y_t D_d K_{2ti}\big|}{h_2^{D_2} f_U(U_i)}$

$= O_p(L_n) O_p\left(\frac{L_{1n}}{h_2}\right) = o_p(n^{-1/2})$

(4b). By A.1 in proof of Theorem 2, we have

$B_{3122} = \sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti}}{h_2^{D_2+1} f_U(U_i)}(\hat{U}_{id} - U_{id})$

$= -\left\{\sum_{d=1}^{D_2}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\Big(U_{ld} + \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)\Big)\right\}\big(1 + O_p(L_{1n})\big)$

$\equiv -\left\{\sum_{d=1}^{D_2}(T_{1d} + T_{2d})\right\}\big(1 + O_p(L_{1n})\big)$.

We show that $T_{1d}, T_{2d} = o_p(n^{-1/2})$.

(i) $T_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3.

- $\theta_n = \sigma_{1n}^2 = \mathrm{E}(\psi_{nitl}|P_i,P_t) = \mathrm{E}(\psi_{nitl}|P_i,P_l) = 0$, as $\mathrm{E}(v_{Y2t}|U_t) = \mathrm{E}(U_{ld}|Z_l) = 0$;
- $\phi_{2n} = \mathrm{E}(\psi_{nitl}|P_t,P_l) = \frac{v_{Y2t}U_{ld}}{h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_i X^*_{2i,j} D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\Big|Z_l,U_t\right) \leq \frac{C|v_{Y2t}U_{ld}|}{h_2}$;
- $\sigma_{2n}^2 \leq \mathrm{E}(\phi_{1n}^2) = O(h_2^{-2})$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1}h_2^{D_2+2})^{-1}\big)$;
- $H_n^{(1)} = 0$; $\quad H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\big((nh_2)^{-1}\big) = o_p(n^{-1/2})$;
- $H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big) = O_p\Big(\big(n^3h_1^{D_1}h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $\quad i=t=l$, $\quad \frac{1}{n^3}\sum_{i=1}^{n}\psi_{niii}$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_2(0) K_1(0)}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(n^2 h_1^{D_1} h_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $\quad i=t\neq l$, $\quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\psi_{niil}$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{l=1\\ i\neq l}}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_2(0) K_{1li}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{ld} = O_p\Big(\big(nh_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $\quad i=l\neq t$, $\quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{niti}$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_1(0)}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{id} = O_p\Big(\big(nh_1^{D_1} h_2\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $\quad i\neq t=l$, $\quad \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\psi_{nitt}$

$$= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{t=1\\ i\neq t}}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} U_{td} = O_p\Big(\big(nh_2\big)^{-1}\Big) = o_p(n^{-1/2}).$$

In sum, we have $T_{1d} = o_p(n^{-1/2})$.

(ii) $T_{2d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_{1li}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}\big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \equiv \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3.

- $\theta_n = \mathrm{E}(\psi_{nitl}|P_i) = \mathrm{E}(\psi_{nitl}|P_l) = \mathrm{E}(\psi_{nitl}|P_i,P_l) = 0$, as $\mathrm{E}(v_{Y2t}|U_t) = 0$;
- $\mathrm{E}(\psi_{nitl}|P_t) = \frac{v_{Y2t}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X^*_{2i,j} D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\big|U_t\Big) \leq \frac{Ch_1^{s_1}|v_{Y2t}|}{h_2}$;
- $\sigma_{1n}^2 \leq \mathrm{E}(\phi_{1n}^2) = O\big(\frac{h_1^{2s_1}}{h_2^2}\big)$;
- $\mathrm{E}(\psi_{nitl}|P_i,P_t) = \frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(K_{1li}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|Z_i\Big) \leq \frac{Ch_1^{s_1}\big|\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti}\big|}{h_2^{D_2+1} f_U(U_i)f_Z(Z_i)}$, $\mathrm{E}(\psi_{nitl}|P_t, P_l) = \frac{v_{Y2t}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X^*_{2i,j} D_d K_{2ti} K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big|U_t,Z_l\Big) \leq \frac{Ch_1^{s_1}|v_{Y2t}|}{h_2}$;
- $\sigma_{2n}^2 \leq C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i,P_t) + \mathrm{E}^2(\psi_{nitl}|P_t,P_l)\big) = O\Big(\frac{h_1^{2s_1}}{h_2^{D_2+2}}\Big)$;
- $\sigma_{3n}^2 = \mathrm{V}(\phi_{nitl}) \leq C\mathrm{E}(\psi_{nitl}^2) = O_p\big((h_1^{D_1-2}h_2^{D_2+2})^{-1}\big)$;
- $H_n^{(1)} = O_p\Big(\big(\frac{\sigma_{1n}^2}{n}\big)^{1/2}\Big) = O_p\big(n^{-1/2}h_1^{s_1}h_2^{-1}\big) = o_p(n^{-1/2})$;
- $H_n^{(2)} = O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big) = O_p\Big(h_1^{s_1}\big(n^2h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$;
- $H_n^{(3)} = O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big) = O_p\Big(\big(n^3h_1^{D_1-2}h_2^{D_2+2}\big)^{-1/2}\Big) = o_p(n^{-1/2})$.

We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $i = t = l$, $\quad i = l \neq t$, $\quad \psi_{nitl} = 0$;

if $i = t \neq l$,
$$\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\psi_{niil}$$
$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq l}}^{n}\sum_{l=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_2(0) K_{1li}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2^{D_2+1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $i \neq t = l$,
$$\frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\psi_{nitt}$$
$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^{n}\sum_{t=1}^{n}\frac{\eta_i X^*_{2i,j} v_{Y2t} D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1} f_U(U_i)f_Z(Z_i)}\big(\Pi_d(Z_t)-\Pi_d(Z_i)\big) = O_p\Big(h_1\big(nh_2\big)^{-1}\Big) = o_p(n^{-1/2}).$$

We have $B_{3122} = o_p(n^{-1/2})$.

$$\text{(4c).}\quad B_{3123} = \sum_{d=1}^{D_2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\eta_i X_{2i,j}^{*}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)D_d K_{2ti}}{h_2^{D_2+1}f_U(U_i)}(\hat{U}_{id}-U_{id})$$
$$= -\left\{\sum_{d=1}^{D_2}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^{*}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\Big(U_{ld}+\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big)\Big)\right\}\big(1+O_p(L_{1n})\big)$$
$$\equiv -\left\{\sum_{d=1}^{D_2}(W_{1d}+W_{2d})\right\}\big(1+O_p(L_{1n})\big).$$

We show that $\sum_{d=1}^{D_2}W_{1d}=\frac{1}{n}\sum_{i=1}^{n}a_{1ni,j}+o_p(n^{-1/2})$, $W_{2d}=o_p(n^{-1/2})$.

$$\text{(i)}\ W_{1d} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\frac{\eta_i X_{2i,j}^{*}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)D_d K_{2ti}K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}U_{ld}\equiv\frac{1}{n^3}\sum_{i=1}^{n}\sum_{t=1}^{n}\sum_{l=1}^{n}\psi_{nitl}$$

If $i\neq t\neq l$, let $U_n=\binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl}=\theta_n+3H_n^{(1)}+3H_n^{(2)}+H_n^{(3)}$ be a U-statistic of degree 3:

