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

© November 2014 Xin Geng
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:

\[ Y_i = \beta_0 + X_2 \beta + m(X_1, Z_1) + \epsilon_i, \quad \text{for } i = 1, \ldots, n \]  

(1)

\[ X_i = \Pi(Z_i) + U_i \]  

(2)

\[ E(U_i|Z_i) = 0, \quad E(\epsilon_i|Z_i, U_i) = E(\epsilon_i|U_i) \]  

(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 \( \epsilon_i \) 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 \( \Pi: \mathbb{R}^{D_1} \rightarrow \mathbb{R}^{D_2} \) is an unknown function. Let \( \text{E}(\cdot) \) denote expectation. Variables \( X_i \) are taken as endogenous in that \( \text{E}(\epsilon_i|X_i) \neq 0 \), and the variables \( Z_i \) are exogenous due to (3).

We are interested in estimating \( \beta \) 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 \( \beta \) 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 $E(\varepsilon_i | X_1, Z, U_i) = E(\varepsilon_i | U_i)$, and $E(X_2 | X_1, Z, U_i) = E(X_2 | Z, U_i) = X_2$. Note that $E(\varepsilon_i | U_i)$ is an unknown function of $U_i$, thus we can denote it by $h(U_i): \mathbb{R}^{D_2} \rightarrow \mathbb{R}$, and using (1), we have:

$$E(Y_i | X_1, Z, U_i) = \beta_0 + X_2\beta + m(X_1, Z) + h(U_i) \quad (4)$$

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 $(\varepsilon_i, U_i)$ be independent of $Z_i$ and $E(U_i) = 0$. 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 $\beta$ 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 $\hat{\beta}$. 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., \( E(\varepsilon_i|Z_i) = 0 \). 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) \( E(\varepsilon_i) = 0 \), the moment restrictions in (3) imply that \( E(\varepsilon_i|Z_i) = 0 \). 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, \( \beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^{D_2} \). Let \( v_i = Y_i - E(Y_i|X_i, Z_i, U_i) \), and rewrite (4) as:

\[
Y_i - X_i \beta - \beta_0 = m(X_i, Z_i) + h(U_i) + v_i, \quad \text{for } i = 1, \cdots, n, \tag{5}
\]

where, by construction, \( E(v_i|X_i, Z_i, U_i) = 0 \).

Note that if \( \beta, \beta_0 \) 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 \( \beta \).

2.1 Moment Conditions

For identification, it is standard to assume that \( E(m(X_i, Z_i)) = E(h(U_i)) = 0 \), since each component in an additive nonparametric model can only be estimated up to an constant. For simplicity, let \( M = (X_i', Z_i')' \). Like in Kim et al.
(1999), define our “instrument” function as
\[ \eta(M_i, U_i) = \frac{f_M(M_i) f_U(U_i)}{\phi(M_i, U_i)} = \eta, \]
where \( f_M \) is the joint marginal density of
\[ M_i = (X_i', Z_i')', \]
\( f_U \) the marginal density of \( U_i \), and \( \phi \) the joint density of \( M_i \) and \( U_i \). The essential reason for choosing such “instrument” function lies in that
\[
E(\eta(M_i, U_i)|M_i) = 1; \quad E(\eta(M_i, U_i)h(U_i)|M_i) = 0.
\]
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 \( \eta \) on both sides of (5), and taking conditional expectations given \( M_i \) and \( U_i \) separately, we have
\[
E(\eta_i(Y_i - X_2i\beta - \beta_0) | M_i) = m(M_i); \quad E(\eta_i(Y_i - X_2i\beta - \beta_0) | U_i) = h(U_i)
\]  \hspace{1cm} (6)
If \( \beta, \beta_0 \) 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 \( \beta \) and \( \beta_0 \). Since \( m(M_i) \) and \( h(U_i) \) can be expressed as conditional expectations containing \( \beta, \beta_0 \) in (6), we can plug them into (5), rearranging, with \( \beta_0 = E(\eta_i(Y_i - X_2i\beta)) \), we have
\[
Y_i^* = X_2i^* \beta + v_i, \quad \text{for} \quad i = 1, \cdots, n,
\]  \hspace{1cm} (7)
where \( Y_i^* \equiv Y_i - E(\eta_i Y_i | M_i) - E(\eta_i Y_i | U_i) + E(\eta_i Y_i) \), and \( X_2i^* \equiv X_2i - E(\eta_i X_2i | M_i) - E(\eta_i X_2i | U_i) + E(\eta_i X_2i) \).
Note that equation (7) provides infinitely many moment conditions to estimate \( \beta \), since by pre-multiplying an arbitrary measurable function \( L(X_{1i}, Z_i, U_i) \), we still have \( E(L(X_{1i}, Z_i, U_i) v_i | X_{1i}, Z_i, U_i) = 0 \). 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 \( \beta \).
In our case, we choose \( L(X_{1i}, Z_i, U_i) = \sqrt{\eta} \). Then, we consider:
\[
\sqrt{\eta_i} Y_i^* = \sqrt{\eta_i} X_2i^* \beta + \sqrt{\eta_i} v_i, \quad \text{for} \quad i = 1, \cdots, n.
\]  \hspace{1cm} (8)
In matrix form we write \( \sqrt{\eta} Y^* = \sqrt{\eta} X_2^* \beta + \sqrt{\eta} v \), where \( Y^* = (Y_1^*, \cdots, Y_n^*)' \), \( X_2^* = (X_{21}^*, \cdots, X_{2n}^*)' \), \( v = (v_1, \cdots, v_n)' \),
\( \sqrt{\eta} = \text{diag} \left\{ \sqrt{\eta_i} \right\}_{i=1}^n \), and \( E(\sqrt{\eta_i} v_i | X_{1i}, Z_i, U_i) = 0 \). Note that by choosing \( \beta_0 = E(\eta_i(Y_i - X_2i\beta)) \) and \( L(X_{1i}, Z_i, U_i) = \sqrt{\eta} \)
\( \sqrt{n} \), we have 
\( \text{E}(\eta Y_i^*|M_i) = \text{E}(\eta Y_i^*|U_i) = \text{E}(\eta X_{2i}^*|M_i) = \text{E}(\eta X_{2i}^*|U_i) = 0 \). These conditions are crucial in establishing the asymptotic properties of our estimator of \( \beta \), 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 \( \gamma_i(M_i) \equiv \text{E}(\eta Y_i|M_i), \gamma_2(U_i) \equiv \text{E}(\eta Y_i|U_i), \gamma_3 \equiv \text{E}(\eta Y_i), \vartheta_1(M_i) \equiv \text{E}(\eta X_{2i}|M_i), \vartheta_2(U_i) \equiv \text{E}(\eta X_{2i}|U_i), \vartheta_3 \equiv \text{E}(\eta X_{2i}), \varphi_{Y_1} = \text{E}(\eta Y_1 - \gamma(M_i), \varphi_{Y_2} = \text{E}(\eta Y_2 - \gamma_2(U_i), \varphi_{X1i} \equiv \eta X_{2i} - \gamma_2(U_i), \varphi_{X2i} \equiv \eta X_{2i} - \gamma_2(U_i) \). Now we have \( \sqrt{n} \eta_i \) as our regressors, and 
\( \text{E}(\sqrt{n} \eta_i) = 0 \). Equation (8) suggests an estimator of \( \beta \) by inserting estimators of \( \sqrt{n} \eta_i \) and \( \sqrt{n} \eta_i \) prior to application of a standard rule, such as no-intercept ordinary least square (OLS) method. Note that by (6), we have \( m(M_i) = \gamma_1(M_i) - \gamma_1(M_i)\beta - \beta_0, \) and \( h(U_i) = \gamma_2(U_i) - \gamma_2(U_i)\beta - \beta_0. \) 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 \( \Phi_0 = \text{E}(\eta X_{2i}X_{2i}^*) \).

### 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

\[
\hat{\Pi}_j(Z_i) = \arg \min_{\theta} \frac{1}{nh_1^2} \sum_{i=1}^{n} (X_{i,j} - \theta)^2 K_1 \left( \frac{Z_i - Z_i}{h_1} \right) \quad \text{for} \ j = 1, \ldots, D_2,
\]

where \( X_{i,j} \) is the \( j^{th} \) element of \( X_i \), \( h_1 > 0 \) is the associated bandwidth, and \( K_1 : \mathbb{R}^{d_1} \to \mathbb{R} \) is a multivariate kernel function. Denote the estimates by \( \hat{\Pi}(Z_i) = (\hat{\Pi}_1(Z_i), \ldots, \hat{\Pi}_{D_2}(Z_i))' \) and calculate the nonparametric residuals \( \hat{U}_i = (\hat{U}_{i1}, \ldots, \hat{U}_{iD_2})' \), where \( \hat{U}_{ij} \equiv X_{i,j} - \hat{\Pi}_j(Z_i) \), for \( j = 1, \ldots, D_2 \) and \( i = 1, \ldots, n \).

2. Obtain Rosenblatt density estimators for \( f_{\hat{U}}, f_M \) and \( \phi \):

\[
\hat{f}_{\hat{U}}(u) = \frac{1}{nh_2^2} \sum_{i=1}^{n} K_2 \left( \frac{\hat{U}_i - u}{h_2} \right), \quad \hat{f}_M(m) = \frac{1}{nh_3^2} \sum_{i=1}^{n} K_3 \left( \frac{M_i - m}{h_3} \right),
\]

\[
\hat{\phi}(m,u) = \frac{1}{nh_4^4} \sum_{i=1}^{n} K_4 \left( \frac{(M'_i - \hat{U}_i') - (m' - u')'}{h_4} \right),
\]
Second, we give the uniform convergence in probability rate of the NW estimator constructed using estimated residuals. In this section, we study the asymptotic properties of the estimators described in the previous section. We first establish

\[ \hat{\beta} = \hat{\beta}_0 + \delta_0, \]

where \( \delta_0 \) is the bias correction. Then we establish the asymptotic normality of the estimator. Finally, we show that the estimator is consistent.}

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

\[
\begin{align*}
\hat{\gamma}(M_i) &= \frac{1}{nh_3} \sum_{i=1}^{n} K_3 \left( \frac{M_i - M_i}{h_3} \right) \hat{\psi} Y_i, \\
\hat{\gamma}_1(M_i) &= \frac{1}{nh_3} \sum_{i=1}^{n} K_3 \left( \frac{M_i - M_i}{h_3} \right) \psi Y_i, \\
\hat{\gamma}_2(U_i) &= \frac{1}{nh_2} \sum_{i=1}^{n} K_2 \left( \frac{U_i - U_i}{h_2} \right) \hat{\psi} Y_i, \\
\hat{\gamma}_3(U_i) &= \frac{1}{nh_2} \sum_{i=1}^{n} K_2 \left( \frac{U_i - U_i}{h_2} \right) \psi Y_i.
\end{align*}
\]

Estimation for expectations \( \gamma_1 \) and \( \gamma_3 \) is trivial, as we can just use the population average with \( \hat{\psi} \) replacing \( \psi \).

i.e., \( \hat{\gamma}_1 = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi} Y_i \), and \( \hat{\gamma}_3 = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi} X_{2i} \). Thus, estimators for \( Y_i^* \) and \( X_{2i}^* \) are given as \( \hat{Y}_i = Y_i - \hat{\gamma}_1(M_i) - \hat{\gamma}_3(U_i) + \hat{\gamma}_1, \quad \hat{X}_{2i} = X_{2i} - \hat{\gamma}_1(M_i) - \hat{\gamma}_3(U_i) + \hat{\gamma}_3 \), for \( i = 1, \ldots, n \).

4. Using the estimators \( \hat{\psi}, \hat{Y}, \hat{X}_2 \) derived in Steps 2 and 3 instead of \( \psi, Y_i^* \) and \( X_{2i}^* \) in (8), we have the no-intercept OLS estimator for \( \beta \):

\[ \hat{\beta} = (X_2' \hat{\psi} X_2)' \hat{\psi} Y, \]

where \( Y = (Y_1, \ldots, Y_n)' \), \( X_2 = (X_{21}, \ldots, X_{2n})' \), and \( \hat{\psi} = \text{diag} \{ \hat{\psi}_i \}_{i=1}^{n} \).

Given \( \beta_0 = E(Y_i - X_{2i} \beta) \) and \( \hat{\beta} \), an estimator for \( \beta_0 \) is \( \hat{\beta}_0 = Y - X_2 \hat{\beta} \), where \( Y = \frac{1}{n} \sum_{i=1}^{n} Y_i \), and \( X_2 \equiv \frac{1}{n} \sum_{i=1}^{n} X_{2i} \). For \( m(M_i) \) and \( h(U_i) \), we have

\[ \hat{m}(M_i) = \hat{\gamma}_1(M_i) - \hat{\gamma}_1(M_i) \hat{\beta} - \hat{\beta}_0, \quad \hat{h}(U_i) = \hat{\gamma}_3(U_i) - \hat{\gamma}_3(U_i) \hat{\beta} - \hat{\beta}_0. \]

### 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., $\int k_i(x)x^jdx = 0$ for $j = 1, \cdots, s_i - 1$, and $\int |k_i(x)||x|^4dx < C$. 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:

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

(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 $\sum_{j=0}^{\frac{1}{2}(s-2)} c_j m_{2(i+j)} = \delta_{0i}$, $0 \leq i \leq \frac{1}{2}(s-2)$ where $\delta_{ij}$ Kronecker’s delta, will give us the desired kernel. For example, $k_4(x) = \left(\frac{1}{2} - \frac{1}{2}x^2\right)(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$, $k_6(x) = \left(\frac{15}{4} - \frac{5}{4}x^2 + \frac{1}{4}x^4\right)(2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$. 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 $r \geq 1$) $k_{2r}(x) = P_{2r-2}(x)$, where $P_{2r} = P_{2r-2} + (-1)H_{2r}(2r!)^{-1}$ and $H_r(x) = xH_{r-1} - (r - 1)H_{r-2}$ is the $r$th 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 \( f_M(M_i), f_Z(Z_i), \phi(M_i, U_i), f_U(Z_i, Z_t), f_U(U_i) \) are uniformly bounded away from zero and infinity.

Assumption A3. \( E(m(M_i)) = E(h(U_i)) = 0, \ E(m^2(M_i)), E(h^2(U_i)) < \infty, \ E(v_i^2|X_{1i}, Z_t, U_i) = \sigma_i^2 < \infty, \ E(U_{ij}^2|Z_t) = \sigma_{ij}^2 < \infty, \ E(v_{x1ij}^2|M_i) = \sigma_{x1ij}^2 < \infty, \ E(v_{x2ij}^2|U_i) = \sigma_{x2ij}^2 < \infty, \ E(v_{y1i}^2|M_i) = \sigma_{y1i}^2 < \infty, \ E(v_{y2i}^2|U_i) = \sigma_{y2i}^2 < \infty, \) and Cramer’s conditions: \( E|X_{2i,j}|^p \leq C^{p-2}p! E|X_{2,i,j}|^2 < \infty, \) \( E[(U_{ij}|^p|Z_t) \leq C^{p-2}p!\sigma_{ij}^2, \) 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 \( d = 1, \cdots, D_2, \Pi_d(Z_t), \phi(M_i, U_i), f_Z(Z_i), m(M_i), \)

\( h(U_i) \in C^\infty. \)

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 \( L_{im} = \left( \frac{\log n}{nh_i} \right)^{1/2} + h_i^n, \) for \( i = 1, \cdots, 4, \) and \( L_n = \sum_{i=2}^4 L_{im}, \) where \( h_i \rightarrow 0 \) as \( n \rightarrow \infty \) and satisfy:
(i) \( h_1 = n^{-\delta}, \) where \( \frac{1}{2n_i} < \delta < \min_{i=2,4} \frac{D_i}{D_1(2^{i-1} + D_i)}, \) and \( \frac{D_i}{D_1} > \max_{i=2,4} \frac{D_i}{D_1} + \frac{1}{2}. \)

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

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

Assumption A5 provides the order of all the bandwidths used in the paper. The fact that, using residual estimates \( \{U_i\}_{i=1}^n \) instead of \( \{U_t\}_{t=1}^n \) 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 \( h_2, h_3, h_5, \) 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 \( \mathcal{G}_Z \subset \mathbb{R}^D, \) we have

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

where \( L_{1n} = \left( \frac{\log n}{nh_1^4} \right)^{1/2} + h_1^{4i}. \) This uniform convergence rate of NW estimator in probability is used throughout this paper. Note that \( f_U(\hat{U}_i) \) and \( \phi(M_i, \hat{U}_i) \) are used to approximate \( f_U(U_i) \) and \( \phi(M_i, U_i) \) in \( \eta_i. \) In Theorem 1, we show that the uniform convergence rate of \( \hat{f}_U(\hat{U}_i) \) to \( f_U(U_i) \) using \( \{U_i\}_{i=1}^n \) 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 \( \phi(M_i, \hat{U}_i). \) All proofs of the theorems are provided in Appendix.