- $\theta_n=\mathrm{E}(\phi_{nitl}|P_i)=\mathrm{E}(\psi_{nitl}|P_t)=\mathrm{E}(\psi_{nitl}|P_i,P_t)=0,\quad\text{as}\quad\mathrm{E}(U_{ld}|Z_l)=0;$
- $\phi_{1n}=\mathrm{E}(\psi_{nitl}|P_l)=\frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X_{2i,j}^{*}D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)\Big|Z_l\Big)\le C|U_{ld}|;$
- $\sigma_{1n}^2\le\mathrm{E}(\phi_{1n})=O(1);$
- $\mathrm{E}(\psi_{nitl}|P_i,P_l)=\frac{\eta_i X_{2i,j}^{*}K_{1li}U_{ld}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\mathrm{E}\Big(D_d K_{2ti}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)\Big|U_i\Big)\le\frac{C\left|\eta_i X_{2i,j}^{*}K_{1li}U_{ld}\right|}{h_1^{D_1}f_U(U_i)f_Z(Z_i)},$
  $\mathrm{E}(\psi_{nitl}|P_t,P_l)=\frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\Big(\frac{\eta_i X_{2i,j}^{*}D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)\Big|U_t,Z_l\Big)\le CU_{ld};$
- $\sigma_{2n}^2\le C\mathrm{E}\big(\mathrm{E}^2(\psi_{nitl}|P_i,P_l)+\mathrm{E}^2(\psi_{nitl}|P_t,P_l)\big)=O\big(h_1^{-D_1}\big);$
- $\sigma_{3n}^2=\mathrm{V}(\phi_{nitl})\le C\mathrm{E}(\psi_{nitl}^2)=O_p\big((h_1^{D_1}h_2^{D_2})^{-1}\big);$
- $H_n^{(1)}=O_p(n^{-1/2}),\quad H_n^{(2)}=O_p\Big(\big(\frac{\sigma_{2n}^2}{n^2}\big)^{1/2}\Big)=O_p\Big(\big(n^2h_1^{D_1}\big)^{-1/2}\Big)=o_p(n^{-1/2});$
- $H_n^{(3)}=O_p\Big(\big(\frac{\sigma_{3n}^2}{n^3}\big)^{1/2}\Big)=O_p\Big(\big(n^3h_1^{D_1}h_2^{D_2+2}\big)^{-1/2}\Big)=o_p(n^{-1/2}).$

We have $U_n=3H_n^{(1)}+o_p(n^{-1/2})$, where $H_n^{(1)}=\frac{1}{n}\sum_{l=1}^{n}\mathrm{E}(\psi_{nitl}|P_l)$. In this case, we need to investigate $H_n^{(1)}$ a little further. Note that $\gamma_2(U_t)-\gamma_2(U_i)=\mathbf{J}\gamma_2(U_i)(U_t-U_i)+\frac{1}{2}(U_t-U_i)'\mathbf{H}\gamma_2(U_{ti})(U_t-U_i)$, where $U_{ti}=\lambda U_i+(1-\lambda)U_t$, for $\lambda\in(0,1)$. Plugging this into $\mathrm{E}(\psi_{nitl}|P_l)$, we have

$$H_n^{(1)} = \frac{1}{n}\sum_{l=1}^{n}\mathrm{E}(\psi_{nitl}|P_l)\equiv\frac{1}{n}\sum_{l=1}^{n}a_{1nl,j}+\frac{1}{n}\sum_{l=1}^{n}b_{1nl,j}$$

$$\text{where}\quad a_{1nl,j} = \frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^{*}D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\mathbf{J}\gamma_2(U_i)(U_t-U_i)\Big|Z_l\right),$$
$$b_{1nl,j} = \frac{U_{ld}}{h_1^{D_1}h_2^{D_2+1}}\mathrm{E}\left(\frac{\eta_i X_{2i,j}^{*}D_d K_{2ti}K_{1li}}{f_U(U_i)f_Z(Z_i)}\frac{1}{2}(U_t-U_i)'\mathbf{H}\gamma_2(U_{ti})(U_t-U_i)\Big|Z_l\right)$$

Since $b_{1nl,j} \le Ch_2|U_{ld}|$, $\mathrm{E}(b_{1nl,j}) = 0$, and $\mathrm{V}\big(\frac{1}{n}\sum_{l=1}^n b_{1nl,j}\big) = O(h_2^2 n^{-1})$, by Chebyshev's Inequality, we have $\frac{1}{n}\sum_{l=1}^n b_{1nl,j} = O_p(h_2 n^{-1/2}) = o_p(n^{-1/2})$, and $H_n^{(1)} = \frac{1}{n}\sum_{l=1}^n a_{1nl,j} + o_p(n^{-1/2})$.

Note that $W_{1d} = \frac{1}{n^3}\binom{n}{3}U_n + o_p(n^{-1/2})$. By exchanging i and l in $H_n^{(1)}$ for future notation convenience, we have

$$\begin{aligned}
\sum_{d=1}^{D_2} W_{1d} &= \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^n\sum_{d=1}^{D_2}\frac{U_{id}}{2h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X_{2l,j}^* D_d K_{2tl} K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}\gamma_2(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\,\Big|\,Z_i\right) + o_p(n^{-1/2}) \\
&\equiv \frac{6}{n^3}\binom{n}{3}\frac{1}{n}\sum_{i=1}^n a_{1ni,j} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n a_{1ni,j} + \left(\frac{6}{n^3}\binom{n}{3} - 1\right)\frac{1}{n}\sum_{i=1}^n a_{1ni,j} + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n a_{1ni,j} + o_p(n^{-1/2}).
\end{aligned}$$

The last equation follows from that $\left(\frac{6}{n^3}\binom{n}{3} - 1\right) = o(1)$, and $\frac{1}{n}\sum_{i=1}^n a_{1ni,j} = O_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $\quad i = t = l, \qquad i = t \neq l, \qquad \psi_{nitl} = 0;$

if $\quad i = l \neq t, \qquad \dfrac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\sum_{l=1}^n \psi_{nitl}$

$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_1(0)}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)U_{id} = O_p\Big(\big(nh_1^{D_1}\big)^{-1}\Big) = o_p(n^{-1/2});$$

if $\quad i \neq t = l, \qquad \dfrac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\sum_{l=1}^n \psi_{nitt}$

$$= \frac{1}{n^3}\sum_{\substack{i=1\\ i\neq t}}^n\sum_{t=1}^n\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_{1ti}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)U_{td} = O_p(n^{-1}) = o_p(n^{-1/2}).$$

In sum, we have $W_{1d} = o_p(n^{-1/2})$.

(ii) $W_{2d} = \dfrac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n\frac{\eta_i X_{2i,j}^* D_d K_{2ti} K_{1li}}{h_1^{D_1}h_2^{D_2+1}f_U(U_i)f_Z(Z_i)}\big(\gamma_2(U_t)-\gamma_2(U_i)\big)\big(\Pi_d(Z_l)-\Pi_d(Z_i)\big) \equiv \dfrac{1}{n^3}\sum_{i=1}^n\sum_{t=1}^n\sum_{l=1}^n\psi_{nitl}.$

If $i \neq t \neq l$, let $U_n = \binom{n}{3}^{-1}\sum_{i\neq t\neq l}\psi_{nitl} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a U-statistic of degree 3.