**Theorem 1.** Under A1-A5, for arbitrary convex and compact subsets \( \mathcal{G}_Z \subset \mathbb{R}^D, \) \( \mathcal{G}_U \subset \mathbb{R}^D, \) and \( \mathcal{G}_M \subset \mathbb{R}^D, \) we have

\[
\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} \left| \hat{f}_U(\hat{U}_i) - f_U(U_i) \right| = O_p(L_{2n}), \quad \sup_{M \in \mathcal{G}_M} \left| \hat{f}_M(M_i) - f_M(M_i) \right| = O_p(L_{3n}),
\]

\[
\sup_{(Z, U, M) \in \mathcal{G}_Z \times \mathcal{G}_U \times \mathcal{G}_M} \left| \phi(M_i, \hat{U}_i) - \phi(M_i, U_i) \right| = O_p(L_{4n}),
\]

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

Note that in Theorem 1 we establish the uniform convergence rate of \( \hat{f}_U(\hat{U}_i) \) and \( \phi(M_i, \hat{U}_i) \) over \( \mathcal{G}_Z \times \mathcal{G}_U \) and \( \mathcal{G}_Z \times \mathcal{G}_U \times \mathcal{G}_M \) 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 \( \mathcal{G}_Z. \) Theorem 1 and A2 together imply
that $|\hat{\eta} - \eta_0| = O_p(L_n)$ uniformly, where $L_n = \sum_{i=2}^{d} L_{in}$, 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 $\mathcal{G}_Z$, $\mathcal{G}_U$ and $\mathcal{G}_M$, we have

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

(15)

Similarly, we have the same uniform convergence rate of $\hat{g}_1(M_i)$, $\hat{g}_1(M_i)$ and $\hat{g}_2(U_i)$, as $\hat{g}_2(U_i)$ 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 $L_{1u} / h_2^2 = o(n^{-1/2})$. This result will help establish the order of elements in $\hat{\beta} - \beta$. Note that

$$
\sqrt{n}(\hat{\beta} - \beta) = \left( \frac{1}{n} \hat{X}_2^2 \hat{\eta} \hat{X}_2 \right)^{-1} \frac{1}{\sqrt{n}} \hat{X}_2^2 \hat{\eta}(\hat{Y} - \hat{X}_2 \beta),
$$

(16)

where $\hat{Y} = Y - \hat{\gamma} = (Y - \gamma) - (\hat{\eta} - \gamma) \equiv Y^* - V_Y$, $\hat{X}_2 = X_2 - \hat{\gamma} = (X_2 - g) - (\hat{\gamma} - g) \equiv X_2^* - V_X$;

$$
\gamma \equiv \gamma(M_i) + \gamma(U_i) - \gamma, \quad g_i \equiv g_1(M_i) + g_2(U_i) - g_3;
$$

$$
\hat{\gamma} \equiv \hat{\gamma}(M_i) + \hat{\gamma}(U_i) - \hat{\gamma}, \quad \hat{g}_i \equiv \hat{g}_1(M_i) + \hat{g}_2(U_i) - \hat{g}_3;
$$

$$
V_{Yi} = \hat{\gamma} - \gamma \equiv V_{Y1i} + V_{Y2i} + V_{Y3i}, \quad V_{Xi} = \hat{g}_i - g_i \equiv V_{X1i} + V_{X2i} + V_{X3i}.
$$

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 $\frac{1}{n} \hat{X}_2^2 \hat{\eta} \hat{X}_2$, and (ii) establish asymptotic normality of the second term $\frac{1}{\sqrt{n}} \hat{X}_2^2 \hat{\eta}(\hat{Y} - \hat{X}_2 \beta)$. 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 $\frac{1}{n} \hat{X}_2^i \hat{\eta} \hat{X}_2 \overset{p}{\rightarrow} \Phi_0$ and $\frac{1}{\sqrt{n}} \hat{X}_2^i \hat{\eta} (Y - \hat{X}_2 \beta) \overset{d}{\rightarrow} \mathcal{N}(\Phi_1 + \Phi_2)$, where $\Phi_0$, $\Phi_1$ and $\Phi_2$ are given in Theorem 3.

**Theorem 3.** Under A1-A5, assuming that matrix $\Phi_0$ exists and is nonsingular, we have

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

(17)

where

$$\Phi_{0(i,j)} = E \left[ \eta_i (X_{2i,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}) (X_{2i,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}) \right];$$

$$\Phi_{1(i,j)} = E \left[ \eta_i^2 (X_{2i,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}) (X_{2i,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}) \right] \sigma_i^2;$$

$$\Phi_{2(i,j)} = E \sum_{\delta=1}^{D_1} \sum_{\delta=1}^{D_2} \left[ (\Pi_{2i}(Z_i) - U_{2i,j} - g_{1j}(M_t) - g_{2j}(U_t) + g_{3j}) D_{\delta} h(U_t) \eta_i | Z_i \right] \left[ (\Pi_{2j}(Z_i) - U_{2j,k} - g_{1k}(M_t) - g_{2k}(U_t) + g_{3k}) D_{\delta} h(U_t) \eta_j | Z_i \right];$$

for $j, k = 1, \ldots, D_{22}$.

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 $\Phi_i$, $i = 1, 2, 3$. By proof of Theorem 3, we have that

$$\frac{1}{n} \hat{X}_2^i \hat{\eta} \hat{X}_2 \overset{p}{\rightarrow} \Phi_0, \quad \frac{1}{\sqrt{n}} \hat{X}_2^i \hat{\eta} \overset{d}{\rightarrow} \mathcal{N}(0, \Phi_1), \quad \frac{1}{\sqrt{n}} \hat{X}_2^i \hat{\eta} (V_{Y_2} - V_{X_2} \beta) \overset{d}{\rightarrow} \mathcal{N}(0, \Phi_2).$$

Hence, it’s easy to show that

$$\hat{\Phi}_0 = \frac{1}{n} \hat{X}_2^i \hat{\eta} \hat{X}_2, \quad \hat{\Phi}_1 = \frac{1}{n} \hat{X}_2^i \hat{\eta} \hat{Y} \hat{\eta} \hat{X}_2, \quad \hat{\Phi}_2 = \frac{1}{n} \hat{X}_2^i \hat{\eta} (V_{Y_2} - V_{X_2} \beta) (V_{Y_2} - V_{X_2} \beta)' \hat{\eta} \hat{X}_2$$

(18)

are consistent estimators for $\Phi_0$, $\Phi_1$ and $\Phi_2$ separately, where $\hat{\nu} \equiv Y - X_2 \hat{\beta} - \hat{m} - \hat{h}$.

Given Theorems 2, 3 and (11), we have the uniform convergence rate of $\hat{m}(M_t)$ and $\hat{h}(U_t)$ at $O_p \left( L_n + \frac{t_{\.0} \alpha}{h_2} \right)$, 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 $m(M_t)$ and $h(U_t)$, or even possibly for $\beta$. 

13
With \( \hat{\beta}, \hat{\beta}_0, \hat{m}(M_i) \) and \( \hat{h}(U_i) \), we can estimate \( m(M_i) \) and \( h(U_i) \) by \( \hat{m}(M_i) \) and \( \hat{h}(U_i) \) using local linear regression:

\[
\begin{align*}
(\hat{m}(M_i), \hat{\delta}(M_i)) &= \arg\min_{m,\delta} \frac{1}{n} \sum_{i=1}^{n} (Y_{i1} - m - (M_i - M_i)^\prime \delta) 2 K_3 \left( \frac{M_i - M_i}{h_3} \right), \\
(\hat{h}(U_i), \hat{\delta}(U_i)) &= \arg\min_{h,\eta} \frac{1}{n} \sum_{i=1}^{n} (Y_{i2} - h - (U_i - U_i)^\prime \eta) 2 K_2 \left( \frac{U_i - U_i}{h_2} \right),
\end{align*}
\]

(19)

where \( Y_{i1} = Y_i - X_{2i} \hat{\beta} - \hat{\beta}_0 - \hat{h}(U_i) \), \( Y_{i2} = Y_i - X_{2i} \hat{\beta} - \hat{\beta}_0 - \hat{m}(M_i) \).

And a second stage estimator for \( \beta \) is given as

\[
\hat{\beta} = (X_2^\prime X_2)^{-1} X_2^\prime \tilde{Y}
\]

(20)

where \( \tilde{Y} \) is \( n \times 1 \) with \( i \)th element \( \tilde{Y}_i = Y_i - \hat{m}(M_i) - \hat{h}(U_i) - \hat{\beta}_0, \) and \( X_2 = (X_{21}^\prime, \cdots, X_{2n}^\prime)^\prime \).

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 \( \hat{\beta}, \hat{m}(\cdot) \) for future reference. Consider the following data generating processes (DGPs):

**DGP 1:** \( Y_i = \text{Ln}(|X_{1i} - 1| + 1) \text{ sgn}(X_{1i} - 1) + X_{2i} \beta + \beta_0 + \epsilon_i \)

**DGP 2:** \( Y_i = \frac{\exp(X_{1i})}{1 + c \exp(X_{1i})} + X_{2i} \beta + \beta_0 + \epsilon_i \)

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 \( X_{2i} = Z_{1i}^2 + Z_{2i}^2 + U_{2i} \). \( \epsilon_i \) and \( U_i = (U_{1i}, U_{2i}) \) are generated as

\[
\begin{pmatrix}
\epsilon_i \\
U_i
\end{pmatrix} \sim NID \left( 0, \begin{pmatrix}
1 & \theta & \theta \\
\theta & 1 & \theta^2 \\
\theta & \theta^2 & 1
\end{pmatrix} \right),
\]

where the values \( \theta = 0.3, 0.6 \), and 0.9 indicate weak, moderate, and strong endogeneity. It is easy to verify that \( E(\epsilon_i|Z_i) = 0, E(U_i|Z_i) = 0 \), and thus \( E(\epsilon_i|U_i, Z_i) = E(\epsilon_i|U_i) = \frac{\theta}{1 + \theta^2} (U_{1i} + U_{2i}) \).
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: $k(x) = \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{\sqrt{5}})I(|x| < \sqrt{5})$, 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 $h_i = \frac{1}{11} \hat{\sigma}(W_i)$, 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 $\beta$, 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 $(\hat{\beta}, \hat{m}(\cdot))$ and $(\tilde{\beta}, \tilde{m}(\cdot))$ 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 ($\theta$) increases. As we predicted, the second stage estimators $(\tilde{\beta}, \tilde{m}(\cdot))$ outperform the first stage $(\hat{\beta}, \hat{m}(\cdot))$ 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.

<table>
<thead>
<tr>
<th>DGP</th>
<th>$n = 100$</th>
<th>$n = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta = 0.3$</td>
<td>$\theta = 0.6$</td>
</tr>
<tr>
<td>B</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>$\hat{\beta}, \hat{m}(-)$</td>
<td>0.065 0.062 0.09 0.66</td>
<td>0.069 0.056 0.089 0.644</td>
</tr>
<tr>
<td>$\tilde{\beta}, \tilde{m}(-)$</td>
<td>-0.029 0.044 0.052 0.397</td>
<td>-0.037 0.044 0.057 0.388</td>
</tr>
</tbody>
</table>

### 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 $\gamma, \beta$, define $\gamma^\beta = \prod_{t=1}^D \gamma_t^\beta_t$, $|\beta| = \sum_{t=1}^D \beta_t$. $D_d f(\gamma) = \frac{\partial}{\partial \gamma_d} f(\gamma)$, $D_{dd}^2 f(\gamma) = \frac{\partial^2}{\partial \gamma_d \partial \gamma_d} f(\gamma)$, $D^\beta f(\gamma) = \frac{\partial^{|\beta|}}{\partial \gamma_1 \cdots \partial \gamma_D} f(\gamma)$. $f'(\gamma)$ and $\mathbf{H} f(\gamma)$ denote the Jacobian and Hessian matrix of $f(\gamma)$. Note that for a scalar function $f(\gamma)$, $f'(\gamma)$ is exactly the transpose of the gradient vector of $f(\gamma)$. $A \times B$ denotes the Cartesian product of two sets $A$ and $B$. $\chi_A$ denotes the indicator function for the set $A$, $P(A)$ denotes the probability of event $A$ in the probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{V}(\cdot)$ 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

$$U_n = \left(\begin{array}{c} n \end{array}\right)^{-1} \sum_{(n,k)} \phi_n(P_{i_1}, \cdots, P_{i_k}),$$

where $\sum_{(n,k)}$ denotes the sum over all subsets $1 \leq i_1 < \cdots < i_k \leq n$ of $\{1, \cdots, n\}$. Now let $\phi_n(z_1, \cdots, z_c) = \mathbb{E}(\phi_n(P_1, \cdots, P_c, P_{c+1}, \cdots, P_k)|P_1 = p_1, \cdots, P_c = p_c)$, $\sigma^2_{\phi_n} = \mathbb{V}(\phi_n(P_1, \cdots, P_c))$ and $\theta_n = \mathbb{E}(\phi_n(P_{i_1}, \cdots, P_{i_k})$. In addition, recursively define $h^{(1)}_n(P_1) = \phi_{1n}(p_1) - \theta_n$, $\cdots$, $h^{(c)}_n(P_1, \cdots, P_c) = \phi_{cn}(p_1, \cdots, p_c) - \sum_{j=1}^{c-1} \sum_{\{c,j\}} h^{(j)}_n(P_{i_1}, \cdots, P_{i_j}) - \theta_n$ for $c = 2, \cdots,$
where $H_n^{(j)}(P_{i_1}, \ldots, P_{i_j}) = \binom{n}{j}^{-1} \sum_{(n,j)} h_n^{(j)}(P_{i_1}, \ldots, P_{i_j})$. The order of $U_n$ can be determined by studying each $H_n^{(j)}$ and $\theta_n$ 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 $\sigma^2_{j_n}$. 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 $|\hat{f}_M(M_i) - f_M(M_i)| = o_p(n)$. 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 $K_{2i} = K_2 \left( \frac{U_i - U_{1_i}}{\beta_2} \right)$, $K_{2i} = K_2 \left( \frac{U_i - U_{2_i}}{\beta_2} \right)$, and other kernels similarly. Since $K_2$ is 4-times partially continuously differentiable, by Taylor Theorem,

$$|\hat{f}_U(\hat{U}_i) - \hat{f}_U(U_i)| = \begin{aligned} \frac{1}{nh_2^2} \sum_{i=1}^n (\hat{K}_{2i} - K_{2i}) \\
\frac{1}{nh_2^2} \sum_{i=1}^n \left( \sum_{|\beta|=1}^3 \frac{H^\beta}{\beta!} D^\beta K_{2i} + \sum_{|\beta|=4} \frac{H^\beta}{\beta!} D^\beta K_2 \left( \frac{U_i - U_{1_i}}{\beta_2} + \lambda H \right) \right) \\
\leq \sum_{i=1}^n |T_i| \end{aligned}$$

where $H \equiv \frac{1}{\beta_2} (\hat{U}_i - U_i) - \frac{1}{\beta_2} (\hat{U}_i - U_i), \lambda \in (0, 1)$.

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

1. $|T_i| \leq \sum_{|\beta|=1}^n \left| \frac{1}{nh_2^2} H^\beta D^\beta K_{2i} \right|
\leq \sum_{d=1}^{D_2} \left( \left| \frac{1}{nh_2^2} \sum_{i=1}^n (\hat{U}_{1d} - U_{1d}) D_{1d} K_{2i} \right| + \left| \frac{1}{nh_2^2} \sum_{i=1}^n (\hat{U}_{2d} - U_{2d}) D_{2d} K_{2i} \right| \right)
\equiv \sum_{d=1}^{D_2} (|T_{i1}| + |T_{i2}|)$.
1.1. \[ |T_{11}| \leq |\hat{\Omega}_d - U_d| \left( (n h_2^{D_2+1})^{-1} \sum_{i=1}^n D_d K_{2i} u \right) \equiv |\hat{\Omega}_d - U_d| C_1(U_i). \] It can be shown that by Lemma 4, \[ E(C_1(U_i)) \to D_d f(U_i), \quad \sup_{U \in \mathcal{G}} |C_1(U_i) - E(C_1(U_i))| = O_p \left( \frac{\log n}{n h_2^{D_2+2}} \right) = o_p(1) \]

Thus \( \sup_{U \in \mathcal{G}} |C_1(U_i)| = o_p(1) \). Note that \( |\hat{\Omega}_d - U_d| = |\hat{\Omega}_d(Z_i) - \Pi_d(Z_i)| \), and by uniform convergence rate of Nadaraya-Watson estimator, we have \( \sup_{Z \in \mathcal{G}} |\hat{\Omega}_d - U_d| = O_p(L_{1n}) \). Consequently, \( |T_{11}| = O_p(L_{1n}) \) uniformly.

1.2. Given \( \hat{\Omega}_d(Z_i) = \frac{1}{n h_1 f(Z_i)} \sum_{l=1}^n K_{1l} x_{i,l} d, \) and \( \hat{f}_Z(Z_i) = \frac{1}{n h_1} \sum_{i=1}^n K_{1i} \), we have

\[ -(\hat{\Omega}_d - U_d) = \hat{\Omega}_d(Z_i) - \Pi_d(Z_i) \]

\[ = \frac{1}{n h_1 f(Z_i)} \sum_{l=1}^n K_{1l} \left( U_{id} + \Pi_d(Z_i) - \Pi_d(Z_i) \right) \]

\[ = \left\{ \frac{1}{n h_1 f(Z_i)} \sum_{i=1}^n K_{1i} \left( U_{id} + \Pi_d(Z_i) - \Pi_d(Z_i) \right) \right\} (1 + O_p(L_{1n})) \] (A.1)

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

\[ |T_{12}| \leq \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{l \neq i} \frac{1}{h_1 h_2^{D_2+1} f(Z_i)} K_{1l} D_d K_{2i} U_{id} \right| \]

\[ + \frac{1}{n^2} \sum_{i=1}^n \sum_{l \neq i} \frac{1}{h_1 h_2^{D_2+1} f(Z_i)} K_{1l} D_d K_{2i} \left( \Pi_d(Z_i) - \Pi_d(Z_i) \right) \left( 1 + O_p(L_{1n}) \right) \]

\[ = |T_{121}| + |T_{122}| (1 + O_p(L_{1n})) \]

\[ |T_{121}| = \left| \frac{1}{n^2} \sum_{i=1}^n \sum_{l \neq i} \frac{1}{h_1 h_2^{D_2+1} f(Z_i)} K_{1l} (0) D_d K_{2i} U_{id} \right| \]

\[ + \frac{1}{2} \left( O \left( n^{-3} + \frac{n}{2} \right) \right) \sum_{l=1}^n \sum_{i=1}^n \frac{1}{h_1 h_2^{D_2+1} f(Z_i)} K_{1l} D_d K_{2i} U_{id} \]

\[ = |E_1 + E_2| \].

We can show that \( |E_1| = O_p \left( (n h_1^2 h_2)^{-1} \right) \) by Markov’s Inequality, and \( |E_2| \leq C |U_n| \), where \( U_n = \left( \sum_{i=1}^n \sum_{l \neq i} \frac{1}{h_1 h_2^{D_2+1} f(Z_i)} U_{id} \right) \). Note that \( H_n^{(1)} = \frac{1}{n} \sum_{i=1}^n h_n^{(1)} (U_i, P_i) \) and \( \frac{1}{n} \sum_{i=1}^n \Phi_m (U_i, P_i) \) is a U-statistic. \( \theta_n = E(\Phi_m) = 0 \) in this case. Given Cramer’s
condition in A3, by Lemma 3, we have $\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} |H_n^{(1)}| = O_p((\log n)/n^{1/2})$. For $H_n^{(2)}$, by Theorem 1 in Yao and Martins-Filho (2013), $H_n^{(2)} = (\sigma^2_{2n}/n^2)^{1/2}O_p(1)$. And $\sigma^2_{2n} \equiv \mathbb{V}(\phi_{nlt}) = \mathbb{E}(\phi^2_{nlt}) \leq 4\mathbb{E}^2(Y^2_{nlt}) = O\left((h^D_1 h^D_2 h^D_{2+2} - 1)^1\right)$. Thus $H_n^{(2)} = (n^2 h^D_1 h^D_2 h^D_{2+2} - 1/2)O_p(1)$ uniformly. In sum, $|T_{121}| = O_p(\left((n h^D_1 h^D_2 - 1)^1 + (\log n)/n^{1/2} + (n^2 h^D_1 h^D_2 h^D_{2+2} - 1/2)\right)) = O_p(L_{1n})$ uniformly by A5.

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

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

In sum,

$$\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} |T_1| = O_p(L_{1n}).$$

2. $|T_2| \leq \sum_{|\beta| = 2} \sup_{Z \in \mathcal{G}_Z} \left|\frac{1}{nh^D_2} \sum_{t=1}^n H^D_1 D^B K_{2t}\right|$, when 1 appears in the $d^h$ and $k^h$ position of $\beta$, we have:

$$\left|\frac{1}{nh^D_2} \sum_{t=1}^n H^D_1 D^B K_{2t}\right| \leq \left| \frac{1}{2nh^D_2} \sum_{t=1}^n \left[ (\hat{U}_{id} - U_{id}) - (\hat{U}_{id} - U_{id}) \right] \left[ (\hat{U}_{ik} - U_{ik}) - (\hat{U}_{ik} - U_{ik}) \right] D^B K_{2t}\right|.$$

Since $\sup_{Z \in \mathcal{G}_Z} |\hat{U}_{ab} - U_{ab}| = O_p(L_{1n})$, for $a = i, j$ and $b = d, k$, we have $|T_2| = O_p\left(\frac{l^3_z}{kh^D_2}\right) \frac{1}{nh^D_2} \sum_{i=1}^n |D^B K_{2t}| = O_p\left(\frac{l^3_z}{kh^D_2}\right) C_2(U_i)$ uniformly. As $\mathbb{E} |C_2(U_i)| = O(1)$ uniformly for $U_i \in \mathcal{G}_U$, we have $\sup_{U \in \mathcal{G}_U} |C_2(U_i)| = O_p(1)$ by Markov’s Inequality. Thus, $\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} |T_2| = O_p\left(\frac{l^3_z}{kh^D_2}\right)$.

3. Similarly, $\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} |T_3| = O_p\left(\frac{l^3_z}{kh^D_2}\right)$.

4. $|T_4|$ is different from $|T_2|$ and $|T_3|$ in that $\sup_{U \in \mathcal{G}_U} |C_4(U_i)| = O_p(1/h^D_2)$, where $C_4(U_i) \equiv \frac{1}{nh^D_2} \sum_{t=1}^n |D^B K_{2t}|$, for any $|\beta| = 4$, thus $\sup_{(Z, U) \in \mathcal{G}_Z \times \mathcal{G}_U} |T_4| = O_p\left(\frac{l^4_z}{kh^D_2}\right)$.

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

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

Uniform order of $|\hat{f}(M, \hat{U}) - \phi(M, U)|$ is derived in the similar way under A5.
\textbf{Theorem 3} \textbf{Proof.} We start with the \( j \)th element of \( \hat{g}_2(\hat{U}) - g_2(U) \). Note that

\[
\hat{g}_2(\hat{U}) - g_2(U) = \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} \hat{K}_{2i} \hat{h}_i X_{2i,j} - g_2(U)
\]

\[
= \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} \hat{K}_{2i} \left\{ (\hat{h}_i - h_i)X_{2i,j} + v_{X2i,j} + ((g_2(U)) - g_2(U)) \right\}
\]

\[
= \left\{ \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} K_{2i} C_{X2i} + \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} J_{\hat{K}_{2i}}(\hat{U} - U - (\hat{U} - U)) C_{X2i}
\right. 
\]

\[
+ \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} R_{ji} C_{X2i} \right\} (1 + O_p(L_{2n}))
\]

\[
\equiv \left( \sum_{k=1}^{3} T_k \right) (1 + O_p(L_{2n})),
\]

where \( R_{ji} \) is the remainder term of a Taylor expansion of \( \hat{K}_{2i} \) at \( \left( \frac{U_i - \hat{U}_i}{h_2} \right) \).

We will show that \( T_1 = O_p(L_m) \), \( T_2 = O_p \left( \frac{L_m}{h_2} \right) \), 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 \( C_{X2i} \). By Theorem 2 and A2, we have that

\[
\sup_{\{Z, U\} \in G_2 \times G_U} |\hat{h}_i - h_i| = O_p(L_{2n} + L_{3n} + L_{4n}) \equiv O_p(L_m).
\]

By Markov’s Inequality, \( |T_{11}| \leq O_p(L_m) \frac{1}{nh_2^D} \sum_{i=1}^{n} |K_{2i}X_{2i,j}| = O_p(L_m) \), since by Lemma 1 and A3 ,

\[
E \left( \frac{1}{nh_2^D} \sum_{i=1}^{n} |K_{2i}X_{2i,j}| \right) = \frac{1}{h_{2}^{D}} E \left( |K_{2i}| \Pi_{2i}(Z_i) + U_i \right)
\]

\[
= \int |K_2(\gamma)| \Pi_{2i}(Z_i) + U_i + h_2 \gamma |f_{UZ}(U_i + h_2 \gamma, Z_i) d\gamma dZ_i
\]

\[
\rightarrow \int |K_2(\gamma)| d\gamma \left( \int |\Pi_{2i}(Z_i)| f_{UZ}(U_i, Z_i) dZ_i + U_i f_{U}(U_i) \right) < \infty.
\]

By Chebyshev’s Inequality, we have \( |T_{12}| = \frac{1}{nh_2^D f_{U}(U)} \sum_{i=1}^{n} |K_{2i}v_{X2i,j}| = O_p \left( (nh_2^D)^{-1/2} \right) = O_p(L_m) \), since \( E(T_{12}) = 0 \), and \( V(T_{12}) = E(T_{12}^2) \leq \frac{C}{nh_2^D} E \left( K_{2i}^2 v_{X2i,j}^2 \right) = O \left( (nh_2^D)^{-1} \right) \).
For $T_{13}$, note that by Taylor Theorem,

$$E(T_{13}) = \frac{1}{h_2^{D_2+1} f_U(U_i)} \sum_{i=1}^{n} \mathbb{E} \left( K_{2i} \left( g_{2j}(U_i) - g_{2j}(U_i) \right) \right)$$

$$= \frac{1}{f_U(U_i)} \int K_{2i}(\gamma) \left( g_{2j}(U_i + h_2 \gamma) - g_{2j}(U_i) \right) f_U(U_i + h_2 \gamma) d\gamma$$

$$= O(h_2^{D_2}),$$

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

2. For $T_2$, we have

$$T_2 = \frac{1}{nh_2^{D_2+1} f_U(U_i)} \sum_{i=1}^{n} J_{2i} \left( \hat{U}_i - U_i - (\hat{U}_i - U_i) \right) C_{X_{2i}}$$

$$= O_p \left( \frac{L_{1n}}{h_2} \right) \sum_{d=1}^{D_2} \sum_{i=1}^{n} \left| \hat{U}_i \right| \sum_{i=1}^{n} D_{d_i} K_{2i} \left( (\hat{U}_i - \eta_i) X_{2i,j} + v_{X_{2i},j} + \left( g_{2j}(U_i) - g_{2j}(U_i) \right) \right)$$

$$= O_p \left( \frac{L_{1n}}{h_2} \right),$$

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

3. $R_{1i}$ is the remainder term of a Taylor expansion of $K_{2i}$ at $\left( \frac{\hat{U}_i - U_i}{h_2} \right)$, thus $R_{1i} = \sum_{|\beta| = 1}^{3} D^\beta K_{2i} H^\beta$

$$+ \sum_{|\beta| = 4}^{3} \frac{1}{4!} D^\beta K_{2i} \left( \frac{\hat{U}_i - U_i}{h_2} \right) H^\beta,$$

where $\left( \frac{\hat{U}_i - U_i}{h_2} \right) \equiv \left( \frac{\hat{U}_i - U_i}{h_2} \right) + \lambda H$, $\lambda \in (0, 1)$, and $H = \frac{1}{h_2} (\hat{U}_i - U_i - (\hat{U}_i - U_i))$. Thus, let $T_3 \equiv \sum_{k=1}^{3} T_{3k}$, with

$$T_{31} = \frac{D_3}{nh_2^{D_2+1} f_U(U_i)} \sum_{i=1}^{n} D_{d_1} K_{2i} \left( (\hat{U}_i - U_i) - (\hat{U}_i - U_i) \right) C_{X_{2i}}$$

$$\leq O_p \left( \frac{L_{1n}}{h_2} \right) \sum_{i=1}^{n} |D_{d_1} K_{2i} C_{X_{2i}}| = O_p \left( \frac{L_{1n}}{h_2} \right)$$

by Lemma 1 and A3. Similarly, $T_{32} = O_p \left( \frac{L_{1n}}{h_2} \right)$. By A1, $T_{33} \leq \sum_{i=1}^{n} |C_{X_{2i}}| = O_p \left( \frac{L_{1n}}{h_2} \right)$. By A5, we can show that $|T_3| = O_p \left( \frac{L_{1n}}{h_2} \right)$ uniformly.
Combining 1-3, we have $\sup_{|Z| \leq 2} |\hat{g}_2(U_i) - g_2(U_i)| = O_p \left( L_n + \frac{L_4}{n^2} \right)$. For $\hat{g}_j(M_i) - g_j(M_i)$, note that

$$\hat{g}_j(M_i) - g_j(M_i) = \frac{1}{nh^2_j f_j(M_i)} \sum_{i=1}^{n} K_{3;i} \hat{\eta}_i X_{2i,j} - g_j(M_i)$$

$$= \left\{ \frac{1}{nh^2_j f_j(M_i)} \sum_{i=1}^{n} K_{3;i} \left( (\hat{\eta}_i - \eta_i) X_{2i,j} + \nu_{1;i,j} + \left( (g_j(M_i) - g_j(M_i)) \right) \right) \right\}(1 + O_p(L_3n)) \quad (A.3)$$

Thus order of $|\hat{g}_2(U_i) - g_2(U_i)|$ can be found similarly to $|T_1|$ in part 1. For $|\hat{g}_3j - g_3j|$, we have

$$\hat{g}_3j - g_3j = \frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_i X_{2i,j} - E(\eta_i X_{2i,j}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{\eta}_i - \eta_i) X_{2i,j} + \frac{1}{n} \sum_{i=1}^{n} \eta_i X_{2i,j} - E(\eta_i X_{2i,j})$$

$$= O_p(L_n) + O_p(n^{-1/2}) = O_p(L_n)$$

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

\[ \square \]

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

$$\hat{Y} - \hat{X}_2 \beta = Y - \gamma - (X_2 - \hat{g}_2) \beta = Y - X_2 \beta - (\gamma - \hat{g}_2 \beta)$$

$$= Y - X_2 \beta - m - \beta_0 - ((\gamma - \hat{g}_2) - (\gamma - g) \beta)$$

$$= \nu - \sum_{k=1}^{3} (V_{yk} - V_{xk} \beta).$$

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

$$\frac{1}{n} \hat{X}_2 \eta \hat{X}_2 = \frac{1}{n} \hat{X}^2 \eta X_2 - \frac{1}{n} X_2 \eta V_2 - \frac{1}{n} X_2 \eta X_2 - \frac{1}{n} V_2 \eta V_2 \equiv \sum_{k=1}^{4} A_k,$$

$$\frac{1}{n} \hat{X}_2 \eta (\hat{Y} - \hat{X}_2 \beta) = \frac{1}{n} \hat{X}_2 \eta \nu - \frac{1}{n} \hat{X}_2 \eta (V_{y1} - V_{x1} \beta) - \frac{1}{n} \hat{X}_2 \eta (V_{y2} - V_{x2} \beta) - \frac{1}{n} \hat{X}_2 \eta (V_{y3} - V_{x3} \beta) \equiv \sum_{k=1}^{4} B_k.$$

**The proof has five steps:**

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

23
(2) We show that $\sqrt{n}B_1 \xrightarrow{d} \mathcal{N}(0, \Phi_1)$.