- $\theta_n = O(h_1^{s_1}) = o_p(n^{-1/2})$;
- $\mathrm{E}(\psi_{nitl}|P_i) = \frac{\eta_i X^*_{2i,j}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \mathrm{E}\Big(D_d K_{2ti} K_{1li} \big(\gamma_2(U_t) - \gamma_2(U_i)\big) \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \Big) \leq \frac{C h_1^{s_1} \left| \eta_i X^*_{2i,j} \right|}{f_U(U_i) f_Z(Z_i)}$,
  $\mathrm{E}(\psi_{nitl}|P_t) = O(h_1^{s_1}), \quad \mathrm{E}(\psi_{nitl}|P_l) = O(h_1^{s_1})$;
- $\sigma_{1n}^2 \leq C\mathrm{E}\Big(\mathrm{E}^2(\psi_{nitl}|P_i) + \mathrm{E}^2(\psi_{nitl}|P_l) + \mathrm{E}^2(\psi_{nitl}|P_t) \Big) = O(h_1^{s_1})$;
- $\mathrm{E}(\psi_{nitl}|P_i, P_t) = \frac{\eta_i X^*_{2i,j} D_d K_{2ti} \left(\gamma_2(U_t) - \gamma_2(U_i)\right)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \mathrm{E}\big(K_{1li} \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big) \big| Z_i \big) \leq \frac{C h_1^{s_1} \left| \eta_i X^*_{2i,j} D_d K_{2ti} \left(\gamma_2(U_t) - \gamma_2(U_i)\right) \right|}{h_2^{D_2+1} f_U(U_i) f_Z(Z_i)}$,
  $\mathrm{E}(\psi_{nitl}|P_i, P_l) = \frac{\eta_i X^*_{2i,j} K_{1li} \left(\Pi_d(Z_l) - \Pi_d(Z_i)\right)}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \mathrm{E}\big(D_d K_{2ti} \big(\gamma_2(U_t) - \gamma_2(U_i)\big) \big| U_i \big) \leq \frac{C \left| \eta_i X^*_{2i,j} K_{1li} \left(\Pi_d(Z_l) - \Pi_d(Z_i)\right) \right|}{h_1^{D_1} f_U(U_i) f_Z(Z_i)}$,
  $\mathrm{E}(\psi_{nitl}|P_t, P_l) = O(h_1^{s_1})$;
- $\sigma_{2n}^2 = O\left(h_1^{2s_1} h_2^{-D_2} + h_1^{2-D_1} + h_1^{2s_1} \right) = O\left(h_1^{2s_1} h_2^{-D_2} + h_1^{2-D_1} \right)$;
- $\sigma_{3n}^2 = O\left(\left(h_1^{D_1-2} h_2^{D_2} \right)^{-1} \right)$;
- $H_n^{(1)} = O_p\left(\left(\frac{\sigma_{1n}^2}{n} \right)^{1/2} \right) = O\left(h_1^{1/2 s_1} n^{-1/2} \right) = o_p(n^{-1/2})$;
- $H_n^{(2)} = O_p\left(\left(\frac{\sigma_{2n}^2}{n^2} \right)^{1/2} \right) = O_p\left(h_1^{s_1} \left(n^2 h_2^{D_2} \right)^{-1/2} + \left(n^2 h_1^{D_1-2} \right)^{-1/2} \right) = o_p(n^{-1/2})$;
- $H_n^{(3)} = O_p\left(\left(\frac{\sigma_{3n}^2}{n^3} \right)^{1/2} \right) = O_p\left(\left(n^3 h_1^{D_1-2} h_2^{D_2+2} \right)^{-1/2} \right) = o_p(n^{-1/2})$.

  We have $U_n = o_p(n^{-1/2})$.

For all other cases, by Markov's Inequality and A5, we have

if $\quad i = t = l, \qquad i = l \neq t, \qquad i = t \neq l, \qquad \psi_{nitl} = 0$;

if $\quad i \neq t = l, \qquad \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq t}}^{n} \sum_{t=1}^{n} \psi_{nitt}$

$$= \frac{1}{n^3} \sum_{\substack{i=1 \\ i \neq t}}^{n} \sum_{t=1}^{n} \frac{\eta_i X^*_{2i,j} D_d K_{2ti} K_{1ti}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_i)} \big(\gamma_2(U_t) - \gamma_2(U_i)\big) \big(\Pi_d(Z_l) - \Pi_d(Z_i)\big)$$

$$= O_p\Big(\frac{h_1}{n} \Big) = o_p(n^{-1/2}).$$

We have $B_{312} = -\frac{1}{n} \sum_{i=1}^{n} a_{1ni,j} + o_p(n^{-1/2})$. For B_{313}, analysis will exactly similar to B_{312}, but note that for the term having order $O_p(n^{-1/2})$ in B_{312}, the corresponding term in B_{313}, denote W'_{1d}, is

$$W'_{1d} \;=\; \frac{1}{n^3} \sum_{i=1}^{n} \sum_{t=1}^{n} \sum_{l=1}^{n} \frac{\eta_i X^*_{2i,j} \big(\gamma_2(U_t) - \gamma_2(U_i)\big) D_d K_{2ti} K_{1lt}}{h_1^{D_1} h_2^{D_2+1} f_U(U_i) f_Z(Z_t)} U_{ld}.$$

The difference here is we have Z_t instead of Z_i, such that $\mathrm{E}(\psi_{nitl}|P_l)=0$ in that $\mathrm{E}(\eta_i X^*_{2i,j}|U_i)=0$. Thus, by the same arguments for the rest of terms, we have $B_{313}=o_p(n^{-1/2})$.

As to B_{32}, analysis is similar to B_{31} given above. For the term having order $O_p(n^{-1/2})$, we can actually combine B_{31} and B_{32} together to work it out. Note that

$$V_{Y2i}-V_{X2i}\beta = \left\{\frac{1}{nh_2^{D_2}f_U(U_i)}\sum_{t=1}^{n}\hat{K}_{2ti}\Big[(\hat{\eta}-\eta_t)(Y_t-X_{2t}\beta)+(v_{Y2t}-v_{X2t}\beta) + \Big(\big(\gamma_2(U_t)-\gamma_2(U_i)\big)-\big(g_2(U_t)-g_2(U_i)\big)\beta\Big)\Big]\right\}\Big(1+O_p(L_{2n})\Big),$$

and the term that is of order $O_p(n^{-1/2})$ involves the third term in bracket, which is $\big(\gamma_2(U_t)-g_2(U_i)\beta-\beta_0\big)-\big(\gamma_2(U_i))-g_2(U_i)\beta-\beta_0\big)=h(U_t)-h(U_i)$. Thus using $\big(h(U_t)-h(U_i)\big)$ instead of $\big(\gamma_2(U_t)-\gamma_2(U_i)\big)$ in W_{1d}, we have $B_3=\frac{1}{n}\sum_{i=1}^n a_{ni}+o_p(n^{-1/2})$, where

$$a_{ni}=\sum_{d=1}^{D_2}\frac{U_{id}}{2h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X^*_{2l}D_dK_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}h(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\Big|Z_i\right).$$

Step 5: Combing orders of B_1,B_2,B_3,B_4, we have $\frac{1}{n}\hat{X}_2'\hat{\eta}(\hat{Y}-\hat{X}_2\beta)=B_{11}+\frac{1}{n}\sum_{i=1}^n a_{ni}+o_p(n^{-1/2})$. Next we investigate $\sqrt{n}(B_{11}+\frac{1}{n}\sum_{i=1}^n a_{ni})$.