(3) We show that $B_2, B_3 = o_p(n^{-1/2})$.

(4) Let $a_{ni} = \sum_{j=1}^{D_2} \frac{U_{id} h_i(u_i)}{\sum_{j=1}^{D_2} K_{ji}(u_i) h_i(u_i)} E(\eta_i X_{2i,j} X_{2i,k}) |Z|$. We show that $B_3 = \frac{1}{2} \sum_{j=1}^{n} a_{ni} + o_p(n^{-1/2})$.

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

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

$$\Phi_{0(\eta)} = E(\eta_i X_{2i,j} X_{2i,k}) = E\left\{ \eta_i (X_{2i,j} - g_{1j}(M_i) - g_{2j}(U_i) + g_{3j}) (X_{2i,k} - g_{1k}(M_i) - g_{2k}(U_i) + g_{3k}) \right\} < \infty,$$

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

(i) $\eta_i$ is uniformly bounded;

(ii) $E|X_{2i,j} X_{2i,k}| \leq \left( E(X_{2i,j}^2) E(X_{2i,k}^2) \right)^{1/2} \leq \infty$ by Cauchy-Schwarz Inequality;

(iii) $E|X_{2i,j} g_{1k}(M_i)| \leq \left( E(X_{2i,j}^2) E(g_{1k}^2(M_i)) \right)^{1/2}$;

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

By the non-singularity of $\Phi_0$ in A3, we have $A_1^{-1} \xrightarrow{p} \Phi_0^{-1}$. And for $\bar{A}_2 = \frac{1}{n} \sum_{i=1}^{n} \eta_i X_{2i}^2 V_{X_{2i},i}$, the $(k, j)^{th}$ element $A_2(k, j) = \frac{1}{n} \sum_{i=1}^{n} \eta_i X_{2i,k} V_{X_{2i},i,j} \leq O_p(L_n) \frac{1}{n} \sum_{i=1}^{n} |\eta_i X_{2i,k}| = O_p(L_n) = o_p(1)$, by Theorem 3. Similarly we have $A_3, A_4 = o_p(1)$. Thus,

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

**Step 2:** Note that $B_1 = \frac{1}{n} \sum_{i=1}^{n} \hat{X}_2 \eta_i \nu_i = \frac{1}{n} \sum_{i=1}^{n} X_{2i}^2 \eta_i \nu_i - \frac{1}{n} \sum_{i=1}^{n} V_{X_{2i},i}^2 \eta_i \nu_i \equiv B_{11} - B_{12}$. By Levy central limit theorem and Cramer-Wold device, we have $\sqrt{n}B_{11} \xrightarrow{d} \mathcal{N}(0, \Phi_1)$, since

(i). $X_{2i}^2 \eta_i \nu_i$ is IID; (ii). $E(X_{2i}^2 \eta_i \nu_i) = 0$; (iii). $E(\nu_i^2 | Z, U_i) = \sigma_i^2$;

(iv). $V(X_{2i}^2 \eta_i \nu_i) = E(X_{2i}^2 \eta_i^2 \nu_i^2 X_{2i}^2) = \sigma_i^2 E(\eta_i^2 X_{2i}^2 X_{2i,k}^2) = \Phi_1 < \infty$, where

$$\Phi_{1(\eta)} = \sigma_i^2 E(\eta_i^2 X_{2i,j} X_{2i,k}^2) = \sigma_i^2 E\left\{ \eta_i^2 (X_{2i,j} - g_{1j}(M_i) - g_{2j}(U_i) + g_{3j}) (X_{2i,k} - g_{1k}(M_i) - g_{2k}(U_i) + g_{3k}) \right\} < \infty.$$

24
For $B_{12}$, the $j$th element can be written as

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

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

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

$$B_{121} = \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( X_{1t,i} - \eta_i \right) \right\} \left( 1 + O_p(L_{3n}) \right) \equiv \left( \sum_{k=1}^{3} B_{121k} \right) \left( 1 + O_p(L_{3n}) \right)$$

where $B_{1211} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( \hat{\eta}_i - \eta_i \right) X_{2t,i}, \quad B_{1212} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( \hat{\eta}_i \right) X_{2t,i}, \quad B_{1213} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( \hat{g}_{1j}(M_i) - g_{1j}(M_i) \right)$.

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

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

$$B'_{1211} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( X_{2t,i} \right) = E_{1n} + E_{2n},$$

$$E_{1n} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3(0)} \left( X_{2t,i} \right), \quad E_{2n} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3ti} \left( X_{2t,i} \right)$$

By Chebychev’s Inequality and that $E(E_{1n}) = 0, V(E_{1n}) = E(E_{1n}^2) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_i v_i K_{3(0)} \left( X_{2t,i} \right)$, we have $E_{1n} = O_p(n^{-1/2}(\eta_3^{D_1})^{-1}) = o_p(n^{-1/2})$. $E_{2n} \leq CU_n$, where $U_n$ is a U-statistic such that $U_n = \left( \frac{n}{2} \right)^{-1}$.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{ij} = \left( \frac{n}{2} \right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_{ij} \psi_{ij}, \quad \psi_{ij} = \frac{\eta_i v_i K_{3ti}}{h_3 f_{M}(M_i)} \eta_i X_{2t,i}.$$  Since $E(\psi_{ij} | M_i) = 0$, we have $\theta_n = 0, \phi_{1n} = E(\psi_{1n} | P_i) = E(\psi_{1n} | P_i) = \frac{\eta_i v_i K_{3ti}}{h_3 f_{M}(M_i)} \eta_i X_{2t,i}$. Thus $\sigma_{2n}^2 = V(\phi_{1n}) \leq E(\phi_{1n}^2) \leq C \sigma_2^2 E \left( \frac{C(M_i)}{f_{M}(M_i)} \right) \leq E(g_{1j}^2(M_i)) < \infty$ by Lemma 1 and A3. By Theorem 1 in Yao and Martins-Filho (2013), $H_{n}^{(1)} = O_p(\left( \frac{C(M_i)}{f_{M}(M_i)} \right)^{1/2}) = O_p(n^{-1/2})$. $\sigma_{2n}^2 = V(\psi_{1n}) \leq CE(\psi_{1n}^2) \leq C \sigma_2^2 E(\eta_3^{D_1} \eta_i X_{2t,i}^2) = O(h_3^{D_1})$ by
Lemma 1 and A3. $H_n^{(2)} = O_P\left((\frac{\sigma^2_n}{n})^{1/2}\right) = O_P(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2})$. In sum, $B_{1211} = O_P(n^{-1/2})O_P(L_n) = o_p(n^{-1/2})$.

(1b. $B_{1212} = \frac{1}{n} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{\eta_{ij}K_{ij}}{n^{2/3}f_M(M_i)}v_{X_{1i,j}} \equiv E_{1n} + E_{2n}$.

$E_{1n} = o_p(n^{-1/2})$ as $E(E_{1n}) = 0$, $V(E_{1n}) = \frac{1}{n^2} E\left(\frac{\eta_{ij}K_{ij}(0)}{n^{2/3}f_M(M_i)}v_{X_{1i,j}}\right) = O(n^{-3}h_3^{-2D_3}) = o_p(n^{-1}).$

$E_{2n} \leq CU_n \equiv C\left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \psi_{nitt}$ with $\psi_{nitt} = \frac{\eta_{ij}K_{ij}}{n^{2/3}f_M(M_i)}v_{X_{1i,j}}$. We analyze each component in $U_n = \Theta_n + 2H_n^{(1)} + H_n^{(2)}$ by Hoeffding’s decomposition in Hoeffding (1961).

\begin{itemize}
  \item $\Theta_n = \sigma^2_{n} = 0$, as $E(V_i|M_i) = E(V_{X_{1i,j}}|M_i) = 0$;
  \item $\sigma^2_{2n} = V(\phi_{nitt}) \leq CE(\psi_{nitt}) \leq \frac{C\sigma^2 \sigma^2_{n}}{h_3^{2D_3}}E(K^{2}_{nitt}) = O(h_3^{-D_3});$
  \item $H_n^{(1)} = 0, H_n^{(2)} = O_P\left((\frac{\sigma^2_n}{n})^{1/2}\right) = O_p(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2}).$
\end{itemize}

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

(1c. $B_{1213} = \frac{1}{n} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{\eta_{ij}K_{ij}}{h_3^{2/3}f_M(M_i)}(g_{1j}(M_i) - g_{1j}(M_j)) \leq CU_n$, where $U_n = \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \psi_{nitt}$ with

$\psi_{nitt} = \frac{\eta_{ij}K_{ij}}{h_3^{2/3}f_M(M_i)}(g_{1j}(M_i) - g_{1j}(M_j))$ is a $U$-statistic of degree 2.

\begin{itemize}
  \item $\Theta_n = E(\phi_{nitt}|P_i) = 0$, as $E(V_i|M_i) = 0$.
  \item $\phi_{nitt} = E(\phi_{nitt}|P_i) = \frac{\eta_{ij}K_{ij}}{n^{2/3}f_M(M_i)}E(K_{nitt}(g_{1j}(M_i) - g_{1j}(M_j))|M_i) \leq \frac{C\sigma^2}{h_3^{2D_3}}E(K_{nitt}|M_i)$.
  \item $\sigma^2_{n} = E(\phi^2_{nitt}) = O(h_3^{-2D_3}) = o(1)$.
  \item $\sigma^2_{2n} = V(\phi_{nitt}) \leq CE(\psi_{nitt}) \leq \frac{C\sigma^2 \sigma^2_{n}}{h_3^{2D_3}}E(K^{2}_{nitt}(g_{1j}(M_i) - g_{1j}(M_j))^2) = O(h_3^{-D_3+2})$.
  \item $H_n^{(1)} = O_p\left((\frac{\sigma^2_n}{n})^{1/2}\right) = o_p(n^{-1/2}), H_n^{(2)} = O_p\left((\frac{\sigma^2_n}{n})^{1/2}\right) = O_p(n^{-1/2}(nh_3^{D_3})^{-1/2}) = o_p(n^{-1/2}).$
\end{itemize}

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 $\frac{1}{nh_2^{2/3}f_M(U_i)} \sum_{i=1}^{n} R_iC_X2i = o_p(n^{-1/2})$ uniformly, by A.2 in Theorem 3, we have

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

26
where \( B_{1221} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \eta_{t} \nu_{t} K_{2t} \mathbf{C}_{X_{2t}} \), \( B_{1222} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} (\hat{U}_i - U_i) \mathbf{C}_{X_{2t}}}{f_U(U_i)} \), and \( B_{1223} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} (\hat{U}_i - U_i) \mathbf{C}_{X_{2t}}}{f_U(U_i)} \).

Similarly to \( B_{1221} \) 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 \( C_{X_{2t}} \), let \( B_{1222} = \sum_{k=1}^{3} B_{1222k} \).

We show that \( B_{1222k} = o_p(n^{-1/2}) \) for \( k = 1, 2, 3 \).

\[(2a) \quad B_{12221} = - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} (\hat{U}_i - U_i) (\hat{\eta}_t - \eta_t) X_{2t,j}}{f_U(U_i)} \leq O_p(L_n) O_p \left( \frac{L_{1n}}{h_2} \right) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{D_2}{h_2^2} \left| \sum_{j=1}^{n} \eta_{t} \nu_{t} X_{2t,j} D_{ij} K_{2t} \right| = O_p(L_n) O_p \left( \frac{L_{1n}}{h_2} \right) = o_p(n^{-1/2}) \]

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

\[(2b) \quad B_{12222} = - \sum_{d=1}^{D_2} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} (\hat{U}_i - U_i)}{h_2^2 + f_U(U_i)^2} \left( \hat{U}_i - U_id \right) = \left\{ \sum_{d=1}^{D_2} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} K_{1ii}}{h_2^2 + f_U(U_i)^2} \left( U_id + (\Pi_d(Z_i) - \Pi_d(Z_i)) \right) \right\} (1 + O_p(L_{1n})) \]

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

\[T_{1d} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \frac{\eta_{t} \nu_{t} \mathbf{J} K_{2t} K_{1ii}}{h_2^2 + f_U(U_i)^2} U_id = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{t=1}^{n} \sum_{i=1}^{n} \psi_{nit} \]

If \( i \neq t \neq l \), let \( U_n = \left( \begin{smallmatrix} 1 \end{smallmatrix} \right)^{-1} \sum_{i \neq t \neq l} \psi_{nit} = \left( \begin{smallmatrix} 1 \end{smallmatrix} \right)^{-1} \sum_{i < l < t} \psi_{nit} \) be a U-statistic of degree 3. We analyze each component in \( \theta_n = \theta_n + 3\mathcal{H}_n^{(1)} + 3\mathcal{H}_n^{(2)} + \mathcal{H}_n^{(3)} \) by Hoeffding’s decomposition in Hoeffding (1961).

\[\cdot \theta_n, E(\psi_{nit} | P_i), E(\psi_{nit} | P_i, P_i) = 0, \quad \text{as} \quad E(v_{it} | M_i, U_i) = E(v_{it} | U_i) = E(U_i | Z_i) = 0;\]

\[\cdot \sigma_{1n}^2 = \sigma_{2n}^2 = 0;\]

\[\cdot \sigma_{3n}^2 = V(\psi_{nit}) \leq CE(\psi_{nit}) = o_p((h_1^3 h_2^2 + 2 + 1)^{-1});\]
\[ T_n^{(1)} = T_n^{(2)} = 0, \quad T_n^{(3)} = O_p\left(\left(\frac{\sigma_n^2}{n^2}\right)^{1/2}\right) = O_p\left((n^3 h_1^D h_2^{D+2})^{-1/2}\right) = o_p(n^{-1/2}). \]

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

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

\[
\begin{align*}
&\text{if } i = t = l, \\
&\frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \psi_{nli} = 1 \sum_{i=1}^{n} \sum_{l=1}^{n} \eta_{i,t,\nu,\nu,2, \nu} D_d K_2(0) K_1(0) U_{id} = O_p\left((n^3 h_1^D h_2^{D+1})^{-1}\right) = o_p(n^{-1/2}); \\
&\text{if } i = t \neq l, \\
&\frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \psi_{nli} = 1 \sum_{i=1}^{n} \sum_{l=1}^{n} \eta_{i,t,\nu,\nu,2, \nu} D_d K_2(0) K_1 l U_{id} = O_p\left((n h_2^{D+1})^{-1}\right) = o_p(n^{-1/2}); \\
&\text{if } i \neq t = l, \\
&\frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \psi_{nli} = 1 \sum_{i=1}^{n} \sum_{l=1}^{n} \eta_{i,t,\nu,\nu,2, \nu} D_d K_2 l_1 U_{id} = O_p\left((n h_2^{D+1})^{-1}\right) = o_p(n^{-1/2}); \\
&\text{if } i \neq t \neq l, \\
&\frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \psi_{nli} = 1 \sum_{i=1}^{n} \sum_{l=1}^{n} \eta_{i,t,\nu,\nu,2, \nu} D_d K_2 l U_{id} = O_p\left((n h_2^{D+1})^{-1}\right) = o_p(n^{-1/2}).
\end{align*}
\]

In sum, we have \( T_{id} \equiv o_p(n^{-1/2}). \)

(ii) \( T_{2d} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \eta_{i,j,\nu,\nu,2, \nu} D_d K_2 l_1 U_{id} (\Pi_d(Z_i) - \Pi_d(Z_l)) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \psi_{nli}. \)

If \( i \neq t \neq l, \) let \( U_n = \left(\begin{array}{c} T \end{array}\right)^{-1} \sum_{i \neq t \neq l} \psi_{nli} = \theta_n + 3 H_n^{(1)} + 3 H_n^{(2)} + H_n^{(3)} \) be a \( U \)-statistic of degree 3.

- \( \theta_n = E(\phi_{nli}|P_i) = E(\psi_{nli}|P_i, P_j) = E(\psi_{nli}|P_i, P_j) = 0, \) as \( E(\psi_i| M_i, U_i) = E(\psi_{X2, j}|U_i) = 0; \)
- \( \sigma_{1n}^2 = 0; \)
- \( \sigma_{2n}^2 \leq CE\left(E^2(\psi_{nli}|P_i, P_j)\right) = O\left((h_1^{1D} h_2^{D+1})^{-1}\right); \)
- \( \sigma_{3n}^2 \leq V(\phi_{nli}) \leq CE(\psi_{nli}) = O_p\left((h_1^{1D} h_2^{D+1})^{-1}\right); \)

28
\[ \cdot H^{(1)}_n = 0, \quad H^{(2)}_n = O_p \left( \frac{\sigma^2_{n,2}}{n^2} \right) = \frac{O_p \left( h_{n,2}^3 (n^2 h_{n,2}^2) \right)}{n^2} = o_p(n^{-1/2}), \quad H^{(3)}_n = O_p \left( \frac{\sigma^2_{n,2}}{n^2} \right) \\
= O_p \left( n^3 h_{n,2}^3 h_{n,2}^2 \right) = o_p(n^{-1/2}). \]

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

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

\[
\text{if } i = t \neq l, \quad \psi_{nil} = 0;
\]

\[
\text{if } i = t \neq l, \quad \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nil}
\]

\[
= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \eta_j \Pi_i \Pi_j \left( \sum_{k=1}^{n} h_{n,2}^3 h_{n,2}^2 \right)
\]

\[
\text{if } i \neq t \neq l, \quad \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nilt}
\]

\[
\text{if } i \neq t \neq l, \quad \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nilt} = O_p(n^{-1/2});
\]

\[
\text{if } i \neq t \neq l, \quad \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nilt} = o_p(n^{-1/2}).
\]

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

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

\[
B_{12223} = -\frac{D_{t,2}}{D_{l,2}} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \eta_j \left( g_{t,j}(U_i) - g_{l,j}(U_i) \right) D_{t,2} K_{t,2i} \left( U_t - U_l \right)
\]

\[
= \left\{ \sum_{d=1}^{D_{t,2}} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \eta_j \left( g_{t,j}(U_i) - g_{l,j}(U_i) \right) D_{t,2} K_{t,2i} \left( U_t + (\Pi_t(Z_i) - \Pi_l(Z_i)) \right) \right\} \left( 1 + O_p(L_{1n}) \right)
\]

\[
\equiv \left\{ \sum_{d=1}^{D_{t,2}} (W_{td} + W_{ld}) \right\} \left( 1 + O_p(L_{1n}) \right).
\]

We show that \( W_{td}, W_{ld} = o_p(n^{-1/2}) \).

(i) \( W_{td} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \eta_i \eta_j \left( g_{t,j}(U_i) - g_{l,j}(U_i) \right) D_{t,2} K_{t,2i} \left( U_t \right) \equiv \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \psi_{nilt} \)

If \( i \neq t \neq l \), let \( U_n = \left( \frac{n}{n-1} \right) \sum_{i \neq l} \psi_{nilt} = \theta_n + 3H^{(1)}_n + 3H^{(2)}_n + H^{(3)}_n \) be a U-statistic of degree 3:

\[
\cdot \theta_n = \sigma^2_{n,2} = E(\psi_{nilt} | P_t, P_l) = E(\psi_{nilt} | P_t, P_l) = 0, \quad \text{as} \quad E(\psi_{nilt} | M_t, U_t) = E(U_t | Z_t) = 0;
\]

\[
\cdot E(\psi_{nilt} | P_t, P_l) = \frac{\eta_i \eta_j \left( g_{t,j}(U_t) - g_{l,j}(U_l) \right) \left( U_t \right) \left( U_t \right)}{h_{n,2}^3 h_{n,2}^2 \left( f_t(U_t) f_l(U_l) \right)} \leq \frac{c}{h_{n,2}^3 h_{n,2}^2 \left( f_t(U_t) f_l(U_l) \right)};
\]

29
\[ \sigma_2^2 \leq \text{CR}(E^2(\psi_{n+1}|P, P_1)) = \mathcal{O}(h_1^{-D_1}); \]

\[ \sigma_3^2 = \text{V}(\phi_{n+1}) \leq \text{CR}(\psi_{n+1}^2) = \mathcal{O}_p(h_1^{D_1} h_2^{D_2} h_3^{-1}); \]

\[ H_n^{(1)} = 0, \quad H_n^{(2)} = \mathcal{O}_p\left(\left(\frac{\sigma_2}{n^2}\right)^{1/2}\right) = \mathcal{O}_p\left(\left(n^2 h_1^{-1}\right)^{-1/2}\right) = \mathcal{O}_p(n^{-1/2}), \quad H_n^{(3)} = \mathcal{O}_p\left(\left(\frac{\sigma_3}{n^3}\right)^{1/2}\right) = \mathcal{O}_p(n^{-1/2}). \]

We have \( U_n = \mathcal{O}_p(n^{-1/2}) \).

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

\[ \text{if } i = t = l, \quad \psi_{n+1} = 0; \]

\[ \text{if } i = t \neq l, \quad \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{n+1} \]

\[ = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i \neq j} \eta_{ij} \frac{U_i U_j D_k K_{2k}^{K_1}}{f_U(U_i) f_Z(Z_i)} (g_{2j}(U_i) - g_{2j}(U_j)) = \mathcal{O}_p\left(\left(n h_1^{-2}\right)^{-1}\right) = \mathcal{O}_p(n^{-1/2}); \]

\[ \text{if } i \neq t, \quad \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{n+1} \]

\[ = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i \neq j} \eta_{ij} \frac{U_i U_j D_k K_{2k}^{K_1}}{f_U(U_i) f_Z(Z_i)} (g_{2j}(U_i) - g_{2j}(U_j)) = \mathcal{O}_p(n^{-1}) = \mathcal{O}_p(n^{-1/2}). \]

In sum, we have \( W_{ld} = \mathcal{O}_p(n^{-1/2}) \).

(ii) \( W_{ld} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i \neq j} \eta_{ij} \frac{U_i U_j D_k K_{2k}^{K_1}}{f_U(U_i) f_Z(Z_i)} (g_{2j}(U_i) - g_{2j}(U_j)) (\Pi_d(Z_i) - \Pi_d(Z_j)) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{n+1}. \)

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

\[ \theta_n = \text{E}(\psi_{n+1}|P) = \text{E}(\psi_{n+1}|P_1) = \text{E}(\psi_{n+1}|P_1, P_1) = 0, \quad \text{as} \quad \text{E}(v_i|U_i, M_i) = 0; \]

\[ \text{E}(\psi_{n+1}|P) = \frac{n^2 \eta_{ij} D_k K_{2k}^{K_1}}{h_1^{D_1} h_2^{D_2} h_3^{D_3}} = \mathcal{O}(1); \]

\[ \sigma_1^2 \leq \text{CR}(E^2(\psi_{n+1}|P)) \leq \mathcal{O}(1); \]

\[ \text{E}(\psi_{n+1}|P_1, P_1) = \frac{n^2 \eta_{ij} D_k K_{2k}^{K_1}}{h_1^{D_1} h_2^{D_2} h_3^{D_3}} \mathcal{O}(1); \]

\[ \text{E}(\psi_{n+1}|P_1, P_1) = \frac{n^2 \eta_{ij} D_k K_{2k}^{K_1}}{h_1^{D_1} h_2^{D_2} h_3^{D_3}} \mathcal{O}(1); \]

\[ \sigma_2^2 \leq \text{CR}(E^2(\psi_{n+1}|P, P_1)) = \mathcal{O}(1); \]

\[ \sigma_3^2 = \text{V}(\phi_{n+1}) \leq \text{CR}(\psi_{n+1}^2) = \mathcal{O}(h_1^{D_1} h_2^{D_2} h_3^{-1}); \]

\[ \sigma_4^2 = \text{V}(\phi_{n+1}) \leq \text{CR}(\phi_{n+1}^2) = \mathcal{O}(h_1^{D_1} h_2^{D_2} h_3^{-1}); \]
Step 3: We first show that $B_4 = o_p(n^{-1/2})$. Note that $-B_4 = \frac{1}{n} \bar{X}_2^2 \eta (V_{Y2} - V_{X3} \beta) = \frac{1}{n} \bar{X}_2^2 \eta V_{Y3} - \frac{1}{n} \bar{X}_2^2 \eta V_{X3} \beta \equiv B_{41} + B_{42}$.

By Theorem 3, we have $|V_{X1}|, |V_{Y3}| = O_p(L_n)$. Thus $B_{41} = V_{Y3} (\frac{1}{n} \sum_{i=1}^n X_{2i} \eta - \frac{1}{n} \sum_{i=1}^n V_{X1} \eta_i) = O_p(L_n) (O_p(n^{-1/2}) + O_p(L_n)) = o_p(n^{-1/2})$ by A5. Similarly, $-B_2 = \frac{1}{n} \bar{X}_2^2 \eta (V_{Y2} - V_{X2} \beta) \equiv B_{21} + B_{22}$, 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 $B_{21} = \frac{1}{n} \sum_{i=1}^n X_{2i} \eta_i V_{Y1i} - \frac{1}{n} \sum_{i=1}^n V_{X1i} \eta_i V_{Y1i} \equiv B'_{21} + o_p(n^{-1/2})$ 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

$$B'_{21,j} = \left( \frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \frac{\eta_i X_{2i,l} K_{3l} \phi_{(Y_{1l})}}{f_{M_i}(M_l)} \right) (1 + O_p(L_{3n}))$$

$$= \left( \sum_{k=1}^3 B_{21k} \right) (1 + O_p(L_{3n}))$$

where $B_{21i} = \frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \frac{\eta_i X_{2i,l} K_{3l}}{f_{M_i}(M_l)} (\hat{\eta} - \eta_i) Y_{1i}$, $B_{212} = \frac{1}{n^2} \sum_{i=1}^n \sum_{l=1}^n \frac{\eta_i X_{2i,l} K_{3l}}{f_{M_i}(M_l)} V_{Y1l}$.
We show that $B_{21k} = o_p(n^{-1/2})$ for $k = 1, 2, 3$.

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

$$B'_{211} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta X_{2i,j}^* K_{3i} \eta Y_i = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta \psi_i \equiv E_{21n} + E_{21n},$$

with

$$E_{1n} = \frac{1}{n^2} \sum_{i=1}^{n} \eta X_{2i}^* \eta Y_i, \quad E_{2n} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} O_{n}.$$ 

By Markov’s Inequality and that $E|E_{1n}| \leq \frac{C}{n h_3^2} E|X_{2i}^* Y_i| = O\left(\left(n h_3^2\right)^{-1}\right)$, we have $E_{1n} = o_p(n^{-1/2})$, $E_{2n} \leq CU_n$, where $U_n$ is a $U$-statistic of degree 2 such that $U_n = \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nii}$.

- $\theta_n = E(\psi_{nii}|P_i) = 0$, as $E(\eta X_{2i}^* | M_i) = 0$;
- $\phi_{in} = E(\psi_{nii}|P_i) = \frac{\eta X_{2i}^*}{h_3} E(K_{3i} \eta Y_i | M_i) = \frac{\eta X_{2i}^*}{h_3} \int K(\psi) \gamma_i (M_i + h_3 \psi) f_M(M_i + h_3 \psi) d\psi$;
- $\sigma_{in}^2 \leq CE(E^2(\psi_{nii}|P_i)) = O(1)$ by Lemma 1 and A3.
- $\sigma_{2n}^2 \leq CE(E^2(\psi_{nii}|P_i)) = \frac{C}{n h_3^2} E\left(\frac{(\eta X_{2i}^*)^2 (\eta Y_i)^2 K_{3i}^{2}}{f_M(M_i)}\right) = O\left(h_3^{-D_i}\right)$;
- $H_n^{(1)} = O_p\left(\left(\frac{n^2}{n^2}\right)^{1/2}\right) = O_p\left(\left(\frac{n^2}{n^2}\right)^{1/2}\right) = O_p\left((n^2 h_3^2)^{-1/2}\right) = o_p(n^{-1/2})$.

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). $B_{212} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta X_{2i,j}^* K_{3i} \eta Y_{11} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nii} = E_{1n} + E_{2n}$, where

$$E_{1n} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nii} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nii} = E_{1n} + E_{2n},$$

where $E_{2n} \leq CU_n$, and $U_n = \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{nii} = \theta_n + 2H_n^{(1)} + H_n^{(2)}$ is a $U$-statistic of degree 2.

- $\theta_n = \sigma_{in}^2 = 0$, as $E(\eta X_{2i}^* | M_i) = E(\psi_{nii}|M_i) = 0$;
- $\sigma_{2n}^2 = V(\phi_{nii}) \leq CE(\psi_{nii}^2) = O(h_3^{-D_i})$;
- $H_n^{(1)} = 0$, $H_n^{(2)} = O_p\left(\left(\frac{n^2}{n^2}\right)^{1/2}\right) = O_p\left((n^2 h_3^2)^{-1/2}\right) = o_p(n^{-1/2})$.