Let $\lambda\in\mathbb{R}^{D_2}$ be a non-stochastic vector such that $\lambda'\lambda=1$. Denote $B_{11}+\frac{1}{n}\sum_{i=1}^n a_{ni}=\frac{1}{n}\sum_{i=1}^n(X^*_{2i}\eta_i v_i+a_{ni})\equiv\frac{1}{n}\sum_{i=1}^n b_{ni}$, and we have $\mathrm{E}(\lambda' b_{ni})=0$ as $\mathrm{E}(X^*_{2i}\eta_i v_i)=\mathrm{E}(a_{ni})=0$, and $\mathrm{E}(\lambda' b_{ni}b_{ni}'\lambda)=\lambda'\mathrm{E}(X^*_{2i}\eta_i^2v_i^2X^{*\prime}_{2i})\lambda+\lambda'\mathrm{E}(a_{ni}a_{ni}')\lambda=\lambda'\Phi_1\lambda+\lambda'\mathrm{E}(a_{ni}a_{ni}')\lambda$. Denote $X_{2i,j}=\Pi_{2j}(Z_i)+U_{2i,j}$, the j^{th} element of a_{ni} can be written as

$$\begin{aligned}
a_{ni,j} &= \sum_{d=1}^{D_2}\frac{U_{id}}{h_1^{D_1}h_2^{D_2}}\mathrm{E}\left(\frac{\eta_l X^*_{2l,j}D_dK_{2tl}K_{1il}}{f_U(U_l)f_Z(Z_l)}\mathbf{J}h(U_l)\Big(\frac{U_t-U_l}{h_2}\Big)\Big|Z_i\right)\\
&= \int\frac{1}{h_1^{D_1}h_2^{D_2}}\Big(\Pi_{2j}(Z_l)+U_{2l,j}-g_{1j}(M_l)-g_{2j}(U_l)+g_{3j}\Big)\sum_{d=1}^{D_2}U_{id}D_dK_{2tl}K_{1il}\mathbf{J}h(U_l)\left(\frac{U_t-U_l}{h_2}\right)\\
&\quad\times\frac{\eta_l(M_l,U_l)}{f_U(U_l)f_Z(Z_l)}f_U(U_t)f_{ZUM}(Z_l,U_l,M_l)dU_tdZ_ldU_ldM_l\\
&= \int\Big(\Pi_{2j}(Z_i-h_1\gamma)+U_{2t,j}-h_2\psi_{2j}-g_{1j}(M_l)-g_{2j}(U_t-h_2\psi)+g_{3j}\Big)\sum_{d=1}^{D_2}U_{id}D_dK_2(\psi)K_1(\gamma)\\
&\quad\times\mathbf{J}h(U_t-h_2\psi)\psi\frac{\eta_l(M_l,U_t-h_2\psi)}{f_U(U_t-h_2\psi)f_Z(Z_i-h_1\gamma)}f_U(U_t)f_{ZUM}(Z_i-h_1\gamma,U_t-h_2\psi,M_l)d\gamma d\psi dU_tdM_l
\end{aligned}$$

$$\to \int \Big(\Pi_{2j}(Z_i)+U_{2t,j}-g_{1j}(M_l)-g_{2j}(U_t)+g_{3j}\Big)\sum_{d=1}^{D_2} U_{id}\left(-D_d h(U_t)\right)\eta_l(M_l,U_t)f_{UM|Z}(U_t,M_l|Z_i)dU_t dM_l$$
$$= -\sum_{d=1}^{D_2}\mathrm{E}\left(\Big(\Pi_{2j}(Z_i)+U_{2t,j}-g_{1j}(M_t)-g_{2j}(U_t)+g_{3j}\Big)D_d h(U_t)\eta_t\Big|Z_i\right)U_{id}.$$

The convergence follows by Lemma 1, A3 and that $\int D_d K_2(\psi)\psi d\psi = (0,\cdots,-1,\cdots,0)'$, where -1 appears on the d^{th} position of the vector. Hence, the $(j,m)^{\text{th}}$ element of $\mathrm{E}(a_{ni}a_{ni}')$ converges to

$$\Phi_{2_{(j,m)}} \equiv \mathrm{E}\Bigg[\sum_{d=1}^{D_2}\sum_{\delta=1}^{D_2}\mathrm{E}\left(\Big(\Pi_{2j}(Z_i)+U_{2t,j}-g_{1j}(M_t)-g_{2j}(U_t)+g_{3j}\Big)D_d h(U_t)\eta_t\Big|Z_i\right)$$
$$\times\,\mathrm{E}\left(\Big(\Pi_{2m}(Z_i)+U_{2t,m}-g_{1m}(M_t)-g_{2m}(U_t)+g_{3m}\Big)D_\delta h(U_t)\eta_t\Big|Z_i\right)\mathrm{E}(U_{id}U_{i\delta}|Z_i)\Bigg]$$

By Lyapunov's central limit theorem, we have $\sqrt{n}\big(B_{11}+\frac{1}{n}\sum_{i=1}^n a_{ni}\big)\xrightarrow{d}\mathcal{N}(0,\Phi_1+\Phi_2)$, provided $\lim_{n\to\infty}\sum_{i=1}^n \mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta}=0$ for some $\delta>0$. Note that by Cr Inequality,

$$\sum_{i=1}^n \mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta} = n^{-\delta/2}\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left|\sum_{j=1}^{D_{22}}\lambda_j a_{ni,j}\right|^{2+\delta} \le n^{-\delta/2}D_{22}^{1+\delta}\sum_{j=1}^{D_{22}}\lambda_j^{2+\delta}\mathrm{E}|a_{ni,j}|^{2+\delta}$$

$$\text{where}\quad \mathrm{E}|a_{ni,j}|^{2+\delta} \to \int\left|\sum_{d=1}^{D_2}\mathrm{E}\Big(\big(\Pi_{2j}(Z_i)+U_{2t,j}-g_{1j}(M_t)-g_{2j}(U_t)+g_{3j}\big)D_d h(U_t)\eta_t\Big|Z_i\Big)\right|^{2+\delta}$$
$$\times\,|U_{id}|^{2+\delta}f_{ZU}(Z_i,U_i)dZ_i dU_i$$
$$\le C\sum_{d=1}^{D_2}\int\left|\mathrm{E}\Big(\big(\Pi_{2j}(Z_i)+U_{2t,j}-g_{1j}(M_t)-g_{2j}(U_t)+g_{3j}\big)\Big|Z_i\Big)\right|^{2+\delta}|U_{id}|^{2+\delta}f_{ZU}(Z_i,U_i)dZ_i dU_i$$
$$<\infty \quad\text{since}\quad \mathrm{E}\big(|U_{id}|^{2+\delta}|Z_i\big)<C<\infty \quad\text{and}\quad \mathrm{E}|X_{2i,j}|^{2+\delta}<\infty.$$

Thus $\lim_{n\to\infty}\sum_{i=1}^n \mathrm{E}\big|n^{-1/2}\lambda' a_{ni}\big|^{2+\delta}=0$ for some $\delta>0$, and we have $\frac{1}{n}\hat{X}_2'\hat{\eta}(\hat{Y}-\hat{X}_2\beta)\xrightarrow{d}\mathcal{N}(0,\Phi_1+\Phi_2)$. From step 1, we have $\left(\frac{1}{n}\hat{X}_2'\hat{\eta}\hat{X}_2\right)^{-1}\xrightarrow{p}\Phi_0^{-1}$. Together, we have