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

32
(3c). \( B_{213} = \frac{1}{n} \sum_{i \neq t} \sum_{i=1}^{n} \frac{\eta_i X_{2i,j} K_{3i}^{0} K_{4i}^{0} \hat{Y} (M_t) - \hat{Y} (M_i)}{h_3 f_{M}(M_t)} \) which has a degree 2 \( U \)-statistic where

\[
\psi_{\text{crit}} \leq C U_n, \text{ where } U_n = \sum_{i \neq t}^{n} \psi_{\text{crit}} \text{ is a } U \text{-statistic of degree 2.}
\]

- \( \theta_n = \text{E}(\psi_{\text{crit}} | P_t) = 0 \), as \( \text{E}(\eta_i X_{2i,j}^0 | M_t) = 0 \);
- \( \phi_t = \text{E}(\psi_{\text{crit}} | P_t) = \frac{\eta_i X_{2i,j}^0}{h_3 f_{M}(M_t)} \text{E}(K_{3i} (\hat{Y}(M_t) - \hat{Y}(M_i)) | M_t) \leq \frac{C h_3^2 \eta_i X_{2i,j}^0}{f_{M}(M_t)} \),
- \( \sigma^2_t \leq \text{E}(\phi_t^2) = O(h_3^2) = o(1) \);
- \( \sigma^2_{2n} = \text{V}(\psi_{\text{crit}}) \leq C E(\psi_{\text{crit}}^2) = O(h_3^{-2}) \);
- \( H_n^{(1)} = O_p \left( \left( \frac{\sigma^2_{2n}}{n} \right)^{1/2} \right) = o_p(n^{-1/2}) \), \( H_n^{(2)} = O_p \left( \left( \frac{\sigma^2_{2n}}{n} \right)^{1/2} \right) = O_p(n^{-1/2}) \).

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 \( -B_3 = \frac{1}{n} \hat{Y}(V y - V x_2 \beta) \equiv B_{31} + B_{32} \). 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 \( B_{31} = \frac{1}{n} \sum_{i=1}^{n} X_{2i} \eta_i V_{Y2i} - \frac{1}{n} \sum_{i=1}^{n} V_{Yi} \eta_i V_{Y2i} \equiv B_{31} + o_p(n^{-1/2}) \). Similar to A.2 given in Theorem 3, by Taylor Theorem, we have

\[
V_{Y2i} = \hat{Y}(\hat{U}_i) - \gamma(\hat{U}_i) = \frac{1}{nh_2^2 f_{\hat{U}}(\hat{U}_i)} \sum_{i=1}^{n} \hat{K}_{2i} \eta_i Y_i - \gamma(\hat{U}_i)
\]

\[
= \left\{ \frac{1}{nh_2^2 f_{\hat{U}}(\hat{U}_i)} \sum_{i=1}^{n} \hat{K}_{2i} \left( (\hat{\eta}_i - \eta_i) Y_i + V y_{2i} + (\gamma(\hat{U}_i) - \gamma(\hat{U}_i)) \right) \right\} \left( 1 + O_p(L_{-2n}) \right)
\]

\[
= \left\{ \frac{1}{nh_2^2 f_{\hat{U}}(\hat{U}_i)} \sum_{i=1}^{n} K_{2i} C_{2i} Y_{2i} + \frac{1}{nh_2^2 f_{\hat{U}}(\hat{U}_i)} \sum_{i=1}^{n} K_{2i} \left( \hat{U}_i - U_i - (\hat{U}_i - U_i) \right) C_{2i} Y_{2i} \right\} \left( 1 + O_p(L_{-2n}) \right)
\]

where \( R_{2i} \) is the remainder term of a Taylor expansion of \( \hat{K}_{2i} \) at \( \left( \frac{U_2 - U_i}{h_2} \right) \).

Similar to the \( T_3 \) term in Theorem 3, we have \( \frac{1}{nh_2^2 f_{\hat{U}}(\hat{U}_i)} \sum_{i=1}^{n} R_{2i} C_{2i} Y_{2i} = o_p(n^{-1/2}) \) uniformly. Thus, we have the \( j \)th element of \( B_{31}' \) as

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

33
where
\[ B_{311} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i) C_{Y2i}, \]
\[ B_{312} = - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i) C_{Y2i}, \]
\[ B_{313} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i) C_{Y2i}, \]
and \( C_{Y2i} = (\hat{h}_i - \eta_i) Y_i + v_{Y2i} + (\gamma_2(U_i) - \gamma_2(U_i)) \).

We will show that \( B_{311} = B_{313} = o_p(n^{-1/2}) \) and \( B_{312} = \frac{1}{n} \sum_{i=1}^{n} a_{1i,j} + o_p(n^{-1/2}) \), where
\[ a_{1i,j} = \sum_{d=1}^{D_2} \frac{U_{id}}{2h_1^d h_2^d} E \left( \frac{\eta_i X_{2i,j} D_d K_{2i}(\hat{U}_i - U_i)}{f_U(U_i)f_Z(Z_i)} J\gamma_2(U_i) \left| U_i \right| \right). \]

The components in \( B_{311} \) are similar to \( B_{121} \) with \( U_i \) replacing \( M_i \), \( \eta_i X_{2i,j} \) replacing \( \eta_i v_i \), \( C_{Y2i} \) replacing \( C_{X1i,j} \), and \( E(\eta_i X_{2i,j}|U_i) = 0 \) 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 \( C_{Y2i} \), we have \(-B_{312} = \sum_{k=1}^{n} B_{312k} \) with
\[ B_{3121} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i)(\hat{h}_i - \eta_i) Y_i, \]
\[ B_{3122} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i) v_{Y2i}, \]
\[ B_{3123} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i)(\gamma_2(U_i) - \gamma_2(U_i)). \]

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

(4a) \( B_{3121} = \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} K_{2i}(\hat{U}_i - U_i) \frac{Y_i}{h_2} \right\} o_p(L_n) \)
\[ \leq O_p(L_n) O_p \left( \frac{L_n}{h_2} \right) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{d=1}^{D_2} \frac{n_0}{h_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} \frac{Y_i D_d K_{2i}}{h_2^2 f_U(U_i)} \]
\[ = O_p(L_n) O_p \left( \frac{L_n}{h_2} \right) = o_p(n^{-1/2}) \]

(4b) By A.1 in proof of Theorem 2, we have
\[ B_{3122} = \sum_{d=1}^{D_2} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} v_{Y2i} D_d K_{2i} (\hat{U}_i - U_i) \]
\[ = \left\{ \sum_{d=1}^{D_2} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i X_{2i,j} v_{Y2i} D_d K_{2i} K_{1i} \left( (U_i + (\Pi_d(Z_i) - \Pi_d(Z_i)) \right) \right\} \left( 1 + O_p(L_n) \right) \]
\[ = \left\{ \sum_{d=1}^{D_2} \left( T_{id} + T_{2d} \right) \right\} \left( 1 + O_p(L_n) \right). \]
We show that $T_{1d}, T_{2d} = o_p(n^{-1/2})$.

(i) $T_{1d} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} U_{id} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \psi_{nil}$

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

. $\theta_n = \sigma_{\psi}^2 = E(\psi_{nil} | P_t, P_l) = E(\psi_{hil} | P_l, P_t) = 0$, as $E(\psi_{Y_2} | U_i) = E(U_{id} | Z_i) = 0$;

. $\phi_{2n} = E(\psi_{nil} | P_t, P_l) = \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} \leq \frac{C|\psi_{Y_2} U_i|}{h_2}$;

. $\sigma_{\phi}^2 \leq E(\phi_{2n}^2) = O(h_2^{-2})$;

. $\sigma_{\psi}^2 = V(\phi_{nil}) = CE(\psi_{nil}^2) = O_p((h_1^2 h_2^2 + 2)^{-1})$;

. $H_n^{(1)} = O_p\left(\left(\frac{\sigma_{\psi}^2}{n^3}\right)^{1/2}\right) = O_p\left((nh_2)\right) = o_p(n^{-1/2})$;

. $H_n^{(3)} = O_p\left(\left(\frac{\sigma_{\psi}^2}{n^3}\right)^{1/2}\right) = O_p\left((nh_2^2 h_2^2 + 2)^{-1/2}\right) = o_p(n^{-1/2})$.

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

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

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

$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} U_{id} = O_p\left((nh_2^2 h_2^2 + 1)^{-1}\right) = o_p(n^{-1/2})$;

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

$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} U_{id} = O_p\left((nh_2^2)^{-1}\right) = o_p(n^{-1/2})$;

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

$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} U_{id} = O_p\left((nh_2^2)^{-1}\right) = o_p(n^{-1/2})$;

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

$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\eta X_{2i,j} \nu_{Y_2} D_d K_{2i} K_{1i} h_{1i} h_{2i} f_U(U_i) f_Z(Z_i)}{h_1^2 h_2^2 + 1} U_{id} = O_p\left((nh_2)^{-1}\right) = o_p(n^{-1/2})$.  

35
In sum, we have $T_{1d} = o_p(n^{-1/2})$.

(ii) \[ T_{2d} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \eta_i X_{2i,j}^{(i)} Y_{2j} D_{d} K_{1li} \left( \Pi_d(Z_i) - \Pi_d(Z_l) \right) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \Psi_{nil} \]

If $i \neq t \neq l$, let $U_n = \left( \sum_{i \neq t \neq l} \right)^{-1} \sum_{i \neq t \neq l} \Psi_{nil} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)}$ be a $U$-statistic of degree 3.

- $\theta_n = E(\Psi_{nil} | P_t) = E(\Psi_{nil} | P_l) = E(\Psi_{nil} | P_t, P_l) = 0$, as $E(\Psi_{nil} | U_i) = 0$;
- $E(\Psi_{nil} | P_t) = \frac{\psi_{y2}}{h_1^l h_2^t} E\left( \eta_i X_{2i,j}^{(i)} D_{d} K_{1li} \left( \Pi_d(Z_i) - \Pi_d(Z_t) \right) | U_i \right) \leq \frac{C(\psi_{y2})}{h_2^t}$;
- $\sigma_i^2 \leq E(\Phi_i^2) = O\left( \frac{1}{h_2^t} \right)$;
- $E(\Psi_{nil} | P_t, P_l) = \frac{\eta_i X_{2i,j}^{(i)} Y_{2j} D_{d} K_{1li}}{h_1^l h_2^t} E\left( K_{1li} \left( \Pi_d(Z_i) - \Pi_d(Z_l) \right) | Z_l \right) \leq \frac{C(\psi_{y2})}{h_2^t} E(\Psi_{nil} | P_t, P_l) = \frac{\psi_{y2}}{h_1^l h_2^t} E\left( \eta_i X_{2i,j}^{(i)} Y_{2j} D_{d} K_{1li} \left( \Pi_d(Z_l) - \Pi_d(Z_i) \right) | U_i, Z_l \right) \leq \frac{C(\psi_{y2})}{h_2^t}$;
- $\sigma_i^2 \leq CE(E^2(\Psi_{nil} | P_t, P_l) + E^2(\Psi_{nil} | P_t, P_l)) = O\left( \frac{1}{h_2^t} \right)$;
- $\sigma_i^2 \leq V(\Phi_i) \leq CE(E^2(\Psi_{nil})) = O_p(h_1^{D_1^2} h_2^{D_2 + 2})^{-1}$;
- $H_n^{(1)} = O_p\left( \frac{\sigma_i^2}{n^3} \right)^{1/2} = o_p(n^{-1/2})$;
- $H_n^{(2)} = O_p\left( \frac{\sigma_i^2}{n^3} \right)^{1/2} = o_p\left( n^{D_2 + 2} h_2^{D_2 + 2} \right)^{-1/2} = o_p(n^{-1/2})$;
- $H_n^{(3)} = O_p\left( \frac{\sigma_i^2}{n^3} \right)^{1/2} = o_p\left( n^{D_1^2} h_2^{D_2 + 2} \right)^{-1/2} = o_p(n^{-1/2})$.

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

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

if $i = t = l$, $i = l \neq t$, $\Psi_{nil} = 0$;

if $i = t \neq l$, \[
\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \Psi_{nil} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \eta_i X_{2i,j}^{(i)} Y_{2j} D_{d} K_{1li} \left( \Pi_d(Z_i) - \Pi_d(Z_l) \right) = o_p\left( h_1 \left( n h_2^{D_2 + 1} \right)^{-1} \right) = o_p(n^{-1/2});
\]

if $i \neq t = l$, \[
\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \Psi_{nil} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \eta_i X_{2i,j}^{(i)} Y_{2j} D_{d} K_{1li} \left( \Pi_d(Z_l) - \Pi_d(Z_i) \right) = o_p\left( h_1 \left( n h_2 \right)^{-1} \right) = o_p(n^{-1/2}).
\]

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

36
\[ B_{3123} = \frac{D_3}{n} \sum_{i=1}^{n} \sum_{l=1}^{n} \eta X_{2i,j} \left( \gamma_i(U_l) - \gamma_j(U_l) \right) D_d K_{2ii} \left( \hat{Q}_{id} - U_{id} \right) \]

\[ = - \left\{ \frac{D_3}{n} \sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{j=1}^{n} \eta X_{2i,j} \left( \gamma_i(U_l) - \gamma_j(U_l) \right) D_d K_{2ii} K_{il} \left( U_{id} + (\Pi_d(Z_l) - \Pi_d(Z_i)) \right) \right\} \left( 1 + O_p(L_{1n}) \right) \]

\[ = - \left\{ \sum_{d=1}^{D_3} (W_{1d} + W_{2d}) \right\} \left( 1 + O_p(L_{1n}) \right). \]

We show that \( \sum_{d=1}^{D_3} W_{1d} = \frac{1}{n} \sum_{i=1}^{n} a_{1ni} + o_p(n^{-1/2}) \), \( W_{2d} = o_p(n^{-1/2}) \).

(i) \( W_{1d} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta X_{2i,j} \left( \gamma_i(U_l) - \gamma_j(U_l) \right) D_d K_{2ii} K_{il} U_{id} \approx \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{niil} \)

If \( i \neq j \neq l \), let \( U_n = \left( \frac{1}{n} \right)^{-1} \sum_{i \neq j \neq l} \psi_{niil} = \theta_n + 3H_n^{(1)} + 3H_n^{(2)} + H_n^{(3)} \) be a U-statistic of degree 3:

- \( \theta_n = E(\psi_{niil}|P_i) = E(\psi_{niil}|P_i = E(\psi_{niil}|P_i, P_l) = 0 \), as \( E(U_{id}|Z_i) = 0 \);
- \( \phi_{tl} = E(\psi_{niil}|P_i, P_l) = \frac{U_{id}}{h_1^2 h_2^2 + f_U(U_l) f_Z(Z_l)} \times \left( \gamma_i(U_l) - \gamma_j(U_l) \right) \mid Z_l \leq C(U_{id}); \)
- \( \sigma^2_n \leq CE(\psi_{niil}) = O(1); \)
- \( \sigma^2_{2n} \leq CE(\psi_{niil}) = O_p\left(h_1^{-1} h_2^{-1}\right); \)
- \( H_n^{(1)} = O_p(n^{-1/2}), \quad H_n^{(2)} = O_p\left(\frac{\sigma^2_n}{n}\right)^{1/2} = O_p\left((n^2 h_1^{-1})^{-1/2}\right) = o_p(n^{-1/2}); \)
- \( H_n^{(3)} = O_p\left(\frac{\sigma^2_n}{n}\right)^{1/2} = O_p\left((n^3 h_1^{-1} h_2^{-2})^{-1/2}\right) = o_p(n^{-1/2}). \)

We have \( U_n = 3H_n^{(1)} + o_p(n^{-1/2}) \), \( H_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} E(\psi_{niil}|P_i). \) In this case, we need to investigate \( H_n^{(1)} \) a little further. Note that \( \gamma_i(U_l) - \gamma_j(U_l) = Y_{2i}(U_l)(U_l - U_i) + \frac{1}{2}(U_l - U_i)^T H_{2i}(U_l)(U_l - U_i), \)

where \( U_{ii} = \lambda U_i + (1 - \lambda) U_l, \) for \( \lambda \in (0, 1) \). Plugging this into \( E(\psi_{niil}|P_i) \), we have

\[ H_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} E(\psi_{niil}|P_i) = \frac{1}{n} \sum_{i=1}^{n} a_{1ni} + \frac{1}{n} \sum_{i=1}^{n} b_{1ni} \]

where

\[ a_{1ni} = \frac{U_{id}}{h_1^2 h_2^2 + f_U(U_l) f_Z(Z_l)} \left( \frac{\eta X_{2i,j} D_d K_{2ii} K_{il}}{f_U(U_l) f_Z(Z_l)} \right) \left( Y_{2i}(U_l)(U_l - U_i) \right) \]

\[ b_{1ni} = \frac{U_{id}}{h_1^2 h_2^2 + f_U(U_l) f_Z(Z_l)} \left( \frac{\eta X_{2i,j} D_d K_{2ii} K_{il}}{f_U(U_l) f_Z(Z_l)} \right) \left( \frac{1}{2} (U_l - U_i)^T H_{2i}(U_l)(U_l - U_i) \right) \]
Since $b_{1n,i} \leq C h_2 |U_{id}|$, $E(b_{1n,i}) = 0$, and $V \left( \frac{1}{n} \sum_{i=1}^{n} b_{1n,i} \right) = O(h_2^2 n^{-1})$, by Chebyshev’s Inequality, we have $\frac{1}{n} \sum_{i=1}^{n} b_{1n,i} = o_p(h_2 n^{-1/2})$, and $H_n^{(1)} = \frac{1}{n} \sum_{i=1}^{n} a_{1n,i} + o_p(n^{-1/2})$.

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

$$\sum_{i=1}^{D_2} W_{1d} = \frac{6}{n^3} \left( \frac{n}{3} \right) \sum_{i=1}^{n} \sum_{d=1}^{D_2} \frac{U_{id}}{2h_1^3 h_2^3} E \left( \frac{\eta X_{i2,i}^e D_{j} K_{2j} K_{1i} i d}{f_U(U_i) f_Z(Z_i)} J_{2}^2(U_i) \left( \frac{U_i - U_l}{h_2} \right) \right) + o_p(n^{-1/2})$$

$$= \frac{6}{n^3} \left( \frac{n}{3} \right) \sum_{i=1}^{n} a_{1n,i} + o_p(n^{-1/2})$$

$$= \frac{1}{n} \sum_{i=1}^{n} a_{1n,i} + o_p(n^{-1/2})$$

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

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

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

if $i = l \neq t$,

$$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{l \neq t} \frac{\eta X_{i2,i}^e D_{j} K_{2j} K_{1i} i d}{f_U(U_i) f_Z(Z_i)} (\gamma_2(U_i) - \gamma_2(U_l)) U_{id} = o_p \left( \left( \frac{n h_1}{D_1} \right)^{-1} \right) = o_p(n^{-1/2});$$

if $i \neq t = l$,

$$= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{l \neq t} \frac{\eta X_{i2,i}^e D_{j} K_{2j} K_{1i} i d}{f_U(U_i) f_Z(Z_i)} (\gamma_2(U_i) - \gamma_2(U_l)) U_{id} = o_p(n^{-1}) = o_p(n^{-1/2}).$$

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

(ii) $W_{2d} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{l \neq t = l} \frac{\eta X_{i2,i}^e D_{j} K_{2j} K_{1i} i d}{f_U(U_i) f_Z(Z_i)} (\gamma_2(U_i) - \gamma_2(U_l)) (\Gamma_d(Z_i) - \Gamma_d(Z_l)) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{l \neq t = l} \psi_{silt}.$

If $i \neq t \neq l$, let $U_n = \left( \frac{n}{3} \right)^{-1} \sum_{i \neq t \neq l} \psi_{silt} = \theta_n + 3 H_n^{(1)} + 3 H_n^{(2)} + H_n^{(3)}$ be a $U$-statistic of degree 3.
\( \theta_n = O(h_n^3) = o_p(n^{-1/2}); \)
\[ \text{E}(\psi_{n+1}|P_t) = \frac{\eta_{X_{2i,j}}}{h_1^3 h_2^3 f(U_i) f_Z(Z_i)} E(D_{dK_{2i}} K_{1ii} (\gamma_2(U_i) - \gamma_2(U_i)) (\Pi_{d}(Z_i) - \Pi_{d}(Z_i))) \leq \frac{1}{h_n^3} \left| \eta_{X_{2i,j}} \right|, \]
\( \text{E}(\psi_{n+1}|P_i) = O(h_n^3), \quad \text{E}(\psi_{n+1}|P_i) = O(h_n^3); \)
\( \sigma_{1n}^2 \leq C \left( E^2(\psi_{n+1}|P_t) + E^2(\psi_{n+1}|P_t) + E^2(\psi_{n+1}|P_i) \right) = O(h_n^3); \)
\[ \text{E}(\psi_{n+1}|P_t, P_t) = \frac{\eta_{X_{2i,j}} D_{dK_{2i}} (\gamma_2(U_i) - \gamma_2(U_i))}{h_1^3 h_2^3 + f(U_i) f_Z(Z_i)} E(K_{1ii}(\Pi_{d}(Z_i) - \Pi_{d}(Z_i))) \leq \frac{1}{h_n^3} \left| \eta_{X_{2i,j}} D_{dK_{2i}} \right|, \]
\[ \text{E}(\psi_{n+1}|P_t, P_t) = \frac{\eta_{X_{2i,j}} D_{dK_{2i}} (\gamma_2(U_i) - \gamma_2(U_i))}{h_1^3 h_2^3 + f(U_i) f_Z(Z_i)} E(D_{dK_{2i}} (\gamma_2(U_i) - \gamma_2(U_i))) \leq \frac{1}{h_n^3} \left| \eta_{X_{2i,j}} D_{dK_{2i}} \right|, \]
\( \text{E}(\psi_{n+1}|P_t, P_t) = O(h_n^3); \)
\( \sigma_{2n}^2 = O \left( h_1^{2s_1} h_2^{2s_2} + h_1^{2s_1} h_2^{2s_2} \right) = O \left( h_1^{2s_1} h_2^{2s_2} + h_1^{2s_1} h_2^{2s_2} \right); \)
\( \sigma_{3n}^2 = O \left( h_1^{3s_1} h_2^{3s_2} \right); \)
\( H_n^{(1)} = O_p \left( \frac{\sigma_{2n}^2}{n} \right)^{1/2} = O \left( h_1^{1/2s_1} n^{-1/2} \right) = o_p(n^{-1/2}); \)
\( H_n^{(2)} = O_p \left( \frac{\sigma_{2n}^2}{n} \right)^{1/2} = O \left( h_1^{1/2s_1} (n^2 h_2^{2s_2})^{-1/2} + (n^2 h_1^{2s_1} h_2^{2s_2})^{-1/2} \right) = o_p(n^{-1/2}); \)
\( H_n^{(3)} = O_p \left( \frac{\sigma_{2n}^2}{n} \right)^{1/2} = O \left( (n^3 h_1^{3s_1} h_2^{3s_2})^{-1/2} \right) = o_p(n^{-1/2}). \)

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

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

if \( i = t = l, \quad i = l \neq t, \quad i = t \neq l, \quad \psi_{n+1} = 0; \)

if \( i \neq t = l, \quad \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{n+1} \]
\[ = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\eta_{X_{2i,j}} D_{dK_{2i}} K_{1ii} (\gamma_2(U_i) - \gamma_2(U_i)) (\Pi_{d}(Z_i) - \Pi_{d}(Z_i)))}{h_1^{3s_1} h_2^{3s_2} f(U_i) f_Z(Z_i)} \]
\[ = o_p \left( \frac{h_1}{n} \right) = o_p(n^{-1/2}). \]

We have \( B_{312} = -\frac{1}{n} \sum_{i=1}^{n} a_{i1,i} + o_p(n^{-1/2}). \) 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

\[ W_{1d}' = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \frac{\eta_{X_{2i,j}} (\gamma_2(U_i) - \gamma_2(U_i)) D_{dK_{1ii}} K_{1ii} (\gamma_2(U_i) - \gamma_2(U_i))}{h_1^{3s_1} h_2^{3s_2} f(U_i) f_Z(Z_i)} U_{1d}. \]
The difference here is we have $Z_t$ instead of $Z_i$, such that $E(\psi_{ni\ell}|P_i) = 0$ in that $E(\eta_iX_{2j}|U_i) = 0$. 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

$$
V_{Y2} - V_{X2}\beta = \left\{ \frac{1}{nh_n^{D_2}f(U_i)} \sum_{i=1}^{n} \hat{K}_{2i} \left[ (\hat{\eta} - \eta_i) (Y_i - X_{2i}\beta) + (v_{Y2} - v_{X2}\beta) \right. \right.
$$

$$
+ \left. \left( (\gamma_2(U_i) - \gamma_2(U_i)) - (g_2(U_i) - g_2(U_i))\beta \right) \right] \left( 1 + O_p(L_{2n}) \right),
$$

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

$$
a_{ni} = \sum_{d=1}^{D_2} \frac{U_{id}}{h_1^n h_2^{D_2}} E \left( \frac{\eta_i X_{2j}^* D_d K_{2j} K_{2i} \eta_j h(U_i) \left( \frac{U_i - U_l}{h_2} \right)}{f(U_i)f(Z_i)} \right),
$$

**Step 5:** Combing orders of $B_1, B_2, B_3, B_4$, we have $\frac{1}{\sqrt{n}} \hat{\lambda}_i (\hat{\gamma} - X_2\beta) = B_{11} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{ni} + o_p(n^{-1/2})$. 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 $E(\lambda' b_{ni}) = 0$ as $E(X_{2j}^* \eta_i v_i) = E(a_{ni}) = 0$, and $E(\lambda' b_{ni} b_{ni}') = \lambda' E(X_{2j}^* \eta_i v_i X_{2j}^* \eta_j) + \lambda' E(a_{ni} a_{ni}') = \lambda' \Phi_1 \lambda + \lambda' E(a_{ni} a_{ni}') \lambda$. Denote $X_{2j,i} = \Pi_{2j}(Z_t) + U_{2j,i}$, the $j^{th}$ element of $a_{ni}$ can be written as

$$
a_{ni,j} = \sum_{d=1}^{D_2} \frac{U_{id}}{h_1^n h_2^{D_2}} E \left( \frac{\eta_i X_{2j}^* D_d K_{2j} K_{2i} \eta_j h(U_i) \left( \frac{U_i - U_l}{h_2} \right)}{f(U_i)f(Z_i)} \right).
$$

$$
= \int \frac{1}{h_1^n h_2^{D_2}} \left( \Pi_{2j}(Z_t) + U_{2j,i} - g_{1j}(M_t) - g_{2j}(U_i) + g_{3j} \right) \sum_{d=1}^{D_2} U_{id} D_d K_{2j} K_{2i} \eta_j h(U_i) \left( \frac{U_i - U_l}{h_2} \right) f(U_i)f(Z_i) f_{ZUM}(Z_t, U_i, M_t) dU_idZ_idU_idM_t
$$

$$
= \int \left( \Pi_{2j}(Z_t - h_1\gamma) + U_{2j,i} - h_2 \psi_{2j} - g_{1j}(M_t) - g_{2j}(U_i - h_2\psi) + g_{3j} \right) \sum_{d=1}^{D_2} U_{id} D_d K_{2j} \left( \psi \right) K_{1}(\gamma)
$$

$$
\times h(U_i - h_2\psi) \psi \frac{\eta_i(M_t, U_i - h_2\psi)}{f(U_i - h_2\psi)f(Z_t - h_1\gamma)} f(U_i)f_{ZUM}(Z_t - h_1\gamma, U_i - h_2\psi, M_t) d\gamma d\psi dU_idM_t.
$$

40
\[
\lim_{d} \Phi_{2,(m)}(Z) = \lim_{d} E \left[ \left( \Pi_{2d}(Z) + U_{2d} - g_{1d}(M_d) - g_{2d}(U_d) + g_{3d} \right) D_{2d} h(U_d) \eta | Z \right] U_{id}.
\]

The convergence follows by Lemma 1, A3 and that \( \int D_{d} K_{2}(\psi) \psi d\psi = (0, \ldots, -1, \ldots, 0) \), where \(-1\) appears on the \(d\)th position of the vector. Hence, the \((j,m)\)th element of \(E(a_{mi}a_{ni})\) converges to

\[
\Phi_{2,(m)}(Z) = E \left[ \sum_{j=1}^{D_{2d}} \left( \Pi_{2j}(Z) + U_{2j} - g_{1j}(M_j) - g_{2j}(U_j) + g_{3j} \right) D_{2j} h(U_j) \eta | Z \right].
\]

By Lyapunov’s central limit theorem, we have \(\sqrt{n}(B_{11} + \frac{1}{n} \sum_{i=1}^{n} d_{mi}) \xrightarrow{d} \mathcal{N}(0, \Phi_{1} + \Phi_{2})\), provided

\[
\lim_{n \to \infty} \sum_{i=1}^{n} E[n^{-1/2} \lambda^t a_{mi}]^{2+\delta} = 0 \text{ for some } \delta > 0.
\]

Note that by Cr Inequality,

\[
\sum_{i=1}^{n} E[n^{-1/2} \lambda^t a_{mi}]^{2+\delta} = \sum_{i=1}^{n} E \left[ \sum_{j=1}^{D_{2d}} \lambda_{j} a_{mi,j} \right]^{2+\delta} \leq \sum_{j=1}^{D_{2d}} \lambda_{j}^{2+\delta} E[a_{mi,j}]^{2+\delta}
\]

where \(E[a_{mi,j}]^{2+\delta} \to \int \left| \sum_{d=1}^{D_{2d}} \left( \Pi_{2j}(Z) + U_{2j} - g_{1j}(M_j) - g_{2j}(U_j) + g_{3j} \right) D_{2j} h(U_j) \eta | Z \right]^{2+\delta} U_{id}^{2+\delta} f_{ZU}(Z, U_j) dZ dU_j \)

\[
\leq C \sum_{d=1}^{D_{2d}} \int \left| \left( \Pi_{2j}(Z) + U_{2j} - g_{1j}(M_j) - g_{2j}(U_j) + g_{3j} \right) | Z \right|^{2+\delta} U_{id}^{2+\delta} f_{ZU}(Z, U_j) dZ dU_j \]

\[
< \infty \quad \text{since} \quad E[U_{id}]^{2+\delta} < C < \infty \quad \text{and} \quad E[X_{2j}]^{2+\delta} < \infty.
\]

Thus \(\lim_{n \to \infty} \sum_{i=1}^{n} E[n^{-1/2} \lambda^t a_{mi}]^{2+\delta} = 0 \) 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

\[
\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Phi_{0}^{-1}(\Phi_{1} + \Phi_{2})\Phi_{0}^{-1}).
\]
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

\[
\int K(\gamma)f(x+h_n\gamma)d\gamma \to f(x)\int K(\gamma)d\gamma \leq C \quad \text{as} \quad n \to \infty
\]

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 \(|\beta| = 0, \cdots, 3\), \(K(x)^{\beta}\) 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

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

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 \(K(x) - K(y) = JK(x^*)(x-y)\), where \(x^* = x + \lambda(y-x), \lambda \in (0,1), \) and \(|D_j K(x^*)| = |k'(x^*)| \prod_{p\neq j} |k(x^*_p)| \leq C\), we have \(|K(x) - K(y)| \leq C \sum_{i=1}^D |x_i - y_i| \leq CD (\sum_{i=1}^D (x_i - y_i)^2)^{1/2} \leq C||x - y||_E\) for some \(C > 0\) by triangular and \(C_r\) Inequality.