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{d}\mathcal{N}\big(0,\Phi_0^{-1}(\Phi_1+\Phi_2)\Phi_0^{-1}\big).$$

□

Lemma 1. *Assume that: a)* $|K(\gamma)| \leq C$ *for all* $\gamma \in \mathbb{R}^D$*; b)* $\int |K(\gamma)| d\gamma < \infty$*; c)* $||\gamma||_E |K(\gamma)| \to 0$ *as* $||\gamma||_E \to \infty$*; d)* $h_n > 0$ *for all n and* $h_n \to 0$ *as* $n \to \infty$*. Let* $f(x) : \mathbb{R}^D \to \mathbb{R}$ *such that e)* $\int |f(\gamma)| d\gamma < \infty$*. Then, for every continuity point x of* $f(x)$*, we have*

$$\int K(\gamma) f_X(x + h_n \gamma) d\gamma \to f_X(x) \int K(\gamma) d\gamma \leq C \quad as \quad n \to \infty$$

Lemma 1 is a standard result. Here we omit the proof.

Lemma 2. *Assume that* $K(x) : \mathbb{R}^D \to \mathbb{R}$ *is a product kernel* $K(x) = \prod_{j=1}^{D} k(x_j)$ *with* $k(x) : \mathbb{R} \to \mathbb{R}$ *such that: a)* $k(x)$ *is continuously differentiable everywhere; b)* $|k(x)||x|^3 \leq C$*, for any* $x \in \mathbb{R}$ *and some* $C > 0$*; c)* $|k'(x)||x|^3 \leq C$*, for any* $x \in \mathbb{R}$ *and some* $C > 0$*. Thus, for any* $|\beta| = 0, \cdots, 3$*,* $K(x)x^\beta$ *satisfies a local Lipschitz condition, i.e., for any* $x \neq y \in A$*, where* $A \subset \mathbb{R}$ *is a bounded convex set, we have*

$$|K(x)x^\beta - K(y)y^\beta| \leq C||x - y||_E, \qquad \text{for some } C > 0.$$

Proof. Note that by a)-c), for any $x \in \mathbb{R}$, we have $|k(x)||x|^i, |k'(x)||x|^i \leq C, i = 0, \cdots, 3$.

(a). $|\beta| = 0$.

Since by mean value theorem $K(x) - K(y) = \mathbf{J}K(x^*)(x - y)$, where $x^* = x + \lambda(y - x)$, $\lambda \in (0, 1)$, and $|D_i K(x^*)| = |k'(x_i^*)| \prod_{p \neq i}^{D} |k(x_p^*)| \leq C$, we have $|K(x) - K(y)| \leq C \sum_{i=1}^{D} |x_i - y_i| \leq CD \left(\sum_{i=1}^{D} (x_i - y_i)^2 \right)^{1/2} \leq C||x - y||_E$ for some $C > 0$ by triangular and C_r Inequality.

(b). $|\beta| = 1$. For any $i = 1, \cdots, D$,

$$\begin{aligned}
|K(x)x_i - K(y)y_i| &= |x_i(K(x) - K(y)) + K(y)(x_i - y_i)| \\
&= |x_i \mathbf{J}K(x^*)(x - y) + K(y)(x_i - y_i)| && \text{by the mean value theorem} \\
&= \left| (x_i D_i K(x^*) + K(y))(x_i - y_i) + \sum_{p \neq i} x_i D_p K(x^*)(x_p - y_p) \right| \\
&\leq C \sum_{i=1}^{D} |x_i - y_i| && \text{by triangular inequality} \\
&\leq C||x - y||_E && \text{by the } C_r \text{ Inequality}
\end{aligned}$$

Mean value theorem is used in the second equation since $k(x)$ is continuously differentiable on the convex set A. And since set A is bounded, there exists a $C \geq 0$ such that $y_i - x_i = \Delta_i$ and $|\Delta_i| \leq C$. Thus $x_i^* \equiv x_i + \lambda(y_i - x_i) = x_i + \lambda\Delta$, and we have $|x_i k'(x_i^*)| = |x_i k'(x_i + \lambda\Delta)| \leq C$ by c).

(c). $|\beta| = 2$. For any $i, j = 1, \cdots, D$,

$$\begin{aligned}
|K(x)x_i x_j - K(y)y_i y_j| &= |x_j(K(x)x_i - K(y)y_i) + K(y)y_i(x_j - y_j)| \\
&\leq |x_j K(x) + x_j y_i D_i K(x^*)|\,|x_i - y_i| + \left|x_j y_i D_j K(x^*) + K(y)y_i\right|\,|x_j - y_j| \\
&\quad + \left|\sum_{p \neq i,j} x_j y_i D_p K(x^*)\right|\,|x_p - y_p| \\
&\leq C\,||x - y||_E
\end{aligned}$$

(d). $|\beta| = 3$. For any $i, j, l = 1, \cdots, D$,

$$\begin{aligned}
\left|K(x)x_i x_j x_l - K(y)y_i y_j y_l\right| &= \left|x_l(K(x)x_i x_j - K(y)y_i y_j) + K(y)y_i y_j(x_l - y_l)\right| \\
&\leq \left|x_i x_j x_l D_i K(x^*) + x_j x_l K(y)\right|\,|x_i - y_i| + \left|x_i x_j x_l D_j K(x^*) + x_l K(y)y_i\right|\,|x_j - y_j| \\
&\quad + \left|x_i x_j x_l D_l K(x^*) + K(y)y_i y_j\right|\,|x_l - y_l| + \sum_{p \neq i,j,l} \left|x_i x_j x_l D_p K(x^*)\right|\,|x_p - x_p| \\
&\leq C\,||x - y||_E
\end{aligned}$$

□

Lemma 3. *Let $\{W_i\}_{i=1}^n$ be a sequence of independent and identically distributed (IID) random variables, $G_n(W_i, w) : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}$ such that: a) $|G_n(W_i, w) - G_n(W_i, w')| \leq B_n(W_i)\|w - w'\|$ for all w, w' and $B_n(W_i) > 0$ with $E(B_n(W_i)) < C < \infty$; b) $E(G_n(W_i, w)) < \infty$ and $E\left(|G_n(W_i, w) - E(G_n(W_i, w))|^p\right) \leq C^{p-2} p! E\left((G_n(W_i, w) - E(G_n(W_i, w)))^2\right) < \infty$ for some $C > 0$ for all $i = 1, 2, \cdots$ and $p = 3, 4, \cdots$. Then, if $S_n(w) = \frac{1}{n}\sum_{i=1}^n G_n(W_i, w)$, for $w \in G_w$ a compact subset of $\mathbb{R}^K$,*

$$\sup_{w \in G_w} |S_n(w) - E(S_n(w))| = O_p\left(\left(\frac{log\, n}{n}\right)^{1/2}\right).$$

Proof. Since G_w is a compact subset of $\mathbb{R}^K$, there exists $w_0 \in \mathbb{R}^K$ such that $G_w \subset B(w_0, r) = \{w \in \mathbb{R}^K : \|w - w_0\| < r\}$.

Thus, for all $w, w' \in G_w$, $\|w - w'\| < 2r$. By the Heine-Borel Theorem, every infinite open cover of G_w contains a finite subcover which we construct as $\{B(w_k, n^{-1/2})\}_{k=1}^{l_n}$ with $w_k \in G_w$ and $l_n < n^{K/2}C$. For $w \in B(w_k, n^{-1/2})$, by condition a), we have

$$|S_n(w) - S_n(w_k)| \quad \leq \quad n^{-1/2}\frac{1}{n}\sum_{i=1}^{n} B_n(W_i) = O_p(n^{-1/2})$$

since $E(B_n(W_i)) < \infty$ and $\{W_i\}_{i=1}^n$ is and IID sequence. Similarly, $|E(S_n(w)) - E(S_n(w_k))| = O(n^{-1/2})$ and using the triangle inequality we have, $|S_n(w) - E(S_n(w))| \leq |S_n(w_k) - E(S_n(w_k))| + O_p(n^{-1/2})$. Since $\left(\frac{n}{logn}\right)^{1/2} n^{-1/2} = o(1)$ it suffices to show that for all $\varepsilon > 0$, there exists a constant Δ_ε such that for $n \geq N$

$$P\left(\left(\frac{n}{logn}\right)^{1/2} \max_{1 \leq k \leq l_n} |S_n(w_k) - E\left(S_n(w_k)\right)| \geq \Delta_\varepsilon\right) \leq \varepsilon.$$

Let $\varepsilon_n = \left(\frac{logn}{n}\right)^{1/2} \Delta_\varepsilon$ and note that

$$P\left(\max_{1 \leq k \leq l_n} |S_n(w_k) - E\left(S_n(w_k)\right)| \geq \varepsilon_n\right) \leq \sum_{k=1}^{l_n} P\left(|S_n(w_k) - E\left(S_n(w_k)\right)| \geq \varepsilon_n\right).$$

Given condition b), and letting $c_n = 4V(G_n(W_i, w_k)) + 2C\varepsilon_n$, by Bernstein's Inequality, we have

$$P\left(\left|\sum_{i=1}^{n} G_n(W_i, w_k) - \sum_{i=1}^{n} E\left(G_n(W_i, w_k)\right)\right| \geq n\varepsilon_n\right) \leq 2\exp\left(-\frac{n\varepsilon_n^2}{c_n}\right) = 2\exp\left(-\frac{\Delta_\varepsilon^2 logn}{c_n}\right) = 2n^{-\frac{\Delta_\varepsilon^2}{c_n}}.$$

Hence, $P\left(\max_{1 \leq k \leq l_n} |S_n(w_k) - E\left(S_n(w_k)\right)| \geq \varepsilon_n\right) \leq 2l_n n^{-\frac{\Delta_\varepsilon^2}{c_n}} < Cn^{K/2 - \frac{\Delta_\varepsilon^2}{c_n}}$. Since, $\varepsilon_n \to 0$ as and $V(G_n(W_i, w_k)) < \infty$, we can choose Δ_ε sufficiently large such that $K/2 - \frac{\Delta_\varepsilon^2}{c_n} < 0$ and

$$P\left(\max_{1 \leq k \leq l_n} |S_n(w_k) - E\left(S_n(w_k)\right)| \geq \varepsilon_n\right) \leq \varepsilon.$$

□

Lemma 4. *Assume that $K(x) : \mathbb{R}^D \to \mathbb{R}$ is a product kernel $K(x) = \prod_{j=1}^{D} k(x_j)$ with $k(x) : \mathbb{R} \to \mathbb{R}$ such that: a) $\int k(x)dx = 1$; b) $|k(x)||x|^{7+c} \to 0$ as $x \to \infty$, for some $c > 0$; c) $k(x)$ is continuously differentiable everywhere, and $|k'(x)||x|^3 \to 0$ as $x \to \infty$. In addition, assume that 1) $\{(X_t, \varepsilon_t)'\}_{t=1,2,\ldots}$ is an independent and identically distributed sequence of random vectors; 2) The joint density of X_t and ε_t is given by $f_{X\varepsilon}(x, \varepsilon) = f_X(x) f_{\varepsilon|X}(\varepsilon|x)$; 3) $f_X(x)$ and all of its partial*

derivatives of order $< s$ are differentiable and uniformly bounded on $\mathbb{R}^D$; 4) $0 < \inf_{x\in\mathscr{G}} f_X(x)$ *and* $\sup_{x\in\mathscr{G}} f_X(x) \leq C$. *Let* $w(X_t - x;x) : \mathbb{R}^D \to \mathbb{R}$ *and* $g(\varepsilon) : \mathbb{R} \to \mathbb{R}$ *be measurable functions. Define*

$$s(x) = \frac{1}{nh_n^D}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x;x)g(\varepsilon_t)$$

where $|\beta| = 0,1,2,3$. *If*

i) $\mathrm{E}\big(|g(\varepsilon_t)|^a \big| X\big) \leq C < \infty$ *for some* $a \geq 2$;

ii) $w(X_t - x;x)$ *satisfies a Lipschitz condition and* $|w(X_t - x,x)| < C$ *for all* $x \in \mathbb{R}^D$;

Then, for an arbitrary compact set $\mathscr{G} \subseteq \mathbb{R}^D$, *we have*

$$\sup_{x\in\mathscr{G}} |s(x) - \mathrm{E}(s(x))| = O_p\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$$

provided that $h_n \to 0$, $nh_n^{D+2} \to \infty$ *and* $\frac{nh_n^D}{\log n} \to \infty$ *as* $n \to \infty$.

Proof. Let $B(x_0,r) = \{x \in \mathbb{R}^D : ||x - x_0||_E < r\}$ for $r \in \mathbb{R}^+$. $\mathscr{G}$ compact implies that there exists $x_0 \in \mathbb{R}^D$ such that $\mathscr{G} \subseteq B(x_0,r)$. Therefore, for all $x,z \in \mathscr{G}$, $||x - z||_E < 2r$. Let $h_n > 0$ be such that $h_n \to 0$ as $n \to \infty$ where $n \in \{1,2,\cdots\}$. For any n, by the Heine-Borel Theorem, every infinite cover for $\mathscr{G}$ contains a finite subcover $\left\{B\left(x^k, C\left(\frac{n}{h_n^{D+2}}\right)^{-1/2}\right)\right\}_{k=1}^{l_n}$ with $x^k \in \mathscr{G}$ and $l_n \leq C\left(\frac{n}{h_n^{D+2}}\right)^{D/2}$. Now let

$$s^{\tau}(x) = \frac{1}{nh_n^D}\sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x;x)g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\leq B_n\}}$$

with $B_1 \leq B_2 \leq \cdots$ such that $\sum_{t=1}^{\infty} B_t^{-a} < \infty$ for some $a > 0$.

$$\sup_{x\in\mathscr{G}} |s(x) - \mathrm{E}(s(x))| \leq \sup_{x\in\mathscr{G}} |s(x) - s^{\tau}(x)| + \sup_{x\in\mathscr{G}} |\mathrm{E}(s(x) - s^{\tau}(x))| + \sup_{x\in\mathscr{G}} |s^{\tau}(x) - \mathrm{E}(s^{\tau}(x))| \equiv T_1 + T_2 + T_3.$$

1. $T_1 = \sup_{x\in\mathscr{G}} \left| (nh_n^D)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x;x)g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|>B_n\}} \right|$. By Chebyshev's Inequality, for $a > 0$,