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

\[
|K(x)x_i - K(y)y_i| = |x_i(K(x) - K(y)) + K(y)(x_i - y_i)|
\]

\[
= |x_iJK(x^*)(x-y) + K(y)(x_i - y_i)| \quad \text{by the mean value theorem}
\]

\[
= \left| (x_iD_j K(x^*) + K(y))(x_i - y_i) + \sum_{p \neq i} x_i D_p K(x^*)(x_p - y_p) \right|
\]

\[
\leq C \sum_{i=1}^D |x_i - y_i| \quad \text{by triangular inequality}
\]

\[
\leq C||x - y||_E \quad \text{by the} \ C_r \text{ Inequality}
\]
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 $x_i^+ = x_i + \lambda(y_i - x_i) = x_i + \lambda \Delta$, and we have $|x_i k'(x_i^+)| = |x_i k'(x_i + \lambda \Delta)| \leq C$ by c).

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

$$|K(x)x_i x_j - K(y)y_i y_j| = |x_i(K(x)x_i - K(y)y_i) + K(y)y_i (x_j - y_j)|$$

\[
\leq |x_i K(x) + x_j y_i D_j K(x^*)||x_i - y_i| + |x_j y_i D_j K(x^*) + K(y)y_i||x_j - y_j|
\]

\[
+ \sum_{p \neq i,j} x_j y_i D_p K(x^*)||x_p - y_p|
\]

\[
\leq C ||x - y||_E
\]

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

$$|K(x)x_i x_j x_l - K(y)y_i y_j y_l| = |x_i(K(x)x_i x_j - K(y)y_i y_j) + K(y)y_i y_j (x_l - y_l)|$$

\[
\leq |x_i x_j x_l D_l K(x^*) + x_j y_i K(y)||x_i - y_i| + |x_i x_j x_l D_j K(x^*) + x_l K(y)y_i||x_j - y_j|
\]

\[
+ |x_i x_j x_l D_l K(x^*) + K(y)y_i y_j||x_l - y_l| + \sum_{p \neq i,j,l} |x_i x_j x_l D_p K(x^*)||x_p - x_p|
\]

\[
\leq C ||x - y||_E
\]

\[\square\]

**Lemma 3.** Let $\{W_i\}_{i=1}^{n}$ be a sequence of independent and identically distributed (IID) random variables, $G_n(W_i, w) : \mathbb{R} \times \mathbb{R}^K \rightarrow \mathbb{R}$ 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 $E((|G_n(W_i, w) - E(G_n(W_i, w))|^p) \leq C^{p-2} p! E((G_n(W_i, w) - E(G_n(W_i, w)))^2) < \infty$ 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 $G_w \subset B(w_0, r) = \{w \in \mathbb{R}^K : ||w - w_0|| < r\}$.  

43
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 \( \{B(w_k, n^{-1/2})\}_{k=1}^l \) 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
\[
|S_n(w) - S_n(w_k)| \leq n^{-1/2} \frac{1}{n} \sum_{i=1}^n B_n(w_i) = O_p(n^{-1/2})
\]
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 \( \left(\frac{n}{\log n}\right)^{1/2} n^{-1/2} = o(1) \) it suffices to show that for all \( \varepsilon > 0 \), there exists a constant \( \Delta_\varepsilon \) such that for \( n \geq N \)
\[
P\left( \left(\frac{n}{\log n}\right)^{1/2} \max_{1 \leq k \leq l_n} |S_n(w_k) - E(S_n(w_k))| \geq \Delta_\varepsilon \right) \leq \varepsilon.
\]
Let \( \varepsilon_n = \left(\frac{\log n}{n}\right)^{1/2} \Delta_\varepsilon \) and note that
\[
P\left( \max_{1 \leq k \leq l_n} |S_n(w_k) - E(S_n(w_k))| \geq \varepsilon_n \right) \leq \sum_{k=1}^{l_n} P(|S_n(w_k) - E(S_n(w_k))| \geq \varepsilon_n).
\]
Given condition b), and letting \( c_n = 4V(G_n(W_i, w_k)) + 2C\varepsilon_n \), by Bernstein’s Inequality, we have
\[
P\left( \left| \sum_{i=1}^n G_n(W_i, w_k) - \sum_{i=1}^n E(G_n(W_i, w_k)) \right| \geq n\varepsilon_n \right) \leq 2\exp\left(-\frac{n\varepsilon_n^2}{c_n}\right) = 2\exp\left(-\frac{\Delta_\varepsilon^2 \log n}{c_n}\right) = 2n^{-\frac{\Delta_\varepsilon^2}{c_n}}.
\]
Hence, \( P\left( \max_{1 \leq k \leq l_n} |S_n(w_k) - E(S_n(w_k))| \geq \varepsilon_n \right) \leq 2l_n n^{-\frac{\Delta_\varepsilon^2}{c_n}} < C n^{K/2 - \frac{\Delta_\varepsilon^2}{c_n}} \). Since, \( \varepsilon_n \to 0 \) as and \( V(G_n(W_i, w_k)) < \infty \), we can choose \( \Delta_\varepsilon \) sufficiently large such that \( K/2 - \frac{\Delta_\varepsilon^2}{c_n} < 0 \) and
\[
P\left( \max_{1 \leq k \leq l_n} |S_n(w_k) - E(S_n(w_k))| \geq \varepsilon_n \right) \leq \varepsilon.
\]

\[\square\]

**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) \( \{(X_i, \varepsilon_i)^{r}_{1,2,...} \} \) is an independent and identically distributed sequence of random vectors; 2) The joint density of \( X_i \) and \( \varepsilon_i \) is given by \( f_{X_i}(x, \varepsilon) = f_X(x) f_{\varepsilon|x}(\varepsilon|x) \); 3) \( f_X(x) \) and all of its partial
derivatives of order \(< s\) are differentiable and uniformly bounded on \(\mathbb{R}^D\), 4) \(0 < \inf f(x) \) and \(\sup f(x) \leq C\). Let \(w(X_t; x) : \mathbb{R}^D \to \mathbb{R}\) and \(g(\varepsilon) : \mathbb{R} \to \mathbb{R}\) be measurable functions. Define
\[
s(x) = \frac{1}{n h_n^D} \sum_{i=1}^n K \left( X_i - x \right) \left( \frac{X_i - x}{h_n} \right)^\beta w(X_t; x) g(\varepsilon_i)
\]
where \(|\beta| = 0, 1, 2, 3\). If
i) \(E(|g(\varepsilon)|^a | X) \leq C < \infty\) for some \(a \geq 2\);
ii) \(w(X_t; x)\) satisfies a Lipschitz condition and \(|w(X_t; x)| < C\) for all \(x \in \mathbb{R}^D\);

Then, for an arbitrary compact set \(\mathcal{G} \subseteq \mathbb{R}^D\), we have
\[
\sup_{x \in \mathcal{G}} |s(x) - E(s(x))| = O_p \left( \frac{\log n}{n h_n^D} \right)^{1/2}
\]
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 \(r \in \mathbb{R}^+\). \(\mathcal{G}\) compact implies that there exists \(x_0 \in \mathbb{R}^D\) such that \(\mathcal{G} \subseteq B(x_0, r)\). Therefore, for all \(x, z \in \mathcal{G}\), \(||x - z||_E < 2r\). 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 \(\mathcal{G}\) 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 \(x^k \in \mathcal{G}\) and \(l_n \leq C \left( \frac{n}{h_n^{D+2}} \right)^{D/2}\). Now let
\[
s^\tau(x) = \frac{1}{nh_n^D} \sum_{i=1}^n K \left( X_i - x \right) \left( \frac{X_i - x}{h_n} \right)^\beta w(X_t; x) g(\varepsilon_i) \chi_{\{|g(\varepsilon_i)| \leq B_u\}}
\]
with \(B_1 \leq B_2 \leq \cdots\) such that \(\sum_{i=1}^\infty B_i^{-a} < \infty\) for some \(a > 0\).

\[
\sup_{x \in \mathcal{G}} |s(x) - E(s(x))| \leq \sup_{x \in \mathcal{G}} |s(x) - s^\tau(x)| + \sup_{x \in \mathcal{G}} |E(s(x) - s^\tau(x))| + \sup_{x \in \mathcal{G}} |s^\tau(x) - E(s^\tau(x))| \equiv T_1 + T_2 + T_3.
\]

1. \(T_1 = \sup_{x \in \mathcal{G}} \left( \frac{nh_n^D}{l_n} \right)^{-1} \sum_{i=1}^n K \left( \frac{X_i - x}{h_n} \right) \left( \frac{X_i - x}{h_n} \right)^\beta w(X_t; x) g(\varepsilon_i) \chi_{\{|g(\varepsilon_i)| \leq B_u\}}\). By Chebyshev’s Inequality, for \(a > 0\),
2. For \( T_n > 1 \), which gives we conclude that for any \( m \), due to uniform bound of \( w \) where the first integral after the inequality is uniformly bounded by i) and by Chebyshev’s Inequality, 

\[
\limsup_{t \to \infty} \left\{ \frac{\mathbb{E}(|g(\epsilon)|^a)}{B_n^a} \right\} < C \sum_{t=1}^{\infty} B_t^{-a} < \infty
\]

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

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

\[
\mathbb{E}(s(x) - s^*(x)) = \frac{1}{m h^B} \sum_{t=1}^{\infty} \int_{|g(\epsilon)| > B_n} K \left( \frac{X_t - x}{h_n} \right) \left( \frac{X_t - x}{h_n} \right)^{\beta} w(X_t - x; x) g(\epsilon_t) f_X(X_t) f(\epsilon_t) dX_t d\epsilon
\]

\[
\leq \int K(\gamma)^{\beta} w(h_n \gamma; x) f_X(x + h_n \gamma) d\gamma \int |g(\epsilon)| f(\epsilon)^{X} |X(\epsilon)| \chi_{|g(\epsilon)| > B_n} d\epsilon
\]

\[
\leq C \int |g(\epsilon)| f(\epsilon) \chi_{|g(\epsilon)| > B_n} d\epsilon
\]

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

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

By Hölder’s Inequality, for \( a > 1 \), we have

\[
\int |g(\epsilon)| f(\epsilon)^{X} |X(\epsilon)| \chi_{|g(\epsilon)| > B_n} d\epsilon \leq \left( \int |g(\epsilon)|^a f(\epsilon)^{\alpha} d\epsilon \right)^{\frac{1}{a}} \left( \int \chi_{|g(\epsilon)| > B_n} f_X(\epsilon) d\epsilon \right)^{\frac{1}{1-a}}
\]

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

\[
\left( \int \chi_{|g(\epsilon)| > B_n} f(\epsilon)^{\alpha} |X| d\epsilon \right)^{\frac{1}{1-a}} = (P(|g(\epsilon)| > B_n | X))^{\frac{1}{1-a}} \leq C \left( \frac{\mathbb{E}(|g(\epsilon)|^a | X)}{B_n^a} \right)^{\frac{1}{1-a}} \leq CB_n^{1-a}.
\]

46
Hence, \( T_2 = O(B_n^{1-a}) \).

3. Rewrite \( T_3 \) as: \( T_3 = \sup_{x \in \mathcal{D}} |s^x(x) - E(s^x(x))| \leq \sup_{x \in \mathcal{D}} |s^x(x) - s^x(x^k)| + \sup_{x \in \mathcal{D}} |E(s^x(x) - s^x(x^k))| \\
+ \max_{1 \leq k \leq n} |s^x(x^k) - E(s^x(x^k))| \equiv T_{31} + T_{32} + T_{33}.

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

\[
|s^x(x) - s^x(x^k)| \leq \frac{1}{nh_n^2} \sum_{i=1}^n \left| K \left( \frac{X_i - x^k}{h_n} \right) \left( \frac{X_i - x}{h_n} \right) \right|^\beta \left| w(X_i - x, x) \right|
+ \left| K \left( \frac{X_i - x^k}{h_n} \right) \left( \frac{X_i - x}{h_n} \right) \right|^\beta \left| w(X_i - x, x - x^k) - w(x, x - x^k) \right| |
\]

\[
\leq \left( \frac{C}{h_n^2 + |x^k - x|} + h_n \frac{C}{h_n^2 + |x^k - x|} \right) \frac{1}{n} \sum_{i=1}^n |g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}|
\]

\[
\leq C \left( \frac{1}{nh_n^2} \right)^{1/2} + h_n \left( \frac{1}{nh_n^2} \right)^{1/2} \frac{1}{n} \sum_{i=1}^n |g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}|,
\]

where the second inequality follows by Lemma 2 and b), i.e., local Lipschitz condition and uniform boundedness of \( K \left( \frac{X_i - x^k}{h_n} \right) \left( \frac{X_i - x}{h_n} \right) \). By the measurability of \( g \) and condition 1) we have that

\( \{ |g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}| \}_{i=1,2,\ldots} \) is IID. By condition i) and Kolmogorov’s law of large numbers (LLN) we have \( \frac{1}{n} \sum_{i=1}^n |g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}| - E(|g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}|) = o_p(1) \) and \( T_{31} \leq C \left( \frac{1}{nh_n^2} \right)^{1/2} \).

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

3.3. \( T_{33} = \max_{1 \leq k \leq n} |s^x(x^k) - E(s^x(x^k))| \). For \( \varepsilon_n = \frac{nD}{\log n} \Delta \) with \( 0 < \Delta \to 0 \) we note that

\[
\mathbb{P} \left( \max_{1 \leq k \leq n} |s^x(x^k) - E(s^x(x^k))| \geq \epsilon_n \right) \leq \sum_{k=1}^n \mathbb{P} \left( |s^x(x^k) - E(s^x(x^k))| \geq \epsilon_n \right).
\]

Let \( s^x(x^k) - E(s^x(x^k)) = \frac{1}{n} \sum_{i=1}^n Z_{in} \) with

\[
Z_{in} = \frac{1}{h_n^2} K \left( \frac{X_i - x^k}{h_n} \right) \left( \frac{X_i - x}{h_n} \right) \left| w(X_i - x^k, x) \right| g(e_i) \chi_{\{|g(e_i)| \leq B_n\}}
- E \left( \frac{1}{h_n^2} K \left( \frac{X_i - x^k}{h_n} \right) \left( \frac{X_i - x}{h_n} \right) \left| w(X_i - x^k, x) \right| g(e_i) \chi_{\{|g(e_i)| \leq B_n\}} \right).
\]

By the bounds on \( |K(x)||x^\beta| \) and \( |w(x)||g(e_i)\chi_{\{|g(e_i)| \leq B_n\}}| \leq B_n \) we have that \( |Z_{in}| \leq Ch_n^{-D}B_n \). By Bernstein’s
where \( c(n) = 2h_n^D V(Z_n) + \frac{2}{3} CB_n \left( \frac{\log n}{nh_n^D} \right)^{1/2} \Delta_c^2 \). Consequently,

\[
P \left( \max_{1 \leq k \leq n} |s^T(x) - E(s^T(x))| \geq \varepsilon_n \right) \leq 2n n^{-\frac{\Delta_c^2}{\varepsilon_n}} \leq 2C \left( \frac{n}{h_n^{D+2}} \right)^{D/2} n^{-\frac{\Delta_c^2}{\varepsilon_n}} = 2C \left( \frac{1}{h_n^{D+2} n^{\frac{2\Delta_c^2}{\varepsilon_n} - 1}} \right)^{D/2}
\]

provided \( \Delta_c^2/D > c(n) \). Hence, given that \( nh_n^{D+2} \rightarrow \infty \) as \( n \rightarrow \infty \) 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 \( B_n \varepsilon_n \rightarrow 0 \), i.e., \( B_n \varepsilon_n = o(1) \), guaranteeing that the second term of \( c(n) \) is \( o(1) \). Furthermore, \( h_n^D V(Z_n) \leq C \) 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 \( \sup_{s \in \Theta} |s(x) - E(s(x))| = O(B_n^{1-a}) + O \left( \left( \frac{\log n}{nh_n^D} \right)^{1/2} \right) \). 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

\[
\sup_{s \in \Theta} |s(x) - E(s(x))| = O_p \left( \left( \frac{\log n}{nh_n^D} \right)^{1/2} \right).
\]
References


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