$P(|g(\varepsilon_t)| > B_t) < \frac{\mathrm{E}(|g(\varepsilon_t)|^a)}{B_t^a} < \frac{C}{B_t^a}$ by i). Consequently,

$$\sum_{t=1}^{\infty} P(|g(\varepsilon_t)| > B_t) < \sum_{t=1}^{\infty} \frac{\mathrm{E}(|g(\varepsilon_t)|^a)}{B_t^a} < C \sum_{t=1}^{\infty} B_t^{-a} < \infty$$

By the Borel-Cantelli Lemma $P\left(\limsup_{t\to\infty}\{|g(\varepsilon_t)| > B_t\}\right) = 0$. Hence, for any $\varepsilon > 0$, there exists an m' such that for all m satisfying $m > m'$ we have $P(|g(\varepsilon_m)| \leq B_m) > 1-\varepsilon$. Since $\{B_t\}_{t=1,2,\cdots}$ is an increasing sequence we conclude that for any $n > m$ we have $P(|g(\varepsilon_m)| \leq B_n) > 1-\varepsilon$. Hence, there exists an N such that for any $n > \max\{N, m\}$ we have that for all $t \leq n$, $P(|g(\varepsilon_t)| \leq B_n) > 1-\varepsilon$ and therefore $\chi_{|g(\varepsilon_t)|>B_n} = 0$ with probability 1, which gives $T_1 = o_{as}(1)$.

2. For T_2, note that by 1) and 2), we have

$$\begin{aligned}
\mathrm{E}(s(x) - s^{\tau}(x)) &= \frac{1}{nh_n^D} \sum_{t=1}^{n} \int \int_{|g(\varepsilon_t)|>B_n} K\left(\frac{X_t - x}{h_n}\right) \left(\frac{X_t - x}{h_n}\right)^{\beta} w(X_t - x; x) g(\varepsilon_t) f_X(X_t) f(\varepsilon_t) dX_t d\varepsilon_t \\
&\leq \int K(\gamma)\gamma^{\beta} w(h_n\gamma; x) f_X(x + h_n\gamma) d\gamma \int |g(\varepsilon)| f_{\varepsilon|X}(\varepsilon|x) \chi_{\{|g(\varepsilon)|>B_n\}} d\varepsilon \\
&\leq C \int |g(\varepsilon)| f(\varepsilon) \chi_{\{|g(\varepsilon)|>B_n\}} d\varepsilon
\end{aligned}$$

due to uniform bound of $w(X_t - x; x)$, $f_X(x)$ and by Lemma 1,

$$\int |K(\gamma)\gamma^{\beta} f_X(x + h_n\gamma)| d\gamma \to |f_X(x)| \int |K(\gamma)\gamma^{\beta}| d\gamma \leq C \quad \text{as} \quad n \to \infty.$$

By Hölder's Inequality, for $a > 1$, we have

$$\int |g(\varepsilon)| f_{\varepsilon|X}(\varepsilon|x) \chi_{\{|g(\varepsilon)|>B_n\}} d\varepsilon \leq \left(\int |g(\varepsilon)|^a f_{\varepsilon|X}(\varepsilon|x) d\varepsilon\right)^{1/a} \left(\int \chi_{\{|g(\varepsilon)|>B_n\}} f_{\varepsilon|X}(\varepsilon|x) d\varepsilon\right)^{1-1/a},$$

where the first integral after the inequality is uniformly bounded by i) and by Chebyshev's Inequality,

$$\left(\int \chi_{\{|g(\varepsilon)|>B_n\}} f_{\varepsilon|X}(\varepsilon|x) d\varepsilon\right)^{1-1/a} = \left(P(|g(\varepsilon)| > B_n|X)\right)^{1-1/a} \leq C \left(\frac{\mathrm{E}\left(|g(\varepsilon)|^a|X\right)}{B_n^a}\right)^{1-1/a} \leq C B_n^{1-a}.$$

Hence, $T_2 = O(B_n^{1-a})$.

3. Rewrite T_3 as: $T_3 = \sup_{x\in\mathscr{G}}|s^\tau(x) - \mathrm{E}(s^\tau(x))| \le \sup_{x\in\mathscr{G}}|s^\tau(x) - s^\tau(x^k)| + \sup_{x\in\mathscr{G}}|\mathrm{E}(s^\tau(x) - s^\tau(x^k))|$

$$+ \max_{1\le k\le l_n}|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \equiv T_{31} + T_{32} + T_{33}.$$

3.1. For $x \in B\left(x^k, C\left(\frac{n}{h_n^{D+2}}\right)^{-1/2}\right)$, we have

$$\begin{aligned}
|s^\tau(x) - s^\tau(x^k)| \le & \frac{1}{nh_n^D}\sum_{t=1}^n \left(\left| K\left(\frac{X_t - x}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^\beta - K\left(\frac{X_t - x^k}{h_n}\right)\left(\frac{X_t - x^k}{h_n}\right)^\beta \right| |w(X_t - x; x)| \right. \\
& \left. + \left| K\left(\frac{X_t - x^k}{h_n}\right)\left(\frac{X_t - x^k}{h_n}\right)^\beta \right| |w(X_t - x; x) - w(X_t - x^k; x^k)| \right) |g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}| \\
\le & \left(\frac{C}{h_n^{D+1}}||x^k - x||_E + h_n \frac{C}{h_n^{D+1}}||x^k - x||_E \right) \frac{1}{n}\sum_{t=1}^n |g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}| \\
\le & C\left(\left(\frac{1}{nh_n^D}\right)^{1/2} + h_n\left(\frac{1}{nh_n^D}\right)^{1/2} \right) \frac{1}{n}\sum_{t=1}^n |g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}|,
\end{aligned}$$

where the second inequality follows by Lemma 2 and b), i.e., local Lipschitz condition and uniform boundedness of $\left| K\left(\frac{X_t - x^k}{h_n}\right)\left(\frac{X_t - x^k}{h_n}\right)^\beta \right|$. By the measurability of g and condition 1) we have that $\{|g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}|\}_{t=1,2,\cdots}$ is IID. By condition i) and Kolmogorov's law of large numbers (LLN) we have $\frac{1}{n}\sum_{t=1}^n (|g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}| - \mathrm{E}(|g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}}|)) = o_p(1)$ and $T_{31} \le C\left(\frac{1}{nh_n^D}\right)^{1/2}$.

3.2. Following similar arguments we have $T_{32} = \mathrm{E}(|s(x) - s(x^k)|) \le C\left(\frac{1}{nh_n^D}\right)^{1/2}$.

3.3. $T_{33} = \max_{1\le k\le l_n}|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))|$. For $\varepsilon_n = \frac{nh_n^D}{\log n}^{-1/2}\Delta_\varepsilon$ with $0 < \Delta_\varepsilon < \infty$ we note that

$$P\left(\max_{1\le k\le l_n}|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \ge \varepsilon_n \right) \le \sum_{k=1}^{l_n} P(|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \ge \varepsilon_n).$$

Let $s^\tau(x^k) - \mathrm{E}(s^\tau(x^k)) = \frac{1}{n}\sum_{t=1}^n Z_{tn}$ with

$$\begin{aligned}
Z_{tn} = & \frac{1}{h_n^D} K\left(\frac{X_t - x^k}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^\beta w(X_t - x^k; x^k) g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}} \\
& - \mathrm{E}\left(\frac{1}{h_n^D} K\left(\frac{X_t - x^k}{h_n}\right)\left(\frac{X_t - x}{h_n}\right)^\beta w(X_t - x^k; x^k) g(\varepsilon_t)\chi_{\{|g(\varepsilon_t)|\le B_n\}} \right).
\end{aligned}$$

By the bounds on $|K(x)||x^\beta|$ and w, $|g(\varepsilon_t)|\chi_{\{|g(\varepsilon_t)|\le B_n\}} \le B_n$ we have that $|Z_{tn}| \le Ch_n^{-D}B_n$. By Bernstein's

Inequality,

$$P(|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \geq \varepsilon_n) \leq 2\exp\left(\frac{-n\frac{\log n}{nh_n^D}\Delta_\varepsilon^2}{2n^{-1}\sum_{t=1}^{n}\mathrm{V}(Z_{tn}) + \frac{2}{3}\frac{CB_n}{h_n^D}\left(\frac{\log n}{nh_n^D}\right)^{1/2}\Delta_\varepsilon^2}\right)$$
$$= 2\exp\left(\frac{-\log n\Delta_\varepsilon^2}{2h_n^D\mathrm{V}(Z_{tn}) + \frac{2}{3}CB_n\left(\frac{\log n}{nh_n^D}\right)^{1/2}\Delta_\varepsilon^2}\right)$$
$$= 2\,n^{-\frac{\Delta_\varepsilon^2}{c(n)}}$$

where $c(n) = 2h_n^D\mathrm{V}(Z_{tn}) + \frac{2}{3}CB_n\left(\frac{\log n}{nh_n^D}\right)^{1/2}\Delta_\varepsilon^2$. Consequently,

$$P\left(\max_{1\leq k\leq l_n}|s^\tau(x^k) - \mathrm{E}(s^\tau(x^k))| \geq \varepsilon_n\right) \leq 2l_n n^{-\frac{\Delta_\varepsilon^2}{c(n)}} \leq 2C\left(\frac{n}{h_n^{D+2}}\right)^{D/2} n^{-\frac{\Delta_\varepsilon^2}{c(n)}} = 2C\left(\frac{1}{h_n^{D+2}n^{\frac{2\Delta_\varepsilon^2}{Dc(n)}-1}}\right)^{D/2}$$
$$< 2C\left(\frac{1}{h_n^{D+2}n}\right)^{D/2}$$

provided $\Delta_\varepsilon^2/D > c(n)$. Hence, given that $nh_n^{D+2} \to \infty$ as $n \to \infty$ the left-hand side of the inequality is $< \varepsilon$ provided $c(n)$ is bounded. To show that $c(n)$ is bounded, we choose B_n such that $B_n\varepsilon_n \to 0$, i.e., $B_n\varepsilon_n = o(1)$, guaranteeing that the second term of $c(n)$ is $o(1)$. Furthermore, $h_n^D\mathrm{V}(Z_{tn}) \leq C$ given condition i) and $\int |K(\gamma)\gamma^{2\beta}|d\gamma < \infty$ for $|\beta| = 0, \cdots, 3$ due to b). Thus, $T_{33} = O\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$.

In sum, we have $T_3 = O\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$.

Combining results from 1 to 3, we have that $\sup_{x\in\mathscr{G}}|s(x) - \mathrm{E}(s(x))| = O(B_n^{1-a}) + O\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$. To show that $B_n^{1-a} = O\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right)$, since $B_n\varepsilon_n = o(1)$ implies that $B_n = o\left(\left(\frac{nh_n^D}{\log n}\right)^{1/2}\right)$, we have

$$\left(\frac{nh_n^D}{\log n}\right)^{1/2} B_n^{1-a} = \left(\frac{nh_n^D}{\log n}\right)^{1/2}\left(\frac{nh_n^D}{\log n}\right)^{(1-a)/2} o(1) = \left(\frac{nh_n^D}{\log n}\right)^{1-a/2} o(1) = o(1),$$

where the last equality follows if $a \geq 2$, which is assumed in i). Thus, we have

$$\sup_{x\in\mathscr{G}}\left|s(x) - \mathrm{E}(s(x))\right| = O_p\left(\left(\frac{\log n}{nh_n^D}\right)^{1/2}\right).$$

□

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795–1843.

Blundell, R., Powell, J. L., 2003. Endogeneity in Nonparametric and Semiparametric Regression Models. Cambridge University Press, Cambridge, UK.

Engle, R. F., Granger, C. W. J., Rice, J., Weiss, A., 1986. Semiparametric Estimates of the Relation Between Weather and Electricity. Journal of the American Statistical Association 81, 310–320.

Hasan, S. A., 2012. Engel Curves and Equivalence Scales for Bangladesh. The Australian National University, ASARC Working Papers 2012-15.

Heckman, N. E., 1986. Spline Smoothing in a Partly Linear Model. Jounal of the Royal Statistical Society Series B, 244–248.

Hoeffding, W., 1961. The strong law of large numbers for U-Statistics.

Hoyle, R. H., 1995. Structural Equation Modeling: Concepts, Issues, and Applications. Vol. 34. Sage Publications, Inc.

Kim, W., Linton, O. B., Hengartner, N., 1999. A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. Journal of Computational and Graphical Statistics 8, 278–297.

Lahiri, K., Schmidt, P., 1978. On the Estimation of Triangular Structural Systems. Econometrica 46, 1217–1221.

Lei, P.-w., Wu, Q., 2007. Introduction to Structural Equation Modeling : Issues and Practical Considerations. Educational Measurement: Issues and Practice 26, 33–43.

Lessmann, C., 2014. Spatial inequality and development - Is there an inverted-U relationship? Journal of Development Economics 106, 35–51.

Li, Q., 1996. On the root-n-consistent semiparametric estimation of partially linear models. Economics Letters 51, 277–285.

Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

Linton, O. B., Hardle, W., 1996. Estimation of additive regression models with known links. Biometrika 83, 529–540.

Martins-Filho, C., Yang, K., 2007. Finite sample performance of kernel-based regression methods for non-parametric additive models under common bandwidth selection criterion. Journal of Nonparametric Statistics 19, 23–62.

Martins-Filho, C., Yao, F., 2012. Kernel-Based Estimation of Semiparametric Regression in Triangular Systems. Economics Letters 115, 24–27.

Newey, W. K., Powell, J. L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equation models. Econometrica 67, 565–603.

Otsu, T., 2011. Empirical likelihood estimation of conditional moment restriction models with unknown functions. Econometric Theory 27, 8–46.

Pagan, A., Ullah, A., 1999. Nonparametric econometrics. Cambridge University Press, Cambridge, UK.

Reiss, P. C., Wolak, F. A., 2007. Chapter 64 Structural Econometric Modeling: Rationales and Examples from Industrial Organization.

Robinson, P. M., 1988. Root-n consistent semiparametric regression. Econometrica 56, 931–954.

Stock, J., 1989. Nonparametric Policy Analysis. Journal of the American Statistical Association 84, 567–575.

Su, L., Ullah, A., 2008. Local polynomial estimation of nonparametric simultaneous equations models. Journal of Econometrics 144, 193–218.

Yao, F., Martins-Filho, C., 2013. An asymptotic characterization of finite degree U-statistics with sample size dependent kernels: applications to nonparametric estimators and test statistics. Communications in Statistics - Theory and methods, forthcoming.